

배아줄기세포에서 트랜스 스플라이싱 전사체의 분석

하홍석·허재원¹·김대수²·박상제·배진한·안궁·윤세은·김희수*

부산대학교 자연과학대학 생명과학과, ¹한국생명공학연구원 국가영장류센터, ²한국생명공학연구원 국가생물자원정보관리센터

Received March 6, 2009 / Accepted April 10, 2009

Analysis of Trans-splicing Transcripts in Embryonic Stem Cell. Ha Hong-Seok, Huh Jae-Won¹, Kim Dae-Soo², Park Sang-Je, Bae Jin-Han, Ahn Kung, Yun Se-Eun and Kim Heui-Soo*. *Department of Biological Sciences, College of Natural Sciences, Pusan National University, ¹National Primate Research Center, KRIBB, ²Korean BioInformation Center, KRIBB - Genetic mutations by gene fusion result from chromosomal rearrangement, trans-splicing, and intergenic splicing. Trans-splicing is a phenomenon in which two pre-mRNAs grow together into one. We analyzed the trans-splicing products in embryonic stem cells. By using bioinformatic tools, 70 trans-splicing transcripts were identified. They are classified into 6 types according to fusion pattern: 5'UTR-5'UTR, 5'UTR-3'UTR, 3'UTR-3'UTR, 5'UTR-CDS, 3'UTR-CDS, CDS-CDS. The fusion products are more abundant in CDS regions than in UTR regions, which contain multiple intron numbers. Chromosome analysis showing gene fusion via trans-splicing indicated that chromosomes 17 and 19 were activated. These data are of great use for further studies in relation to fusion genes and human diseases.*

Key words : Fusion gene, trans-splicing, bioinformatics, genetic mutation, chromosome analysis

서 론

두 개 이상의 유전자가 융합이 되는 유전자 융합(gene fusion)에 관한 연구는 이미 오래전부터 이뤄지고 있었다. 이러한 유전자 융합을 일으키는 기작으로 염색체의 재배열(chromosomal rearrangement)과 트랜스 스플라이싱(trans-splicing), 유전자간 스플라이싱(intergenic splicing) 등이 있다.

가장 많은 연구가 이뤄진 염색체의 재배열에 의한 유전자 융합은 암과 밀접한 연관을 가진다[1,13,10,17,18,23]. 염색체 전좌(translocation)로 인해 Burkitt 림프종이 야기되는데, 그러한 이유는 c-MYC과 BCL2의 발현 조절의 변화에 의한 것이다[6,7,16]. 비슷한 전좌가 만성 골수성 백혈병에서 일어났는데, 여기서 BCL-ABL의 융합 단백질이 발견되었다[20]. 고형암의 경우에는 염색체의 이상에 의해서 retinoblastoma의 RB, Wilms 종양의 WT1, 결장암의 DCC처럼 암억제 유전자의 발현이 변화됨을 알 수 있었다[15]. 1번 염색체의 TPR에 의해서 7번 염색체의 c-MET이 활성화에 의해 융합 전사체(chimeric transcripts)가 만들어진다는[22]. 이후에 갑상선 암에서 염색체의 재배열에 의해서 RET/PTC 융합 전사체가 발견되는 등[18] 많은 융합 전사체가 발견되었다.

반면에 인간에서의 트랜스 스플라이싱 연구는 상대적으로 적게 진행되었다. 스플라이싱은 전사기작 과정 중에 인트론 영역을 제거하고 필요한 엑손 영역만을 골라내는 중요한 기

작이다. 하나의 pre-mRNA를 이용한 기작을 시스 스플라이싱(cis-splicing)이라고 부르며, 스플라이세오솜(spliceosome)에 의해서 형성된다. 트랜스 스플라이싱은 두개의 pre-mRNA를 이용하여 성숙한 mRNA를 만드는 기작이다. 주로 트리파노소마, 선충, 식물과 조류의 엽록체, 식물의 미토콘드리아에서 발견되었다[2]. 선충과 simaian virus 40의 RNA를 COS 세포에 트랜스펙션을 시키는 실험을 통해서, 진핵생물과 포유동물에서 트랜스 스플라이싱 기작이 보존되어 있음을 알게 되었다[4]. 이후에 포유동물 세포에서도 트랜스 스플라이싱 현상이 탐지되었다[3,19,21,24]. 그리고 생쥐의 간 COT유전자에서 트랜스 스플라이싱의 결과 중 하나인 엑손 반복(exon repetition)을 확인하고, 이를 in vitro 상에서 증명하는 실험도 진행되었다[5].

정상적인 조직에서는 적은 양의 융합 전사체의 발현이 관찰된다[9]. 정상 조직에서는 염색체의 재배열이 적기 때문이다. 따라서 트랜스 스플라이싱을 연구하는 과학자들은 정상적인 조직에서 발견되는 전사체에 대하여 주목하게 되었다. 그 중 하나로 자궁 기질 암(human endometrial stromal sarcomas, ESSs)의 7번 염색체와 17번 염색체의 전좌에 의해서 형성된 JAZF1-JJAZ1의 융합 전사체와 유사한 전사체가 전좌가 전혀 일어나지 않은 정상의 인간 자궁 내막 기질 세포(human endometrial stromal cell line, HESC)에서 발견되었다[11]. 이러한 사실을 고려해 볼 때, 정상조직에서 발견되는 융합 전사체는 트랜스 스플라이싱에 의해 형성될 가능성이 높다는 것을 시사한다. 따라서 우리는 현재 등록된 mRNA 데이터에서 배아줄기세포에서 발견되는 융합 전사체를 통해 트랜스 스플라이싱 전사체에 대한 분석을 하였다.

*Corresponding author

Tel : +82-51-510-2259, Fax : +82-51-581-2962

E-mail : khs307@pusan.ac.kr

재료 및 방법

트랜스 스플라이싱 전사체 동정

NCBI (<http://www.ncbi.nlm.nih.gov/>) 데이터베이스에서 EST와 Human Full length cDNA의 염기 서열 정보를 확보하였다. eVOC ontologies와 cDNA Library의 정보를 토대로 정상 조직을 선택하였으며, 그 중에서도 염색체 이상에 의한 데이터를 줄이기 위해서, 배아줄기세포 유래의 전사체만을 추출하였다. BLAST (100 bp,97%)를 통해서 두 개의 유전자의 염기 서열을 가진 전사체만을 분류하였다. 이후에 유전자간 선택적 스플라이싱에 의한 전사체를 제거하기 위해 다른 염색체상에 존재하는 유전자의 융합 사례만을 추출하고, 이들의 스플라이싱 구간을 확인하였다.

두 개의 유전자의 융합 영역 분석

BLAST를 통해서 전사체의 두 영역을 구분하여 각각의 서열을 UCSC (<http://genome.ucsc.edu/>)의 BLAT 프로그램에 넣어서 기존의 유전자와의 위치 관계를 확인하였다. 이때 매치되는 염기 서열의 방향을 고려하여, 정확히 융합된 위치를 규명하였다.

융합 유전자와 염색체간의 관계 분석

두 개의 유전자의 염색체 정보를 매치시키고, 융합 전사체의 두 영역 중에 한 영역이 유전자군과 유사하여 염색체 정보가 여러 개일 경우에는 가장 서열의 유사성이 높은 경우를 대상으로 하였다.

결 과

배아줄기세포에서 발견하는 트랜스 스플라이싱 전사체 데이터를 두 개의 유전자의 융합 되는 형태에 따라서, 5'UTR-5'UTR, 5'UTR-3'UTR, 3'UTR-3'UTR, 5'UTR-CDS, 3'UTR-CDS, CDS-CDS의 6개의 유형으로 분류하여 동정하였다(Fig. 1). 두 유전자의 융합되는 영역을 독립적으로 분석해 보면, CDS 영역이 100개로 많이 융합되고 있는 현상을 쉽게 인지할 수 있었으며, 3'UTR과 5'UTR의 경우는 30개, 10개로 탐지되었다(Fig. 2).

융합되는 유전자의 염색체상의 위치분석을 시도하였다. 17번과 2번 염색체가 4건으로 가장 많이 융합되었음을 알 수 있었다. 17번과 19번 염색체, 10번과 19번 염색체, 2번과 8번 염색체가 3건으로 그 다음을 차지했다(Fig. 3A). 핵형분석에 의한 다양한 염색체 분석 결과, 유전자의 융합되는 빈도가 1번에서 12번까지는 평균 6.5건의 비율로 비교적 많은 현상으로 탐지되고 있었으나, 13번에서 Y염색체까지는 평균 2.8건으로 탐지되었다. 염색체 17번과 19번의 경우에는 각각 15건, 17건으로 두드러지게 높은 빈도수를 보였다(Fig. 3B).

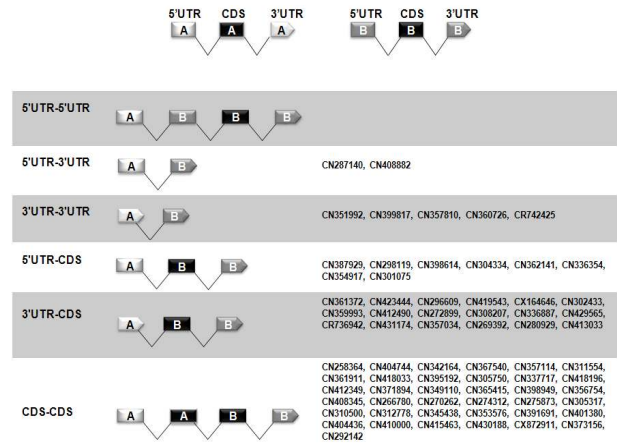


Fig. 1. Identification of trans-splicing transcripts and classification by 6 types structure (5'UTR-5'UTR, 5'UTR-3'UTR, 3'UTR-3'UTR, 5'UTR-CDS, 3'UTR-CDS, CDS-CDS).

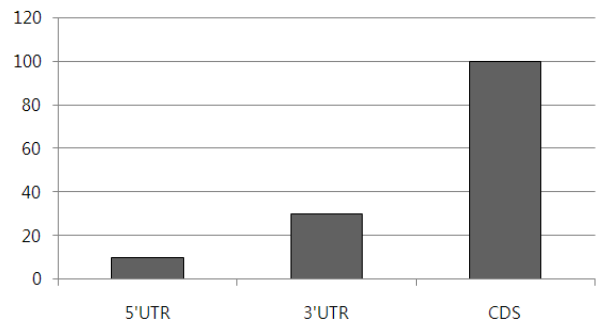


Fig. 2. Frequency of fusion region by independent events.

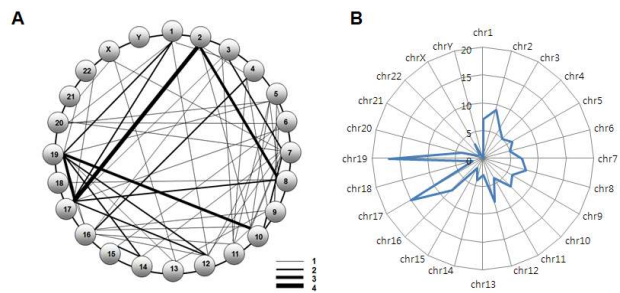


Fig. 3. Schematic diagrams of inter-chromosomal relationship (A) and chromosomal distributions of independent events (B).

고 찰

유전자의 융합 위치와 빈도의 관계

CDS, 3'UTR, 5'UTR 순으로 융합의 빈도수와 조합이 명확하게 다음을 알 수 있었다. 이러한 현상이 나타나는 가장 큰 이유는 엑손의 크기나 개수의 영향으로 생각된다. CDS는 인트론 개수가 많기 때문에[8], 더 많은 트랜스 스플라이싱이 일어날 확률을 제공할 수 있으리라 사료된다. 즉, 엑손유전자의

사이에 있는 인트론의 5'에 있는 donor site와 3'에 위치하고 있는 acceptor site의 개수의 차이에 의해 트랜스 스플라이싱이 일어날 확률이 높음을 시사한다. 인간 유전자의 3'UTR은 5'UTR보다 평균적으로 5배 정도 크지만, 인트론의 수는 36배 정도 적은 편이다[14]. 이러한 점을 감안하면, 인트론의 수와 더불어 융합하는 구간의 크기도 중요한 요인으로 사료된다.

융합되는 유전자의 염색체상 위치 분석

인간 염색체의 핵형에 의한 융합유전자의 염색체상의 위치를 상호 나열하여 분석해 본 결과 염색체의 크기가 다소 영향력을 주고 있음을 시사하였다. 즉, 전반적으로 염색체의 크기가 큰 1번에서 12번 염색체까지 빈도수가 고르게 분포하고 있었으며, 17번과 19번 염색체를 제외하고는 13번 이후의 상염색체와 두 개의 성염색체는 낮은 빈도를 보였다(Fig. 3). 유전체가 크면, 그곳에 위치한 유전자의 수도 상대적으로 많기 때문에 트랜스 스플라이싱이 일어날 확률이 높아진다. 하지만 염색체별 유전체의 크기가 번호와 거의 비례하는 점과 실제 성염색체인 X염색체는 6, 7번 정도의 크기임을 감안했을 때, 크기가 절대적인 요소라고 하기는 어렵다. 특히 17번과 19번 염색체의 두드러지게 높은 수치는 또 다른 요소가 작용하였을 것으로 추측된다. 염색체 각각에는 다양한 스플라이세오솜에 관련된 유전자들이 존재한다. 이들은 일반적으로 알려진 시스템 스플라이싱뿐만 아니라 트랜스 스플라이싱에도 영향을 주는 것으로 알려져 있다[12]. 따라서 17번 염색체에 존재하는 PRPF8, EFTUD2, SFRS1, THOC4, DHX8, DDX5, GEMIN4, EIF4A3, DDX42, SFRS2, POLR2A, DHX33 12개의 유전자, 19번 염색체에 존재하는 SF3A2, C19orf29, C19orf43, U2AF2, HNRNPM, SNRPA, SNRNP70, SNRNP2, PTBP1, MORG1, GEMIN7, PRPF31, SF4, U2AF1L4, XAB2, HNRNPUL1, HNRNPL, LOC100130932, DDX39, POLR2I, POLR2E, GTF2F1, DHX34 23개의 유전자들은 트랜스 스플라이싱을 담당할 수 있는 후보 유전자로 시사된다.

요약

유전자의 융합으로 인한 돌연변이는 염색체 재배열, 트랜스 스플라이싱, 유전자간 스플라이싱으로 인하여 야기된다고 알려져 있다. 우리는 두 개의 서로 다른 유전자의 pre-mRNA의 융합으로 인하여 만들어지는 트랜스 스플라이싱의 전사 산물에 관심을 가져, 인간의 태아 줄기 세포에서 이러한 돌연변이 양상을 분석하였다. 배아줄기세포의 mRNA에서 트랜스 스플라이싱 전사체 70개를 탐지해 내고, 이들의 융합되는 패턴에 따라 5'UTR-5'UTR, 5'UTR-3'UTR, 3'UTR-3'UTR, 5'UTR-CDS, 3'UTR-CDS, CDS-CDS의 6개의 유형으로 분류하여 분석하였다. 두 유전자의 융합되는 영역은 UTR영역보다 CDS에서 풍부하였는데, 이러한 이유는 많은 인트론 수로 인해 야기

되는 것으로 추정된다. 융합되는 유전자의 염색체상의 위치분석 결과, 17번과 19번 염색체가 융합유전자의 활성화를 나타내었다. 이러한 연구결과는 향후 융합유전자와 인간의 질병 연구에 크게 기여할 것으로 사료된다.

감사의 글

이 논문은 2007년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음 (KRF-2007-521-C00213).

References

- Bernard, O., N. Lecoite, P. Jonveaux, M. Souyri, M. Mauchauffe, R. Berger, C. J. Larsen, and D. Mathieu-Mahul. 1991. Two site-specific deletions and t(1;14) translocation restricted to human T-cell acute leukemias disrupt the 5' part of the tal-1 gene. *Oncogene* 6, 1477-1488.
- Bonen, L. 1993. Trans-splicing of pre-mRNA in plants, animals, and protists. *FASEB J.* 7, 40-46.
- Breen, M. A. and S. J. Ashcroft. A truncated isoform of Ca²⁺/calmodulin-dependent protein kinase II expressed in human islets of Langerhans may result from trans-splicing. *FEBS Lett.* 1997 409, 375-379.
- Bruzik, J. P. and T. Maniatis. 1992. Spliced leader RNAs from lower eukaryotes are trans-spliced in mammalian cells. *Nature* 360, 692-695.
- Caudevilla, C., D. Serra, A. Miliar, C. Codony, G. Asins, M. Bach, and F. G. Hegardt. 1998. Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. *Proc. Natl. Acad. Sci. USA* 95, 12185-12190.
- Cleary, M. L., S. D. Smith, and J. Sklar. 1986. Cloning and structural analysis of cDNAs for bcl-2 and a hybrid bcl-2/immunoglobulin transcript resulting from the t(14;18) translocation. *Cell* 47, 19-28.
- Hockenbery, D., G. Nunez, C. Milliman, R. D. Schreiber, and S. J. Korsmeyer. 1990. Bcl-2 is an inner mitochondrial membrane protein that blocks programmed cell death. *Nature* 348, 334-346.
- Hong, X., D. G. Scofield, and M. Lynch. 2006. Intron size, abundance, and distribution within untranslated regions of genes. *Mol. Biol. Evol.* 23, 2392-2404.
- Janz, S., M. Potter, and C. S. Rabkin. 2003. Lymphoma- and leukemia-associated chromosomal translocations in healthy individuals. *Genes Chromosomes Cancer* 36, 211-223.
- Kourlas, P. J., M. P. Strout, B. Becknell, M. L. Veronese, C. M. Croce, K. S. Theil, R. Krahe, T. Ruutu, S. Knuutila, C. D. Bloomfield, and M. A. Caligiuri. 2000. Identification of a gene at 11q23 encoding a guanine nucleotide exchange factor: evidence for its fusion with MLL in acute myeloid leukemia. *Proc. Natl. Acad. Sci. USA* 97, 2145-2150.
- Li, H., J. Wang, G. Mor, and J. Sklar. 2008. A neoplastic gene fusion mimics trans-splicing of RNAs in normal hu-

- man cells. *Science* **321**, 1357-1361.
12. Lidie, K. B. and F. M. van Dolah. 2007. Spliced leader RNA-mediated trans-splicing in a dinoflagellate, *Karenia brevis*. *J. Eukaryot Microbiol.* **54**, 427-435.
 13. Mitelman, F., B. Johansson, and F. Mertens. 2007. The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* **7**, 233-245.
 14. Mignone, F., C. Gissi, S. Liuni, and G. Pesole. 2002. Untranslated regions of mRNAs. *Genome Biol.* **3**, e0004.1-0004.9.
 15. Mikkelsen, T. and W. K. Cavenee. 1990. Suppressors of the malignant phenotype. *Cell Growth Differ.* **1**, 201-207.
 16. Nishikura, K., A. ar-Rushdi, J. Erikson, R. Watt, G. Rovera, and C. M. Croce. 1983. Differential expression of the normal and of the translocated human c-myc oncogenes in B cells. *Proc. Natl. Acad. Sci. USA* **80**, 4822-4826.
 17. Pardanani, A., R. P. Ketterling, S. R. Brockman, H. C. Flynn, S. F. Paternoster, B. M. Shearer, T. L. Reeder, C. Y. Li, N. C. Cross, J. Cools, D. G. Gilliland, G. W. Dewald, and A. Tefferi. 2003. CHIC2 deletion, a surrogate for FIP1L1-PDGFR α fusion, occurs in systemic mastocytosis associated with eosinophilia and predicts response to imatinib mesylate therapy. *Blood* **102**, 3093-3096.
 18. Pierotti, M. A., M. Santoro, R. B. Jenkins, G. Sozzi, I. Bongarzone, M. Grieco, N. Monzini, M. Miozzo, M. A. Herrmann, A. Fusco, I. D. Hay, G. D. Porta, and G. Vecchio. 1992. Characterization of an inversion on the long arm of chromosome 10 juxtaposing D10S170 and RET and creating the oncogenic sequence RET/PTC. *Proc. Natl. Acad. Sci. USA* **89**, 1616-1620.
 19. Shimizu, A. and T. Honjo. 1993. Synthesis and regulation of trans-mRNA encoding the immunoglobulin epsilon heavy chain. *FASEB. J.* **7**, 149-154.
 20. Shtivelman, E., B. Lifshitz, R. P. Gale, and E. Canaani. 1985. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. *Nature* **315**, 550-554.
 21. Sullivan, P. M., P. Petrusz, C. Szpirer, and D. R. Joseph. 1991. Alternative processing of androgen-binding protein RNA transcripts in fetal rat liver. Identification of a transcript formed by trans splicing. *J. Biol. Chem.* **266**, 143-154.
 22. Testa, J. R., M. Park, D. G. Blair, A. Kalbakji, K. Arden, and G. F. Vande Woude. 1990. Analysis by pulsed field gel electrophoresis reveals complex rearrangements in two MET alleles in a chemically-treated human cell line, MNNG-HOS. *Oncogene* **5**, 1565-1571.
 23. Tomlins, S. A., D. R. Rhodes, S. Perner, S. M. Dhanasekaran, R. Mehra, X. W. Sun, S. Varambally, X. Cao, J. Tchinda, R. Kuefer, C. Lee, J. E. Montie, R. B. Shah, K. J. Pienta, M. A. Rubin, and A. M. Chinnaiyan. 2005. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-648.
 24. Vellard, M., A. Sureau, J. Soret, C. Martinerie, and B. Perbal. 1992. A potential splicing factor is encoded by the opposite strand of the trans-spliced c-myc exon. *Proc. Natl. Acad. Sci. USA* **89**, 2511-2515.