Journal of Computational Intelligence in Bioinformatics.
ISSN: 0973-385X Volume 1 Number 1 (2008) pp. 55–63

http://www.ripublication.com/jcib.htm

UTR Mutation Analysis

V. Niranjan[1], R. Mahmood[1], Arpit Saxena[2] and A. KalaiVani[3]

[1]Department of Biotechnology and Bioinformatics,
Kuvempu University, Jnana Sahyadri,
Shankaraghatta, Karnataka 577451, India
[2]Guru Gobind Singh Indraprastha University, New Delhi, India
[3]Department of Pharmacoinformatics,
Sastra University, Thanjore, Tamilnadu, India

Abstract

One of the objectives of molecular biologists is to determine the types and amounts of various proteins. The presence of proteins, in turn, depend upon the concentration of mRNA from which it is synthesized, the frequency at which it is translated and the stability of mRNA or protein itself. The regulatory elements control the stability of the corresponding mRNA in the cytoplasm and the rate of its translation into proteins. Moreover they play an important role in regulating the cellular locations of some mRNAs. One of the most important regulatory elements is untranslated regions, 5' and 3' ends of mRNA which affects the translation efficiency of mRNA and thereby determines the synthesis of protein encoding it. A mutation in UTR either at 5' or 3' or both ends would be reflected in the form of change in the concentration of protein. And this deviation from the normal concentration may be implicated in a disorder of human body and it is for this reason, the present work included their analysis. The pubmed was completely screened to identify an X-chromosome related UTR mutations in the human genome. The genes showing these mutations were annotated for gene name, mutation type and position of mutations. It was discovered that mutations at any end in 5' or 3' result in disease. In UTR mutations 46 were found to be located on X-chromosome. Out of these 5' end was found to be associated with 38 mutations and 13 mutations were found at 3'end. Some mutations could be mapped to both 5' and 3' ends. The type of disease depends upon the gene in which mutation is occurring. Thus a detailed analysis of UTR mutations revealed that disease could be caused not only because of mutations in the protein coding regions but also at the regions which are not translated and those regions could be further analyzed for the probability that these are disease causing regions.

Introduction

The levels of proteins do not always remain constant and may change in response to different environmental conditions. Consequently, a protein may be found in high or low concentrations depending upon the cellular environment. The presence of a protein, in turn, depends upon the concentration of mRNA from which it is synthesized, the frequency at which it is translated and the stability of mRNA or protein itself. The regulatory elements which control the expression of a gene and hence production of proteins are present in the non-coding part of the genome and not the protein coding part/sequence. These elements control the stability of the corresponding mRNA in the cytoplasm and the rate of its translation into proteins. Moreover, they also play an important role in regulating the cellular locations of some mRNAs so that newly synthesized protein is concentrated where it is needed in the cell. Thus, the study of these regulatory elements becomes significant in order to comprehend the mechanisms underlying the control of translation of mRNA. One of the most important regulatory elements are untranslated regions at the 5' and 3' ends of mRNA which effect the translation efficiency of mRNA and thereby determine the synthesis of the protein encoded by it. A mutation in UTR either at 5' or 3' or both ends would be reflected in the form of change in the concentration of a protein and this deviation from the normal concentration may be implicated in a disorder of human body and it is for this reason, the present work included their analysis. (Encyclopedia of Life Sciences-3 PRIME UTR Mutations and Human Disorders, 5 PRIME UTRs and Regulation).

These are sections of RNA before and after its start and stop sequences respectively that are not translated. They are formed from template DNA strand from which the RNA is transcribed. These regions are known as the 5'UTR and 3'UTR due to the fact that DNA and RNA run from 5' to 3' and this region is at the end of RNA sequence. These sequences do not code for proteins. However, their importance lies in the fact that they may, by their varying affinity for certain enzymes promote or inhibit the relative stability of RNA molecule. 3' UTR is a region of DNA which is transcribed into mRNA and becomes the 3'end of the message, but which does not contain protein coding sequences, the sequences between stop codon and polyAtail is considered to be 3'UTR. The 3'UTR may affect the translation efficiency of mRNA. (Voltmer-Irsch S *et al*., 2007; Rajkowitsch L *et al*., 2004; Müller C *et al*., 2003; Mendrysa SM *et al*., 2001; Kocarek TA *et al*., 2000; Osman F *et al*., 1999; A region of gene (DNA) which is transcribed into mRNA becoming the 5'UTR is the portion of DNA starting from cap and extending to the base just before the AUG translation initiation codon. Though it is not translated, it is believed to have sequences which alter translation efficiency of mRNA or which affect stability of mRNA

The stability of different mRNAs in the cytoplasm varies widely with majority having very short half-lives. These mRNAs that encode proteins are expressed for very short durations of time and they generally have copies of the sequence AUUUA in the 3'UTR (Anant S, Davidson NO 2000). It is believed that these AU rich elements stimulate degradation of an mRNA and therefore UTRs determine the survival time mRNA in cytoplasm and hence its translation(Zhao D et al., 2005; Lai WS, Blackshear PJ 2001; Lai WS et al., 2000; Brown CR et al., 1995). It has been found that eukaryotic cells can regulate the rate of degradation or translation of some

mRNAs by interaction with specific RNA-binding proteins of the sequences present at 3'UTR (Vasudevan S, Steitz JA 2007; Meijer HA et al., 2007; Misquitta CM et al., 2006; Nagaoka K et al., 2006; The sequence at 3'UTR also helps in localization of RNA to specific sites in the cytoplasm by binding to certain cytoskeleton elements

Several aspects of structure of UTRs have been revealed by the completion of human genome project. Base composition of 5' and 3' UTR is 60%, it is 40% in the case of 3'UTR. The average length of 5'UTR ranges between 100 to 200 nucleotides and is found to be roughly constant in different taxonomic classes. The length of 3'UTR is much more variable and ranges between 300 in plants and fungi to 800 nucleotides in human and vertebrates (Belancio VP *et al*., 2007; Shabalina SA *et al.,* 2003; Liu GD *et al.,* 2003; Croft KE *et al.,* 2003; Pesole G *et al.,* 2001; Pesole G *et al.,* 1997; Pesole G *et al.,* 1994).

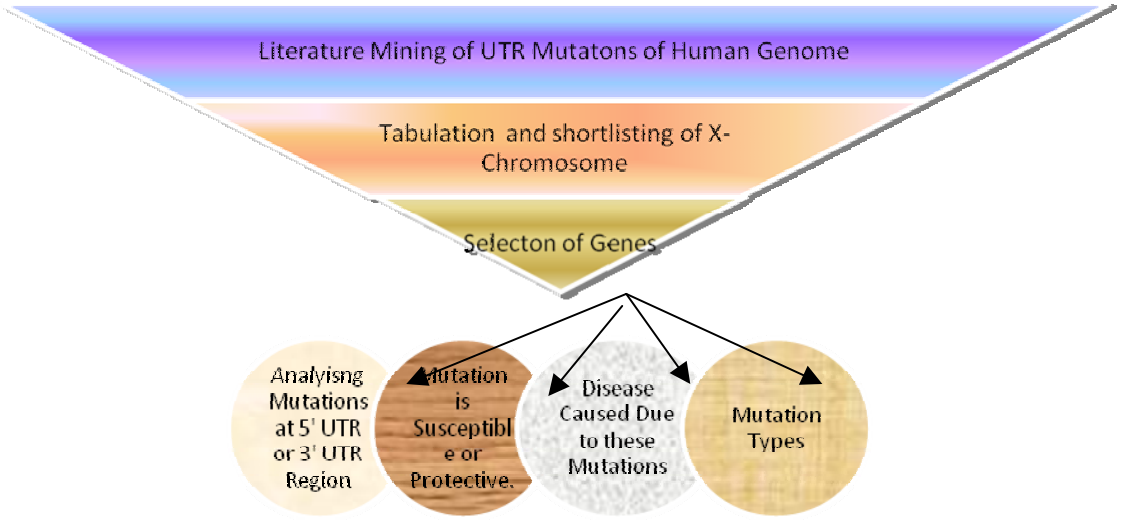
The genomic region corresponding to the UTRs of an mRNA may contain introns more frequently in the 5' than 3'UTR. Alternative UTRs can be formed from the use of different transcription start sites, polyadenylation sites or splice donor or accepted sites (Lee JY et al., 2007; Gissi C et al., 2006; Costessi L et al., 2006; Tran P et al., 2002; Blasi P et al., 2002). It has been found that they vary in abundance within a tissue, a developmental stage or in case of a disease and thereby may be considered for further analysis as they effect the patter of gene expression(Carninci P et al., 2006; Sobocki T et al., 2006; Gale CP, Grant PJ 2004; Tran P et al., 2002; Sakate R et al., 2007; Trinklein ND et al., 2007; Fusco F et al., 2006).

A mutation is defined as a heritable change in the genotype of an organism. It was considered that a mutation in the gene at the protein coding region would result in a change in the protein encoded by that gene. But recent evidences has highlighted the importance of mutations in the non coding region such as untranslated regions whose sequences are not translated into proteins which result in several disorders in the human body. UTR play a cardinal role in the regulation of gene expression at the level of mRNA they control the half life of RNAs and thereby determine their stability in the cytoplasm(Clement JQ *et al*., 2001; Jones TR, Cole MD 1987; Khalili K, Weinmann R 1984). The time duration for which RNA stays in cytoplasm would be reflected in the form of type of protein formed as a result of RNA translation. Thus, any change or mutation in the UTR sequences would later affect its function of controlling mRNA stability and hence an anomaly in the type of protein formed. Any change in protein expression in a cell is in turn instrumental in causing a disease or a disorder in the body. Therefore a detailed analysis of UTR mutations is imperative in order to procure information of various reasons underlying the occurrence of diseases. UTR mutations in regions of X-chromosome give an insight into various X-linked disorders and present this region as a disease causing region (Reamon-Buettner SM *et al*., 2007; Katano Y *et al*., 2007; Entezam A *et al*., 2007; López de Silanes I *et al*., 2007; Coutinho AM *et al*., 2007).

Materials and Method

UTR mutations play a very important role in the occurrence of diseases. Literature mining of UTR mutations in the human genome was carried out with NCBI pubmed database and tabulated. Mutations occurring in the X-chromosome were short listed

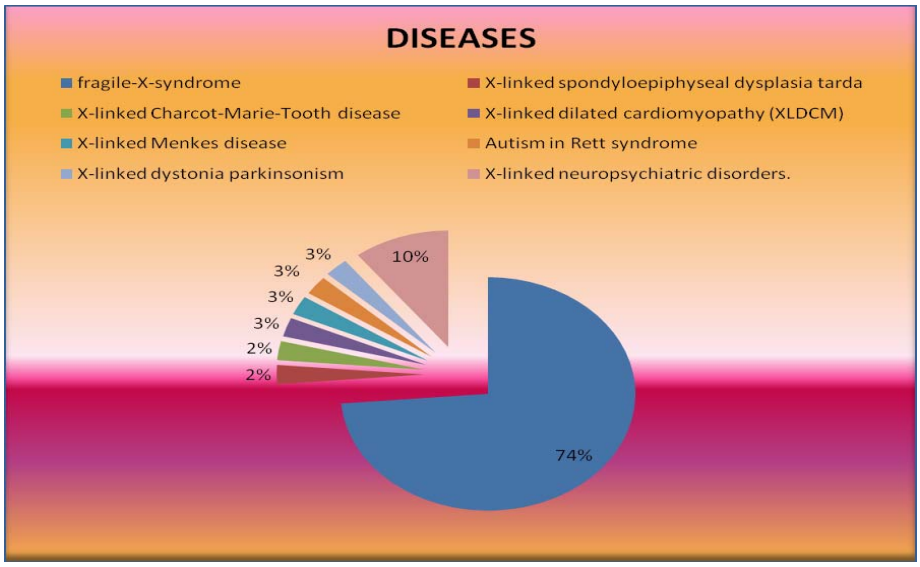
and the associated genes were selected. These were then analyzed as indicated in the flow chart below.



Methodology for analysis of UTR mutations for Human X-chromosome

Result and Discussion

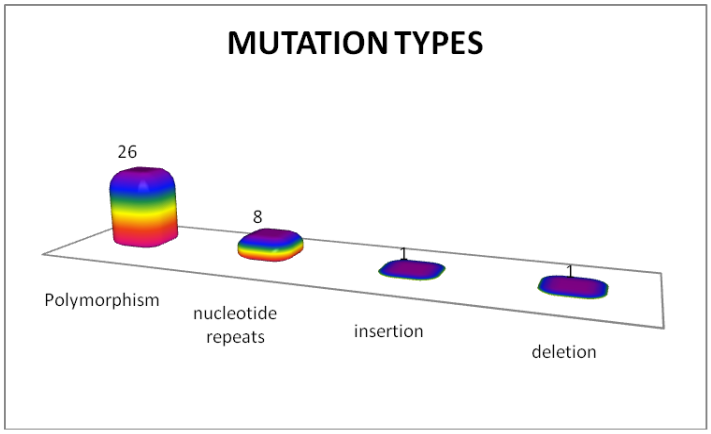
Compiling the UTR mutations in human X-chromosome on their disease association is shown in the figure



Percentage of UTR mutation in x linked disorder

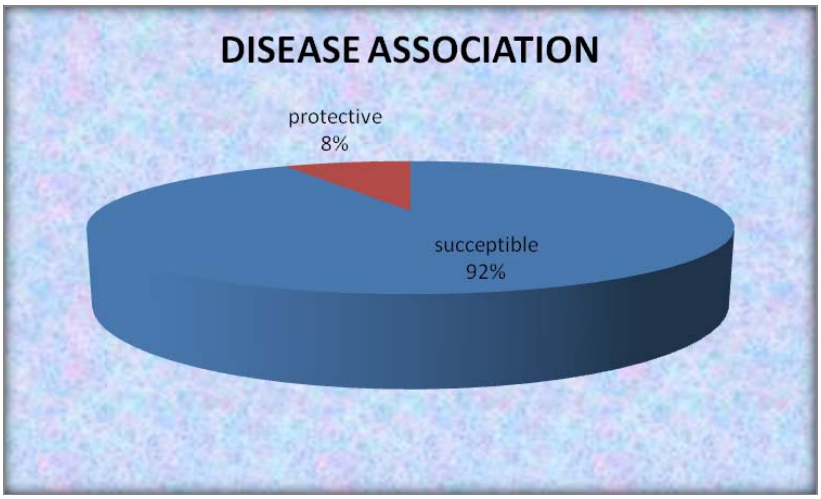
Around 8 types of diseases could be linked to the X-chromosome which occurs in different regions of RNA and various mutation types . Fragile-X-Syndrome is the most predominant of all the diseases, occurring at the highest percentage due to mutations in the X-chromosome. Other diseases include X-Linked neuropsychiatric disorders, X-Linked Parkinsonism, Autism in Rett syndrome, Menkes disease, Charcot-Marie-Tooth disease and Spondyloepiphyseal dysplasia tarda.

The types of mutations were also studied which are shown in the Fig below.



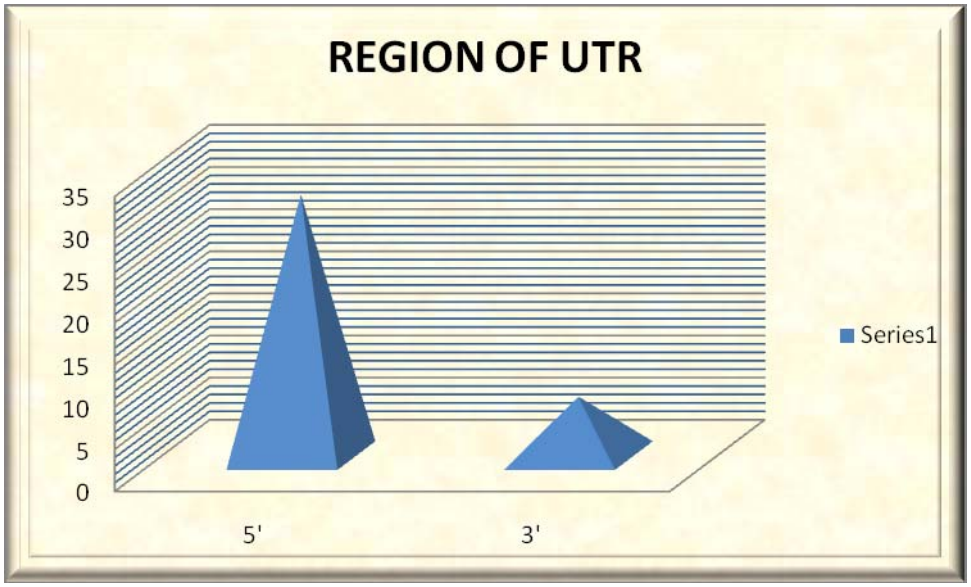
Mutation types

There are 4 different types of mutations occur in the X-Chromosome. Majority of these mutations are nucleotide polymorphisms and the remaining are small scale mutations. They occur in the different regions of UTR and in different genes. Diseases vary according to the gene in which the mutation has occurred.



Protective and susceptible range in disease

It gives an account of the UTR mutations on X-chromosome delineating whether or not these mutations are implicated in causing diseases. It was found that only 8% of these mutations were involved in protective or stabilizing functions in the genome whereas a substantial percentage of 92% were instrumental in causing various X-linked disorders.



Total number of UTR mutation in 3’ & 5’ region of genes in X -chromosome

It represents a comparative analysis of the number of diseases associated with the 5’ UTR and 3’ UTR of the X-chromosome. It was discovered that most of the diseases occur at the 5’ UTR. The reason might be that the mutations are more at the 5’UTR as compared to the 3’ UTR region and the number of genes located in the 5’ UTR region is more than the 3’ UTR which easily undergo mutations , thereby associating more diseases with the former

A comparative study between the nature of the mutations and the region of the UTR was done and it was found out that most of the mutations were susceptible in causing diseases. Almost all the mutations at the 5’ UTR region belong to this category. Most of the mutations at the 3’ UTR region were also susceptible except two which were found to be protective (Fig 68).

Conclusion

The major breakthrough that was found on analyzing the disease caused due to a mutation at the 3’ or the 5’ UTR. It was found that the X- chromosome has maximum number of UTR mutation followed by Chromsome1. Fragile X syndrome mainly caused due to UTR mutations and the fragile site in Xq22.1 may be due to UTR mutations. Fragile X syndrome mainly caused by a mutation called ‘polymorphism’

and the frequency of other mutations varied with the type of disorder. It was also discovered that more than one type of mutations in a gene could cause one disorder. It was discovered that most of the mutations occurred at the 5' region and the most frequent form of these mutations was polymorphism. Other types of mutations included nucleotide tandem repeats, insertions, deletions; in decreasing order of occurrence. UTR analysis threw light on the fact and further confirmed that diseases could result from mutations not only in the coding region of a gene but also the untranslated regions.

References

[1] Duncan K et.al (2006). Sex-lethal imparts a sex-specific function to UNR by recruiting it to the msl-2 mRNA 3' UTR: translational repression for dosage compensation ,Genes Dev. Feb 1;20(3):368-79.

[2] Saxena A et.al (2006). Lost in translation: translational interference from a recurrent mutation in exon 1 of MECP2. J Med Gcenet.Jun;43(6):470-7. Epub Sep 9.

[3] Mahajan A et.al (2004). Molecular characterization of hemophilia B in North Indian families: identification of novel and recurrent molecular events in the factor IX gene. Haematologica. Dec;89(12):1498-503.

[4] Voltmer-Irsch S et.al (2007). Regulatory mechanisms governing the oocyte-specific synthesis of the karyoskeletal protein NO145.,J Cell Sci. Apr 15;120(Pt 8):1412-22. Epub Mar 20.

[5] Rajkowitsch L et.al (2004). Reinitiation and recycling are distinct processes occurring downstream of translation termination in yeast. J Mol Biol. Jan 2;335(1):71-85

[6] Müller C et.al (2003). 3'UTRs of glutathione peroxidases differentially affect selenium-dependent mRNA stability and selenocysteine incorporation efficiency. Biol Chem. Jan;384(1):11-8.

[7] Mendrysa SM et.al (2001). Characterization of the 5' and 3' untranslated regions in murine mdm2 mRNAs. Gene. 2001 Feb 7;264(1):139-46.

[8] Kocarek TA et.al (2000). Post-transcriptional regulation of rat CYP2E1 expression: role of CYP2E1 mRNA untranslated regions in control of translational efficiency and message stability. : Arch Biochem Biophys. Apr 1;376(1):180-90.

[9] Osman F et.al (1999). A cis-acting element in the 3'-untranslated region of human TNF-alpha mRNA renders splicing dependent on the activation of protein kinase PKR. Genes Dev. Dec 15;13(24):3280-93.

[10] Zhao D et.al (2005). Novel signaling stimulated by arsenite increases cholesterol metabolism through increases in unphosphorylated steroidogenic acute regulatory (StAR) protein. Mol Cell Endocrinol. Feb 28;231(1-2):95-107

[11] Lai WS and Blackshear PJ. 2001 Jun 22,Interactions of CCCH zinc finger proteins with mRNA: tristetraprolin-mediated AU-rich element-dependent

mRNA degradation can occur in the absence of a poly(A) tail. J Biol Chem. Jun 22;276(25):23144-54. Epub Mar 28.

[12] Lai WS et.al (2000). Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to Au-rich elements and destabilization of mRNA. J Biol Chem. Jun 9;275(23):17827-37.

[13] Brown CR et.al (1995). Herpes simplex virus trans-regulatory protein ICP27 stabilizes and binds to 3' ends of labile mRNA J Virol. Nov;69(11):7187-95

[14] Belancio VP et.al (2007). Requirements for polyadenylation at the 3' end of LINE-1 elements Gene. Apr 1;390(1-2):98-107. Epub 2006 Aug 18.

[15] Shabalina SA et.al (2003). Patterns in interspecies similarity correlate with nucleotide composition in mammalian 3'UTRs.,Nucleic Acids Res. Sep 15;31(18):5433-

[16] Liu GD et.al (2003). Full-length genome of wild-type hepatitis A virus (DL3) isolated in China. World J Gastroenterol. Mar;9(3):499-504.

[17] Croft KE et.al (2003). Macronuclear molecules encoding actins in spirotrichs J Mol Evol. 2003 Mar;56(3):341-50.

[18] Pesole G et.al (2001). Structural and functional features of eukaryotic mRNA untranslated regions. Gene. Oct 3;276(1-2):73-81.

[19] Pesole G et.al (1997). Structural and compositional features of untranslated regions of eukaryotic mRNAs. Gene. Dec 31;205(1-2):95-102.

[20] Pesole G et.al (1994). Sequence analysis and compositional properties of untranslated regions of human mRNAs. Gene. 1994 Mar 25;140(2):219-25.

[21] Lee JY et.al (2007). PolyA_DB 2: mRNA polyadenylation sites in vertebrate genes,Nucleic Acids Res. Jan;35(Database issue):D165-8.

[22] Gissi C et.al (2006). Huntingtin gene evolution in Chordata and its peculiar features in the ascidian Ciona genus,BMC Genomics. Nov 8;7:288

[23] Costessi L et.al (2006). Brain-specific promoter and polyadenylation sites of the beta-adducin pre-mRNA generate an unusually long 3'-UTR. Nucleic Acids Res. 2006 Jan 9;34(1):243-53. Print 2006.

[24] Tran P et.al(2002). Multiple transcription start sites and alternative splicing in the methylenetetrahydrofolate reductase gene result in two enzyme isoforms. Mamm Genome. Sep;13(9):483-92

[25] Blasi P et.al (2002). Structure of human succinic semialdehyde dehydrogenase gene: identification of promoter region and alternatively processed isoforms. Mol Genet Metab. Aug;76(4):348-62.

[26] Gale CP and Grant PJ (2004). The characterisation and functional analysis of the human glyoxalase-1 gene using methods of bioinformatics. Gene. Oct 13;340(2):251-60

[27] Fusco F et.al (2006). Multiple regulatory regions and tissue-specific transcription initiation mediate the expression of NEMO/IKKgamma gene. Gene. Nov 15;383:99-107. Epub Aug 2.

[28] Sakate R et.al (2007). Mapping of chimpanzee full-length cDNAs onto the human genome unveils large potential divergence of the transcriptome. Gene. Sep 1;399(1):1-10. Epub Apr 20.

[29] Trinklein ND et.al (2007). Integrated analysis of experimental data sets

reveals many novel promoters in 1% of the human genome. Genome Res. Jun;17(6):720-31

[30] Clement JQ et.al (2001). Localization and stability of introns spliced from the Pem homeobox gene. J Biol Chem. May 18;276(20):16919-30. Epub Feb 23

[31] Jones TR and Cole MD (1987). Rapid cytoplasmic turnover of c-myc mRNA: requirement of the 3' untranslated sequences. Mol Cell Biol. Dec;7(12):4513-21.

[32] Khalili K and Weinmann R (1984). Actin mRNAs in HeLa cells. Stabilization after adenovirus infection. J Mol Biol. Dec 25;180(4):1007-21.

[33] Reamon-Buettner SM et.al (2007). Mutations in the 3'-untranslated region of GATA4 as molecular hotspots for congenital heart disease (CHD). BMC Med Genet. Jun 25;8:38.

[34] Katano Y et.al (2007). Association with 5'-untranslated region and response to interferon in chronic hepatitis C. Hepatogastroenterology.Apr-May;54(75):854-7.

[35] Entezam A et.al (2007). Regional FMRP deficits and large repeat expansions into the full mutation range in a new Fragile X premutation mouse model. Gene. Jun 15;395(1-2):125-34. Epub Mar 16.

[36] López de Silanes I et.al (2007). Aberrant regulation of messenger RNA 3'-untranslated region in human cancer. Cell Oncol.;29(1):1-17.

[37] Coutinho AM et.al (2007). MECP2 coding sequence and 3'UTR variation in 172 unrelated autistic patients. Am J Med Genet B Neuropsychiatr Genet. Jun 5;144(4)