



Non-coding DNA – a brief review

Anandakumar Shanmugam¹, Arumugam Nagarajan², Shanmughavel Pramanayagam^{3*}

¹Postdoctoral fellow, Dept of biotechnology, Bhupat and Jyoti mehta school of Bio sciences, Indian institute of Technology, Chennai –600 036, India.

²Under Graduate Lab, Bhupat and Jyoti mehta School of bio sciences, Dept of Biotechnology, Indian Institute of Technology, Chennai –600 036, India.

³Computational Biology Laboratory, Department of Bioinformatics, Bharathiar University -Coimbatore-641 046, Tamilnadu, India.

ARTICLE INFO

Article history:

Received on: 27/03/2017

Accepted on: 11/07/2017

Available online: 30/10/2017

Key words:

Non-coding regions, comparative genomics, functional genomics, functional annotation, *in vivo*.

ABSTRACT

In addition to the coding segments, genomes of all organisms are made of several highly conserved non-protein coding regions. Biochemical analysis by isolating non-coding regions from cells, tissues or whole organism studies are powerful tools for their identification. In lieu of this, identifying and annotating these regions using comparative and functional genomics approaches should be a high priority. Understanding and identifying their location and what these segments are composed of would pave way for functional annotation. Large scale functional genomics approaches help to identify novel genes and allow to hypothesize its *in vivo* function systematically in turn aid in annotating the conserved regions obtained from comparative genomics at the sequence level. In this review, we survey all non-coding regions, their importance and their functional roles newly discovered.

1. INTRODUCTION

In the 70s a number of laboratories converged in revealing coding and non-coding regions in nuclear DNAs [1-4]. Today one goal of genetic projects is to systematically localize all these regions, characterize their function [5] and get a comprehensive understanding of how they act together. But the estimated number of genes continues to fluctuate, the reason being the genes are either not well – clear or easily identifiable. The current available data is depends on three techniques such as cDNA cloning and EST (Expressed Sequence Tag) sequencing of polyadenylated messenger RNAs [6, 7]. Conserved coding exon identification by comparative analysis of genome [8] Prediction of genes by computationally [9]. These techniques are most excellent for highly expressed, phylogenetically conserved protein coding regions and large data sets, except for one class of genes – the non-coding regions of the genome. The non-coding regions are segments of DNA that do not comprise a gene and do not code for a protein [10]. These regions sometimes referred to as "junk DNA" are interspersed throughout the genome. The non-coding regions get transcribed but are neither translated nor directly involved with the process of translation and hence no functional protein is produced. However, some of these regions are thought to have known biological function. Certain regions

also produce transcripts that are involved directly in RNA processing and translation, rather than being expressed into messenger RNAs that encode proteins. Its includes transfer RNAs (tRNA), ribosomal RNAs (rRNA), small nuclear RNAs (snRNA), small nucleolar RNAs (snoRNA), etc., The number and diversity of these non-coding regions remain essentially unknown, even after the completion of many genome sequences. Some questions like the number of non-coding genes in a genome, their importance, their function inside a cell and whether these large set of genes have gone undetected because of their inability to be translated into proteins have always remained a mystery [11] To address such questions, development of new systematic gene discovery approaches specifically aimed at non-coding regions is of utmost necessity. The idea that these classes of genes have remained undetected is stimulating, if not skeptical. Non-coding regions are almost absent in bacterial genomes but makes up as much as 90% or more of the genome in higher organisms [12-14] Modern comparative genomics studies suggest that genetic variations between any two allied species are more because of the modifications in the non-coding regions slightly than in the protein-coding genes. Transcription – the process which results in the formation of RNA molecules is the primary way in which genetic information affects a cell's function. Studies that employ metabolic labeling of newly synthesized RNA indicated that a vast proportion of nuclear DNA was actually transcribed and bulk of this heterogeneous RNA (hnRNA) never access the cytoplasm and hence did not get decoded into proteins [15-18]. Novel genomics and RNomic research have cross verified these findings [19, 20]

* Corresponding Author
Email: shanmughavel@buc.edu.in

Understanding the phenomena of RNAi and the discovery of post – transcriptional gene silencing has brought the functional significance of non-coding DNA sequences, especially the non-coding RNA's into light. Recently, researchers have elucidated the roles of non-coding regions in causing diseases in humans [21]. Alteration in these region are coupled with disease propensity, but accurately how these functional changes is ambiguous [22]. There are several segments of non-coding regions including : Cis- and Trans-regulatory elements, Introns, Non-coding functional RNA, Pseudo genes, Repeat sequences, Telomeres, transposons and viral elements. These regions assumed to be “junk” seem to be responsible for a varied number of disorders in humans. Understanding what these segments are composed of and annotating them would help in understanding their specific roles in the genome.

2. Non-coding RNA

Non-coding RNA's are functional RNA molecules that do not get translated into a protein. These regions are thought to be involved in many cellular processes as in gene regulation and disease states. These RNA's that does not act as messenger RNAs (mRNAs), ribosomal RNAs (rRNAs) or transfer RNAs (tRNAs) has been given by various names where the term small RNAs (sRNAs) has been dominant in bacteria, whereas the term non-coding RNAs (ncRNAs) has been predominant in eukaryotes. These ncRNA's vary in size greatly. For example, regions of 21–25 nucleotides length form the large family of microRNAs (miRNAs). Those miRNAs are involved in regulation of development in *Caenorhabditis elegans*, *Drosophila*, and mammals [23-27] those that are 100-200 nucleotides in length are sRNAs commonly found as translational regulators in bacterial cells [28, 29] and those that are >10,000 nucleotides in length are involved in gene silencing in higher eukaryotes [30-32]. Thus, ncRNAs vary greatly in function and some RNAs is below:

3. RNA Processing and Modification

Processing of the 5' end of precursor tRNA and some rRNA's is done by the catalytic ribonuclease P (RNase P) RNA, which is found in all organisms (Frank and Pace 1998). For splicing of pre-mRNAs in eukaryotes, small nuclear RNAs (snRNAs) are important [33]. Small nucleolar RNAs (snoRNAs) direct the methylation of 2'-O-ribose sugar (C/D-box type) and pseudouridylation (H/ACA-box type) of rRNA, tRNA, and ncRNAs by base-pairing with sequences close to the sites has to be altered [25]. The lead RNAs (gRNAs) are available in kinetoplasts to regulate insertions or deletions of uridine residues into mRNA as described by base-pairing methods. [34, 35].

4. Regulation of mRNA Stability and Translation

The process of translation is repressed by *C. elegans* lin-4 and let-7, the first discovered miRNAs. This suppression is because of the formation of base pairs with the 3' end of target mRNAs. Numerous recently identified miRNAs are likely to act in a similar fashion. It is convincing that a few of these miRNAs

target mRNAs for degradation similarly like RNA interference process where the small interfering RNAs (siRNAs) which are processed and amplified from exogenously added, double-stranded RNA leads to gene suppression [36].

5. Other Functions of ncRNAs

For a variety of ncRNAs, functions have already been conceptualized, however scores of the cellular roles remains elusive. For example, the small cytoplasmic Y RNAs which are associated with the Ro autoantigen leave the researchers enigmatic even after many years of study [37]. Although new approaches need to be developed, previous approaches in decrypting the functions of ncRNAs paves the way to answer the question for elucidating the functions of these regions. For genetically malleable organisms, knock – out or over – expression studies can be done for whole – genome expression patterns or for screening the differences in phenotypes (such as viability). Identification of whole component of these regions for better understanding of their functions cause a big challenge for the researchers working on ncRNAs. The functions of numerous ncRNAs were well-known by the biochemical identification of related proteins, and the expansion of more systematic methods for characterizing ncRNA-linked proteins should be fruitful. As computational biology tools expand and improve, comparative genomics approaches should serve as the central avenue for determining the functions of ncRNAs.

6. CIS - Regulatory Elements

These are non-coding DNA sequences that are present in or near a gene and required for the proper expression of genes, usually containing binding sites for transcription factors. A classic example of a cis – acting element is the SECIS (SElenoCysteine Insertion Sequence) element, which is an RNA element having a stem – loop structure of 60 nucleotides in length, channelizes the cell to translate the “UGA” codon as “Selenocysteine” rather than translating into a stop codon.

The cis-acting elements may be classified into four types:

- Promoter: This is the DNA element where transcription initiation begins.
- Enhancer: This element can enhance the transcription process upon binding with the transcription factors (TFs). The TFs that bind to enhancers are known as transcription activators.
- Silencer: This element can repress the transcription process upon binding with the transcription factors (TFs). The TFs binding silencers are referred to as repressors.
- Response Element: These are the identification sites for certain transcription factors. They are mostly positioned within 1 kb from the upstream of the transcriptional start site.

For the transcriptional regulation of gene expression, binding of transcription factors to cis-regulatory elements or

transcription factor binding sites (TFBSs) is mandatory. Hence, discovering the cis- regulatory elements has brought into vital research challenge for few years [38, 39]. The conventional wet – lab experiments for the investigation of these elements is expensive and impractical on large number of genes. As more genome sequences become accessible, computational approaches provide an alternate low – cost efficient method to deal with large data sets [40]. Many approaches have been projected to spot the putative cis – regulatory elements through varied algorithms. For predicting precision of the algorithms, cis-regulatory elements derived computationally are compared with known TFBSs from databases and/or literature. This along with Gene Ontology (GO) annotations helps in validating or associating the purported elements with biological functions [41].

7. Trans - Regulatory Elements

These are diffusible proteins, which modify gene expression by binding to the cis-acting sequences [42-44]. Some of the general properties of different trans-acting factors are:

- Considered to be the subunits of RNA polymerase
- Helps in stabilization of initiation complex by binding to RNA polymerase
- Binds at specific sequences to few promoters and required for initiation of transcription and considered as the “positive regulators” of gene expression.

7.1 Introns

After transcription, the introns are detached by the process of “splicing” leaving only the “exons” which further get translated into proteins. Introns consist of large stretches of DNA whose biological functions are in the process of being elucidated. All higher organisms have introns, and the more complex organisms own a higher part of introns. These indicate that they play functions ranging from minor to major. Recent research elucidates a high level of conservation in some introns, indicating that they have some useful function. Many studies show the involvement of introns in cancer causation either directly or indirectly. For example, introns are may be involved in the transcriptional regulation of apoprotein B, E, and A-11 and few introns in regulating neoplasm developments [45]. Conclusions relating introns and cancer are not plausible for varied reasons including studies on determining if intronic mutations responsible for tumour progression or due to collateral or unrelated damage.

7.2 Pseudogenes

Pseudogenes are non – functional DNA sequences resembling functional genes. The first pseudogene was reported in 1977 and since then a large number of these genes have been reported and described in humans and many other species [46]. There are two accepted processes during which pseudogenes may arise:

- Duplication: These genes also known as “unprocessed pseudogenes” get modified due to mutations, insertions

and deletions which usually results in the loss of gene function at the transcription or translation level (or both).

- Retrotransposition: These genes also termed as “processed pseudogenes” are located on different chromosomes from their functional counterparts. These segments lack introns and regulator genes and are flanked by direct repeats. This occurs due to the activity of long interspersed nuclear element (LINE), which mimics a genomic virus. LINEs follow their own way of making DNA copies of themselves to get integrated into the genome.

7.3 Telomeres

Telomeres are complex, essential, protective nucleoprotein structures located at the ends of eukaryotic chromosomes required for genome stability maintained by the special reverse transcriptase, telomerase. In most eukaryotic cells, telomeres has repetitive GT-rich sequences, which mostly made up of double-stranded DNA (dsDNA) which terminate with 3’ single-stranded tails, called as G-tails. Until recently, these regions were considered as transcriptionally “silent, but now are recognized to be transcribed into non-coding RNA molecules called TERRA. The function(s) of TERRA in telomere metabolism is now the subject of extreme study given the implications these may have on fundamental biological processes such as aging and cancer [47]. In human cells, telomeres are bound and protected by a “shelterin” complex-a six- subunit protein complex consisting of TRF1, TRF2, RAP1, POT1, TPP1 and TIN2. This complex binds specifically to telomeric DNA and regulates telomere function [48]. TRF1 and TRF2 bind specifically to dsDNA telomeric repeats wherein TRF2 is important for repressing ATM activation at telomeres and have been shown to interact with both ATM and Chk2 kinases and thus might play a role in the inhibition of ATM – dependent check point establishment by interfering with the contact between ATM and Chk2 [49, 50]. The DNA repair proteins like Apollo and MRN complex associate with TRF2 hypothesizing that their activities may be regulated by TRF2 at the telomeric region [51, 52]. POT1 binds to G- tails and is important for repressing ATR activation at telomeres [53]. This complex also promotes insertion of 3’ telomeric G-tails into the dsDNA thereby generating a “t-loop” structure. This hides the telomeric 3’ ends from checkpoint proteins and DNA repair. The mammalian shelterin complex is also thought to promote insertion of 3’ telomeric G-tails into the dsDNA portion of telomeres to generate a “t-loop” structure, in that way hiding telomeric 3’ ends from DNA repair and check point proteins [54, 55].

7.4 Transposons

Transposons also identified as transposable elements (TEs) are DNA sequences that translocate from one genomic locus to another. First identified by Barbara McClintock, these elements

are found in almost all organisms in great numbers. For example, TEs make up roughly 50% of the human genome and up to 90% of the maize genome [56]. These elements were once considered as “selfish” DNA but recent studies show their vitality in host genomes [57-59] as epitomized by immunoglobulin V (D) J rearrangement [60, 61], telomere maintenance [62], horizontal gene transfer [55, 56], and intron origination [63]. Besides jumping, the behaviour of these transposons depends on where it lands in the genome. Landing inside a gene may result in a mutation, for example, insertions of L1 into the factor VIII gene caused hemophilia [64].

Studies later on suggested that L1 was present in the APC genes in colon cancer cells but not in the APC genes in healthy cells in the identical individuals. It strongly confirms that L1 transposes in somatic cells in mammals, and this element might play a causal function in disease development [65]. Most TE's appear silent, i.e., no phenotypic effect is seen nor they translocate themselves around the genome. Some are inactive and are unable to move from chromosomal location to another because they have deleterious mutations in them. In spite of being intact and able to move, a few transposons are kept inactive by epigenetic defence mechanisms such as DNA methylation, chromatin remodeling, and miRNAs [66]. Transposon silencing also occurs in the plant genus *Arabidopsis*. Studies reveal that these plants contain more than twenty different mutator transposon sequences, wherein these sequences are silenced or methylated in wild-type plants. If one of the methylating enzymes becomes defective, then these transposons are transcribed [67]. All of these transposable elements are not detrimental. Transposons have the ability to understand the evolution of genomes by expediting translocation of genomic sequences, shuffling of exons and repair of double stranded breaks.

The ability of transposons to enhance genetic diversity, together with the capacity of the genome to reduce most TE activity, results in a balance that makes transposable elements a crucial part of evolution and genome regulation in all organisms that carrying these sequences [66].

8. Structural Variants

Structural variants are modifications in an organism's chromosome due to insertions, deletions, inversions, translocations and duplications commonly referred to as copy – number variants (CNV's) which are usually >1Kb in length [68]. If present at >1% in a population a CNV may be referred to as copy number polymorphism (CNP). Most of these variants lead to genetic disorders. For example, the Charcot-Marie Tooth (CMT) disease, a autosomal dominant disease associated with a gene dosage effect due to an inherited DNA rearrangement was the first to be elucidated in 1991.

In most cases, this disease was associated with a tandem duplication of length 1.5Mb in 17p11.2-p12 propitiated by flanking segmental duplications which surrounds the PMP22 gene. The disease phenotype results from having three copies of the normal gene. The mutual product of the recombination, a single

copy of the PMP22 gene and results in the clinically distinct hereditary neuropathy with liability to pressure palsies [69].

9. Simple and Tandem Repeats

Tandem repeats are pervasive repetitive patterns of nucleotides that occur adjacent to each other [70-72] varying in length from 2 – 50 base pairs in length. By identification of these patterns, an individual's genetic profile can be created. This method is widely used in forensic analysis by using the fingerprints of the culprit. Estimates from the human genome Sequencing project indicate that such repeats make up ~3% of the sequenced human genome. For example, most of the humans have nearly 30 CCG•CGG repeats in the 5' UTR of their FMR1 gene [73]. However, population studies in Caucasians, the only population for which significant data exist, indicate that 246 to 468 females has 55–200 repeats and 3717 to 8918 males has 200 to >1000 repeats at this locus [74].

10. Segmental Duplications

Segmental duplications are repetitive segments of the genome that are 1 to >400 kb in length and 90–100% sequence identity [75]. The evidences show that paralogous segmental duplications, also called as low-copy repeats (LCRs), may be significant determinants of genomic plasticity in all eukaryotes [76-78]. Computational analysis revealed that these duplications occur frequently in the genome wide analysis of nine model eukaryotes even after concealing of the repetitive elements. Inverted repeats have the ability to form secondary structural elements like palindromes or stem-loops which can recombine and lead to chromosomal rearrangements. This contrivance predisposes many genetic disorders in humans bestowing to tumorigenesis and a vital role in primate karyotype evolution.

11. CONCLUSION

As researchers sift through the long-neglected introns and intergenic stretches of DNA, comparative genomics tools will only help in functional annotation but also in identifying the genetic source of diseases but also in identification of appropriate drugs and understanding of the pharmacological effects of these segments. In cases where organisms vary distantly, the most conserved genes are used, such as rRNA genes (16s rRNA) are found to play vital roles. These segments have been conserved for over millions of years such that they do not seem diverse that they all seems equally unrelated, but cannot give sufficient "resolution" to determine the relationship between closely related species as they may be almost or completely identical. Thus, using comparative and functional genomics, studies on non coding segments seems plausible in the near future.

12. ACKNOWLEDGMENTS

We thank Mr. Aslam Basha technical Officer, Dept of Biotechnology, IIT – Madras, Chennai-600 036 for his timely help.

Financial Support and Sponsorship: The authors kindly acknowledge the financial support from UGC SC/ST Post doctoral fellowship award No: F.PDFSS-2014-15-SC-TAM-9117).

Conflict of Interests: There are no conflicts of interest.

13. REFERENCES

- Volpe P., Eremenko T., Preferential methylation of regulatory gene sequences in *HeLa* cells. *FEBS Lett.* 1974; 44: 121-126.
- Breathnach R., Mandel J.L. and Chambon P., Ovalbumin gene is split in chicken DNA. *Nature.* 1977; 270: 314-319.
- Berget S.M., Moore C., Sharp P.A., Spliced segments at the 5' terminus of Adenovirus 2 late mRNA. *Proc. Nat. Acad. Sci. USA,* 1977; 74: 3171-3175.
- Chow L.T.R., Gelinas R.E., Broker T.R., Roberts R.J., An amazing sequence arrangement at the 5' ends of *Adenovirus 2* messenger RNA. *Cell.* 1977; 12: 1-8.
- Cacciamani T., Virgili S., Centurelli M., Bertoli E., Eremenko T. and Volpe P., (2002) Specific methylation of the CpG-rich domains in the promoter of the human tissue transglutaminase gene. *Gene,* 2002; 297: 103-112.
- Liang F., Holt I., Pertea G., Karamycheva S., Salzberg S. L., Quackenbush J., Gene index analysis of the human genome estimates approximately 120,000 genes. *Nat Genet.* 2000; 25:239-40.
- Ewing B., Green P., Analysis of expressed sequence tags indicates 35,000 human genes. *Nat.Genet.*, 2000; 25: 232-234.
- Roest Crollius H., Jaillon O., Dasilva C., Bouneau L., Fischer C., Fizames C., et al., Estimate of human gene number provided by genomewide analysis using *Tetraodon nigroviridis* DNA sequence. *Nature Genetics.* 2000; 25: 235-238.
- Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., et al., The sequence of the human genome. *Science.* 2001; 29: 1304-1351.
- Alexander R.P., Fang G., Rozowsky J., Snyder M., Gerstein M.B., Annotating non-coding regions of the genome. *Nature Reviews Genetics.*, 2010; 11: 559-71.
- Feng Jun, Li Guang, Wang Yi-Quan. Research progress of conserved non-coding elements in metazoan. *HEREDITAS.*, 2013; 35: 35-44.
- Eddy S.R., Computational genomics of noncoding RNA genes. *Cell.* 2002; 109: 137-140.
- Gottesman S., Stealth regulation: biological circuits with small RNA switches. *Genes Dev.* 2002; 16: 2829-2842.
- Mattick J.S., RNA regulation: a new genetics? *Nat Rev Genet.* 2004; 5: 316-323.
- Britten R.J., Davidson E.H., Gene regulation for higher cells: a theory. *Science.* 1969; 165: 349-357
- Britten R.J., Davidson E.H., Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q Rev Biol.* 1971; 46: 111-138.
- Goldstein L., Trescott O.H., Characterization of RNAs that do and do not migrate between cytoplasm and nucleus. *Proc Natl Acad Sci USA.* 1970; 67: 1367-1374.
- Calvet J.P., Pederson T., Secondary structure of heterogeneous nuclear RNA: two classes of doublestranded RNA in native ribonucleoprotein. *Proc Natl Acad Sci USA.* 1977; 74: 3705-3709
- Frith M.C., Pheasant M., Mattick J.S., The amazing complexity of the human transcriptome. *Eur J Hum Genet.*, 2005; 13: 894-897.
- Greally J.M., Genomics: Encyclopaedia of humble DNA. *Nature.*, 2007; 447: 782-783.
- Anandakumar S., Suda parimala R., Shanmughavel P., Functional annotation of introns in mitochondrial genome - a brief review. *Mitochondrial DNA.* 2014;20: 1-4.
- Ernst J et al., Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature.* 2011;473: 43-49.
- Lagos-Quintana M., Rauhut R., Lendeckel W., Tuschl T., Identification of Novel Genes Coding for Small Expressed RNAs *Science.*, 2001; 294: 853-858.
- Lee R.C., Ambros V., An Extensive Class of Small RNAs in *Caenorhabditis elegans.* *Science.*, 2001; 294: 862-864.
- Grosshans H., Slack F.J., Micro-RNAs: small is plentiful *Cell Biol.*, 2002; 156: 17-21.
- Wassarman K.M., Zhang A., Storz G., Small RNAs in *Escherichia coli.* *Trends Microbiol.*, 1999; 7: 37-45.
- Altuvia S., Wagner E. G. H., Switching on and off with RNA. *Proc. Natl. Acad. Sci. U.S.A.*, 2000; 97: 9824-9826.
- Erdmann V.A., Szymanski M., Hochberg A., de Groot N., Barciszewski J., Non-coding , mRNA-like RNAs database Y2K. *Nucleic Acids Res.*, 2000; 28: 197-200.
- Avner P., Heard E., Making sense (and antisense) of the X inactivation center. *Nature Rev. Genet.*, 2001; 98: 10025-10027.
- Storz G., An Expanding Universe of Noncoding RNAs *Science.*, 2002; 296:1260-3.
- Will C.L., Lu-hrmann R., Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.*, 2001; 13: 290-301.
- Kable M.L., Heidmann S., Stuart K.D., Sequence bias in edited kinetoplastid RNAs. *Trends Biochem. Sci.*, 2000; 6: 1492-1497.
- Simpson L., Thiemann O. H., Savill N.J., Alfonzo J. D., Maslov D.A., Evolution of RNA editing in trypanosome mitochondria. *Proc. Natl. Acad. Sci. USA.*, 2000; 97: 6986-6993.
- Nishikura K.A., Short primer on RNAi: RNA-directed RNA polymerase acts as a key catalyst. *Cell.*, 2001; 107: 415-8.
- Chen X., Quinn A.M, Wolin S.L., Ro ribonucleoproteins contribute to the resistance of *Deinococcus radiodurans* to ultraviolet irradiation. *Genes Dev.*, 2000; 14: 777-782.
- FitzGerald P.C., Shlyakhtenko A., Mir A.A., Vinson C., Clustering of DNA sequences in human promoters. *Genome Res.*, 2004; 14: 1-13.
- Xie X., Lu J., Kullbokas E.J., Golub T., Mootha V., Lindblad-Toh K., Lander E.S.,et al., Systematics discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature.*, 2005; 434: 338-345.
- Boris Lenhard, Albin Sandelin, Luis Mendoza, Pär Engström, Niclas Jareborg and Wyeth W Wasserman Identification of conserved regulatory elements by comparative genome analysis. *Journal of Biology.* 2003; 2-13.
- Wei Shi, Wanlei Zhou and Dakang Xu. Identifying cis-regulatory elements by statistical analysis and phylogenetic footprinting and analyzing their coexistence and related gene ontology. *Physiol Genomics.* 2007; 31: 374-384.
- Platt T., Transcription termination and the regulation of gene expression. *Annu Rev Biochem.* 1986; 55: 339-72.
- Hentze M.W., Kühn L.C., Molecular control of vertebrate iron metabolism: mRNA- based regulatory circuits operated by iron, nitric oxide, and oxidative stress. *Proc. Natl. Acad. Sci. U.S.A.*, 1996; 93: 8175-82.
- Hirotsune S., Yoshida N., Chen A., Garrett L., Sugiyama F., et al., An Expressed Pseudogene Regulates the messenger-RNA Stability of Its Homologous Coding Gene. *Nature.* 2003; 426: 91-6.
- Text Book of Evolutionary analysis, (4th ed.)Columbia: Pearson Prentice Hall, 2007.p. 235-48.
- Corrado C., Transcriptional regulation of telomeric non-coding RNA Implications on telomere biology, replicative senescence and cancer. *RNA Biology.* 2010; 7: 18-22.
- Denchi E.L., de Lange T., Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature.* 2007; 448: 1068-71.
- Wu L., Multani A.S., He H., Cosme-Blanco W., Deng Y., Deng J.M., *et al.*, Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. *Cell.* 2006; 126:49-53.
- Bae N.S., Baumann P.A., RAPI/TRF2 complex inhibits nonhomologous end-joining at human telomeric DNA ends. *Mol Cell.* 2007; 26: 323-34.
- Hockemeyer D., Daniels J.P., Takai H., de Lange T. Recent expansion of the telomeric complex in rodents: Two distinct POT1 proteins protect mouse telomeres. *Cell.* 2006; 12: 63-77.

49. Churikov D., Price C.M., Pot1 and cell cycle progression cooperate in telomere length regulation. *Nat Struct Mol Biol.*, 2008; 15: 79–84.
50. Griffith J.D, Comeau L., Rosenfield S., Stansel R.M., Bianchi A., Moss H., de Lange T. Mammalian telomeres end in a large duplex loop. *Cell.* 1999; 97: 503–14.
51. Poulet A., Buisson R., Faivre-Moskalenko C., Koelblen M., Amiard S., Montel F., *et al.*, TRF2 promotes, remodels and protects telomeric Holliday junctions. *EMBO J.* 2009; 28: 641–51.
52. San Miguel P. Nested retrotransposons in the intergenic regions of the maize genome. *Science.* 1996; 274: 765–768.
53. Kazazian Jr. H.H. Mobile elements: drivers of genome evolution. *Science.* 2014; 303:1626-1632.
54. Brookfield J.F., The ecology of the genomedmobile DNA elements and their hosts. *Nat. Rev. Genet.*, 2005; 6: 128-136.
55. Biemont C., Vieira C., Genetics: junk DNA as an evolutionary force. *Nature.*, 2006; 443: 521-524.
56. Zhou L., Mitra R., Atkinson P.W., Hickman A.B., Dyda F., Craig N.L., Transposition of hAT elements links transposable elements and V(D)J recombination. *Nature.*, 2004; 432: 995-1001.
57. Kapitonov V.V., Jurka J., RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.*, 2005; 3: 0998 – 1011.
58. Eickbush T.H., Telomerase and retrotransposons: which came first. *Science.*, 1997; 277: 911–912.
59. Silva J.C., Kidwell M.G., Horizontal transfer and selection in the evolution of P elements. *Mol. Biol. Evol.*, 2000; 17: 1542-1557.
60. Diao X., Freeling M., Lisch D., Horizontal transfer of a plant transposon. *PLoS Biol.*, 2006; 4: 0119-0128.
61. Roy S.W., The origin of recent introns: transposons? *Genome Biol.*, 2004; 5: 251.
62. Kazazian H., Haemophilia. A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature.*, 1988; 332: 164–166.
63. Miki Y., Disruption of the APC gene by a retrotransposal insertion of L1 sequence in colon cancer. *Cancer Research.*, 1992; 52: 643–645.
64. Pray L., Transposons: The jumping genes. *Nature Education.* 2008; 1: 1-6.
65. Miura A. Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. *Nature.*, 2001; 411: 212–214.
66. Freeman J.L., Perry G.H., Feuk L., Redon R., McCarroll S.A., Altshuler D.M., *et al.*, Copy number variation: new insights in genome diversity. *Genome Res.*, 2006; 16: 949-61.
67. Lupski J.R, de Oca-Luna R.M., Slaugenhaupt S., Pentao L., Guzzetta V., *et al.*, DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell.* 1991; 66: 219-32.
68. Riggins G.J., Lokey L.K., Chastain J.L., Leiner H.A., Sherman S.L., Wilkinson K.D., *et al.*, Human genes containing polymorphic trinucleotide repeats. *Nat. Genet.* 1992; 2: 186–191.
69. Epplen C., Melmer G., Siedlaczek I., Schwaiger F.W., Maueler W., Epplen, J.T., On the essence of “meaningless” simple repetitive DNA in eukaryote genomes. *EXS.* 1993; 67: 29–45.
70. Subirana J.A., Messegueur X., Structural families of genomic microsatellites. *Gene.* 2008; 408: 124–132
71. Strom C.M., Crossley B., Redman J.B., Buller A., Quan F., Peng M., *et al.*, Molecular testing for Fragile X Syndrome: Lessons learned from 119,232 tests performed in a clinical laboratory. *Genet. Med.* 2007; 9: 46–51.
72. Crawford D.C., Acuna J.M., Sherman S.L., FMR1 and the fragile X syndrome: Human genome epidemiology review. *Genet. Med.*, 2001; 3: 359–371.
73. Bailey JA., Yavor A.M., Massa H.F., Trask B.J., Eichler E.E., Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.*, 2001; 11: 1005–1017.
74. Wang Y., Leung FC., Long inverted repeats in eukaryotic genomes: recombinogenic motifs determine genomic plasticity. *FEBS Lett.*, 2006; 580: 1277–1284.
75. Warburton P.E., Giordano J., Cheung F., Gelfand Y., Benson G., Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res.*, 2004; 14: 1861–1869.
76. Achaz G., Netter P., Coissac E., Study of intrachromosomal duplications among the eukaryote genomes. *Mol Biol Evol.*, 2001; 18: 2280–2288.
77. Samonte R.V., Eichler E.E., Segmental duplications and the evolution of the primate genome. *Nat Rev Genet.*, 2002; 3: 65–72.

How to cite this article:

Shanmugam A, Nagarajan A, Pramanayagam S. Non-coding DNA – a brief review. *J App Biol Biotech.* 2017; 5 (05): 42-47.