Journal of Scientific Research
Banaras Hindu University, Varanasi

Vol. 52, 2008 : 53-60
ISSN : 0447-9483

NONCODING DNA IS ESSENTIAL FOR ORIGIN AND EVOLUTION OF BIOLOGICAL COMPLEXITY

S. C. Lakhotia

Department of Zoology, Banaras Hindu University, Varanasi.

Central Dogma

The remarkable progress in biological sciences during the past five decades following the unraveling of DNA structure has been largely based on the so-called "central dogma of molecular biology" which outlines a formal basis to understand the flow of genetic information from genes to phenotype. However, too strong a faith in the "central dogma" and the consequent "protein- centric" role of genes denigrated large chunks of genomic DNA, which do not code for proteins, as "junk " or "selfish" or "parasitic". Paradoxically, however, the increase in biological complexity with evolution is associated with increase in the proportion of noncoding DNA. Thus noncoding DNA, almost non- existent in bacteria, can make up as much as 90% or more of the genome in higher organisms like mammals.

Analysis of genomes of a wide variety of eukaryotes makes it clear that the genetic differences between any two related species are mostly due to changes in the "noncoding" DNA rather than in the protein-coding genes. Thus although the human genome carries ~25 fold more DNA than the fruit fly, the protein-coding genes appear to be less than 2-old greater. Thus, one can ask if the noncoding DNA in our genomes is really "junk", whose accumulation to such high proportions reflects "failure" of natural selection, or is it necessary for evolution of biological complexity. Evolutionary biologists have long argued that the increase in biological complexity is not due to a greater variety of proteins but due to more complex regulatory circuits that allow generation of more complex structures and organizations in spite of diversity of proteins. As a welcome change, recent years have attracted a very fast-growing interest in the diversity of roles played by the noncoding genomic DNA sequences and their transcripts. Some of these are summarized in the following (see Figure 1).

Sequences that Regulate Expression of Genes are Key to Biological Complexity

Multi-cellular eukaryotes require genes to be expressed in a regulated manner in time as well as space. Characterization of promoters and other regulatory elements shows that the cis-and trans-regulatory DNA sequences are often longer than the transcribed parts which they regulate. Alterations in regulation of genes can be brought about, rather rapidly, by small changes in the base sequences and thus create new (or eliminate existing) target sites for binding

of the transcription factors. Similar changes can be brought about by insertion/mobilization of transposable elements. Since the increasing biological complexity has to be associated with more complex regulatory networks, one of the very important functions of the noncoding DNA sequences is to provide for plasticity in regulation of transcriptional activity of protein-coding and other genes.

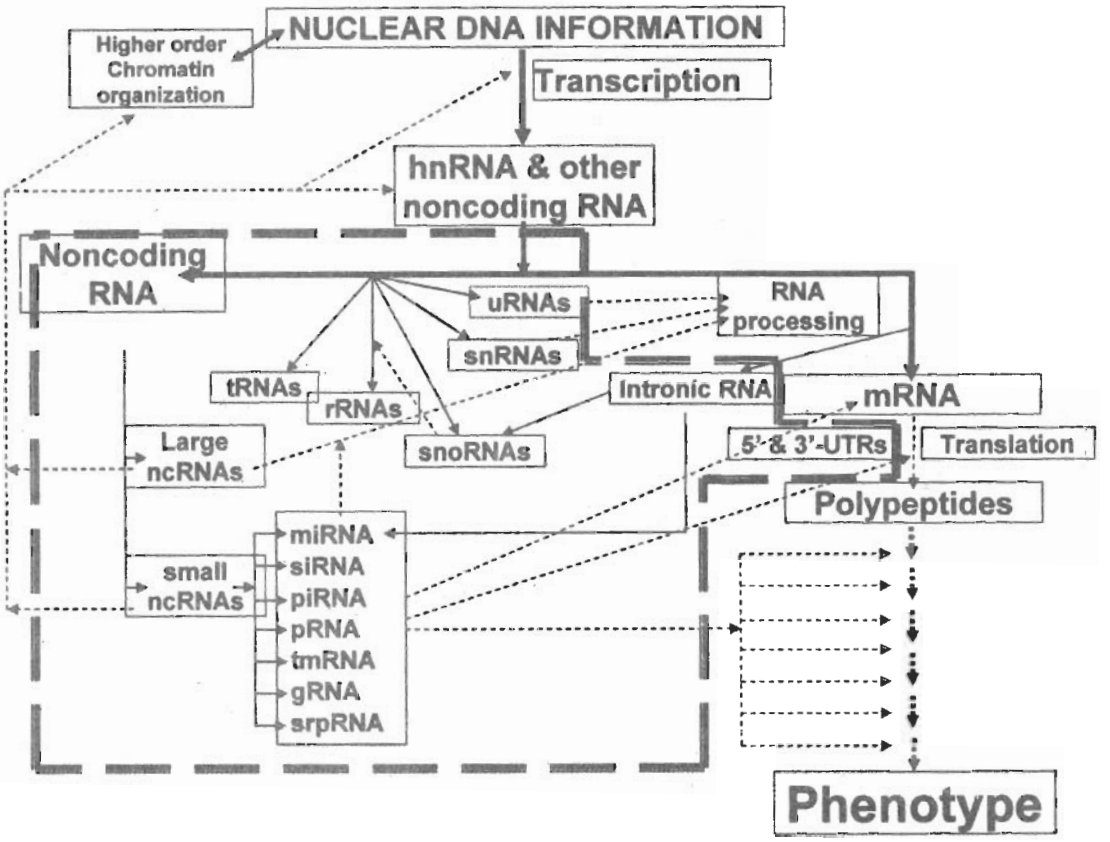

Figure 1. Complex regulation of nuclear DNA information by the diverse species of noncoding RNAs in cells. The broken double lined area includes the various classes of noncoding RNAs. Solid arrows indicate production of different transcripts or polypeptides while the broken arrows indicate regulatory interactions of the various noncoding RNA species at different levels. The double-headed arrow indicates regulatory changes in chromatin organization. The broken thick arrows from 'Polypeptides" to "Phenotype" reflect the multiple steps that result in the final phenotype.

The physical organization of the nuclear DNA, in the form of chromatin, in the 3-dimensional nuclear volume appears to reflect a cell-type specific architecture. It is clear that the cell type specific regulation of sets and sub-sets of genes is hierarchically regulated by the way the chroamtin is organized and remodeled. We do not yet know how much of the genomic DNA is required for regulation at chromatin packaging level.

The Protein Coding Genomic Regions also contain Functionally Significant Noncoding Components

Of the 3.2GB genomic DNA in humans, only about 1.2GB accounts for protein-coding gene or gene related sequences and of this only ~48MB is the actual coding part. The remaining "protein-coding DNA" is comprised of pseudogenes, gene fragments, introns, 5'- and 3'- UTRs (untranslated regions), promoter/regulatory regions etc. Some of these noncoding regions, specially the introns and the UTRs have very significant roles in generation of protein diversity (through alternative splicing of introns) and in regulating the half-lives and locations of the transcripts in cells. In many cases, the length of DNA sequences present as UTR and/or introns is much larger than the protein coding exonic sequences. A particularly remarkable example of the great potential of introns in generating protein diversity is the *Dscam* gene of *Drosophila* which has the potential to generate as many as ~38000 varieties of proteins because of alternative splicing and this potential diversity appears to be significant in neural specificity and axon guidance.

Noncoding RNA Genes have Vital Functions

The part of genome that is transcribed but produces RNA species that are neither translated nor involved in any way with the process of translation is the more intriguing and often the more ignored component of the genome. It has been known for several decades that the RNA species present in a eukaryotic nucleus are very diverse (that is why these were designated as hnRNA or heterogeneous nuclear RNA in 1960s) and that a majority of these never leave the nucleus. Till recently, these RNA species did not receive the deserved attention because of strong influence of the concept of selfish or junk DNA. An increasing number of noncoding RNA species are now shown to be essential for very basic and vital functions in cells as noted below.

Noncoding RNA species can regulate the chromatin organisation and transcriptional states of parts or of entire chromosomes : The phenomenon of dosage compensation evolved in several organisms with heterogametic sex to compensate for the difference in dosages of the sex-chromosome-linked genes between the homo-and hetero-gametic sexes. Dosage compensation in mammals is achieved by activation of one of the two X-chromosomes in somatic cells of females, while in *Drosophila* activity of the X-linked genes in males is enhanced to the combined level of two X-chromosomes in females. It is significant that in spite of such opposing operative mechanisms, noncoding RNA species are the key players that determine the activity level of the X-chromosome.

In mammalian females, the long and nucleus-limited noncoding RNA of the Xist, or its homologue, is produced only by one of the two X-chromosomes and the Xist transcripts spread in cis along the entire length of the X-chromosome. This "painting" of the one-chromosome in female cells provides the platform for binding of other proteins etc. to keep that particular

X-chromosome in an inactive state. On the other hand in the case of *Drosophila*, the single X-chromosome of males becomes hyperactive because noncoding RNAs like Rox 1 and Rox 2 "paint" the entire X-chromosome and thereby allow the assembly of protein complexes (the Msl complex) for "hyperactivity" of the single X-chromosome in male cells. Comparable functions of different species of noncoding RNAs are known to regulate activities of several "imprinted" chromosome regions/loci in mammalian genomes.

Noncoding RNA may regulate RNA processing by regulating activity of different proteins : A large variety of proteins are required for processing and transport of the different protein-coding transcripts synthesized in nucleus. In keeping with dynamism of cellular activities, the RNA processing activities also have to be equally dynamic, which in turn requires the RNA processing and transporting proteins to toggle between active and inactive states. In addition, under conditions of sudden environmental stresses, cells need to quickly reprogramme their RNA synthesis and processing activities, Studies on stressed cells have shown that different classes of RNA processing proteins are sequestered in distinct nuclear compartments or speckles. Studies in our laboratory with *Drosophila* cells have shown that a species of noncoding RNA, the hsrω-n (heat shock RNA omega-nuclear) transcripts, is essential for organizing the omega speckles, which are involved in storage/sequestration of members of hnRNP (heterogeneous nuclear RNA binding proteins) family and other related proteins in normal as well as stressed cells. During development, this noncoding RNA is expressed, in a regulated manner, in almost all cells types of *Drosophila*. In keeping with such vital functions of this noncoding RNA, studies in our laboratory have revealed that its over-as well as under-expression has severe consequences for the organism.

The satellite III noncoding transcripts in human cells appear to carry out somewhat similar function in stressed cells. Heat shock or other cellular stress induced transcripts of the satellite III sequences sequester a variety of RNA processing proteins, RNA polymerase II and heat shock transcription factor etc. in the form of stress granules in human cells.

In both these instances, the hsrω-n or the satellite III noncoding transcripts serve an important function through sequestering specific classes of proteins when they are not required to be active. Recently, it has been shown that binding of a distinct noncoding RNA with heat shock transcription factor (HSF) is essential in stressed cells for activation of the HSF and consequent activation of the stress-induced genes. It is highly likely that many other such noncoding RNA species that regulate activity of variety of proteins remain to be discovered.

Many small noncoding RNAs are Important Riboregulators : Role of RNA as riboregulators has received considerable attention in recent years. A large variety of small RNA species (21-23 nucleotide or slightly longer) are now known to affect gene expression at the levels of chromatin structure, RNA degradation (RNAi) and translation. All these processes obviously have very significant roles in establishing the biological complexity of multicellular organisms. Major categories of these riboregulators are noted below.

MicroRNAs (miRNA) and short-interfering RNAs (siRNA) are the smallest functional RNAs ranging in size from 19 to 25 nucleotides. These regulate gene activity through RNA interference at multiple levels like chromatin organization, transcription, post-transcriptional processing, stability and translation of mRNA etc. Their multiple roles allow these short RNAs to integrate the regulatory networks and thus play significant role in the origin and evolution of biological complexity. *Effe .ce RNA* (eRNA) and *mirtrons* refer to noncoding miRNAs derived from intronic sequences of protein coding or other noncoding genes.

Piwi-interacting RNAs (piRNAs) or *repeat-associated short interfering RNAs* (rasiRNAs) are germ cell-specific 26-31 nucleotide long RNA molecules that silence the mobile genetic elements and repetitive sequences in germ cells. They may have additional roles as well.

Promoter RNAs (pRNA) associate with the promoter regions of genes and control RNA-directed epigenetic remodelling and transcriptional silencing by direct binding of the antisense strand of pRNAs either to DNA or to a sense-stranded RNA corresponding to the promoter.

tmRNAs (tRNA-like and mRNA-like RNA), also known as *10Sa RNA* or *ssRa RNA*, are complex, dual-functional, small, stable RNA species in bacteria that mimic both a tRNA and mRNA. They recognize and recycle the stalled ribosomes.

Guide RNAs (gRNA) carry out RNA editing in certain organisms/genes and are part of the editosome.

Signal recognition particle RNA or the *4.55 RNA* in eukaryotes and bacteria, respectively, is a major structural component of the signal recognition particle (SRP) in the cytoplasm that binds to the mRNAs of proteins destined for secretion.

Concluding Remarks

Understanding gene function in terms of protein synthesizing activity has been a major achievement of modern Biology. Such studies clearly establish that the effective proteome in any species is much larger than that originally encoded by the genome due to the versatile processes like alternative splicing, trans- splicing, RNA-editing, post-translational modifications of proteins etc. Yet, the enormous diversity in structure and organization of the multicellular organisms cannot be explained only in terms of the proteome. Elucidation of functions of many noncoding RNA species in recent years has established that the "noncoding RNome" has a significant role in establishing and maintaining biological complexity. Unraveling of such unexpected functions emphasizes the need for more in depth studies on the noncoding DNA component in any genome. It will be erroneous to ignore this as "junk" or "selfish" DNA. Like in the primitive "RNA-world", the ribonucleic acid molecules can be, and indeed are, functionally versatile even in the modern "DNA-world".

REFERENCES

Anderson, P. (2006) : Kedersha, N. RNA granules. *J. Cel. Biol.*; Vol. 172, pp. 803-808.

Arya, R., Mallik, M., Lakhotia, S. C. (2007) : Heat shock genes-integrating cell survival and death. *J. Biosci.*, Vol. 32, pp. 595- 610.

Bai, X., Larschan, Kwon, S. Y., Badenhost, P., Kuroda, M. I. (2007) : Regional control of chromatin organization by noncoding roX RNAs and the NURF remodeling complex in *Drosophila melanogaster*, *Genetics*, Vol. 176, pp. 1491-1499.

Costa, F. F. (2005) : Non-coding RNAs : new players in eukaryotic biology, *Gene.*, Vol. 357, pp. 83-84.

Deng, X., Meller, V. H. (2006) : Non-coding RNA in fly dosage compensation, *Trends Biochem Sci.*, Vol.31, pp. 526-532.

Frith, M. C., Pheasant, M., Mattick, J. S. (2005) : The amazing complexity of the human transcriptome, *Eur. J. Human Genet.*, Vol. 13, pp. 894- 897.

Greally, J. M. (2007) : Genomics : encyclopaedia of humble DNA. *Nature*, Vol. 447, pp. 782-783.

Han, J., Kim, D., Morris, K. V. (2007) : Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells, *Proc. Natl. Acad. Sci.*, USA; Vol. 104, pp.12422-12427.

Hattori, D., Demir, E., Kim, H. W., Viragh, E., Zipursky, S. L., Dickson, B. J. (2007) : Dscam diversity is essential for neuronal wiring and self-recognition, *Nature*, Vol. 449, pp. 223-227.

Jolly, C., Lakhotia, S. C. (2006) : Human sat III and *Drosophila* hsr omega transcripts : a common paradigm for regulation of nuclear RNA processing in stressed cells. *Nucleic Acids Res;* Vol 34. pp. 5508-5514.

Jolly, C., Metz, A., Govin, J., Vigneron, M., Turner, B. M., Khochbin, S. and Vourc'h C. (2004) : Stress-induced transcription of satellite III repeats, *J. Cell Biol.*, Vol. 164, pp. 25-33.

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., Stadler, P. F., Hertel, J., Hackermuller, J., Hofacker, I. L, Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., Tammana, H., Gingeras, T. R. (2007) : RNA maps reveal new RNA classes and a possible function for pervasive transcription, *Science*, Vol. 316, pp. 1484-1488.

Kloc, M., Wilk, K., Vargas, D., Shirato, Y., Bilinski, S., Etkin, L. D. (2005) : Potential structural role of non-coding and coding RNAs in the organization of the cytoskeleton at the vegetal cortex of Xenopus oocytes, *Development*, Vol. 132, pp. 3445-3457.

Lakhotia, S. C. (1996) : RNA polymerase II dependent genes do not code for protein, *Ind. J. Biochem. Biophys*; Vol. 33, pp. 93-102.

Lakhotia, S. C. (1999) : Non-coding RNAs : versatile roles in cell regulation, *Curr. Sci.*, Vol. 77, pp.479-480.

Lakhotia, S. C. (2003) : The noncoding developmentally active and stress inductible *hsrω* gene of *Drosophila melanogaster* integrates post-transcrptional processing of other nuclear transcripts. In : Barcizewski, J., Erdmann, V. A. (Eds.) Noncoding RNAs : Molecular Biology and Molecular Medicine, New York, Kluwer Academic/Plenum Publishers, pp. 202-219.

Liwang, Q. I., Xinmin, L. I., Shougong, Z., Daochang, A. N. (2006) : Genetic regulation by non-coding RNAs, *Sci. in China : Ser C, Life Sci.*, Vol. 49, pp. 201-217.

Matera, A. G., Terns, R. M., Terns, M. P. (2007) : Non-coding RNAs : lessons from the nuclear and small nucleolar RNAs, *Nat. Rev. Mol. Cell. Biol.*, Vol. 8, pp. 209-220.

Mattick, J. S. (2004) : RNA regulation, a new genetics ?, *Nat. Rev. Genet.*, Vol. 5, pp. 316-323.

Mayr, E. (1970) : Populations, Species, and Evolution : An Abridgement of Animal Species and Evolution, Harvard University Press, Cambridge (USA),

Michalak, P. (2006) : RNA world–the dark matter of evolutionary genomics, *J. Evol Biol.*, Vol. 19, pp. 1768-1774.

Miska, E. A. (2005) : How micro RNAs control cell division, differentiation and death. *Curr. Opin. Genet. Dev.*, Vol. 15, pp. 563- 568.

Mullard, A. (2007) : Hotair rises, new roles for non-coding RNAs, *Nat. Rev. Mol. Cell. Biol.*, Vol. 8, pp. 601.

Nishioka, K., Hirose, S. (2007) : Fpigenetic regulations of chromatin dynamics : a role of noncoding RNAs in epigenetic activation of *Drosophila* Hox genes, *Int. Cong. Ser.*, Vol. 1299, pp. 239-241.

Nudler, E. (2006) : Elipping riboswitches, *Cell*, Vol. 126, pp. 19-22.

Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M., Lai, E. C. (2007) : The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila, Cell.*, Vol. 130, pp. 89-100.

Pennisi, E. (2007) : DNA study forces rethink of what it means to be a gene, *Science*, Vol. 316, pp. 1556-1557.

Prasanth, K. V., Rajendra, T. K., Lal, A. K., Lakhotia, S. C. (2000) : Omega speckles–a novel class of nuclear speckles containing hnRNPs associatedwith noncoding hsr-omega RNA in *Drosophila*, *J. Cell. Sci.*, Vol. 113, pp. 3485-3497.

Prasanth, K. V., Spector, D. L. (2007) : Eukaryotic regulatory RNAs, an answer to the 'genome complexity' conundrum, *Genes Dev.*, Vol. 21, pp. 11-42.

Shamovsky, I., Ivannikov, M., Kandel, E. S., Gershon, D. and Nudler, E. (2006) : RNA-mediated response to heat shock in mammalian cells, *Nature*, Vol. 440, pp. 556-560.

Storz, G., Altuvia, S., Wassarman, K. M. (2005) : An abundance of RNA regulators. Annu *Rev. Biochem*, Vol. 74, pp. 199-217.

The ENCODE Project Consortium (2007) : Identification and analysis of functional elements in 1% of the human genmoe by the ENCODE pilot project. *Nature*; Vol. 447, pp. 799-816.

Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., Zamore, P. D. (2006) : Distinct small RNA pathway silences selfish genetic elements in the germline, *Science*, Vol. 313, pp. 320-324.

Wray, G. A. (2007) : The evolutionary significance of cis-regulatory mutations, *Nat. Rev. Genet.*, Vol. 8, pp. 206-216.

Wutz, A., Gribnau, J. (2007) : X inactivation Xplained, Cur. Opin. Genet. Dev., Vol. 17, pp. 1-7.

www.biobases.ibch.poznan.pl/ncRNA.

Yang, P. K., Kuroda, M. I. (2007) : Noncoding RNAs and intranuclear positioning in monoallelic gene expression, *Cell*, Vol. 128, pp. 777-786.

Zaratiegui, M., Irvine, D. V., Martienssen, R. A. (2007) : Noncoding RNAs and gene silencing, *Cell*, Vol. 128, pp. 763-776.