

Commentary

Central Dogma, Selfish DNA and Noncoding RNAs: A Historical PerspectiveSUBHASH C LAKHOTIA* *Cytogenetics Laboratory, Department of Zoology, Banaras Hindu University, Varanasi 221 005, India*

(Received on 29 January 2018; Revised on 13 February 2018; Accepted on 15 February 2018)

Major discoveries like the one gene-one enzyme hypothesis, demonstration of DNA as the genetic material and finally the elucidation of the double helical structure of DNA in 1940s and early 1950s set the stage for emergence of molecular biology. Parallel cell biological studies during this period also indicated a correlation between rate of protein synthesis in a cell and the amount of cytoplasmic RNA. Following the proposal of George Gamow, a physicist, about the triplet genetic code and possible involvement of RNA in the transfer of information from DNA to proteins, Crick proposed the ‘central dogma of molecular biology’ to suggest the paths of information transfer between nucleic acids and proteins, with the limitation that the information cannot flow back from protein to nucleic acids. With emphasis on proteins as the central phenotypic determinants and the continuing enigma of heterochromatin, which largely appeared to be ‘gene desert’, enriched in repetitive DNA sequences and claimed to be inert in transcription, the many observations in 1960s of a large variety of heterogeneous nuclear RNAs remained ignored. Curiosity in the nuclear RNAs that do not see the face of cytoplasm appeared to be quelled by concepts of ‘selfish’ or ‘junk’ DNA in the early 1980s, notwithstanding the fact that active transcription of typical constitutive heterochromatin regions and repetitive and other noncoding DNA sequences was well demonstrated in 1960s and 1970s. With a few exceptions like the *hsw* and *roX* transcripts in *Drosophila* and the *Xist* RNAs in mammals, the noncoding RNAs remained largely ignored for nearly two decades. The discovery of RNA interference and sequencing of different eukaryotic genomes, including the human genome, led to revisits to possible significance of noncoding RNAs (ncRNAs) in the new millennium. The occasional identification of ncRNAs in early 2000s has in recent years transformed into a ‘tsunami’, resulting in concepts of ‘selfish’ or ‘junk’ DNA themselves becoming junk. There is now increasing realization that the subtle and large phenotypic effects of heterochromatin and the existence of diverse nucleus-limited RNAs reported through painstaking genetic and biochemical studies that were undertaken before molecular biology had grown fully, can be largely related to the enormous diversity of short and long ncRNAs now known to be produced by all genomes. Although Crick’s proposal of Central Dogma was only about the directions of information transfer, its mis-interpretation due to the great emphasis on the central roles of proteins and the reductionist linear approach of molecular biology that led to widespread belief in concepts of ‘selfish’ or ‘junk’ DNA, delayed the appreciation of multi-dimensional roles that ncRNAs actually play in maintaining homeostasis in complex biological networks.

Keywords: Selfish DNA; Junk DNA; hnRNA; Heterochromatin; *hsw*; *roX*; *Xist***Introduction**

Biological inheritance has been known to human society since times immemorial. However, a definitive understanding of the principles underlying the biological inheritance remained largely unknown till Mendel presented results of his experiments on inheritance of several phenotypes in the sweet pea in 1865. As is well known, Mendel’s findings remained ignored till their rediscovery in 1900, following which

the progress in genetics has been very rapid, with profound impact on biology in general. As expected from the basic philosophy of progress in research in science, the field of genetics and understanding of the mechanism of gene action that finally results in phenotype of the individual, have witnessed diverse, and often opposing, views as the experimental approaches evolved, and continue to evolve, to provide details of gene action at ever-increasing finer details.

*Author for Correspondence: E-mail: lakhotia@bhu.ac.in

The present review briefly examines history of development of these concepts and how misunderstanding and/or mis-interpretation of some concepts thwarted the appreciation of great functional significance of the noncoding RNAs in biological organization.

Proteins as Determinants of Phenotype

Following the establishment of Mendel's laws of inheritance and their relations with chromosomal behaviors during meiosis, further characterization of the Mendelian factor or 'gene' followed two different approaches. One led by Morgan (Morgan *et al.*, 1915) attempted to characterize the organization of gene without worrying about its physico-chemical features (Morgan, 1935). The other approach was more of biochemical and physiological and as Goldschmidt (Goldschmidt, 1916) stated 'Those who approached it from the physiological-chemical side all seem to agree that the unit-factors are to be compared in some way to enzymes (Loeb, Robertson, Moore, Bateson, Riddle, etc.) or expressed more generally "that the hereditary factor . . . is a determiner for a given mass of certain ferments"' (Loeb and Chamberlain, 1915)'. Role of proteins, especially the enzymes, as the mediators of gene action was more firmly established by Beadle and Tatum's (Beadle and Tatum, 1941) classical studies on mutations that affected the nutritional requirements in *Neurospora*. This was soon followed by the demonstration that DNA is the bearer of genetic information from one generation to the next (Avery *et al.*, 1944; Hershey and Chase, 1952). Unfolding of the double helical structure of DNA in 1953 (Watson and Crick, 1953), dramatically altered the course of biological studies. Subsequent developments that led to emergence of the 'central dogma of molecular biology', unraveling of the genetic code, understanding of the mechanism of protein synthesis using the RNA template and development of recombinant DNA technology etc led to the most exciting phase in biology in recent times. Most of these studies have been based on the view that it is the proteins that are the key intermediaries between the genetic information in DNA base sequence and the phenotype that emerges and that the RNA acts either as the intermediate between DNA and protein (the messenger or mRNA) or as components (ribosomal and tRNAs) of the platform where amino acids are assembled as guided by the mRNA template to make

proteins. Consequently, a typical and most common definition of gene learnt by all students of biology for the past 60 years has been that a gene is a segment of DNA that encodes a protein. Such beliefs were also interpreted to imply that the DNA segments that are not directly involved in synthesis of proteins are of little consequence for the organism (Doolittle and Sapienza, 1980; Ohno, 1980; Orgel and Crick, 1980). However, contrary to such beliefs that were prevalent during the last few decades of 20th century, the present century has witnessed a remarkable reversal of the negative views about the noncoding DNAs present in all organisms, which has led to an unprecedented interest in the diverse noncoding RNAs during the past two decades.

Establishing RNA as Intermediary Between DNA and Proteins

Notwithstanding Garrod's classical studies on Inborn Errors of Metabolism (Garrod, 1909), the above noted genetic and biochemical approaches to understand Mendelian inheritance and the genotype-phenotypes relationships remained separate. It was only in the late 1930s that the two approaches began to merge (Allen, 2015; Strauss, 2016). The coming together of genetic and biochemical approaches was primarily catalyzed by studies on mutations that affect pigmentation in flowers and eye colour in insects and the biochemical pathways that the different mutations affect (Beadle and Ephrussi, 1936; Caspari, 1933 ; Lawrence and Price, 1940). The real breakthrough in genotype-phenotype relationship was the 'one gene-one enzyme' concept that emerged from Beadle and Tatum's classical studies on *Neurospora* (Beadle and Tatum, 1941). It is notable that around that time the structure of proteins was not very well known. The idea that proteins are linear polymers of amino acids had, however, gained wide acceptance by 1950 (Strauss, 2016). The elegant demonstration that DNA functions as the genetic material in bacteria and viruses (Avery *et al.*, 1944; Hershey and Chase, 1952) and the constancy of DNA content in different cell types of eukaryotes (Swift, 1950) laid to rest the view held by many that proteins functioned as the genetic material.

With improving understanding of protein structure and realization that the specificity of enzyme activity depends upon its 3-dimensional structure,

Pauling suggested “genes serve as the templates on which are molded the enzymes which are responsible for the chemical characters of the organism” (Pauling, 1948). Similarly, Haurowitz (Haurowitz, 1950), taking cue from Pauling’s (Pauling, 1940) earlier suggestion of antibody production on the antigen template, believed “synthesis of antibodies is a special case of protein synthesis, but it is certainly not different from the synthesis of the other proteins such as normal serum globulin”. Identification of sickle cell anemia as a ‘molecular disease’ (Pauling *et al.*, 1949) established a causal link between an abnormal molecule and its pathological consequences and also showed that a gene could alter the physical properties of a protein rather than determining only its presence or absence. Further studies on mechanisms of antibody formation and the finding that they may be present even without an antigen led to weakening of the ‘three-dimensional template theory’ of protein synthesis. The later demonstration that a single gene mutation can change one amino acid of the hemoglobin polypeptide chain in sickle cell anemia (Ingram, 1957) finally demolished the theory that DNA structure acted as a direct template for the 3-dimensional structure of proteins (Strasser, 2006).

The idea that RNA is in some way related to synthesis/abundance of proteins in a cell arose indirectly from independent studies by embryologists and cytologists/cell biologists in 1940s, who examined distribution of different macromolecules in different cell types and developmental stages using cytochemical, histochemical and biochemical approaches. These studies indicated a general correlation between the abundance of RNA and proteins in different cell types that were presumably actively synthesizing proteins (Brachet, 1947; Caspersson and Schultz, 1940; Thorell, 1947). Biochemical studies combined with metabolic labeling of proteins with radioactive amino acids (Holloway and Ripley, 1952) revealed a parallel increase in RNA content of the reticulocytes and the amount of radioactive L-leucine incorporated into proteins. Similarly, the quantity of RNA in pancreas, liver, and kidney was found to be correlated with their vastly different rates of protein synthesis (Allfrey *et al.*, 1953). Although these studies indicated some relation between RNA and proteins in cells, its implications were not clear at that time.

The discovery of double-helix structure of DNA (Watson and Crick, 1953) prompted synthesis of the disparate findings relating genes and proteins on one hand, and cellular contents of RNA and rates of protein synthesis on the other. Gamow, a theoretical physicist and cosmologist, who first mooted the idea of ‘big-bang’, was also the first to propose the idea of the triplet genetic code (Segrè, 2000; Watson, 2003). Interestingly, Gamow’s initial proposal (Gamow, 1954), akin to Pauling’s theory mentioned above, suggested DNA to directly participate in assembly of the linear array of amino acids with three adjacent nucleotides providing a specific shaped ‘hole’ in which a given amino acid could fit in. Gamow’s speculations evolved rapidly. In one of his hand-written letters to Crick in 1954, he suggested “this is why we find 2 strands one to keep code. the other to be transformed to RNA which sneaks to cytoplasm and makes protein”. In another hand-written letter (see Watson, 2003) to Crick on 15th October 1954, Gamow elaborated his views about origin of RNA as follows: “DNA could make RNA by either chemical conversion or acting as template. The former is ruled out by isotope experiments and so we have to decide whether the two strand or one strand stage of DNA is active. If a one strand stage operates it likely does it by a DNA like base pairing mechanism. The answer in this case is trivial. Even if it proved possible to stick a RNA strand in a two stranded structure, of that I am sceptical. Moreover I don’t like the idea of the two strands having to separate since we have no reason to believe it would be viable. Still more I cant see any answer of why ribonucleotides & deoxyribonucleotides wouldn’t occasionally got mixed up and thus cause mistakes in replication process”. In a subsequent paper Gamow and Yèas (Gamow and Yèas, 1955), while considering the vital role of RNA in protein synthesis as suggested by empirical evidence, stated: “It appears rational to assume that the sequence of amino acids characterizing a given protein is uniquely determined by the sequence of nucleotides in the ribonucleic acid molecule. While RNA is a polymer of four different nucleotides, proteins are polymers of 20 different amino acids. Since it is possible to form 20 kinds of triplets from four different elements, this suggests that each of the 20 amino acids is determined by a triplet of nucleotides, taken without regard to order”. Gamow also assembled a ‘RNA-Tie Club’ to foster interest in the

role of RNA as an intermediary between DNA, the genetic storage molecule in cell nucleus, and protein synthesis in cytoplasm (a semi-scientific account of Gamow's remarkable interest in genes, genetic code and RNA can be found in Watson, 2003).

Crick proposed the first scheme of 'central dogma' in his lecture at a symposium on the Biological Replication of Macromolecules, organized by the Society for Experimental Biology at University College London on 19 September 1957. The 'central dogma' concerned the flow of genetic information from DNA to RNA to protein, with the possibility of flow of information from RNA to DNA also left open. A major feature of the central dogma was: "Once information has got into a protein it can't get out again", where information "means the sequence of the amino acid residues, or other sequences related to it" (Crick, 1958).

The proposed involvement of RNA in transfer of genetic information from DNA to proteins stimulated many studies on RNA, accompanied by development of novel methods and approaches for such studies. The terms 'transcription' and 'mRNA' were introduced by Jacob and Monod (Jacob and Monod, 1961) along with the 'operon model' of gene regulation in bacteria. They stated: "According to modern concepts, the deoxynucleotide sequence which constitutes a gene participates in two distinct chemical processes. In the first, for which the term replication should be reserved, free deoxyribonucleotides are linearly assembled by specific base-pairings, forming an identical sequence or replica of the original sequence; the second process, which we shall call transcription, allows the gene to perform its physiological function, i.e., to specify the molecular structure of a certain protein or polypeptide chain. Transcription does not appear to be a direct process, since it most probably involves the formation of an intermediate as carrier of the genetic information. Two stages may then be distinguished in transcription, the first of which is presumably closely similar to replication, involving, however, ribonucleotides instead of deoxynucleotides, and resulting in an RNA "transcript" of the original DNA sequence. In the second transcription stage, the RNA transcript in turn directs the assembly of amino acids into the polypeptide". The 'second transcription stage' was subsequently termed 'translation', a term already used

by Gamow (Gamow, 1954). The above statement of Jacob and Monod (Jacob and Monod, 1961) thus identified the physiological function of the gene to be "to specify the molecular structure of a certain protein or polypeptide chain". Much of the subsequent studies on RNA started with this belief, although with respect to the ribosomal and tRNAs, that were quickly identified in the 1960s, it was realized that they do not 'code' for polypeptides but help in the translation of information in mRNA into the polypeptide chain.

The large number of studies in the 1960s that addressed questions relating to site of synthesis of RNA in cells, kinetics of transport and turnover of different RNAs, size variability, similarities between nuclear and cytoplasmic RNAs etc., employed a variety of novel techniques like cell fractionation, metabolic labeling with radioactive ribonucleotides, transcriptional inhibition, cellular autoradiography, sedimentation coefficient measurements, solution hybridization etc (reviewed in Darnell Jr, 1968; Harris, 1963; Weinberg, 1973). The techniques like cell fractionation, sedimentation co-efficient measurements of RNA molecules, solution hybridization of nucleic acids, RNA sequencing etc that were used in these studies (Harris, 1963) involved intensive manual hard work with hazardous reagents and would appear, when compared with the contemporary approaches, crude and indirect. Yet, these were remarkably advanced for those days and helped lay the foundation of the emerging molecular biology. As may be expected, studies on RNA biology in the early 1960s evoked many questions and speculations, several of which did not stand as later studies provided better insights. The question whether the cytoplasmic RNA was derived from the rapidly labeled nuclear RNA or was synthesized directly in cytoplasm was debated as noted by Harris (1963) at the end of an extensive review: "many instances of the transfer of information from nucleus to cytoplasm will prove to be simply the transfer to the cytoplasm of enzymes or other proteins made in the nucleus; and many of the elaborate schemes which have been evolved to explain an immediate genetic control over cytoplasmic protein synthesis will be reduced to a consideration of the mechanisms by which synthesis of protein is repressed and derepressed in the nucleus—a mechanism in which the labile nuclear RNA may well play a crucial role. However, this scheme, although it could indeed be a close approximation to

certain aspects of protein synthesis in the cell, provides no mechanism for the genetic determination of base sequence in the cytoplasmic RNA; and even if cytoplasmic RNA can be replicated in the cytoplasm it is difficult to avoid the conviction that its base sequence must ultimately be genetically determined. But whether the mechanism of this genetic determination is a continuous flow of RNA molecules from nucleus to cytoplasm, a passage of RNA to the cytoplasm at one particular stage of the cell cycle, or a synthesis of cytoplasmic RNA in association with the nuclear membrane is a problem that may continue to tax our ingenuity for some time to come.”

One of the early studies that indicated existence of mRNA, like that proposed to exist in bacteria (Jacob and Monod, 1961), was that of Scherrer *et al.* (1963) in which they showed that i) the RNA from HeLa cells can stimulate in vitro protein synthesis by *E. coli* extracts, ii) an unstable fraction of RNA exists, and iii) that a small fraction of the total RNA has DNA-like base ratios. In view of these, they suggested that messenger RNA does indeed exist in animal cells as well. The same study (Scherrer *et al.*, 1963) also identified 45S RNA and raised the possibility that this may be “a complex of several ribosomal RNA molecules” and that this implied “a mechanism for accurately dividing the large molecule into smaller ones of specific sizes”.

With further studies, several of the ambiguities got resolved, as evident from the more definitive statements in later reviews on RNA (Darnell Jr, 1968; Weinberg, 1973). The general picture that emerged was that in eukaryotes, RNA was synthesized in nucleus and transported to cytoplasm for directing protein synthesis and the nuclear synthesized RNA included precursors to mRNAs, ribosomal RNAs and tRNAs, the later also named as adaptor or soluble RNAs. Initially, nuclear RNA fraction was classified as D-RNA (DNA-like RNA) and ribosomal RNA (Bolton and McCarthy, 1962). The D-RNA was also named as giant RNA or messenger-like RNA, but was soon commonly renamed as heterogeneous nuclear RNA or hnRNA (Darnell Jr, 1968) because of the wide range of sedimentation coefficients (20S to 100S) of this fraction of nuclear RNA.

It is significant that many studies in the 1960s that examined the synthesis and turnover of nuclear

RNA (Edstrom, 1964; Goldstein and Trescott, 1970; Shearer and McCarthy, 1967; Soeiro *et al.*, 1968) found that a significant fraction of the rapidly labeled hnRNAs was not transported to the cytoplasm but was degraded within the nucleus (Darnell Jr, 1968; Harris, 1963; Weinberg, 1973). The nature and significance of such nucleus restricted hnRNA species remained unknown and unexplored at that time.

Studies from James Bonner’s lab (Holmes and Bonner, 1974; Holmes *et al.*, 1972) reported that many of the large (up to 15-30kb) hnRNAs in mammalian cells comprised interspersed mid-repetitive and single copy sequences. Since a majority of the nucleotides of nuclear RNA were found to never reach the cytoplasm, these results also raised the possibility that many of these hnRNAs, which remain confined to the nucleus, may not contain structural (protein-coding) information (Holmes and Bonner, 1974; Holmes *et al.*, 1972).

A controversial claim, initially originating from James Bonner’s lab (Huang and Bonner, 1965) and later also reported by a few other labs (reviewed by Weinberg, 1973), was that some small (<40 nucleotides) RNAs remain covalently associated with histones and, therefore, were named as ‘chromosomal RNA’. However, in view of contradictory results from different labs and some claims about non-reproducibility of the Bonner lab experiments, Weinberg in his review of nuclear RNAs concluded that the existence of cRNA as a separate class of nuclear RNA is questionable (Weinberg, 1973).

Cytogenetic Studies on Chromatin Organization and Gene Expression

The cytogenetic approach to understand organization and functions of genes progressed somewhat independently of the biochemical and molecular approaches catalyzed by the above discussed developments about DNA-RNA-protein relationship. One of the significant cytogenetic findings was that a variable component of nuclear chromatin remains condensed all through the cell division cycle in most tissues, and hence was named ‘heterochromatin’ (Heitz, 1928). Cytogenetic and later molecular studies revealed that this chromatin component was largely devoid of ‘Mendelian’ genes, replicated in late S, was enriched in highly and mid-repetitive sequences, and the condensed chromatin, as inferred through cellular

autoradiography studies, was generally transcriptionally silent (Brown, 1966; Cooper, 1959; Craig, 2005; Gatti and Pimpinelli, 1992; Lakhotia, 2017a, 2017b; Shah *et al.*, 1973).

Parallel studies on the nuclear content of DNA in different organisms revealed that the haploid DNA content (C-value) in different species, including related ones, varied significantly without any apparent correlation with specific phenotypic attributes, and that the C-value even in the species with least DNA content was much more than required to code for the estimated number of protein-coding genes in that species. Such unexplainable variations in C-values in diverse organisms lead to formulation of the C-value paradox (Gall, 1981; Thomas, 1971; Zuckerkandl, 1976). Although the basic increase in the C-value in more evolved eukaryotic groups appeared to correlate with their biological complexity and this was suggested to be related to the requirements of more complex regulatory circuits (Britten and Davidson, 1969; Mayr, 1970), the unusually large variations in the C-value in members of same taxonomic group, however, defied a simple explanation.

Heterochromatin and the C-value paradox, when viewed from the protein-centric view of the genotype-phenotype relationship, have continued to puzzle cytogeneticists as well as molecular biologists. Although many cytogenetic studies had documented specific as well as general phenotypic effects of different heterochromatin regions in diverse organisms (Brown, 1966; Cooper, 1959; Shah *et al.*, 1973) and some heterochromatin regions and repetitive DNA sequences were also shown to be transcriptionally active (Hess and Meyer, 1968; Holmes and Bonner, 1974; Lakhotia and Jacob, 1974), the underlying mechanisms and their significance remained vague. Consequently, an emergent view among molecular biologists in 1970s was that bulk of the repeat sequence rich heterochromatin regions may not serve any specific functions in eukaryotes.

Was the ‘Central Dogma of Molecular Biology’ Wrong?

The populist view of central dogma appears to place proteins as the crucial mediators of the information readout in the genotype-phenotype relationship. However, the primary suggestion in the central dogma was to indicate the plausible paths of flow of genetic

information between DNA, RNA and proteins (Crick, 1970). The major emphasis in the central dogma was that the information once delivered to proteins cannot go back to nucleic acids (Crick, 1958, 1970). Although there were some early suggestions that information can flow back from proteins to nucleic acids (Cook, 1977; Mekler, 1967), the idea of ‘reverse translation’ did not get experimental support and thus has not been taken seriously. The paths for flow of information from DNA to protein as enunciated in the central dogma did not specifically state that only proteins, whose structure is dictated by the flow of information from DNA/RNA, had the primary or exclusive roles in manifestation of the phenotype. In fact, Crick in his fuller version of the central dogma stated “the real question to ask is, how much extra information is required, in addition to DNA and the code, to make a particular cell work?” (Crick, 1970). Yet, it began to be commonly believed that only the segments of genome that produce mRNAs and those RNAs that are parts of the translational machinery are of consequence in determining the phenotype. Such common perception that gene function is manifest entirely through proteins perhaps gained widespread belief when Watson stated “DNA makes RNA makes proteins’ in his popular book, ‘Molecular Biology of the Gene’ (Watson, 1965).

The widespread perception that DNA makes RNA makes proteins heightened the enigmas associated with the nuclear RNAs that did not see the face of cytoplasm and the heterochromatin associated DNA sequences. Not finding suitable answers to these vexing issues and following the linear logic of reductionist molecular biology, proposals for existence of ‘selfish’ or ‘junk’ DNA in eukaryotic genomes were mooted independently by several groups in 1980 (Doolittle and Sapienza, 1980; Ohno, 1980; Orgel and Crick, 1980). Accordingly, the noncoding unique, repetitive and transposon DNA sequences were considered to be ‘selfish’ or ‘junk’. Despite some debate about these proposals in the 1980s (e.g., see Bernardi, 1983; Gould, 1982; Hickey, 1982; Shuter *et al.*, 1983), these viewpoints were widely accepted so that ‘selfish’ and ‘junk’ DNAs became standard topics in most text-books in genetics, molecular biology and biotechnology.

Use of the word ‘dogma’ in Crick’s proposal was unusual and not justified from the viewpoint of

basic philosophy of science research. He (Crick, 1988) himself explained the reasons for this usage as follows: “I called this idea the central dogma, for two reasons, I suspect. I had already used the obvious word hypothesis in the sequence hypothesis, and in addition I wanted to suggest that this new assumption was more central and more powerful. ... As it turned out, the use of the word dogma caused almost more trouble than it was worth. Many years later Jacques Monod pointed out to me that I did not appear to understand the correct use of the word dogma, which is a belief that cannot be doubted. I did apprehend this in a vague sort of way but since I thought that all religious beliefs were without foundation, I used the word the way I myself thought about it, not as most of the world does, and simply applied it to a grand hypothesis that, however plausible, had little direct experimental support.”

The basic tenets of the ‘central dogma of molecular biology’ thus remain valid, but the very wide acceptance of its mis-interpreted view that DNA makes RNA makes protein as the ‘dogma’ and of the existence of ‘selfish’ and/or ‘junk’ DNA in eukaryotes as the explanation for the C-value paradox and nucleus-limited RNAs, led to the widespread neglect of large parts of genome and its transcriptional outputs.

The Populist View of Central Dogma and Concepts of Selfish and Junk DNA Stimulated Research on Gene Expression but Thwarted Progress in ncRNA Biology

The paths of genetic information flow in cells as envisaged in the central dogma indeed transformed biological research, leading to the birth and subsequent growth of molecular biology. Emergence of recombinant DNA and many other powerful methods for molecular biological studies galvanized researches directed to understanding of the organization, expression and regulation of protein-coding genes in diverse eukaryotes. These in turn led to the remarkable developments in bioinformatics, expansion of internet and many high-throughput technologies and their biotechnological applications. However, studies on those parts of the nuclear genome that were not directly involved in synthesis of one or the other proteins and the noncoding transcripts were not only not pursued actively, but were often discouraged. As

a consequence, much of the excitement in 1960s about the nuclear transcripts that do not move to cytoplasm and thus do not code for proteins, and transcription of repetitive sequences and heterochromatic regions cooled down for lack of support. Thus, this area of RNA biology made limited progress during later decades of the last century. Studies on functional significance of heterochromatin, repetitive sequences and of the diverse transposons, common in all genomes, were mostly driven by the view that transposons were ‘parasites’ and ‘selfish’, and the condensed state of heterochromatin primarily served to keep the transposons in a repressed state to protect the genome against DNA damage and deleterious genome rearrangements that accompany the transposition events.

The protein-centric view of genotype-phenotype relationship received indirect support also from the recombinant DNA methods that were emerging in early 1970s. Cloning, sequencing, expression and alterations of specific DNA sequences found meaning and purpose in terms of the activities and properties of the encoded proteins, leading finally to the enormously popular band-wagon of recombinant DNA technology and genetic engineering based biotechnology. Amidst the rush for cloning, characterizing and altering different proteins for understanding gene-protein functions, and for diverse biotechnological applications and gene-therapy, some basic issues of biology, unfortunately, got widely ignored. For example, studies directed to identify mutations that may be responsible for a given inherited human disease mostly emphasized those affecting the coding potential while others were generally left as SNPs. Similarly, the basic tenet of evolution that tweaking of regulatory circuits rather than evolving new proteins is an important driver of evolution (Britten and Davidson, 1969; Mayr, 1970) was often not the main point of focus in molecular phylogenetic studies. The diverse nucleus-limited hnRNAs also were not actively pursued although their existence continued to be recorded but often ignored as ‘run-on’ transcripts of little consequence (Diaz *et al.*, 1981; Diaz and Gall, 1985).

Resurgence of Noncoding RNAs in the ‘Genomics Era’

Notwithstanding the general lack of support for

research on noncoding RNAs, some noncoding RNAs were shown, during the last two decades of 20th century, to have remarkable functions. These included i) multiple noncoding transcripts of the *93D* or *hsr ω* gene in *Drosophila*, some of whose transcripts appeared to regulate the dynamics of hnRNPs and some other RNA binding proteins (Garbe *et al.*, 1986; Lakhotia, 1996; Lakhotia and Mukherjee, 1982), ii) Xist transcripts associated with inactivation of one of the two X chromosomes in somatic cells of female mammals (Brockdorff *et al.*, 1991; Brown, 1991), and iii) the roX transcripts responsible for hyperactivation of the single X-chromosome in male *Drosophila* (Amrein and Axel, 1997; Meller *et al.*, 1997). These and a few other noncoding RNAs that were reported in diverse organisms laid the foundation for the appreciation that some noncoding RNAs transcribed by the RNA polII can perform significant functions without undergoing the act of translation (Lakhotia, 1996).

My lab reported two interesting findings about the *93D/hsr ω* gene in *Drosophila* in 1982. One was that its singular transcriptional activation following benzamide treatment of larval salivary glands did not result in synthesis of any new protein which suggested that these transcripts may not encode a protein (Lakhotia and Mukherjee, 1982) and second, that this gene was conserved in a wide range of *Drosophila* species since in each of the examined species, one of the heat shock induced loci also responded in a singular manner to benzamide treatment (Lakhotia and Singh, 1982). The essentiality of this noncoding gene for survival of flies was established using a genetic approach (Mohler and Pardue, 1984). The noncoding nature of the *93D/hsr ω* gene was soon confirmed by its cloning and sequencing in different *Drosophila* species (Garbe *et al.*, 1986; Ryseck *et al.*, 1985). Its noncoding nature and the apparent rapid divergence in its base sequence in related species (Garbe *et al.*, 1986; Ryseck *et al.*, 1985) appeared to match the properties of typical ‘selfish’ DNA. However, taking into consideration the rapid sequence divergence of this gene and the earlier reported accumulation of several nuclear non-histone proteins at this gene locus in stressed cells (Saumweber *et al.*, 1980), I suggested (Lakhotia, 1987; Lakhotia, 1989) that the structure, rather than the base sequence of this gene’s noncoding transcripts may be important and that such RNAs have some regulatory functions

through binding with diverse proteins. The protein-binding functions of this gene’s large nucleus limited RNAs were confirmed by discovery of the omega speckles, which we suggested to regulate the availability of diverse hnRNPs and some other RNA-binding proteins in normal and stressed cells (Lakhotia *et al.*, 1999; Prasanth *et al.*, 2000). Such a role of noncoding RNAs in dynamically regulating the availability of regulatory nuclear proteins appeared to be widespread (Jolly and Lakhotia, 2006; Lakhotia, 2011; Lakhotia, 2012; Lakhotia, 2016). More recently, a variety of nuclear sub-organelles have been found in different organisms to contain noncoding ‘architectural RNAs’ to organize the membrane-less phase-separated entities that harbor specific sets of proteins and regulate their dynamicity under different conditions (Chujo and Hirose, 2017; Chujo *et al.*, 2016; Lakhotia, 2017a; Lakhotia *et al.*, 2012). Xist and roX, the other long noncoding RNAs were, on the other hand, established as examples of long noncoding RNAs that impact chromosome scale chromatin organization (Keller and Akhtar, 2015; Lakhotia, 2015; Valencia and Wutz, 2015).

The other experimental approach in 1990s that catalyzed a widespread excitement in noncoding RNAs was the demonstration that phenomena like quelling or transgene co-suppression (Cogoni *et al.*, 1996), post-transcriptional gene silencing, RNA interference or RNAi etc (Cogoni and Macino, 2000; Fire *et al.*, 1998) were dependent upon small noncoding RNAs (Hannon, 2002; Schramke and Allshire, 2004). Another development that re-kindled interest in noncoding RNAs was the revelations emanating from sequencing of whole genomes in different organisms, including the human genome. The genome data graphically demonstrated that the actual protein-coding component in any eukaryotic genome is only a tiny fraction of the total nuclear DNA, being less than 2% in the human genome (Szymanski and Barciszewski, 2002; Zuckerkandl, 2002).

In 1995, I was invited to give a talk at the ‘TRendys’ meeting at Jawaharlal Nehru University, New Delhi. A primary objective of this series of meetings was to highlight areas of research that are likely to become trendy in the coming years. Being convinced that the views about the existence of ‘selfish’ or ‘junk’ DNA are largely due to ignorance, I predicted in this ‘TRendys’ meeting that noncoding

RNAs would soon become trendy. In the published review, that was based on this talk on RNA polII dependent noncoding RNAs (Lakhotia, 1996), I stated “While some of these transcripts may function as ribozymes or as anti-sense regulators, most others may function more directly through their specific protein-binding properties, Since RNA is believed to be the first “living” molecule, it is very likely that some genes even today function only through this class of molecules. It is expected that instead of being ignored as examples of “selfish DNA”, a more positive search for their functions will help unravel the significance of this novel class of genes”. This indeed has happened. The diverse regulatory and other functions of heterochromatin can now be examined in terms of the large variety of small and long noncoding RNAs generated by and associated with this enigmatic component of eukaryotic genomes (Bierhoff *et al.*, 2014; Chuong *et al.*, 2017; Felden and Paillard, 2017; Lakhotia, 2017a, 2017b; Sawyer and Dunder, 2017). The presence of chromosomal RNAs, reported in 1960s but quickly debunked in 1970s (see above), may find parallels in the roles of RNAs like the Xist and roX transcripts and in the many other small and long noncoding RNAs that associate with chromatin and affect its structural and functional organization (Bell *et al.*, 2017; Johnson *et al.*, 2017; Lakhotia, 2016, 2017a, 2017b; Sridhar *et al.*, 2017; Velazquez Camacho *et al.*, 2017).

The new found interest in noncoding RNA since the beginning of this century is indeed remarkable. What started as a trickle at the turn of this millennium has now become a ‘tsunami’ with enormously large numbers of original research papers and reviews on ncRNAs of various sizes and remarkably diverse functions in normal development and disease being published every month.

Concluding Remarks

Progress in science, or for that matter in any discipline of learning, depends upon provocative hypotheses, which researchers attempt to disprove and till proven wrong, are accepted as the working principle underlying a given phenomenon. The ‘central dogma’ proposed in late 1950s was indeed a bold and provocative hypothesis, which triggered and catalyzed the birth of the new era of molecular biology. However, uncertainty about the existence of reverse

transcription in Crick’s scheme of information transfer delayed a wider acceptance of presence of reverse transcriptase in eukaryotic cells and its possible roles in their lives. The name ‘central dogma’ and the somewhat mis-interpreted importance only of proteins, which led, primarily out of ignorance, to the widely held conviction that those parts of genome that do not encode a protein were ‘selfish’ or ‘junk’ and, therefore, inconsequential, had a negative impact on gaining a holistic view of eukaryotic genome organization and its functions. In this context, it is interesting to note that till about 1910, Morgan, who is justifiably recognized as one of the founding pillars of post-Mendelian genetics and thus of contemporary biology, did not believe in Darwinian natural selection or chromosomal and Mendelian theories of inheritance (Allen, 2015; Benson, 2001; Gilbert, 1978). Morgan’s scorn for the state of Mendelian genetics is apparent in the first few lines with which he starts his paper on “What are “factors” in Mendelian explanations?” (Morgan, 1909): “In the modern interpretation of Mendelism, facts are being transformed into factors at a rapid rate. If one factor will not explain the facts, then two are invoked; if two prove insufficient, three will sometimes work out. The superior jugglery sometimes necessary to account for the result, may blind us, if taken too naïvely, to the common-place that the results are often so excellently “explained” because the explanation was invented to explain them. We work backwards from the facts to the factors, and then, presto! explain the facts by the very factors that we invented to account for them. I am not unappreciative of the distinct advantages that this method has in handling the facts. I realize how valuable it has been to us to be able to marshal our results under a few simple assumptions, yet I cannot but fear that we are rapidly developing a sort of Mendelian ritual by which to explain the extraordinary facts of alternative inheritance. So long as we do not lose sight of the purely arbitrary and formal nature of our formulae, little harm will be done; and it is only fair to state that those who are doing the actual work of progress along Mendelian lines are aware of the hypothetical nature of the factor-assumption. But those who know the results at second hand and hear the explanations given, almost invariably in terms of factors, are likely to exaggerate the importance of the interpretations and to minimize the importance of the facts”. However, as his own work with

Drosophila progressed, he reversed his views rapidly within the next 5-6 years (Morgan *et al.*, 1915) and thus laid the foundation for much of the modern experimental biology. It would be an interesting issue for the historians and social scientists to find reasons as to why the mis-interpretation of ‘central dogma’, and the concepts of ‘selfish/junk’ DNA continued to hold sway for decades especially when the number of researchers was several orders of magnitude greater and they had much more powerful experimental tools available to them than in Morgan’s

times? Perhaps the linear reductionist approach and the race to publish and find applications, prevalent in recent times, were some of the contributory factors.

Acknowledgements

Research on the *hsr ω* noncoding gene in my laboratory is currently supported by a COE-II grant from the Department of Biotechnology (Govt. of India). I also thank the Indian National Science Academy (New Delhi) for the Senior Scientist grant.

References

- Allen G E (2015) How many times can you be wrong and still be right? T. H. Morgan, evolution, chromosomes and the origins of modern genetics *Science & Education* **24** 77-99
- Allfrey V, Daly M M and Mirsky A E (1953) Synthesis of protein in the pancreas *J Gen Physiology* **37** 157-175
- Amrein H, Axel R (1997) Genes expressed in neurons of adult male *Drosophila* *Cell* **88** 459-469
- Avery O T, MacLeod C M and McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III *J Exp Med* **79** 137-158
- Beadle G W and Ephrussi B (1936) The differentiation of eye pigments in *Drosophila* as studied by transplantation *Genetics* **21** 225-247
- Beadle G W and Tatum E L (1941) Genetic control of biochemical reactions in *Neurospora* *Proc Natl Acad Sci USA* **27** 499-506
- Bell J C, Jukam D, Teran N A, Risca V I, Smith O K, *et al.* (2017) Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts *bioRxiv* 118786
- Bernardi G (1983) Genome instability and the selfish DNA issue *Folia Biol* **29** 82-92
- Benson K (2001) T. H. Morgan’s resistance to the chromosome theory *Nature Rev Genet* **2** 469-474
- Bierhoff H, Postepska-Igielska A and Grummt I (2014) Noisy silence: non-coding RNA and heterochromatin formation at repetitive elements *Epigenetics* **9** 53-61
- Bolton E T and McCarthy B J (1962) A general method for the isolation of RNA complementary to DNA *Proc Natl Acad Sci USA* **48** 1390-1397
- Brachet J (1947) Nucleic acids in the cell and the embryo In: *Symp Soc Exp Biol* 207-224
- Britten R J and Davidson E H (1969) Gene regulation for higher cells: a theory *Science* **165** 349-357
- Brockdorff N, Ashworth A, Kay G F, Cooper P, Smith S, *et al.* (1991) Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome *Nature* **351** 329-331
- Brown S D (1991) XIST and the mapping of the X chromosome inactivation centre *Bio essays* **13** 607-612
- Brown S W (1966) Heterochromatin *Science* **151** 417-425
- Caspari E (1933) Uber die wirkung eines pleiotropen gens bei der mehlmotte *Ephestia kuhniella* Zeller *W Roux Arch Entw Mech* **130** 355-381
- Caspersson T, Schultz J (1940) Ribonucleic acids in both nucleus and cytoplasm, and the function of the nucleolus *Proc Natl Acad Sci USA* **26** 507-515
- Chujo T and Hirose T (2017) Nuclear bodies built on architectural long noncoding RNAs: unifying principles of their construction and function *Mol Cells* **40** 889-896.
- Chujo T, Yamazaki T and Hirose T (2016) Architectural RNAs (arcRNAs): A class of long noncoding RNAs that function as the scaffold of nuclear bodies *Biochim Biophys Acta (BBA)-Gene Regulatory Mechanisms* **1859** 139-146
- Chuong E B, Elde N C and Feschotte C (2017) Regulatory activities of transposable elements: from conflicts to benefits *Nature Reviews Genetics* **18** 71
- Cogoni C, Ireland J T, Schumacher M, Schmidhauser T J, Selker E U, *et al.* (1996) Transgene silencing of the *al-1* gene in vegetative cells of *Neurospora* is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation *The EMBO journal* **15** 3153
- Cogoni C and Macino G (2000) Post-transcriptional gene silencing across kingdoms *Curr Opin Genet Develop* **10** 638-643
- Cook N D (1977) The case for reverse translation *J Theor Biol* **64** 113-135

- Cooper K W (1959) Cytogenetic analysis of major heterochromatic elements (especially Xh and Y) in *Drosophila melanogaster*, and the theory of "heterochromatin" *Chromosoma* **10** 535-588
- Craig J M (2005) Heterochromatin—many flavours, common themes *Bioessays* **27** 17-28
- Crick F (1970) Central dogma of molecular biology *Nature* **227** 561-563
- Crick F H C (1958) On protein synthesis. In: Symp Soc Exp Biol 138-163
- Crick F H C (1988) What mad pursuit: A personal view of scientific discovery. Basic Books
- Darnell Jr J E (1968) Ribonucleic acids from animal cells *Bacteriological Reviews* **32** 262
- Diaz M O, Barsacchi-Pilone G, Mahon K A and Gall J G (1981) Transcripts from both strands of a satellite DNA occur on lampbrush chromosome loops of the newt *Notophthalmus* *Cell* **24** 649-659
- Diaz M O and Gall J G (1985) Giant readthrough transcription units at the histone loci on lampbrush chromosomes of the newt *Notophthalmus* *Chromosoma* **92** 243-253
- Doolittle W F and Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution *Nature* **284** 601-603
- Edstrom J E (1964) Chromosomal RNA and other nuclear RNA fractions. In: The role of chromosomes in development. (Locke M Ed) pp 137-152 Academic Press
- Felden B and Paillard L (2017) When eukaryotes and prokaryotes look alike: the case of regulatory RNAs *FEMS Microbiology Reviews* fux038
- Fire A, Xu S, Montgomery M K, Kostas S A, Driver S E, *et al.* (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans* *Nature* **391** 806-811
- Gall J G (1981) Chromosome structure and the C-value paradox *J Cell Biol* **91** 3s-14s
- Gamow G (1954) Possible relation between deoxyribonucleic acid and protein structures *Nature* **173** 318-318
- Gamow G and Yēas M (1955) Statistical correlation of protein and ribonucleic acid composition *Proc Natl Acad Sci USA* **41** 1011-1019
- Garbe J, Bendena W, Alfano M and Pardue M (1986) A *Drosophila* heat shock locus with a rapidly diverging sequence but a conserved structure *J Biol Chem* **261** 16889-16894
- Garrod A E (1909) Inborn errors of metabolism. Oxford University Press
- Gatti M and Pimpinelli S (1992) Functional elements in *Drosophila melanogaster* heterochromatin *Annu Rev Genetics* **26** 239-276
- Gilbert S F (1978) The embryological origins of the gene theory *J History Biol* **11** 307-351
- Goldschmidt R (1916) Genetic Factors and Enzyme Reaction *Science* **43** 98-100
- Goldstein L and Trescott O H (1970) Characterization of RNAs that do and do not migrate between cytoplasm and nucleus *Proc Natl Acad Sci USA* **67** 1367-1374
- Gould S J (1982) Darwinism and the expansion of evolutionary theory *Science* **216** 380-387
- Hannon G (2002) RNA interference *Nature* **418** 244-251
- Harris H (1963) Nuclear ribonucleic acid. In: Progress in Nucleic Acid Research and Molecular Biology (Davidson J N, Cohn W E Ed) pp 19-59 Academic Press
- Haurowitz F (1950) Chemistry and Biology of Proteins. Academic Press Inc
- Heitz E (1928) Das heterochromatin der Moose I *Jb wiss Bot* **69** 762-818
- Hershey A D and Chase M (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage *J Gen Physiol* **86** 39-56
- Hickey D A (1982) Selfish DNA: a sexually-transmitted nuclear parasite *Genetics* **101** 519-531
- Hess O and Meyer G F (1968) Genetic activities of the Y chromosome in *Drosophila* during spermatogenesis *Adv Genet* **14** 171-223
- Holloway B W, Ripley S H (1952) Nucleic acid content of reticulocytes and its relation to uptake of radioactive leucine in vitro *J Biol Chem* **196** 695-701
- Holmes D S and Bonner J (1974) Interspersion of repetitive and single-copy sequences in nuclear ribonucleic acid of high molecular weight *Proc Natl Acad Sci USA* **71** 1108-1112
- Holmes D S, Mayfield J E, Sander G and Bonner J (1972) Chromosomal RNA: Its Properties *Science* **177** 72-74
- Huang R-C and Bonner J (1965) Histone-bound RNA, a component of native nucleohistone *Proc Natl Acad Sci USA* **54** 960-967
- Ingram V M (1957) Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin *Nature* **180** 326-328
- Jacob F, Monod J (1961) On the regulation of gene activity. In: Cold Spring Harbor Symposia on Quantitative Biology. Cold Spring Harbor Laboratory Press, 193-211
- Johnson W L, Yewdell W T, Bell J C, McNulty S M, Duda Z, *et al.* (2017) RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin *eLife* **6** e25299

- Jolly C and 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 Research* **34** 5508-5514
- Keller C I and Akhtar A (2015) The MSL complex: juggling RNA-protein interactions for dosage compensation and beyond *Curr Opin Genet Dev* **31** 1-11
- Lakhotia S C (1987) The 93D heat shock locus in *Drosophila*: A review *J Genetics* **66** 139-157
- Lakhotia S C (1989) The 93D heat-shock locus of *Drosophila melanogaster* - modulation by genetic and developmental factors *Genome* **31** 677-683
- Lakhotia S C (1996) RNA polymerase II dependent genes that do not code for protein *Indian J Biochem Biophys* **33** 93-102
- Lakhotia S C (2011) Forty years of the 93D puff of *Drosophila melanogaster* *J Biosciences* **36** 399-423
- Lakhotia S C (2012) Long non-coding RNAs coordinate cellular responses to stress *WIREs RNA* **3** 779-796
- Lakhotia S C (2015) Divergent actions of long noncoding RNAs on X-chromosome remodelling in mammals and *Drosophila* achieve the same end result: dosage compensation *J Genetics* **94** 575-584
- Lakhotia S C (2016) Non-coding RNAs have key roles in cell regulation *Proc Indian Natn Sci Acad* **82** 1171-1182
- Lakhotia S C (2017a) From heterochromatin to long noncoding RNAs in *Drosophila*: Expanding the arena of gene function and regulation. In: Long Non Coding RNA Biology (Rao M R S Ed) pp 75-118 Springer Nature Singapore Pte Ltd
- Lakhotia S C (2017b) Non-coding RNAs demystify constitutive heterochromatin as essential modulator of epigenotype *The Nucleus* **60** 299-314
- Lakhotia S C and Jacob J (1974) EM autoradiographic studies on polytene nuclei of *Drosophila melanogaster*. II. Organization and transcriptive activity of the chromocentre *Exp Cell Res* **86** 253-263
- Lakhotia S C, Mallik M, Singh A K and Ray M (2012) The large noncoding *hsr ω -n* transcripts are essential for thermotolerance and remobilization of hnRNPs, HP1 and RNA polymerase II during recovery from heat shock in *Drosophila Chromosoma* **121** 49-70
- Lakhotia S C and Mukherjee T (1982) Absence of novel translation products in relation to induced activity of the 93D puff in *Drosophila melanogaster Chromosoma* **85** 369-374
- Lakhotia S C, Ray P, Rajendra T K and Prasanth K V (1999) The non-coding transcripts of hsr-omega gene in *Drosophila*: do they regulate trafficking and availability of nuclear RNA-processing factors? *Current Science* **77** 553-563
- Lakhotia S C and Singh A K (1982) Conservation of the 93D puff of *Drosophila melanogaster* in different species of *Drosophila Chromosoma* **86** 265-278
- Lawrence W J C, Price J R (1940) The genetics and chemistry of flower colour variation *Biological Reviews* **15** 35-57
- Loeb J and Chamberlain M M (1915) An attempt at a physico-chemical explanation of certain groups of fluctuating variation *J Exp Zool Part A: Ecological Genetics and Physiology* **19** 559-568
- Mayr E (1970) Populations, Species, and Evolution. An abridgment of Animal Species and Evolution. Cambridge (USA): Harvard University Press
- Mekler L B (1967) Mechanism of biological memory *Nature* **215** 481-484
- Meller V H, Wu K H, Roman G, Kuroda M I and Davis R L (1997) roX1 RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system *Cell* **88** 445-457
- Mohler J and Pardue M L (1984) Mutational analysis of the region surrounding the 93d heat shock locus of *Drosophila melanogaster Genetics* **106** 249-265
- Morgan T H (1909) What are "factors" in Mendelian explanations? *American Breeders Association Reports* **5** 365-368
- Morgan T H (1935) The relation of genetics to physiology and medicine *Scientific Monthly, July* 313-328
- Morgan T H, Sturtevant A H, Muller H J and Bridges C B (1915) The mechanism of Mendelian heredity. Henry Holt
- Ohno S (1980) Gene Duplication, Junk DNA, Intervening sequences and the universal signal for their removal *Rev Brasil Genet* **III** 99-114
- Orgel L E and Crick F H (1980) Selfish DNA: the ultimate parasite *Nature* **284** 604-607
- Pauling L (1940) A theory of the structure and process of formation of antibodies *Journal American Chemical Society* **62** 2643-2657
- Pauling L (1948) Nature of forces between large molecules of biological interest *Nature* **161** 707-709
- Pauling L, Itano H A, Singer S J and Wells I C (1949) Sickle cell anemia, a molecular disease *Science* **110** 543-548
- Prasanth K, Rajendra T, Lal A and Lakhotia S (2000) Omega speckles - a novel class of nuclear speckles containing hnRNPs associated with noncoding hsr-omega RNA in *Drosophila J Cell Sci* **113 Pt 19** 3485-3497
- Ryseck R P, Walldorf U and Hovemann B (1985) Two major RNA products are transcribed from heat-shock locus 93D of *Drosophila melanogaster Chromosoma* **93** 17-20

- Saumweber H, Symmons P, Kabisch R, Will H and Bonhoeffer F (1980) Monoclonal antibodies against chromosomal proteins of *Drosophila melanogaster*: establishment of antibody producing cell lines and partial characterization of corresponding antigens *Chromosoma* **80** 253-275
- Sawyer I A and Dundr M (2017) Chromatin loops and causality loops: The influence of RNA upon spatial nuclear architecture *Chromosoma* **126** 1-17
- Scherrer K, Latham H and Darnell J E (1963) Demonstration of an unstable RNA and of a precursor to ribosomal RNA in HeLa cells *Proc Natl Acad Sci USA* **49** 240-248
- Schramke V and Allshire R (2004) Those interfering little RNAs! Silencing and eliminating chromatin *Current Opinion Genetics Development* **14** 174-180
- Segrè G (2000) The Big Bang and the genetic code *Nature* **404** 437
- Shah V, Lakhota S C and Rao S R V (1973) Nature of heterochromatin *J Sci Ind Res* **32** 467-480
- Shearer R and McCarthy B (1967) Evidence for ribonucleic acid molecules restricted to the cell nucleus *Biochemistry* **6** 283-289
- Shuter B J, Thomas J E, Taylor W D and Zimmerman A M (1983) Phenotypic correlates of genomic DNA content in unicellular eukaryotes and other cells *The American Naturalist* **122** 26-44
- Soeiro R, Vaughan M and Warner J (1968) The turnover of nuclear DNA-like RNA in HeLa cells *J Cell Biol* **39** 112-118
- Sridhar B, Rivas-Astroza M, Nguyen T C, Chen W, Yan Z, *et al.* (2017) Systematic mapping of RNA-chromatin interactions in vivo *Current Biology* **27** 602-609
- Strasser B J (2006) A world in one dimension: Linus Pauling, Francis Crick and the central dogma of molecular biology *Hist Phil Life Sci* **28** 491-512
- Strauss B S (2016) Biochemical genetics and molecular biology: the contributions of George Beadle and Edward Tatum *Genetics* **203** 13-20
- Swift H H (1950) The desoxyribose nucleic acid content of animal nuclei *Physiological Zoology* **23** 169-198
- Szymanski M and Barciszewski J (2002) Beyond the proteome: non-coding regulatory RNAs *Genome Biology* **3** 1-6
- Thomas C A (1971) The genetic organization of chromosomes *Annu Rev Genetics* **5** 237-256
- Thorell B (1947) The relation of nucleic acids to the formation and differentiation of cellular proteins. In: Cold Spring Harbor Symposia on Quantitative Biology. Cold Spring Harbor Laboratory Press, 247-255
- Valencia K and Wutz A (2015) Recent insights into the regulation of X-chromosome inactivation *Adv Genomics Genet* **5** 227-238
- Velazquez Camacho O, Galan C, Swist-Rosowska K, Ching R, Gamalinda M, *et al.* (2017) Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation *eLife* **6** e25293
- Watson J D (1965) Molecular biology of the gene, Benjamin
- Watson J D (2003) Genes, girls and Gamow. Oxford University Press
- Watson J D and Crick F H (1953) Molecular structure of nucleic acids *Nature* **171** 737-738
- Weinberg R A (1973) Nuclear RNA metabolism *Annu Rev Biochem* **42** 329-354
- Zuckerklund E (1976) Gene control in eukaryotes and the c-value paradox "excess" DNA as an impediment to transcription of coding sequences *J Mol Evol* **9** 73-104
- Zuckerklund E (2002) Why so many noncoding nucleotides? The eukaryote genome as an epigenetic machine *Genetica* **115** 105-129.