Supplemental material (Bandea CI. On the concept of biological function, junk DNA and the gospels of ENCODE and Graur *et al*., *bioRxiv*: http://biorxiv.org/)

Supplement 1. Excerpt from: Bandea, CI. A Protective Function for Noncoding, or Secondary DNA, *Med Hypotheses (1990) 31:33-4.*

"*Introduction*

A high percentage of eukaryotic DNA, more than 99% in some species, makes no apparent phenotypic contribution (for reviews see refs 1, 2). The following mechanisms have been proposed for the origin of this noncoding, or secondary DNA: (I) the 'accidental' duplication of some genomic sequences during DNA replication and recombination, (II) the spread of some genomic sequences through retro-transposition, and (III) the spread of proviruses and transposons.

It is not known if secondary DNA has accumulated simply because its rate of deletion has been lower than that of origin, or because individuals possessing secondary DNA have a selective advantage.

A protective role for secondary DNA

I propose that secondary DNA functions as a sink for the integration of proviruses, transposons and other inserting elements, thereby   protecting coding sequences from insertional inactivation or alteration of their expression. Assuming a random site for integration, a genome containing a high percentage of secondary DNA is less prone to insertional damage. This protective mechanism would be more efficient if inserting elements were targeted to secondary DNA. A putative targeting mechanism is homologous recombination. As mentioned, secondary DNA contains sequences that originated from inserting elements. Perhaps, these sequences can direct recombination with homologous sequences from actively transposing parental and related elements.

In addition to its hypothetical protective function, secondary DNA has probably influenced the evolution of cells and their genome. For instance, to accommodate secondary DNA, cells have changed their metabolism (e.g. nucleotides metabolisms) and structure (e.g. nuclear volume). Secondary DNA may have also influenced the rate of evolution by increasing genome fluidity (3).

The protective function of secondary DNA would make evolutionary sense only if insertions occur at a relatively high frequency. The presence of numerous inserted elements in eukaryotic genomes is evidence that insertions in the germ line are common (reviewed in refs 3, 4). Insertions at the somatic level, however, probably occur at much higher frequency in a multicellular organism. In the absence of a protective mechanism, these insertions could lead to high incidents of neoplasmic transformation by altering the expression of some cellular genes. Secondary DNA may also prevent viruses from becoming oncogenic by reducing the probability that host genes having oncogenic potential are incorporated into viral genome.

The advantage in having a protective mechanism against insertions is perhaps best illustrated by organisms that do not have much secondary DNA. Some bacteria and their prophages have coevolved specific sequences where integration occurs, thereby protecting genic sequences.

Conclusion

This paper proposes that secondary DNA was selected as a protective mechanism against insertional damage and suggests that experimental genomic integration of sequences homologous to those of actively inserting elements would enhance the protective function of secondary DNA.

References

1. Dover G A, Flavell RB (eds). Genome Evolution. Academic, London, 1982.
2. Cavalier-Smith T. (ed). The Evolution of genome Size. Wiley, New-York, 1985.
3. Weiner AM, Deininger PL, Efstratiadis A. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annual Review of Biochemistry* 55: 631, 1986.
4. Baltimore D. Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. *Cell* 40:481, 1985."

Supplement 2. A series of 5 comments* on: Pennisi E. ENCODE Project Writes Eulogy for Junk DNA. *Science*, 337: 1159-1161, 2012.

*comments in Science are limited to 2500 characters.

Multiple eulogies for junk DNA?
(Claudiu Bandea; Submitted on 09/11/2012)

Before writing a eulogy for junk DNA (jDNA), we need to know more about it. So what is jDNA?

All genomic sequences that code for proteins and functional RNA, or are involved in regulating gene expression (e.g. promoter elements) are functional DNA (fDNA). However, there are many other sequences that are functional, such as those participating in DNA replication, chromosome organization, etc.

By definition, jDNA is non-functional. However, by its bare presence in the genome, jDNA gets replicated and can undergo recombination, transcription and transposition, and it can be targeted by diverse DNA binding proteins.

ENCODE has been a logical follow up of the Human Genome project, which found that less than 2% of our genome codes for proteins and functional RNAs. Even by including generous estimates of regulatory sequences, the fDNA has been considered just a fraction of the genome; the rest, 90% or more, remained jDNA.

ENCODE has challenged all that, by suggesting that 80% or more of the human genome is fDNA. Accordingly, most of jDNA has evaporated. Whether this interpretation of the data, which involved a change in the definition of fDNA, was a hasty decision that reflects poorly on an otherwise remarkable project remains to be seen.

Here, I want to point out that a previous eulogy for jDNA was penned more than two decades ago (1), when it was proposed that jDNA functions as a sink for the integration of proviruses, transposons and other inserting elements, thereby protecting fDNA from inactivation or alteration of its expression.

Considering that at least 50% of the human genome is composed of transposable elements, and that the rate of their transposition is very high, this protective mechanism makes evolutionary sense. The evolution of alternative protective mechanisms against insertion mutagenesis such as specific integration sites in species that have little jDNA, (e.g. Bacteria) is strong evidence for this selective pressure. However, this pressure enters a new dimension in humans and other multicellular species, in which the number of integration events in somatic cells (including those by retroviruses) that would lead to cancer would be enormous without a protective mechanism. This model is fully consistent with the current data, makes evolutionary sense, and, statistically, is a fact.

1. Bandea CI. A protective function for noncoding, or secondary DNA. *Med. Hypoth.,* 31:33-4. 1990.

Evolutionary questions about junk DNA
(Claudiu Bandea; Submitted on 09/26/2012)

As declared many decades ago by Dobzhansky: "Nothing in biology makes sense except in the light of evolution". Accordingly, the ultimate arguments and data for, or against a functional role of 'junk DNA' (jDNA), might rely on evolutionary principles.

One of the startling conclusion of the ENCODE project was that 80% of the human genome is functional. In a previous comment (*Multiple eulogies for junk DNA?*), I brought forward an old hypothesis proposing that jDNA functions as a sink for the integration of proviruses, transposons and other inserting elements, thereby protecting the genes and their regulatory elements from inactivation or alteration of their function.

The rationale behind this hypothesis on the evolution of genome size and the protective function of jDNA was based on evolutionary data and principles. In order to explore this or any other theory on the potential functions of genomic DNA, it is critical to point out that functional DNA (fDNA) can fulfill two categories of functions: (i) informational functions,

which are based primarily on the specific sequence, and (ii) 'structural' functions, which are more or less independent of the sequence.

The informational DNA (iDNA) includes sequences that code for proteins, functional RNAs and regulatory elements, such as promoters, enhancers, and origins of replication. The 'structural' DNA (sDNA) can have organizational, 'mechanical,' or 'spacing' functions (e.g. DNA in centromeres and in regions located between iDNA sequences), or protective functions against insertional mutagenesis (presumably, all "junk DNA").

In exploring hypotheses on the potential functions jDNA, it is essential to consider its evolutionary origin. Approximately 50% of the human genome is composed of transposable elements (TEs), including thousands of human endogenous retroviruses; it is very likely, also, that much of the rest of jDNA are remnants of TEs that have lost their TE-sequence signatures.

Therefore, from an evolutionary perspective, the central issue is whether jDNA has accumulated simply because its rate of deletion has been lower that of its amplification (in this case, jDNA might be thought of being parasitic in nature), or because the host organisms possessing jDNA have a selective advantage (in this case, jDNA is symbiotic)? In other words, is the evolution of genome size and jDNA under host selective constrains, or not?

Evolutionary constrains on genome size evolution: the Hummingbird Case
(Claudiu Bandea; Submitted on 09/27/2012)

In a previous comment, *Evolutionary questions about junk DNA*, I asked: is the evolution of genome size and junk DNA (jDNA) under host selective constrains?

One of the main lines of evidence supporting the hypothesis that jDNA serves as a defense mechanism against insertional mutagenesis was the evolution of alternative protective mechanisms such as specific integration sites in organisms that have little jDNA (e.g. Bacteria).

Indeed, the evolution of specific integration sites is strong evidence for the high selective pressure imposed by insertional elements on their hosts; it should be pointed out, however, that from an evolutionary perspective, the evolution of specific integration sites was a co-evolutionary event, as both the hosts and the inserting elements benefited from this mechanism.

Unlike Bacteria, which have been under strong selective pressure to maintain a small genome, other organisms, which for various reasons could relax this constrain, used a different strategy for preventing insertion mutagenesis: accumulation of an increasing number of transposable elements including endogenous viruses in their genome in order to serve as a defensive mechanism against insertional mutagenesis. Likely, during its evolution, this novel protective mechanism was enhanced by the presence of additional protective features, such as (sequence-directed) preferable DNA islands for the integration of inserting elements, which relied on the activity of recombination machineries, including homologous recombination.

Although the process of generating protective DNA was based on the amplifying activity of inserting elements, the sites or regions of the host's genome in which this DNA accumulated (i.e. was evolutionary 'fixed'), and its overall quantity, were under evolutionary constraints by the host.

Whereas the constraints regarding the sites for the accumulation of protective DNA were relatively strict, those on the size of the genome and the amount of jDNA were more relaxed, but at work. Probably, the best example of these host evolutionary constrains at work, even in complex organisms such as vertebrates, is the Hummingbird Case: the hummingbirds have the smallest genome size, not only among birds but all tetrapods, apparently because of the selective force imposed by their high metabolic demands associated with powered flight.

Junk DNA and cancer: an extended and unexpected association
(Claudiu Bandea; Submitted on 09/28/2012)

The role of endogenous or exogenous inserting elements, such as retroviruses, in causing cancer in humans and many other species is well established. What is not established, however, is that 'junk DNA' (jDNA), which contains many cancer-inducing inserting elements, might also protect against cancer; it might be a case of 'fighting fire with fire'!

One of the main lines of evidence supporting the hypothesis that jDNA serves as a defense mechanism against insertional mutagenesis was the evolution of alternative protective mechanisms, such as specific integration sites, in organisms that have little jDNA (e.g. Bacteria)(1). Indeed, the evolution of these alternative protective mechanisms is strong evidence for the high selective pressure imposed by insertional elements on their hosts.

However, this selection pressure takes a new dimension in organisms in which insertional mutagenesis can lead to cancer. In humans, for example, given the enormous number of somatic cells and their high turnover rate during the reproductive span, the number of insertion events that would potentially lead to cancer in the absence of protective mechanisms would be evolutionarily drowning (1).

One of the major objectives of the ENCODE project was to help 'connect' the human genome with health and disease. Considering that jDNA might be one of the main protective mechanism against insertional mutagenesis leading to cancer, this hypothetical model on the protective function of jDNA aligns with the ENCODES's objective and with those of the next wave of genomic projects addressing human diseases.

Notably, this hypothetical model can be fully addressed, both analytically and experimentally; for example, transgenic mice carrying DNA sequences homologous to infectious retroviruses such as murine leukemia viruses (MuLV) might be more resistant to cancer induced by experimental MuLV infections as compared to controls.

1. Bandea CI. A protective function for noncoding, or secondary DNA. *Med. Hypoth*., 31:33-4. 1990.

C-value paradox and junk DNA enigma: case solved?
(Claudiu Bandea; Submitted on 09/28/2012)

The concept of 'junk DNA' (jDNA) has its roots in the C-value paradox, which designates the fact that the haploid genome size (C-value) varies enormously among eukaryotic species, and that the C-value doesn't correlate with the presumed number of genes as expected from the apparent organismal complexity. For example, the size of genomes in vertebrates and animals varies by a factor of more than 300 and 3000, respectively.

The C-value paradox led to the paradigm that only a relatively low percentage of the genome is occupied by functional DNA; the rest, over 90% in some species, such as humans, constitutes non-functional, or jDNA.

Part of the enigma surrounding the C-value paradox and the evolutionary origin of jDNA was solved by the discovery that it's composed primarily of transposable elements, including endogenous viruses. However, the puzzle at the center of C-value paradox and jDNA paradigm has remained: (i) is jDNA functional, and has it been maintained because the host organisms possessing it have a selective advantage (i.e. is jDNA symbiotic)?; or (ii) is jDNA non-functional, and has it accumulated simply because its rate of deletion has been lower than the amplification (i.e. is jDNA is parasitic)?

As previously pointed out (1), by its bare presence in the genome, jDNA has an effect on cellular physiology (e.g. nucleotide metabolism; division rate), structure (e.g. nuclear and cellular size), and genome 'fluidity' (e.g. increased recombination versatility and evolutionary co-option of jDNA). Also, by its bare presence in the genome, jDNA gets replicated and it can, for example, undergo transposition and transcription, or it becomes non-specific target for diverse DNA binding proteins, as shown in the ENCODE project. However, these features and correlates do not solve C-value paradox and 'junk DNA' enigma; on the contrary, they confuse these issues.

As argued in the original hypothesis (1) and in the comments I posted here, by serving as a defense mechanism against insertional mutagenesis, which in humans and many other multicellular species leads to cancer, all jDNA is functional. Expectedly, as an adaptive defense mechanism, the amount of protective DNA varies from one species to another based on the insertional mutagenesis activity and evolutionary constrains on genome size.

1. Bandea CI. A protective function for noncoding, or secondary DNA. *Med. Hypoth*., 31:33-4. 1990.