**Student: Marta Sáez Moreno**

**Teacher: Virginia Zamora Gómez**

**School: IES Antonio Gala**

**Question 2: How can you explain human complexity when we have so few protein coding genes, e.g. about 5,000 less than a cucumber?**

A genome is an organism's complete set of DNA which contains the genetic information needed to develop and control the functioning of every organism (1). DNA molecules have twisted strands which are made up of nucleotide bases. They are adenine (A), thymine (T), guanine (G) and cytosine (C) (2). The human genome has around 3 billion of these pairs; the 23 pairs of chromosomes have about 50,000,000 to 300,000,000 base pairs which carry the instructions for making proteins (1).

Before our DNA was sequenced, people thought we had 100,000 genes but on April 14, 2003, after the successful human completion of the genome project, it was revealed that humans have approximately 20,000 to 25,000 genes (3). Furthermore, the Human Genome Project Information page said, in 2004, that "this lower estimate came as a shock to many scientists because counting genes was viewed as a way of quantifying genetic complexity" (4). In fact *E. coli*, a simple bacterium, has about 4,300 genes; *C. elegans*, a little worm, has 19,000 genes; *Arabadopsis thaliana* has 28,000 genes and *Oryza sativa*, rice plant, upwards of 50,000 genes (4). The discovery of these species number of genes did no longer sustain the old believe, that stated complexity and number of genes were totally related to each other [adding, MS].

This apparent paradox was resolved after developing a reasonable understanding of the non-coding elements of the genome (4) versus the coding sequences [addition; MS]. On one hand "a protein-coding gene consists of a promoter followed by the coding sequence, the protein and then a terminator". The promoter is a base-pair sequence that specifies where transcription begins. The terminator is a sequence that specifies the end of the mRNA transcript. The coding sequence contains the information for the polypeptide chain specified by the gene (5) so that it is the DNA piece that will be translated into a protein with an ultimate specific cellular function [addition; MS].

On the other hand non-protein-coding genes or ncRNAs (non-coding RNAs) are portions of DNA, within the gene sequence, that are not translated into a protein but still play a very important role in protein translation. These genes are important for their potential to regulate gene expression (6). "The potential importance of ncRNAs is suggested by the observation that the complexity of an organism is highly correlated with its number of ncRNA genes" and that in the human genome there is only a 2 or 3% of genetic sequences that will become proteins (6).

A gene can have many base pairs but not all of them will turn into a protein. In fact these protein-coding genes are known as exons, because they will be translated into molecules with a biological function. The non-protein coding sequences are inserted in portions of DNA called introns and will develop a varied range of RNA molecules that will control gene expression and, consequently, the synthesis of proteins (3) "Both exons and introns are transcribed to yield a long primary RNA transcript. Introns are then removed by splicing to form the mature mRNA" (7). This mRNA is the one leading to proteins while introns will develop a varied range of ncRNA [adding; MS].

To sum up, I have to point out that, nowadays, it is widely acknowledged that non-coding sequences are the elements of our genome making the difference to explain complexity (3). Reaching this conclusion could have never been possible if our entire genome had not been sequenced along with other species to compare them all [adding; MS]. The scientific community doesn´t have any reason to correlate gene number and complexity (3) and has a new research field: to discover what ncRNAs are responsible for within gene expression control [adding; MS].

References list

(1) National Human Genome Research Institute. Frequently Asked Questions. http://www.genome.gov/11006943

(2) Echevarría. C. et al. 2012. Biology and Geology. 4 Secondary Education. Anaya publisher. Madrid. Spain.

(3) The Tech Museum of Innovation. Stanford University. 2013. The Human Genome is More and Less Than We Expected to Find.

The Human Genome is More and Less Than We Expected to Find | Understanding Genetics

(4) Ryan Gregory T. 2007. Human gene number: surprising (at first) but not paradoxical. Genomicron. http://www.genomicron.evolverzone.com/2007/05/human-gene-number-surprising-at-first/

(5) The Biology place. Pearson Education.

http://www.phschool.com/science/biology_place/biocoach/transcription/procodgn.html

(6) Perkins, DO. et al. 2005. Expanding the 'central dogma': the regulatory role of non-protein-coding genes and implications for the genetic liability to schizophrenia.

http://www.ncbi.nlm.nih.gov/pubmed/15381925

(7) Cooper. GM. 2000. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; "The Complexity of Eukaryotic Genomes". The Complexity of Eukaryotic Genomes - The Cell - NCBI Bookshelf