Implications of non-coding RNA on biology and evolution: from first hominids to modern humans - Review

Gabriel Dorado ¹, Fernando Luque ², Plácido Pascual ³, Inmaculada Jiménez ⁴, Francisco Javier S. Sánchez-Cañete ⁵, Patricia Raya ⁶, Jesús Sáiz ⁷, Adela Sánchez ⁷, Teresa E. Rosales ⁸, Víctor F. Vásquez ⁹, Pilar Hernández ¹⁰

Abstract

Massive genomic/transcriptomic sequencing has revealed a shocking paradox: pervasive or spurious transcription. Although such event is unwanted in principle, some of such transcripts may scape degradation, being further selected by evolution, with fascinating consequences on biology, including our brain development and what made us humans. Indeed, non-coding RNA are involved in many regulatory processes, across the central dogma of molecular biology, and even epigenetics events. Interestingly, that is partially accomplished regulating the expression and function of small RNA, like miRNA. More strikingly, non-coding RNA are involved in neuron physiology and brain neurogenesis, including outgrowth or neuron projections, synaptic functions and translation in synapses. Besides, noncoding RNA can be exported-imported between cells, through exosome vesicles. Surprisingly, some non-coding RNA are indeed translated into micropeptides, which may be involved in brain development. All that allows the remarkable cognitive power of the human brain. Unfortunately, this exquisite development, that made us humans, is specially prone to internal and external perturbations. They may generate neurodevelopmental, neurodegenerative and neuropsychiatric disorders, to which humans are more prone than other primates.

Keywords: second-generation sequencing, SGS, third-generation sequencing, TGS, next-generation sequencing, NGS, ancient RNA, aRNA.

Resumen

La secuenciación genómica/transcriptómica masiva ha revelado una paradoja impactante: la transcripción generalizada o espuria. Aunque tal evento es no deseado en principio, algunos de tales transcritos pueden escapar de la degradación, siendo seleccionados por la evolución, con consecuencias fascinantes en biología, incluido el desarrollo de nuestro cerebro y lo que nos hizo humanos. De hecho, los ARN no codificantes están involucrados en muchos procesos reguladores, a través de todo el dogma central de la biología molecular, e incluso en eventos epigenéticos. Curiosamente, ello se logra parcialmente regulando la expresión y la función de ARN pequeños, como los miARN. Más sorprendentemente, los ARN no codificantes están involucrados en la fisiología de las neuronas y la neurogénesis cerebral, incluyendo las excrecencias o proyecciones neuronales, funciones sinápticas y traducción en sinapsis. Además, el ARN no codificante puede exportarse-importarse entre células, a través de vesículas de exosomas. Sorprendentemente, algunos ARN no codificantes se traducen en micropéptidos, que pueden estar involucrados en el desarrollo del cerebro. Todo eso permite el notable poder cognitivo del cerebro humano. Desafortunadamente, este desarrollo exquisito, que nos hizo humanos, es especialmente propenso a perturbaciones internas y externas. Así, pueden generarse trastornos del neurodesarrollo, neurodegenerativos y neuropsiguiátricos, a los cuales los humanos somos más propensos que otros primates.

Palabras clave: secuenciación de segunda generación, SSG, secuenciación de tercera generación, STG, secuenciación de próxima generación, SPG, ARN antiguo, ARNa.

Introduction

We have recently reviewed the fascinating topic of what made us humans; the evolution from first hominids to modern humans (Dorado et al, 2018), within the interesting interaction of bioarchaeology and molecular biology (Dorado et al, 2007-2019). In short, duplication, repair and conversion of Notch Homolog 2 (*NOTCH2*) genes into Notch Homolog 2 N-terminal-Like (*NOTCH2NL*) ones were involved in such remarkable transformation (Fiddes et al, 2018; Suzuki et al, 2018). The consequence was the expansion of the brain cortex. Unfortunately, that was also related to recurrent neurodevelopmental diseases, to which humans are specially prone, when compared to other animals.

But there is more. It has been also found that other biological changes during organic evolution in the planet Earth may have also contributed to make us humans. That involves the central dogma of molecular biology: DNA makes RNA that makes proteins, albeit –surprisingly– not as initially thought or conceived, as explained below.

Historical evolution of the concept of gene

Gregor Johann Mendel (1865) discovered the inheritance laws of genetic factors generating the phenotype. But, what is the chemical basis for such

remarkable phenomenon? Friedrich Miescher (1871) found that there was a specially acidic compound inside the nucleus (nuclein) of the lymphocytes found in sputa from tuberculosis patients. But, did it correspond to the Mendel's inheritance factors? Frederick Griffith (1928) found that it was possible to transform "rough" (R) harmless pneumococcus into "smooth" (S) virulent ones that killed mice.

At the time, it was thought that such transforming principle should be made of proteins, since they are rich in conformations and functions. Yet, Oswald Theodore Avery, Colin Munro MacLeod and Maclyn McCarty (1944) found something unexpectedly shocking: the transforming molecule was DNA, which, at the time, was considered a rather useless repetitive compound that, therefore, could not contain genetic information. They did not know that it is possible to "construct the world" with just two variants, as computers do working in binary mode (0 & 1). Furthermore, Alfred Day Hershey and his technician Martha Cowles Chase (1952) confirmed such hypothesis, making it a validated theory.

The next step was to elucidate the 3D structure of the DNA. Erwin Chargaff (1950) found that the number of purines (A + G) equaled the amount of pyrimidines (C + T), albeit the reason for that was a mystery at the time. Then came Francis Crick and James D. Watson which, using unpublished X-ray diffraction results of DNA of Rosalind Elsie Franklin, made by her PhD student Raymond George Gosling (1953), proposed the double helix structure for such molecule.

On the other hand, the concept of gene has changed over the time (Figure 1). Charles Robert Darwin and his friend and colleague Alfred Russel Wallace (1858) proposed a revolutionary hypothesis: the origin of species by natural selection. The term "gene" was coined by Wilhelm Johannsen (1905).

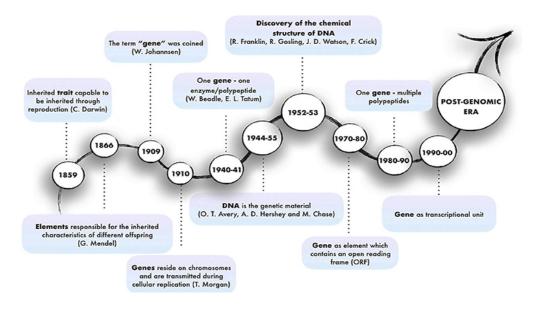


Figure 1.- Historical evolution of the concept of gene. One gene was considered the genetic material encoding one enzyme; and then, a transcriptional unit. But later on, massive genome/transcriptome sequencing brought the paradox of spurious RNA transcription, as described below. © Frontiers Media (Cipriano and Ballarino, 2018).

Thomas Hunt Morgan (1910) found that the genes are within the chromosomes, being duplicated and transmitted during cellular replication. George Wells Beadle and Edward Lawrie Tatum (1941) proposed the popular one geneone enzyme hypothesis. Later on, the gene was associated to one or several Open Reading Frames (ORF), encoding one or several polypeptides, respectively, and eventually, it was considered a transcriptional unit (Cipriano and Ballarino, 2018). But all that knowledge has been shaken with the discovery of pervasive or spurious transcription, as described below.

The paradox of the spurious RNA transcription

It was initially thought that transcription only took place for: i) protein-encoding genes, copied into messenger RiboNucleic Acid (mRNA); ii) ribosomal DNA (rDNA), generating ribosomal RNA (rRNA); and iii) transfer DNA (tDNA), producing transfer RNA (tRNA). But the latest high-performance sequencing platforms that we have reviewed (Dorado et al, 2007, 2008, 2013, 2015, 2016) have revealed surprising facts (Mattick, 2011, 2012; Gomes et al, 2019): i) ~1% of the human genome generates protein-coding transcripts (~20,000 genes; similar to nematodes, which have ~1,000 cells); and yet ii) at least 93% of the genome (maybe all of it) undergoes pervasive or spurious transcription (Figure 2).

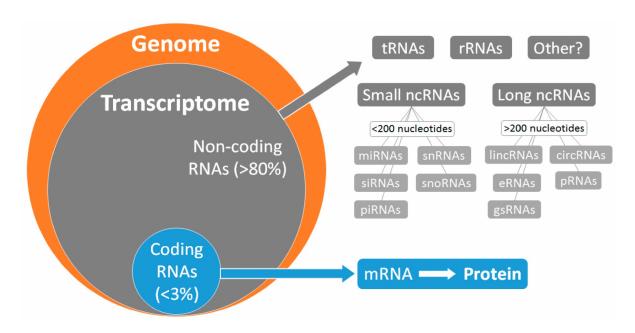


Figure 2.- Generalized transcription. Most transcription corresponds to non-coding RNA. © MDPI (Gomes et al, 2019).

There is also a relationship between coding and noncoding RNA across the central dogma of molecular biology (Guennewig and Cooper, 2014), as shown below (Figure 3).

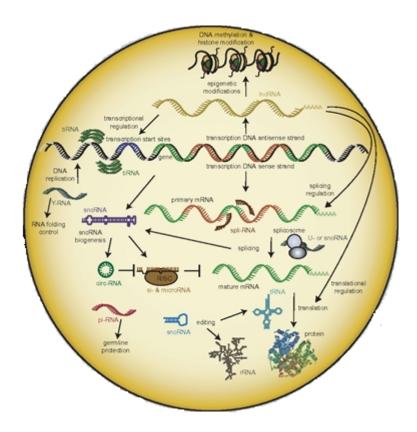


Figure 3. Transcription of coding and non-coding RNA. Recent research has shown interesting implications of some non-coding RNA, in relation to coding ones. © Elsevier (Guennewig and Cooper, 2014).

Yet, such generalized transcription represents a huge energetic load for cells, which have been selected through evolution over 3.8 milliard years of biological evolution, to optimize resources as much as possible. So, what is driving such "unwanted" transcription? Simply, the cellular regulation mechanisms are not 100% efficient, due to the probabilistic nature of molecular interactions. Paradoxically, that allows life. In other words, should the molecular events be 100% precise, controlled, regulated and efficient, life would not exist. Not even the Universe would exist if anomalies would not exist at the quantum level and beyond. Uncertainties and chaos have allowed the existence of everything, including life. Then, chance and necessity have shaped the biodiversity, through the struggle for limited resources, survival of the fittest, reproductive selection and evolution of species, as we have reviewed (Dorado et al, 2018).

Therefore, since control over cellular processes with 100% efficiency is not possible, pervasive or spurious transcription arises, representing a significant energetic load for the cell. Both prokaryotes and eukaryotes (including their mitochondria and chloroplasts) try to degrade the "unwanted" transcribed RNA, using RiboNucleases (RNases) and exosome complexes (not to be confused with exosome vesicles, as indicated below), respectively. Indeed, the uncontrolled pervasive or spurious transcription may cause stop of growth and death in prokaryotes, as well as diseases like cancer in eukaryotes. But, again, no cellular process is 100% efficient. As a consequence, some of such spurious RNA may

scape their scheduled degradation, being selected by evolution, with fascinating results. Indeed, non-coding RNA (ncRNA) can interact with DNA, RNA and peptides (oligopeptides and polypeptides like proteins). Such pleiotropic versatility allows a sophisticated and fine-tuned level of gene expression regulation, across the central dogma of molecular biology, as shown below.

That is not surprising when taking into account that RNA not only carries genetic information, but behaves as an enzyme; that is, has catalytic activity. Indeed, it is thought that life began on planet Earth as an RNA world (Darnell and Doolittle 1986). Thus, the involvement of non-coding RNA in multicellularity and cellular differentiation along the biological evolution (Hart and Goff, 2016). In other words, the ancient small RNA (sRNA) system could have originated as a transcriptional and post-transcriptional supervisor. But it further evolved, eventually becoming controlled by non-coding RNA working as a new layer of regulation (Barry, 2014; Guennewig and Cooper, 2014).

Implications of spurious RNA transcription in brain development

The coding transcriptome –and thus, the proteome– has remained largely constant through evolution, when compared to the non-coding transcriptome (Barry, 2014). Indeed, there is a striking positive correlation between the organic complexity and the non-coding RNA abundance and diversity (Barry, 2014; Guennewig and Cooper, 2014). Such non-coding RNA expansion encompasses both small and long families (which also include circular RNA). Interestingly, such positive correlation is found also in relation to the brain size and –more significantly– the cognitive evolution (Figure 4). Indeed, prokaryotes have a high percentage of coding genes (up to 99.5%) in their genomes, being significantly reduced in protists (10 to 75%) and metazoans (animals; 1 to 27%), and specially in primates (1 to 2%).

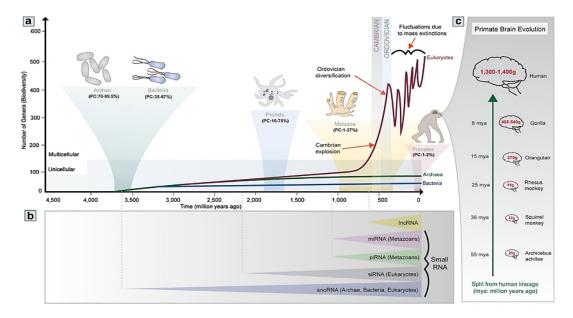


Figure 4.- Non-coding RNA and organic complexity. The graph shows the interesting positive correlation between the increase of non-coding RNA and organic complexity, with special relevance for brain size and cognitive evolution. © Springer & Elsevier (Barry, 2014; Guennewig and Cooper, 2014).

The non-coding RNA are involved in many regulatory events, including epigenetic ones (Figure 5), like chromatin modification and remodeling, splicing, RNA editing, transcription and translation (Zimmer-Bensch, 2019).

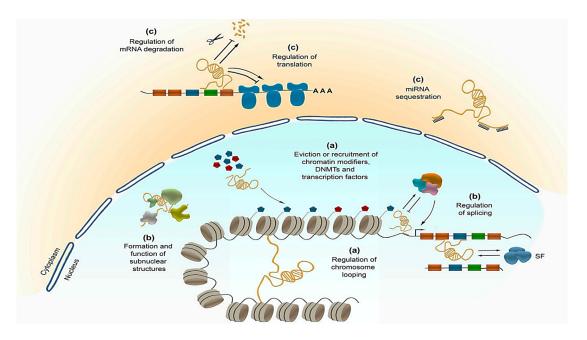


Figure 5.- Regulation of the central dogma of molecular biology by non-coding RNA. Such functionality involves transcriptional (a) and post-transcriptional (b) events in the nucleus, as well as translational and post-translational ones in the cytoplasm (c). © MDPI (Zimmer-Bensch, 2019).

Interestingly, such control of gene expression is partly mediated regulating the expression and function of evolutionary conserved effector sRNA, like microRNA (miRNA) (Barry, 2014), as shown below (Figure 6).

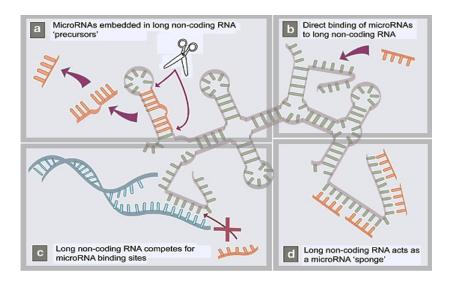


Figure 6.- Interactions of different non-coding RNA involved in regulation of gene expression. Long non-coding RNA (IncRNA) may: i) generate (a) and ii) bind some (b) or iii) many (d) miRNA, behaving as "sponges" regulating their activity, as well as iv) compete with them on binding sites (c). © Springer (Barry, 2014).

As indicated above, it has been found that non-coding RNA are involved in brain development, including growth of projections from the cell body of neurons (known as neurites or neuronal processes) and synapses (Figure 7). Reports about that include: i) enhancer-associated lncRNA (eRNA); ii) transcriptional and post-transcriptional modulations after depolarization, involved in neuronal plasticity; and iii) repression of translation of mRNA into peptides in synapses (Zimmer-Bensch, 2019).

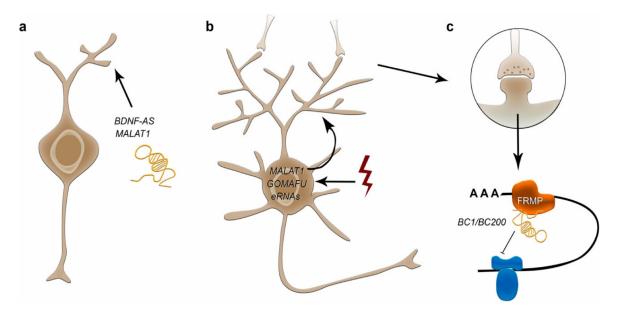


Figure 7.- Involvement of non-coding RNA in neuron physiology. Different activities have been described, like outgrowth or neuron projections **(a)**, synaptic functions **(b)** and translation in synapses **(c)**. © MDPI (Zimmer-Bensch, 2019).

Therefore, non-coding RNA activities modulate the brain neurogenesis, as shown below (Figure 8). Neurons, intermediate progenitors and basal radial glia cells are produced by stem cells (known as radial glia), located in the Ventricular Zone (VZ). Interestingly, the SubVentricular Zone (SVZ) of mice is significantly expanded into inner (iSVZ) and outer (oSVZ) zones in humans. Post-mitotic neurons migrate into the Cortical Plate (CP). Not surprisingly, the human cerebral cortex is highly folded, greatly increasing its area. Different non-coding RNA are involved in the following events of the neural progenitor cells: i) drive differentiation; ii) regulate the self-renewal versus differentiation balance; and iii) control the differentiation of basal delaminating cells, through turnover regulation.

It must be taken into account that the non-coding RNA can be exported-imported between cells, through exosome vesicles; not to be confused with exosome complexes, as indicated above (Thery, 2011). Such intercellular communication is a powerful mechanism enhancing post-synaptic properties, allowing the particular cognitive power of the human brain (Zimmer-Bensch, 2019).

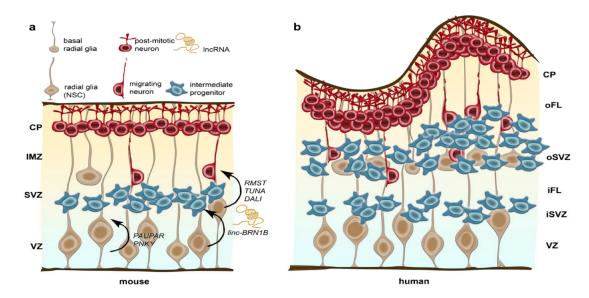


Figure 8.- Involvement of non-coding RNA in brain neurogenesis. Examples of mouse **(a)** and human **(b)** cerebral cortex are shown. © MDPI (Zimmer-Bensch, 2019).

Unfortunately, such exquisite sophistication and complexity that made us humans is also prone to a plethora of internal and external perturbations. Sometimes, they can trigger further useful evolutionary changes, as we have recently reviewed in relation to the evolution from first hominids to modern humans (Dorado et al, 2018). But, most likely –as happens with mutations— they are deleterious, generating neurodevelopmental, neurodegenerative and neuropsychiatric disorders, like autism and schizophrenia. All that highlights the unique human brain complexity and fragility (Mattick, 2011, 2012; Barry, 2014; Guennewig and Cooper, 2014; Zimmer-Bensch, 2019).

Concluding remarks and future prospects

It is considered that the human brain contains more than 85 milliard neurons, interconnected with 10-fold more synapses (Guennewig and Cooper, 2014). Yet, it seems that such complexity alone did not made us humans. The nucleic-acid sequencing advancements in the last years, to which we have contributed (Lario et al, 1997), are uncovering a mind-blowing scenario. It is now considered that noncoding RNA has played a central role in the development of organismal complexity, in general, as well brain size and -most significantly- cognitive evolution that made us humans, in particular. Such new knowledge is challenging previous conceptions about the biological evolution responsible for the human origin. It seems now that epigenetics have played, and are playing, a relevant role in gene expression regulation, in general, and the brain, in particular. Some have even proposed it as a mechanism contributing to a plastic and dynamic inheritance (Mattick, 2011, 2012). The future possibility to directly sequence ancient RNA (aRNA), that we have reviewed (Dorado et al, 2016), takes special relevance in this scenario. Even more, although the so-called non-coding RNA were initially thought as not encoding peptides (as their name implies), it has been recently found that some of them really encode functional short peptides (micropeptides). More importantly, some of them

are involved in brain development (Zimmer-Bensch, 2019). Future research perspectives on these areas are fascinating!

Acknowledgements. Supported by "Ministerio de Economía y Competitividad" (MINECO grant BIO2015-64737-R) and "Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria" (MINECO and INIA RF2012-00002-C02-02); "Consejería de Agricultura y Pesca" (041/C/2007, 75/C/2009 and 56/C/2010), "Consejería de Economía, Innovación y Ciencia" (P11-AGR-7322 and P12-AGR-0482) and "Grupo PAI" (AGR-248) of "Junta de Andalucía"; and "Universidad de Córdoba" ("Ayuda a Grupos"), Spain.

References

- Barry G (2014): Integrating the roles of long and small non-coding RNA in brain function and disease. *Mol Psychiatry* 19:410-416.
- Cipriano A, Ballarino M (2018): The ever-evolving concept of the gene: the use of RNA/protein experimental techniques to understand genome functions. *Front Mol Biosci* 5: 20 (16 pp).
- Darnell JE, Doolittle WF (1986): Speculations on the early course of evolution. *Proc Natl Acad Sci USA* 83: 1271-1275.
- Dorado G, Jiménez I, Rey I, Sánchez-Cañete FJS, Luque F, Morales A, Gálvez M, Sáiz J, Sánchez A, Rosales TE, Vásquez VF, Hernández P (2013): Genomics and proteomics in bioarchaeology Review. *Archaeobios* 7: 47-63.
- Dorado G, Luque F, Pascual P, Jiménez I, Sánchez-Cañete FJS, Pérez-Jiménez M, Raya P, Gálvez M, Sáiz J, Sánchez A, Rosales TE, Vásquez VF, Hernández P (2015): Second-generation nucleic-acid sequencing and bioarchaeology Review. *Archaeobios* 9: 216-230.
- Dorado G, Luque F, Pascual P, Jiménez I, Sánchez-Cañete FJS, Pérez-Jiménez M, Raya P, Sáiz J, Sánchez A, Martín J, Rosales TE, Vásquez VF, Hernández P (2016): Sequencing ancient RNA in bioarchaeology Review. *Archaeobios* 10: 103-111.
- Dorado G, Luque F, Pascual P, Jiménez I, Sánchez-Cañete FJS, Raya P, Sáiz J, Sánchez A, Rosales TE, Vásquez VF (2017): Clustered Regularly-Interspaced Short-Palindromic Repeats (CRISPR) in bioarchaeology Review. *Archaeobios* 11: 179-188.
- Dorado G, Luque F, Pascual P, Jiménez I, Sánchez-Cañete FJS, Raya P, Sáiz J, Sánchez A, Rosales TE, Vásquez VF, Hernández P (2018): Evolution from first hominids to modern humans: philosophy, bioarchaeology and biology Review. *Archaeobios* 12: 69-82
- Dorado G, Luque F, Pascual P, Jiménez I, Sánchez-Cañete FJS, Raya P, Sáiz J, Sánchez A, Rosales TE, Vásquez VF, Hernández P (2019): Bioarchaeology to bring back scents from extinct plants Review. *Archaeobios* 13: 66-75.

- Dorado G, Rey I, Rosales TE, Sánchez-Cañete FJS, Luque F, Jiménez I, Gálvez M, Sáiz J, Sánchez A, Vásquez VF (2009): Ancient DNA to decipher the domestication of dog Review. *Archaeobios* 3: 127-132.
- Dorado G, Rey I, Rosales TE, Sánchez-Cañete FJS, Luque F, Jiménez I, Morales A, Gálvez M, Sáiz J, Sánchez A, Hernández P, Vásquez VF (2010): Biological mass extinctions on planet Earth Review. *Archaeobios* 4: 53-64.
- Dorado G, Rosales TE, Luque F, Sánchez-Cañete FJS, Rey I, Jiménez I, Morales A, Gálvez M, Sáiz J, Sánchez A, Vásquez VF, Hernández P (2011): Ancient nucleic acids from maize Review. *Archaeobios* 5: 21-28.
- Dorado G, Rosales TE, Luque F, Sánchez-Cañete FJS, Rey I, Jiménez I, Morales A, Gálvez M, Sáiz J, Sánchez A, Vásquez VF, Hernández P (2012): Isotopes in bioarchaeology Review. *Archaeobios* 6: 79-91
- Dorado G, Sánchez-Cañete FJS, Pascual P, Jiménez I, Luque F, Pérez-Jiménez M, Raya P, Gálvez M, Sáiz J, Sánchez A, Rosales TE, Vásquez VF, Hernández P (2014): Starch genomics and bioarchaeology Review. *Archaeobios* 8: 41-50.
- Dorado G, Vásquez V, Rey I, Luque F, Jiménez I, Morales A, Gálvez M, Sáiz J, Sánchez A, Hernández P (2008): Sequencing ancient and modern genomes Review. *Archaeobios* 2: 75-80.
- Dorado G, Vásquez V, Rey I, Vega JL (2007): Archaeology meets Molecular Biology Review. *Archaeobios* 1: 1-2.
- Fiddes IT, Lodewijk GA, Mooring M, Bosworth CM, Ewing AD, Mantalas GL, Novak AM, VanDenBout A, Bishara A, Rosenkrantz JL, Lorig-Roach R, Field AR, Haeussler M, Russo L, Bhaduri A, Nowakowski TJ, Pollen AA, Dougherty ML, Nuttle X, Addor MC, Zwolinski S, Katzman S, Kriegstein A, Eichler EE, Salama SR, Jacobs FMJ, Haussler D (2018): Human-specific *NOTCH2NL* genes affect Notch signaling and cortical neurogenesis. *Cell* 173: 1356-1369.
- Gomes CPC, Agg B, Andova A, Arslan S, Baker A, Barteková M, Beis D, Betsou F, Bezzina-Wettinger S, Bugarski B, Condorelli G, da Silva GJJ, Danilin S, de Gonzalo-Calvo D, Buil A, Carmo-Fonseca M, Enguita FJ, Felekkis K, Ferdinandy P, Gyongyosi M, Hackl M, Karaduzovic-Hadziabdic K, Hellemans J, Heymans S, Hlavackova M, Hoydal MA, Jankovic A, Jusic A, Kardassis D, Kerkela R, Kuster GM, Lakkisto P, Leszek P, Lustrek M, Maegdefessel L, Martelli F, Novella S, O'Brien T, Papaneophytou C, Pedrazzini T, Pinet F, Popescu O, Potocnjak I, Robinson E, Sasson S, Scholz M, Simionescu M, Stoll M, Varga ZV, Vinciguerra M, Xuereb A, Yilmaz MB, Emanueli C, Devaux Y (2019): Catalyzing transcriptomics research in cardiovascular disease: The CardioRNA COST Action CA17129. *Noncoding RNA* 5: 31 (13 pp).
- Guennewig B, Cooper AA (2014): The central role of noncoding RNA in the brain. *Int Rev Neurobiol* 116: 153-194.

- Hart RP, Goff LA (2016): Long noncoding RNAs: central to nervous system development. *Int J Dev Neurosci* 55:109-116.
- Lario A, González A, Dorado G (1997): Automated laser-induced fluorescence DNA sequencing: equalizing signal-to-noise ratios significantly enhances overall performance. *Analytical Biochemistry* 247: 30-33.
- Mattick JS (2011): The central role of RNA in human development and cognition. *FEBS Lett* 585: 1600-1616.
- Mattick JS (2012): Rocking the foundations of molecular genetics. *Proc Natl Acad Sci USA* 109: 16400-16401.
- Suzuki IK, Gacquer D, VanHeurck R, Kumar D, Wojno M, Bilheu A, Herpoel A, Lambert N, Cheron J, Polleux F, Detours V, Vanderhaeghen P (2018): Human-specific *NOTCH2NL* genes expand cortical neurogenesis through Delta/Notch Regulation. *Cell* 173: 1370-1384.
- Thery C (2011): Exosomes: secreted vesicles and intercellular communications. *F1000 Biol Rep* 3: 15 (8 pp).
- Zimmer-Bensch G (2019): Emerging roles of long non-coding RNAs as drivers of brain evolution. *Cells* 8: 1399 (24 pp).

