# Unveiling human non-random genome editing mechanisms activated in response to chronic environmental changes.

# I. Where might these mechanisms come from?

**Loris Zamai**

Department of Biomolecular Sciences, University of Urbino Carlo Bo; Urbino;

National Institute for Nuclear Physics (INFN)-Gran Sasso National Laboratory (LNGS), Assergi, L'Aquila, Italy.

**Corresponding author:** Loris Zamai, Department of Biomolecular Sciences, via Ca' le Suore 2, University of Urbino Carlo Bo, 61029 Urbino, Italy.

Tel. (+39) 0722 304319; fax (+39) 0722 304319; e-mail: loris.zamai@uniurb.it

**Running Headline**: Non-random genome editing mechanisms

**Key words:** environment, virus, pollutant, evolution, exaptation, stem cells, transposons, LINE-1, Alu, mutagenic enzymes, APOBEC, ADAR, ORF2p, cancer, Eco-Evo-Devo, symbiosis, ecological genomics, environmental stress, genetic recombination, biological plasticity, hypermutation, epigenetics, fractal systems, natural selection.

**List of Abbreviations**

single nucleotide polymorphisms, (SNP); copy number variations, (CNV); stress induced mutagenesis, (SIM); ecological evolutionary developmental biology, (Eco-Evo-Devo); somatic hypermutation, (SHM); immunoglobulin, (Ig); activation-induced deaminase (AID); CRISPR, (clustered, regularly interspaced, short, palindromic repeats); Cas, (CRISPR-associated); CRISPR RNAs, (crRNAs); Cascade, (CRISPR-associated complex for antiviral defense); apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like, (APOBEC); heavy-chain class-switch recombination, (CSR); cytosine, (C); thymidine, (T); interferons, (IFNs); endogenous retroviruses, (ERVs); long-terminal repeats, (LTR); long interspersed nuclear elements, (LINEs); short interspersed nuclear elements, (SINEs); open reading frame 1 and 2 proteins, (ORF1p and ORF2p); single-stranded DNA, (ssDNA); Protein, (P); guanosine, G; adenosine, A; inosine, (I); adenosine deaminase enzymes acting on RNA (ADARs); switch, (S); androgen receptor, (AR); androgen response elements, (AREs); dihydrotestosterone (DHT).

**Summary**

This article challenges the notion of the randomness of mutations in eukaryotic cells by unveiling stress-induced human non-random genome editing mechanisms. To account for the existence of such mechanisms, I have developed molecular concepts of the cell environment and cell environmental stressors and, making use of a large quantity of published data, hypothesized the origin of some crucial biological leaps along the evolutionary path of life on Earth under the pressure of natural selection, in particular, 1) virus-cell mating as a primordial form of sexual recombination and symbiosis; 2) Lamarckian CRISPR-Cas systems; 3) eukaryotic gene development; 4) antiviral activity of retrotransposon-guided mutagenic enzymes and finally; 5) the exaptation of antiviral mutagenic mechanisms to stress-induced genome editing mechanisms directed at "hypertranscribed" endogenous genes. Genes transcribed at their maximum rate (hypertranscribed), yet still unable to meet new chronic environmental demands generated by "pollution", are inadequate and generate more and more intronic retrotransposon transcripts. In this scenario, RNA-guided mutagenic enzymes (e.g. AID/APOBECs), which have been shown to bind to retrotransposon RNA-repetitive sequences, would be surgically targeted by intronic retrotransposons on opened chromatin regions of the same "hypertranscribed" genes. RNA-guided mutagenic enzymes may therefore "Lamarkianly" generate single nucleotide polymorphisms (SNP) and copy number variations (CNV), as well as transposon transposition and chromosomal translocations in the restricted areas of hyperfunctional and inadequate genes, leaving intact the rest of the genome. CNV and SNP of hypertranscribed genes may allow cells to surgically explore a new fitness scenario, which increases their adaptability to stressful environmental conditions. Like the mechanisms of immunoglobulin somatic hypermutation, non-random genome editing mechanisms may generate several cell mutants, and those codifying for the most environmentally-adequate proteins would have a survival advantage and would therefore be Darwinianly selected. Non-random genome editing mechanisms represent a link between environmental changes and biological novelty and plasticity, and provide a molecular basis to reconcile gene-centered and "ecological" views of evolution.

## Contents

## 1. Waiting for beneficial random mutations within the vast eukaryotic cell genome: is it like "waiting for Godot"?

***1.1. Random mutations: a weak point in the modern synthesis of evolution.*** The modern synthesis of evolution is a gene-centered theory based on Darwinian natural selection and Mendelian genetics. Darwinian selection is a good mechanism for excluding inadequate genomes/organisms, and Mendelian genetics ensures allele inheritance from generation to generation, while limiting the appearance of new characteristics. These mechanisms are not effective in generating biological variability and novelties, which are believed to depend on a relatively constant rate of random mutations. However, random mutations (e.g. during DNA replication) may generate not only advantageous or neutral effects, but also detrimental ones, ultimately yielding many aborted alternatives that can only be managed by large highly proliferating populations such as viruses and bacteria. Viruses and bacteria can take advantage of random mutations not only because of their rapid proliferation, but also because of their low DNA content, which limits the number of random detrimental mutations per biological entity. Indeed, there is a strong inverse correlation between the mutation rate during DNA replication and genome size across all replication systems, suggesting that any increase in DNA content and complexity requires a reduction in deleterious mutations (**Holmes 2011**). Hence, the lowest error rates per nucleotide are reported for eukaryotes, while the highest mutation rates are observed in viroids, short single-stranded circular RNA without a protein coat, which were likely the first virus-like structures and survivors of the hypothetical "RNA world" stage in the evolutionary history of life on Earth (**Durzyńska and Goździcka-Józefiak 2015**). This evidence suggests that error rates have been progressively reduced over evolutionary time and biological complexity (**Holmes 2011**), raising the possibility that new evolutionary mechanisms, able to generate novelties, have been developed during the evolution of increasingly complex organisms. Indeed, bacteria, which have a larger genome (10 to 100 times larger) than viruses (**Holmes 2011**), have developed stress induced mutagenesis (SIM) mechanisms (**Galhardo et al. 2007; Fitzgerald et al. 2017**), which are induced by environmental stresses and represent a quasi-Lamarckian phenomenon (**Koonin and Wolf 2009**). The SIM increase of random genomic changes is mainly mediated by error-prone DNA duplicating and/or repair systems, which allow bacteria to accelerate mutation rates and therefore to enhance their ability to adapt to environmental challenges (**Galhardo et al. 2007; Fitzgerald et al. 2017**). Indeed, SIM, in which the probability of multimutation leaps may be non-negligible, has been hypothesized to be an evolvable adaptive strategy (**Katsnelson et al. 2019**). However, along the vast genome of "slow-proliferating" eukaryotic cells, it is very unlikely that random mutations occurring during DNA duplication and/or repair could produce mutations that are advantageous (or neutral) to adaptation without also producing detrimental ones, and therefore a myriad of unproductive eukaryotic attempts. This is likely the reason why they were selected with the lowest mutation rate per genome replication. It is therefore difficult to explain how only random mutations could have led to the sophisticated molecular entanglements of eukaryotic cells and to the huge and expanding number of species found on Earth today. Indeed, random mutations represent a weak point in the modern synthesis of evolution, and mathematicians have predicted the existence of non-random mutation mechanisms (see **Wright 2000**).

***1.2. The complex protein-protein steric interactions of eukaryotic cells originate from the translation of unrelated genes: the needs of gene editing feedback mechanisms.*** The complexity of the molecular signals that regulate the life of eukaryotic cells is well known. Interactions among proteins, from the cell surface to the cell nucleus, function on the basis of complex steric interactions regulated by weak inter- and intra-molecular bonds. Variable affinities among different inter- and intra-molecular domains generate a complex web of allosteric communications that makes the extraordinary "phenomenon" of eukaryotic cell life possible. Interestingly, the interaction affinities of cellular and extracellular molecules vary according to context. For

example, immunoglobulins usually bind to pathogens at high affinities, while TcR-MHC binding works at intermediate affinities, and sometimes even a loss of recognition can be advantageous, as in the case of CCR5-Delta32 mutation, which provides resistance to human immunodeficiency virus type I (**Agrawal et al. 2004**). How can these recognition affinities be adjusted to the needs of the cell and organism?

Importantly, unlike polycistronic prokaryotic genes, most of the sophisticated entanglements among eukaryotic cell molecules originate from the translation of unrelated genes, which do not "know" that their products will sterically interact with each other after translation in such a complex and functional fashion! This aspect cannot be accounted for by the induction of indiscriminate random mutations along the vast eukaryotic DNA molecule, pointing to the existence of gene editing feedback mechanisms able to surgically generate mutations on the genes of unfit proteins, leaving the rest of the genome intact. The death of cells bearing inadequate protein versions and the survival of cells bearing the fittest mutation(s) might then determine the "selection" of the inheritable version(s). Indeed, yet unknown genome manipulation mechanisms typically found in adaptive immunity have been presumed (see **Koonin and Krupovic 2015**).

***1.3. Environmental changes drive organisms to evolve rapidly: the evolvability trait was likely selected early on in the evolution of life.*** In contrast to lifeless planets that have relatively stable environmental conditions, over the last four billion years, the biological components present on Earth have "stimulated" each other, producing enormous and progressive environmental changes (environmental plasticity), in addition to the changes brought about suddenly by meteorite impacts, earthquakes and volcanic eruptions (**Pope et al. 1998, Kerr 2013**). Soon after the expansion of life on Earth, biological structures were likely poisoned by their own metabolic waste products ("niche construction"), driving these structures to develop evolvability mechanisms that made it possible to plastically and rapidly adapt to and manage increasing concentrations of environmental pollutants of biological origin or suddenly appearing abiotic compounds. In Earth's changeable environment, the evolvability trait was likely a crucial characteristic for the survival of life, and it was probably selected early on as life expanded. Thus, it is likely that environmental pollutants pushed biological species to plastically change into new species through selected evolvability mechanisms, producing new biological pollutants that gradually generated new niches, and hence, new adaptations, in a continuous evolutionary process. In this "never-ending history", some biotic, but also abiotic pollutants (i.e. $O_2$ was a plant pollutant) became fundamental bricks for subsequent biological structures, and abiotic components progressively became biotic ones.

***1.4. The open questions of the "ecological" and punctuated equilibrium views of evolution.*** The need to incorporate ecology and development into modern evolutionary theory is highlighted by ecological evolutionary developmental biology (Eco-Evo-Devo, see **Gilbert et al. 2015**) and the extended evolutionary synthesis (see **Laland et al. 2014**). Indeed, these theories show how environmental changes contribute to shape the production of a range of phenotypes (called developmental plasticity) thus promoting biological adaptation. Subsequently, environmentally-induced phenotypes are selected by environmental conditions and ultimately "accommodated" into the genome, becoming inheritable traits (a process called genetic accommodation). Indeed, the dynamic interplay between environment and biological entities yields not only developmental plasticity (environmentally induced phenotypes), but also developmental symbiosis between organisms and microorganisms, leading to the generation of "microbial organs" and "holobionts" (see **Gilbert et al. 2015**). On the other hand, developing phenotypes affect their niches (niche construction) driving additional biological variation. Notably, when developmental plasticity generates new metabolic pathways, new biological waste products are progressively accumulated. Such products may not be toxic for a long time until their increasing concentrations begin to interfere with biological activities, becoming toxic. However, toxic concentrations of waste products require a rapid biological adaptation, which can occur through the

generation of species able to manage the "new" waste products. Therefore, cyclic phases of rapid speciation (expression of developmental plasticity) followed by longer periods of species stability are expected. Indeed, based on fossil records, the theory of punctuated equilibrium proposes that biological evolution is not constant, but rather, is characterized by rapid accelerations with speciations (clades) followed by long periods of stasis (**Gould and Eldredge 1993**).

In line with this theory, two recent well-known events come to mind: the chromosomal speciation in mice observed in Seveso, Italy in the wake of a major industrial accident involving dioxin contamination (**Garagna et al. 1997**) and the rapid development of the black form of the British peppered moths in areas with heavy industrial pollution (**Arien et al. 2016**). It is important to note that the rapid biological evolution that occurred in these cases was associated with deep environmental changes, suggesting a cause-effect relationship. Of note, under stress conditions, the mutation rate is substantially elevated (e.g. under SIM) and the probability of multiple mutations and saltational evolution (a sudden multi-mutation leap potentially leading to speciation) becomes more likely to occur (**Katsnelson et al. 2019**).

However, it is difficult to imagine that random mutations and selection could have generated such rapid organism adaptations to these deep environmental changes. Unfortunately, the molecular mechanisms underlying the genetic variations of such environmentally induced traits are still largely unknown, and researchers are still working to unveil the molecular rules governing the interactions between the genetic adaptive traits of organisms and environmental niches. This is the missing link between ecological and gene-centered views of evolution.

***1.5. Why should eukaryotic organisms survive viruses?*** Unlike viruses and bacteria, eukaryotic cells in organisms possess a large DNA sequence and transmit their genes to their progeny relatively slowly. During evolution, rapidly evolving viruses have repeatedly threatened the survival of seemingly slowly evolving eukaryotic organisms. Despite these continuous viral challenges, organisms have not only been able to survive but also to expand until the present day.

It is well known that during viral infection, viruses inject their genetic material into host cell genomes. Viral gene translation then provides instructions for their rapid replication and the building of their components. Once mature viruses are formed, they break open the cell and move on to infect other cells. Moreover, error-prone virus replication rapidly generates a repertoire of viral variants that enhance the ability of viruses to adapt and elude cell and organism immune controls. So how is it possible that eukaryotic organisms did not succumb to these rapidly adapting microorganisms and actually managed to expand explosively? Have they developed potent mechanisms able to counter the rapid evolution of viruses "in real time"?

It is likely that complex eukaryotic organisms developed after prokaryotes and viruses, and therefore, in their presence. This means that eukaryotic cells were selected to be able to co-evolve with and manage rapidly proliferating and adapting microorganisms. Indeed, mitochondria and chloroplasts are examples of microorganisms which have not only been managed but also exploited by eukaryotic cells. In this regard, Eco-Evo-Devo teaches us that symbiosis between eukaryotic organisms and their persistent microorganisms seems to be a rule rather than an exception (see **Gilbert et al. 2015**).

However, it would appear that known mechanisms of adaptation are unable to explain the above-mentioned observations regarding eukaryotic organisms. I therefore hypothesize that eukaryotic cells may possess potent, yet to be discovered mechanisms of adaptation to respond to environmental challenges. Such mechanisms would be able to non-randomly choose a few specific genes in which to "Lamarckianly" generate mutations (new acquired characteristics), and subsequently, Darwinian selection would determine the survival and

inheritability of only the advantageous mutations. In this regard, there is a mechanism in the immune system that perfectly matches these characteristics: the somatic hypermutation (SHM) of immunoglobulins (Igs) that occurs in memory B cells (**McHeyzer-Williams et al. 2011**).

***1.6. Immunoglobulin somatic hypermutation as a mechanism model for eukaryotic cell evolution.*** This process starts when B cells produce antibodies with a low affinity for pathogenic antigens. They are inadequate to eradicate the pathogen and its chronic presence (environmental challenge) induces chronic transcription of Ig genes and proliferation of B cells bearing low affinity antibodies. Starting from a single B cell, this mechanism, mediated by the activation-induced deaminase (AID) mutagenic enzyme, generates a range of B cell mutants bearing slightly different Ig gene versions (genetic plasticity) coding for different immunoglobulin variants (phenotypic plasticity). Among the different cell mutants, those producing antibodies with the highest affinity for new and persistent pathogenic antigens are selected (**McHeyzer-Williams et al. 2011**), ultimately eradicating the pathogen from the host organism and increasing the chance of organism survival. Of note, this is a genome editing mechanism additional to those (four) already described by Koonin and colleagues (see **Koonin et al. 2019**).

Interestingly, this mechanism represents an example of "developmental plasticity", in which a single cell genome can generate different cell phenotypes as a function of intra-organismal "environmental" conditions (the pathogen presence). In this case, immunoglobulin gene "plasticity", activated in response to the chronic presence of pathogenic antigens, provides the raw material for the environmental selection of phenotypes, ultimately yielding new inheritable and "environmentally-independent" versions of Ig genes, what Waddington called the genetic assimilation of acquired characters (**Waddington 1959**).

## 2. Are viral nucleic acid insertions in the prokaryotic cell genome harmful or useful to prokaryotic cells? The example of CRISPR-cas systems.

***2.1. Selection of virus-cell mating pairs: a primordial form of "sexuality" and symbiosis.*** At a certain point after life appeared, among the myriad of primordial cells and viruses, it is likely that some of them started to bind to one another through surface molecule interactions, giving rise to specific viral tropism. Subsequently, some viruses acquired the ability to insert new genetic material into prokaryotic cells, similarly to how spermatozoa insert their genetic material into oocytes (see **Fig.1**). Indeed, viruses usually have a preferential prokaryotic or eukaryotic tissue target, a species-specific viral tropism and persistent interaction with their cell host, which represents a primordial form of mating and "sexual" recombination. However, an uncontrolled insertion of nucleic acid material produces cell gene disruptions, and lethally infected cells also stop spreading the virus. Among the myriad of aborted attempts, the pairs of cells and viruses that accidentally generated viral nucleic acid insertions with low cell toxicity had a better chance of survival, which gave them a selective advantage. A controlled horizontal insertion of new genetic material into the cell genome during viral infection could be a rich source of new genes and new nucleotide sequences for host cells. From this point of view, current specialized mechanisms of organismal sexual reproduction would represent selected (among myriads of attempts) improvements in primordial cell-virus interactions, producing nucleic acid recombination (variability/novelty) with a reduced toxicity. It is plausible that in a changeable environment, like that on Earth, the viral provision of new biological tools enhanced cell adaptability, giving infected cells a better chance of survival than uninfected ("sterile") cells. Indeed, these mechanisms may not only have conferred an advantage to cells, but also allowed viruses to survive within their host-cells, representing a sort of cell-virus agreement, a symbiosis, naturally selected among countless unproductive struggles. In line with this hypothesis is the evidence that symbiosis and co-evolution of organism hosts with their microbial communities are the rule and

not the exception (**Gilbert et al. 2015**). How eukaryotic organisms began to interact with a specific set of microorganisms, while avoiding dangerous microorganisms is unclear (**Gilbert et al. 2015**); however, a selective mechanism similar to the one described above (between bacteriophages and prokaryotic cells) can be hypothesized.

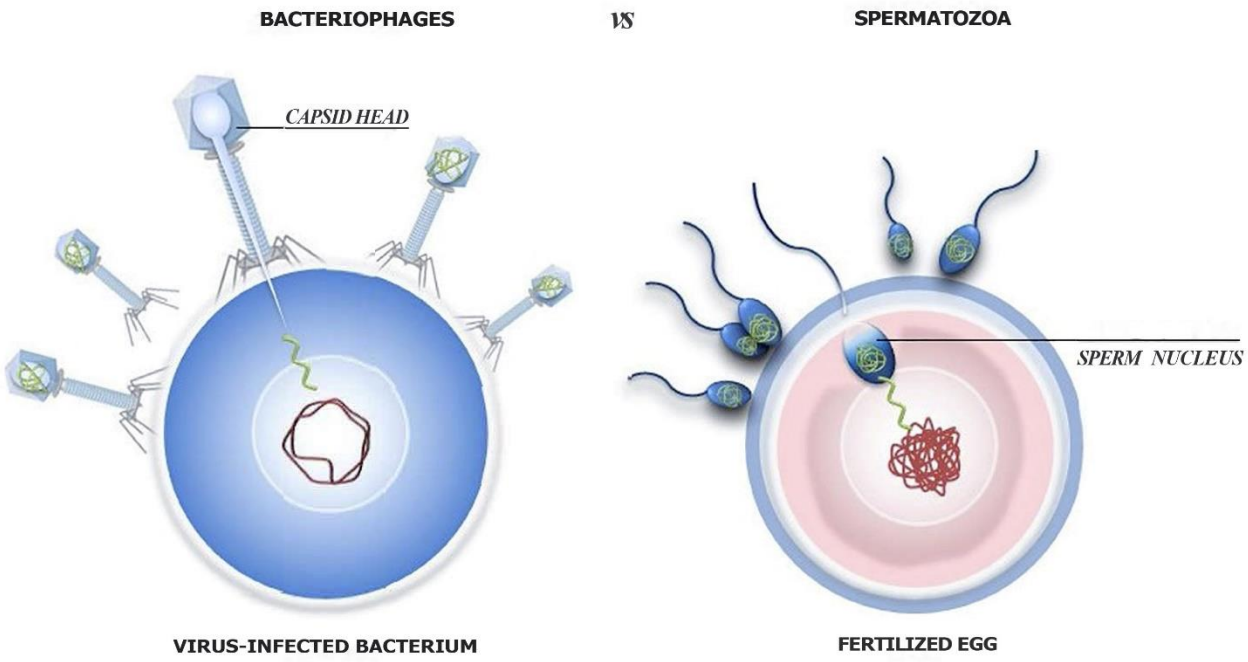


***Fig.1. Virus-cell mating pairs: a primordial form of "sexuality".*** Bacteriophages and male gametes have the ability to insert new genetic material (novelties) into prokaryotic cells and oocytes, respectively. (for reference, see the text).

***2.2. The meanings of virus fragments in Lamarckian CRISPR-cas systems.*** During the evolution of the interplay between prokaryotic cells and viruses, an important improvement in prokaryotes took place when they developed compartments specialized in storing viral nucleic acids, the CRISPR (clustered, regularly interspaced, short, palindromic repeats) - Cas (CRISPR-associated) systems. From a functional point of view, CRISPR-Cas systems constitute adaptive immune organelles segregated far from the cell genome, which have been discovered in most archaeal species and half of bacteria (**Wright et al. 2016, Koonin et al. 2017**). During the first bacteriophage infection, the foreign DNA is processed by cell restriction enzymes in viral fragments. Some of these fragments, called spacers, are integrated within the repeats of CRISPR arrays of prokaryotic cells. The spacer acquisition, also known as adaptation, allows the cells to store viral sequences and to retain a memory of the viruses, providing a heritable record of previous viral exposures. In the case of reinfection, the CRISPR array is transcribed into long CRISPR RNAs (pre-crRNAs) of alternating repeat–spacer sequences and then "spliced" into short CRISPR RNAs (crRNAs). The crRNAs form complexes with Cas endonucleases, which are encoded in an adjacent operon. Repeats are necessary for crRNA binding to Cas proteins, while spacers give the specificity to target Cas endonucleases against viral nucleic acids for their destruction (**Nishimasu et al. 2014**). Spacers contain a 20-nucleotide guide sequence able to bind to the re-entered foreign DNA in a complementary sequence-specific manner (**Nishimasu et al. 2014**), and the target specificity is likely assured by the uniqueness of the viral sequences, which are not expressed in the host cell genome. Regarding the control of CRISPR-Cas activation, evidence suggests that their transcription is elicited by cell envelope stresses, typically generated during viral infection (see **Ratner et al. 2015**). This link is functional because it allows both the integration of viral nucleic acid into CRISPR-Cas systems during the first virus invasion and the transcription of repeat-spacers and Cas endonucleases following virus rechallenge. Recently, it has also been observed that some Cas subunits form a complex with crRNA (called Cascade, CRISPR-associated complex for antiviral defense), which binds the promoter region of the Cas operon and functions as a transcriptional repressor, inhibiting Cas gene transcription through a negative feedback loop (**He et al. 2016**). Upon viral infection, Cascade complex redistribution towards the viral genome would relieve transcriptional repression, allowing a rapid transcriptional response to the reinfection (**He et al. 2016**).

From a structural standpoint, the CRISPR-Cas systems consist of two, usually adjacent components, the CRISPR locus, composed of alternating repeat–spacer non-coding sequences and a polycistronic gene, coding for a series of Cas proteins (**Wright et al. 2016, Koonin et al. 2017**). It should be noted that the CRISPR locus is both transcribed and "spliced", but not translated into proteins. CRISPR-Cas systems call to mind a primordial form of the "intron-exon" gene structure, typical of eukaryotic cells, in which the CRISPR locus represents the prototype of introns. The simple structure of CRISPR-Cas systems, formed by an operon and repetitive sequences, points to its viral origin. Indeed, the manner of the integration of new viral fragments is reminiscent of viral integrases and transposases (**Nishimasu et al. 2014**), and several findings suggest that prokaryotic Cas endonucleases were derived from the "domestication" of viral transposases (see **Jangam et al. 2017 and Koonin et al. 2019**). In line with these observations, bacteriophages with their own CRISPR-Cas systems have been described (**Seed et al. 2013**), raising the possibility of horizontal transpositions of CRISPR-Cas systems from viruses to prokaryotic cells and further pointing to the existence of symbioses between bacteriophages and prokaryotes. Interestingly, both horizontal gene transfers and CRISPR defense systems have been associated with a (quasi)Lamarckian mechanisms of inheritance (**Koonin and Wolf 2009, Haerter and Sneppen 2012**). Indeed, these types of acquired genes/spacers depend on the environment in which the bacteria grow. When challenged by viruses, bacteria non-randomly acquire appropriate DNA "novelties" from those viruses,

allowing them to either integrate viral tools or more quickly respond to virus rechallenge. Therefore, the transferred genes/spacers confer a selective advantage for the growth of the bacteria in that precise environment, exemplifying both the acquired characteristics of Lamarck and the environmentally-induced traits of "ecological" views of evolution.

Finally, viral nucleic acid insertion into prokaryotic cells is an effective cellular tool against viruses, immunizing cells against virus reinfections through a "boomerang response". Because the stable integration of viral non-coding and inactive sequences into the cell genome gives cells a strong selective advantage, these specialized systems to fragment and store copies of viral nucleic acid sequences likely supplanted the mere destruction of viral nucleic acids. Importantly, viral nucleic acid sequences are inserted in a "cell-safe" fashion, far from the cell genome, suggesting that they represent a cellular "strategy" rather than an ongoing threat posed by viruses.

## 3. Viral nucleic acid insertions from the segregated CRISPR-Cas sites of prokaryotes then spread into coding regions of the eukaryotic cell genome. Are they harmful or useful to eukaryotic cells? The hypothesis of retrotransposon-guided human APOBEC antiviral activity.

***3.1. Antiviral activity of the APOBEC family.*** In humans, members of RNA-guided apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) mutagenic enzymes have been shown to inactivate viruses, resembling RNA-guided Cas endonucleases (**Salter et al. 2016, Stavrou and Ross 2015**, **Zhang et al. 2010**). The APOBECs belong to the activation-induced cytidine deaminase gene family that originates from the ancestral activation-induced deaminase (AID) gene responsible for producing a highly diverse repertoire of antibodies through the somatic hypermutation (SHM) and heavy-chain class-switch recombination (CSR) of immunoglobulins (**McHeyzer-Williams et al. 2011**, **Salter et al. 2016, Senavirathne et al. 2015**). A cytidine deaminase domain is responsible for APOBEC cytosine to uracil base modification, finally leading to cytosine to thymidine (C to T) DNA editing (**Salter et al. 2016**). In humans, the APOBEC family consists of 11 different gene products, and some of them have been found to promote extensive (C to T) point mutations on the nascent DNA of some retroviruses during reverse transcription, ultimately leading to virus inactivation or in some cases, to immune escape variants (see **Salter et al. 2016, Stavrou and Ross 2015**). Their expression is upregulated by interferons (IFNs), one of the first/innate immune cell protein responses, and this feature links "innate" IFN production to APOBEC-mediated antiviral activity (**Salter et al. 2016, Stavrou and Ross 2015**). Interestingly, APOBEC proteins have a well-defined tissue-specific expression and therefore function (**Salter et al. 2016**), suggesting that they were developed to counter different tissue-specific viruses to protect distinct tissue cells from virus infections. Hence, they may represent the human counterpart of prokaryotic Cas endonucleases. However, APOBECs are active not only against viruses but also against the threat of some endogenous viral sequences, namely the retrotransposable elements (**Salter et al. 2016, Stavrou and Ross 2015, Chiu et al. 2006, Stenglein and Harris 2006, Bogerd et al. 2006, Muckenfuss et al. 2006, Kinomoto et al. 2007, Hulme et al. 2007, Wissing et al. 2011, Roy-Engel 2012, Horn et al. 2014, Feng et al. 2017**).

***3.2. Endogenous viral fossils: eukaryotic transposons*.** Transposable elements and transposon-like repetitive elements, originally defined as "junk molecules", due to their characteristic repetitive sequences, account for almost half of the human genome (**Mills et al. 2007**). The vast majority (about 98%) of human transposable elements are represented by retrotransposons, which move, duplicating from one genomic position to another, through an RNA copy intermediate (**Koito and Ikeda 2013**). There are three main kinds of retrotransposons: endogenous retroviruses (ERVs), characterized by long-terminal repeats (LTR), and long and short interspersed nuclear elements without LTR (non-LTR), called LINEs and SINEs, respectively (**Salter et al. 2016, Mills et al.**

**2007**). ERV and LINE autonomous elements encode specific enzymes to transpose, whereas SINE non-autonomous "parasitic" elements ("cheaters") use the molecular machinery of the autonomous elements. The majority of ERVs are unable to infect or to retrotranspose due to accumulated mutations; however, during evolution they dispersed numerous IFN-inducible enhancers of innate immunity in mammalian genomes (**Chuong et al. 2016**). Therefore, some of these retroviruses function as regulatory sequences able to induce transcription of adjacent IFN-induced genes involved in innate immunity (**Chuong et al. 2016**) such as the APOBEC genes. The most common mammalian LINE, LINE-1 elements, encode 2 open reading frame proteins (ORF1p and ORF2p), which mediate not only the retrotransposition of LINE-1 itself, but also of SINE and the reverse transcription of cellular mRNAs to generate intron-lacking retropseudogenes (**Roy-Engel 2012**). The ORF1p RNA binding protein possesses a nucleic acid chaperone activity, while ORF2p multifunctional protein with endonuclease and reverse transcriptase activities is responsible for RNA-guided integration of new copies of retrotransposons and retropseudogenes into the genome (**Roy-Engel 2012**). Moreover, since LINE-1 lacking the ORF1p coding sequence strongly increases ORF2p-mediated Alu retrotransposition (see Figure 2A in **Hulme et al. 2007**), ORF1p, when bound to its LINE-1 RNA binding site, it may also have a LINE-1-translational-repressor activity, sterically blocking complete LINE-1 ribosomal translation and ORF2 protein synthesis (a negative translational feedback loop). It is thus possible that, similarly to the Cascade complex in CRIPR-Cas systems (see above), in (stressful) conditions inducing an "excess" of LINE-1 transcription, ORF1p redistribution towards this surplus of "viral" transcribed elements (and consequently its sequestration) would relieve translational repression, allowing a rapid ORF2p translation and consequently an increase in transposon transposition. However, the majority of the several hundred thousand copies of LINE-1 are truncated transpositionally inactive copies (**Mills et al. 2007**). Among SINE, Alu sequences are the most successful elements in the human genome; however, they do not encode proteins, and the vast majority are transpositionally inactive elements (**Mills et al. 2007, Roy-Engel 2012**). They are derived from the evolutionarily conserved 7SL RNA viral sequence, a component of the signal recognition particle involved in protein secretion (**Li et al. 1982, Mills et al. 2007, Roy-Engel 2012**). The different Alu subfamily members contain two 7SL-derived (left and right) Alu domains and the 3' flanking unique genomic sequence (**Li et al. 1982, Bach et al. 2008, Mills et al. 2007, Roy-Engel 2012**).

Retrotransposons are sequences of viral origin which, unlike spacers in CRISPR systems, are considered parasitic DNA sequences dispersed into the eukaryotic genome, whose activity must be tightly controlled to maintain genome host integrity. Indeed, ERV human retroelements contain viral DNA which codes for viral proteins with potential infectivity, and non-LTR elements can potentially damage host genes during their transposition (**Wildschutte et al. 2016, Roy-Engel 2012, Mills et al. 2017, Salter et al. 2016**). On the other hand, some transposons behave like "tools for host genome engineering", successfully involved in both immune systems and natural genome editing mechanisms (see **Koonin and Krupovic 2015, Koonin et al. 2019**).

In humans, the variety of APOBEC proteins is thought to be crucial in countering the genotoxic threat generated by endogenous retroelements (**Salter et al. 2016, Hulme et al. 2007, Roy-Engel 2012**). Indeed, the expansion of the APOBEC family during primate evolution coincides with a decrease in transposon activity (**Salter et al. 2016**). However, the elevated genotoxic activity of APOBECs is also well known (**Salter et al. 2016**), curiously suggesting that the APOBEC response could be even more dangerous than the transposon activation.

The presence of fossil forms of previously inserted viral sequences with accumulated mutations raises the question of why sophisticated eukaryotic cells possess such a huge amount of apparently useless and potentially harmful viral DNA. Are transposons intrinsically selfish as genes are hypothesized to be? Or have retrotransposons developed other functions in addition to retrotransposition?

***3.3. The retrotransposon-guided APOBEC enzyme hypothesis.*** Importantly, unlike virus and ERV inactivation, the inhibition of the retrotransposition of LINE and SINE produced by APOBECs is usually independent of their mutagenic function (**Chiu et al. 2006, Stenglein and Harris 2006, Bogerd et al. 2006, Muckenfuss et al. 2006, Kinomoto et al. 2007, Hulme et al. 2007, Wissing et al. 2011, Horn et al. 2014, Feng et al. 2017**). Rather, it seems to be mediated by the binding and sequestration of retrotransposons away from their genomic integration target sites (**Chiu et al. 2006, Hulme et al. 2007**), which express the consensus sequences of retrotransposition for target primed reverse transcription (**Roy-Engel 2012**). Finally, both RNA-guided Cas endonucleases and APOBEC mutagenic enzymes may represent immune effectors which can bind RNA nucleic acid fragments of viral origin, repeat–spacer sequences and transposable elements, respectively. Indeed, APOBECs show specificity for individual LINE-1 and Alu transposable elements, and specific binding sites between some retrotransposons and specific APOBECs have been described (**Hulme et al. 2007**, **Bach et al. 2008, Chiu et al. 2006, Zhang et al. 2010, Stenglein and Harris 2006, Feng et al. 2017**), raising the possibility that some retrotransposable elements, like repeat–spacers in CRISPR-Cas systems, could represent RNA-guides for APOBECs in human eukaryotic cells.

***3.4. How have endogenous viral sequences spread into the coding regions of the eukaryotic cell genome?*** Unlike prokaryotic spacers, which are segregated into CRISPR-Cas immune organs, eukaryotic transposons spread into the eukaryotic genome. Indeed, Alus, the most successful retroelements in humans, are mainly present in “sensitive” intronic and intergenic DNA regions (**Medstrand et al. 2002**). How and why have transposable elements been integrated into transcribed regions of the eukaryotic cell genome?

It has already been suggested with regard to CRISPR-cas systems that viral nucleic acid integration into the cell genome gives the cell a survival advantage and likely represents a cellular strategy to control the viral threat. Can this also be true in the case of hypothetical retrotransposon-APOBEC systems?

During the first viral challenge, eukaryotic cells transcribe specific genes in response to the viral threat. It is conceivable that the viral fragment insertion could occur more easily in unfolded open state regions of the chromatin with single-stranded DNA (ssDNA), i.e. accessible transcribed DNA regions, during viral infection. Random viral fragment insertions usually produce disruption of the transcribed gene and cell death, which also stops spreading the virus. However, if the viral fragment insertions maintain the integrity of the proteins translated from the “infected” genes, the cell can survive. Notably, this can be achieved when the step sequence for integration is reversed after transcription by a transcript processing event that completely removes viral fragment insertions, a mechanism that is reminiscent of the intron-exon splicing process developed in eukaryotic cells. From this standpoint, eukaryotic RNA splicing would represent the reverse reaction of the viral nucleic acid insertion, and would be capable of restoring the original gene sequence without the viral sequence insertions. Finally, these intron-like viral insertions would fragment genes without affecting the structure and function of the translated proteins. This could be the origin of the endogenous viral sequence spread into the coding regions of the eukaryotic cell genome. Since this type of gene fragmentation has been positively selected in eukaryotic cells, it must have been advantageous to cell survival.

***3.5. Antiviral activity of retrotransposon-guided APOBECs.*** Drawing a comparison with CRISPR-Cas systems raises the question of whether this intron-like viral nucleic acid insertion is functional against the next viral entry. Importantly, a new threat of previously experienced viruses will activate the same sequential responses and transcriptions of the same genes transcribed during the first infection. However, in this second response, in addition to the DNA coding sequences, the inserted viral nucleic acid (intron-like) fragments will also be transcribed. As in the case of repeat-spacers for Cas endonucleases, the ability of viral RNA fragments (retrotransposons) to bind to APOBEC mutagenic enzymes, if simultaneously transcribed, would quickly guide

APOBECs against viral sequences. Indeed, APOBEC3G has been shown to associate with 7SL RNA and Alu elements (**Chiu et al. 2006, Bach et al. 2008, Zhang et al. 2010**). A conserved motif of APOBEC3G is important for the interaction with these RNA viral sequences and association with them has been shown to mediate the targeting of antiviral cytidine deaminases against retroviral complexes (**Bach et al. 2008, Zhang et al. 2010**), suggesting that some Alu/Alu-like elements could represent RNA-drivers of APOBECs against exogenous retroviruses. The conserved repetitive sequences on Alu domains (**Mills et al. 2007, Bach et al. 2008**), like repeats for Cas endonucleases, would allow the binding with APOBEC3G, while unique 3' sequences (**Roy-Engel 2012**), like spacers, would ensure target specificity. The cell-safe viral DNA fragment integration into eukaryotic cell genes would be a useful mechanism to protect the cells from subsequent reinfection. Therefore, the ability to store a copy of viral fragments (transposons) within the intronic regions would give the cell an advantage through a "boomerang response" similar to that described for CRIPR-Cas systems. Finally, the viral nucleic acid dispersion into intronic and intergenic regions was likely selected because it was helpful against virus reinfection, becoming a cell "strategy" to counter new viral threats.

***3.6. The eukaryotic gene development hypothesis*.** It is likely that this mechanism of viral fragment insertion occurred initially in transcribed polycistronic prokaryotic genes and was the basis for eukaryotic intron development, but possibly also for eukaryotic promoter and enhancer regions. Indeed, several response elements for transcription factors of eukaryotic introns, promoters and enhancers are repetitive sequences embedded in Alu elements (**Polak and Domany 2006**), all elements of viral origin disseminated within genic and intergenic regions of the eukaryotic cell genome. Notably, some viral nucleoproteins have an intrinsic ability to associate with their viral nucleic acid sequences. It is thus possible that during viral infections, some viral nucleoproteins could bind viral repetitive elements previously inserted upstream from the gene start site and in intercistronic regions. These viral sequences could therefore function as specific cell sensors for viral products. In fact, it can be hypothesized that, like eukaryotic transcription factors, the steric hindrances created by some viral nucleoproteins could help to keep the gene "open", enhancing the transactivation of previously "infected" genes with intronic (anti)viral sequences. In the course of evolution, we can observe a progressive reduction in the time required to generate a response to counter external threats. Indeed, response speed is fundamental to survival. This mechanism of viral insertion would yield both an enhanced and more rapid response to viral threats, and this could account for its positive selection. From this point of view, retropseudogenes inserted into the genome by the LINE-1 machinery would represent the first step in gene duplication and copy number variation just before the insertion of the promoter and intron regions by exogenous or endogenous viral fragment insertions.

***3.7. Cell integration of viral elements: the richness of foreigner-migrant cultures following unproductive struggles*.** As mentioned above, there is similarity of function between some viral nucleoproteins and transcription factors, suggesting that some eukaryotic transcription factors might derive from the editing of some viral nucleoprotein genes inserted into the cell genome. In this regard, there is evidence indicating that some transcription factors derive from the "domestication" of viral transposases (see **Feschotte 2008**). Transposases are viral nucleoproteins that mediate the insertion of the viral nucleic acid from which they are encoded (for example ORF2p for LINE-1 elements). Domestication of viral proteins occurs when viral coding sequences are "exapted" to generate functional host proteins (**Miller et al. 1997, Volff 2006**). But what could be the advantage that led to the integration and editing of viral coding sequences?

There is no reason to synthesize viral-like proteins in the absence of the virus; however, the endogenous synthesis of viral-like proteins makes sense if linked to viral infection. Such synthesis is an amplifying mechanism of viral perception at low loads, eventually increasing the speed of cell response, which is likely the reason for its initial positive selection. Indeed, several viral proteins have been co-opted by their cell host to

restrict viral infection (see **Jangam et al. 2017, Koonin et al. 2019**). Among the cell proteins able to restrict viral spread, APOBEC family members have a crucial role. Their intrinsic ability to bind to viral sequences suggests that they might have developed from viral nucleoproteins, which are different from those generating transcription factors.

Finally, cell-domesticated viral coding sequences and proteins, as well as viral non-coding sequences would likely have been, at least initially, exploited by host cells as a boomerang against the same viruses from which they derive. However, the endogenous viral sequences, which represent almost half of the cell genome, were likely inserted and domesticated against ancient viruses that no longer exist. For this reason, the existence of this enormous and seemingly useless or even dangerous graveyard of viral zombies within the sophisticated eukaryotic cell genome remains unclear. Have they developed any other functions over time that may be of use even today?

It is known that APOBECs evolved in primates (see **Salter et al. 2016**), concomitantly with LINE-1 elements (see **Roy-Engel 2012**), over 100 million years ago from the AID gene, which appeared more than 500 million years ago (see **Salter et al. 2016)**. On the other hand, both Alu amplification (see **Roy-Engel 2012**) and APOBEC expansion (see **Salter et al. 2016**) occurred between 55 and 35 million years ago; based on their time of expansion, Alu should have binding sites for the "younger" APOBECs, while LINEs should have binding sites for the ones that evolved earlier. Taking advantage of co-evolution of the APOBEC family members and some retrotransposons in primates, it is possible to hypothesize that the expanding range of mutagenic enzymes were tools developed to counter the burst of several types of ancient tissue-specific retroviruses (**Koito and Ikeda 2013**) and the cellular integration of Alu and LINE subfamily retroelements, for which APOBECs have conserved binding domains, may be the result of retrovirus "domestication". Importantly, these ancient retroelements, acting in concert with APOBECs, would still be useful to counter newly evolved viruses, such as HIV (**Bach et al. 2008, Zhang et al. 2010**). From this standpoint, even the integration of HIV fragments and defective HIV-1 provirus producing novel HIV protein-coding RNA transcripts (**Imamichi et al. 2016**) could represent not only a "Trojan Horse" reservoir of HIV, but also products of virus domestication and new biotechnological tools to combat HIV itself and/or future viruses. Hence, it is possible to imagine a new generation of vaccines that aim to insert specific HIV-engineered transposons with an appropriate unique sequence recognizing conserved HIV sequences and conserved repeats for APOBEC binding. To target them into genes that are transcribed early during viral entry in CD4 positive cells, transposons could be loaded into HIV-like capsids/exosomes with affinity for CD4 molecules and able to produce cell membrane stresses typical of viral infection in CD4 positive cells. This type of insertion would be functional because it would first allow the integration of the engineered transposons into specific genes of HIV CD4 target cells during vaccination and their subsequent intronic transcription in the case of HIV virus infection.

In conclusion, natural selection, following countless unproductive cell-virus struggles, may have led to a cell-virus agreement, a symbiosis, in which old viruses became fundamental components of the immune cell organ at the cell's disposal against newly evolved viral threats. Similarly, the microbial communities in symbiosis with their host organism, are integral to numerous host processes including immune system development (see **Gilbert et al. 2015**).

Therefore, in this scenario, natural selection, rather than viral domestication and exploitation, rewarded the mechanisms of microbial "education" and "integration" at different levels of organization. The splicing mechanism described above also appears to be in line with this hypothesis of viral tool integration. There are several lines of evidence suggesting that the catalytic center of the spliceosome is composed of RNA, i.e., it is a ribozyme (**Lee and Rio 2015**), a small catalytic RNA found in a number of RNA virus genomes and viroids (**Scott**

**and Klug 1996**, **Durzyńska and Goździcka-Józefiak 2015**). Through these mechanisms, the viral migrants became relatively "sedentary" viruses, still alive in the cell environment, raising the possibility that cells represent a community generated by sequential integrations of viral versatile operon modules and viral repetitive targeting sequences with regulatory functions.

**4. Exaptation of human retrotransposon-APOBEC systems: from antiviral activity to chronic stress-induced site-specific genome editing.**

***4.1. Transposon- and APOBEC-mediated genome editing mechanisms: a possible link with environmental stress conditions.*** Both CRISPR-Cas and retrotransposon-APOBEC systems would appear to be cell immune systems developed to counter viruses, which use viral nucleic acid fragments to guide immune effector enzymes. However, unlike Cas endonucleases, APOBECs mediate viral mutagenesis, suggesting that in eukaryotic cells, viral genome editing was positively selected over its fragmentation. A new cell mechanism is positively selected only if it produces improvements in cell fitness, so new adaptive abilities should be expected in retrotransposon-APOBEC systems. Notably, unlike genome-dispersed transposons, segregated CRISPR systems are protective against "viral toxicity", yet they limit viral nucleic acid insertion, and thus cell variability. Interestingly, under stress conditions, an archaea species has been observed to activate a primordial form of sexual interaction, a species specific horizontal form of nucleic acid transmission (**Ajon et al. 2011**). This primitive sexual mechanism might have been selected to compensate for a reduction in viral operon module insertion (variability), and notably, it is activated under stress-inducing environmental (ultraviolet rays) conditions, i.e. when an increase in genetic variability is functional to solve the problems imposed by the new environmental conditions.

Interestingly, APOBEC mutagenic functions are directed not only against viral sequences, but also against human endogenous (viral and non-viral) DNA, a feature responsible for both human evolution and human cancer generation (**Pinto et al. 2016, Salter et al. 2016, Chan and Gordenin 2015**). Indeed, some APOBEC members (e.g. antiviral APOBEC3G) have been shown to contribute to the human genome evolution through site-directed mutations of transcribed regions and regulatory elements (**Pinto et al. 2016**), while others (e.g. APOBEC3B) have been associated with carcinogenesis (**Salter et al. 2016, Chan and Gordenin 2015**), suggesting an exaptation of their anti-viral mechanism to the genome editing mechanism. Similar effects have been described for transposons (**Platt et al. 2018, Piacentini et al. 2014, Mourier et al. 2014**), indicating that both transposons and APOBECs have developed new beneficial as well as harmful functions through genome editing mechanisms. Indeed, roles of the CRISPR-Cas machinery in directing processes other than immune functions, such as endogenous gene regulation and genome evolution by self-targeting genome editing, and CRISPR-guided transposition, have been observed (**see Westra et al. 2014, Dimitriu et al. 2019, Koonin et al. 2019, Klompe et. al 2019, Strecker et al. 2019**), providing additional support for the idea that retrotransposon-guided mutagenic enzymes might represent the eukaryotic versions of prokaryotic CRISPR-Cas systems. Indeed, the manner of prokaryotic CRISPR/RNA-guided transpositions (see **Dimitriu et al. 2019, Klompe et. al 2019, Strecker et al. 2019**) is reminiscent of human/eukaryotic LINE-1 transposition mediated by RNA-guided ORF2 multidomain ("fusion") protein (see **Roy-Engel 2012**), a transposition mechanism that preceded that of SINE non-autonomous ("parasitic/cheating") elements. The functional similarities of CRISPR-Cas to retrotransposon-APOBEC systems imply their evolutionary "convergent" relationship, although at different (prokaryotic and eukaryotic) levels of organization. In any case, positive selection of mutagenic genome editing over nucleic acid fragmentation mechanisms must have improved eukaryotic cell fitness.

Indeed, dynamic DNA editing processes produce a series of events including gene point mutations, gene duplications and chromosomal aberrations that ultimately lead to novel phenotypic traits (phenotypic plasticity)

and enhanced ability to face stressful environments (**Piacentini et al. 2014, Gilbert et al. 2015**). However, these events are crucial not only to the organism but also to cancer adaptation and perpetuation, which ultimately leads to organism death. Hence, we can observe that similar effects produce completely different outcomes! Intriguingly, DNA editing events are clustered non-randomly in localized DNA areas of elevated recombination activity, the so-called hotspot genomic regions, in which single nucleotide polymorphisms (**SNP**), copy number variations (**CNV**) and chromosomal translocations are generated. Local increases in mutation density (occurring even in a single generation) produce different levels of cell and organism heterogeneity/plasticity and have been associated with primate evolution as well as with human cancers (**Pinto et al. 2016, Chan and Gordenin 2015, Fitzgerald et al., 2017**), suggesting that both involve common mechanisms of non-random site specific genome editing. In prostate cancer cells, genotoxic stress has been shown to induce non-random chromosomal translocations mediated by upregulation of LINE-1-encoded ORF2 protein and AID (see **Lin et al. 2009**), highlighting the importance of transposons and AID/APOBECs as agents of genome remodeling and cancer development in stressful environmental conditions. Moreover, it has been revealed that transposon insertions are involved in both leukemic and non-leukemic chromosomal translocations (**Jeffs et al. 1998, Lozynskyi and Lozynska 2006, Song et al. 2016,** see **Platt et al. 2018**), suggesting that "transposon-driven" translocations are not themselves sufficient for cell malignant transformation. In this regard, chromosomal translocations are responsible not only for cancer development (**Platt et al. 2018, Lin et al. 2009**), but also for "Robertsonian speciation" (**Garagna et al. 1997**), suggesting that cancer development or cell/organism adaptation might also be related to the environmental context. Indeed, transposon expression is activated by different environmental stress factors (**Platt et al. 2018, Piacentini et al. 2014, Mourier et al. 2014**, **Fitzgerald et al., 2017**), and the stress-response hypothesis of Barbara McClintock proposed that the expression of mobile elements was a cell genomic reaction to stressful environmental conditions (**McClintock 1984**).

***4.2. Stress-induced transposon mobilization: the intronic origin hypothesis*.** It is known that major environmental changes can affect transposon silencing and subsequent transposon activity has been associated with genetic variability, genomic instability and cancer (**Platt et al. 2018, Piacentini et al. 2014, Mourier et al. 2014**). For these reasons, transposons have been defined as parasitic elements with their own autonomous ability to reactivate under stress conditions and to threaten genome integrity. However, the molecular links between environmental stress conditions and transposon upregulation or derepression are elusive. There is no clear evidence of a specific mechanism for transposon transcription mediated by stress-induced regulatory proteins. On the other hand, we now know that the majority of transposons are present in the intronic and regulatory regions. The vast majority (about 90%) of human genes contain transposable elements in their introns and more than a million transposons are present in the introns of the human genome (**Zhang et al. 2011, Platt et al. 2018**). Since the regulatory regions are not transcribed, a main source of transposable elements might be the introns of genes induced to transcription in response to stress conditions. In this regard, many stable intron products have been discovered (**Hesselberth 2013**). It has been well documented that after the splicing of transcribed genes, some introns generate circular molecules, lariat-intermediates and lariat-intron products (**Hesselberth 2013**). These intron-derived products present internal phosphodiester bonds (branchpoint), which protect them from digestion by exoribonucleases (**Hesselberth 2013**). Indeed, hundreds of stable lariat intronic RNAs have been found in the cytoplasm of different vertebrate cells, including human cells (**Talhouarne and Gall 2018**). These stable transcripts derive from short introns coming from genes of widely different cellular functions (**Talhouarne and Gall 2018**) and, as would be expected from functional RNA species, their regulation is tightly controlled (**St Laurent et al. 2012**). Indeed, during inflammatory stress conditions different stable intronic transcripts are regulated differently, even when they are derived from the same locus, suggesting that they represent a reservoir of RNAs with yet unknown functions (**St Laurent et al. 2012**). Lariat-introns are subsequently cleaved to linear form by the lariat-

debranching enzyme, DBR1, before their digestion or function. Interestingly, retroelement transposition requires the lariat-debranching enzyme (**Karst et al. 2000, Cheng and Menees 2004**), suggesting that retrotransposon intermediates are contained in some lariat-introns and the lariat-debranching process is necessary for their retrotransposition. However, stress-induced transposon transcripts do not produce a proportional increase in the transposition rate (**Piacentini et al. 2014**), suggesting that intronic transposons might have developed functions other than transposition, possibly activated by cells under environmental stress conditions. Although the biological significance of cytoplasmic stable intronic RNAs remains obscure, one simple hypothesis is that cytoplasmic lariat intronic RNAs interact directly with RNA-binding proteins such as RNA-guided APOBEC enzymes. Finally, I propose that, under stress conditions, at least part of the transposons may come from the introns of genes transcribed in response to stressful environmental conditions. However, what do the environment and environmental stress mean for a cell?

***4.3. Molecular concepts of the cellular environment and cell environmental stress conditions: an "ecological" view of cells and tissues.*** Excluding physical agents, the environment for a cell in an organism is mainly represented by cell "exogenous" molecules that come into contact and sterically interact with molecules of the cell, hereafter exogenous environmental molecules. Exogenous molecules that come in close proximity to a cell and do not sterically interact with its components are "invisible" to the cell, and since they do not interfere with cellular functions, they are not directly part of the cell environment. Notably, exogenous environmental molecules can be tissue specific, since they may or may not interact with molecules expressed by a specific cell tissue, depending on three-dimensional structures of both exogenous molecules and cellular proteins. Exogenous environmental molecules can be autonomous molecules but also molecules present on the surface of neighboring cells or extracellular components. Therefore, a cell interacting through its surface molecules with neighboring cells constitutes an environment for those neighboring cells (similar concepts can be transferred among organisms). When a specific tissue sterically interacts with an exogenous molecule, it independently yields plastic phenotypic adaptation to exogenous molecule perturbation and "niche construction" that may affect other tissues interacting with it, ultimately generating a cascade involving the whole organism. In normal environmental conditions, exogenous environmental molecules fluctuate around normal concentration ranges through assumption, storage and release (homeostatic) mechanisms, which tend to limit their fluctuation ranges through negative feedback loops. Likewise, cellular proteins fluctuate around optimal concentration ranges as a function of environmental conditions.

However, when specific cell components interact with new, never-experienced exogenous molecules, those molecules can interfere with cell biological functions, generating tissue-specific cell perturbation and stress. Such molecules are usually defined as pollutants or stressors and they can be biotic or abiotic. The binding of pollutants to cellular structures can be tissue specific, and it can also be dependent on a specific protein polymorphism. Another important consideration is the concentration of exogenous environmental molecules; there are normal fluctuations of molecule concentrations to which the cell is able to respond through homeostatic mechanisms, and abnormal ones, either excessive or insufficient, which produce cell stress. In addition, the concentration of exogenous molecules in a specific cell tissue may depend on the organ of entrance in the organism. For example, air pollutants interact at high doses with pneumocytes, while food pollutants interact at high doses with enterocytes and hepatocytes, which is probably why smokers are more likely to develop lung cancer than other malignancies. Interestingly, high concentrations of molecules (pollutants or non-pollutants) induce low-affinity (aspecific) bindings to cellular molecules, which do not occur at low concentrations. Notably, low-affinity bindings are induced by high concentrations not only of exogenous environmental molecules but also of cellular proteins, because in both cases the probability of their interaction increases. Finally, both environmental pollutants and molecule concentrations out of the normal range (either

excessive or insufficient) cause significant changes in the cell environment and cell stress level that may be either temporary or chronic.

***4.4. Cell homeostatic mechanisms control optimal cell protein concentration in normal environmental conditions.*** Cells and organisms have developed specialized homeostatic systems able to fine tune the amount of protein in cells as well as the amount of cells within an organ according to the environmental context. There is an optimal range of expression for each cell protein; a protein excess or deficiency produces cell dysfunction and cell stress. Changes in environmental conditions produce cell perturbations and cell responses seek a new balance (adaptation). Dynamic balances through homeostatic mechanisms regulate optimal protein expressions as a function of environmental conditions. Homeostatic responses are usually epigenetic mechanisms, which, operating through transcription factors, repressors and enzymes, modulate protein expression and stability, ultimately regulating their amount, persistence and functionality at transcriptional and/or post-transcriptional levels. Cell homeostatic mechanisms adjust the protein amount, function and stability according to the cell functional needs, typically through negative feedback regulatory loops. For example, **Figure 2**(**A,B,C**) shows a homeostatic mechanism involved in the protein expression of a signal protein at a transcriptional level. When a protein (**P**) is expressed at low levels compared to the cell's functional needs (**Fig.2A**), the transcription factor ( ) that regulates its transcription is free to induce its gene transcription. The subsequent translation into proteins increases the protein amount and, by binding with a second protein (**P1**), the pathway functional activity in which the protein (**P**) is involved (see **Fig.2B**). To fine-tune protein expression, cells have developed negative feedback mechanisms through repressors ( ) able to inhibit the activity of transcription factors ( ): the higher the expression of the protein (**P**) bound with its partner protein (**P1**) along their signal pathway, the higher the gene repression ( ), which inhibits gene transcription. Finally, lowering protein synthesis, protein degradation produces both low levels of protein (**P**) and repressors (**Fig.2C**), similar to the initial condition (**Fig.2A**). The cycle restarts and the amount of protein (**P**) fluctuates around the optimal needs for cell functionality according to the environmental conditions. Usually, cell epigenetic homeostatic responses counteract normal environmental fluctuations through negative feedback loops that tend to reduce external environmental molecular changes around a normal fluctuating balance. Normal environmental fluctuations induce either an increase or a decrease in gene transcription and protein expression and stability. However, a major problem arises when a new environmental context induces cell homeostatic responses which, though reaching their maximum activity, are not sufficient to meet the new cell needs. This occurs when the demand surpasses the potential ability to synthesize and stabilize a specific protein, i.e. the cell does not find the solution/balance among the epigenetic mechanisms. If the new environmental condition is temporary, it may produce a temporary cell stress; however, when the environmental changes are persistent and/or increasing, chronically stressed cells have to activate different mechanisms in order to survive.

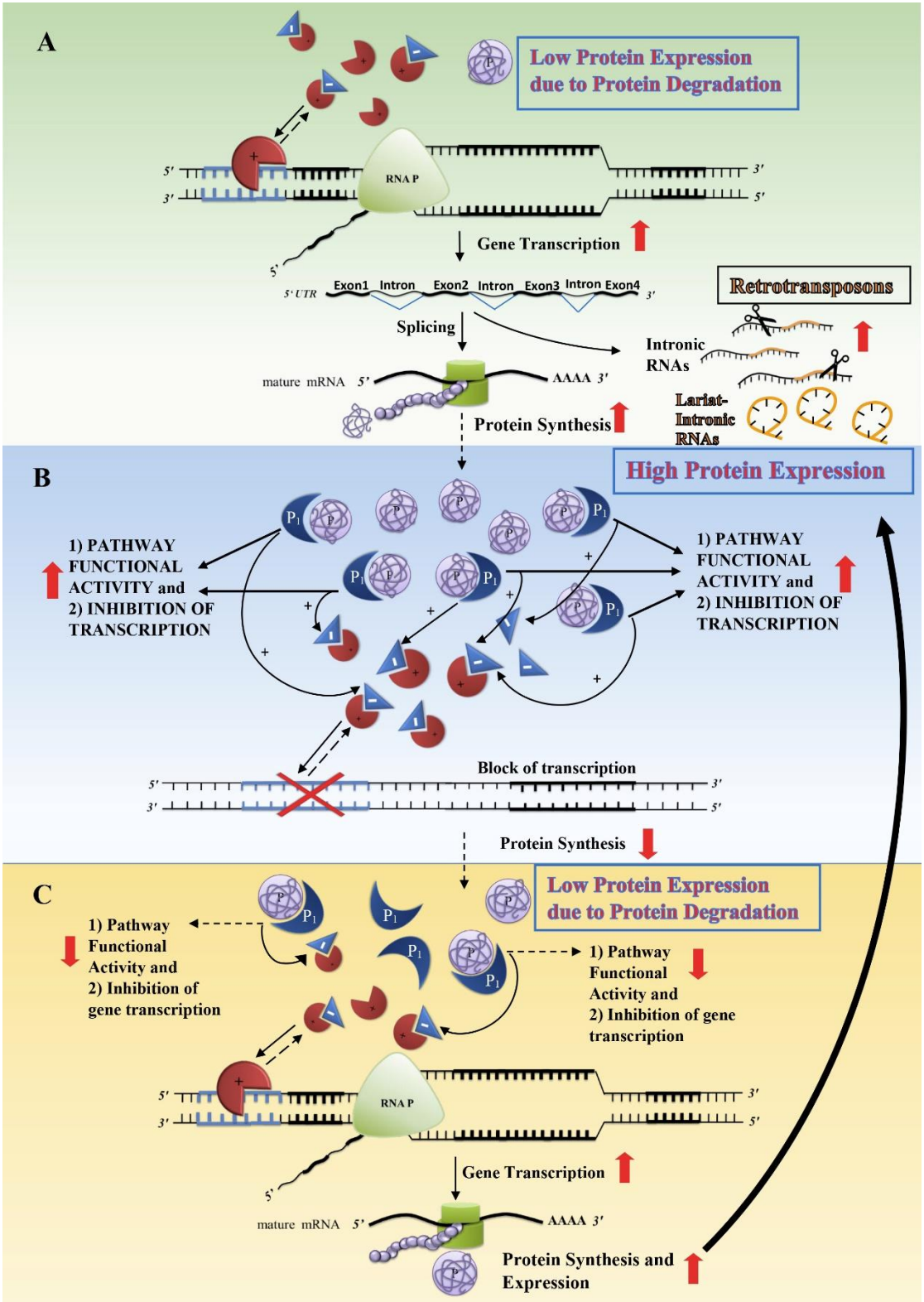


***Fig.2A,B,C. Cell homeostatic mechanisms control optimal cell protein concentration in normal environmental conditions.*** The example in the figure depicts a homeostatic mechanism involved in the protein expression of a signal protein at a transcriptional level. (for reference, see the text; **abbreviations and symbols: RNA P**: RNA polymerase; **P**: protein; : transcription factor; : repressor)**.**

***4.5. Putative cell mechanisms of eukaryotic non-random genome editing in response to chronic stress conditions: chronic stress-induced gene duplication and site-specific mutagenesis.*** I have already described how cell homeostatic responses counteract normal environmental fluctuations and a new "epigenetic" balance is rapidly attained (see **Fig.2**). However, in the case of never-experienced and chronic environmental conditions, for which the cell does not possess an "epigenetic" solution, i.e. the protein requirements exceed the homeostatic possibilities, the cell homeostatic responses are chronically pushed to their limits.

As already suggested, chronic cell stress can be generated by new exogenous molecules that permanently interfere with cellular components or when the cell is constantly exposed to abnormal (either excessive or insufficient) concentrations of "known" exogenous molecules. Persistent environmental stressors continue to perturb cell functions until cell adaptation or cell death occurs. But what is the sequence of events that could lead to cell adaptation in chronic stress conditions?

Sudden and intense stress conditions (e.g. under volcanic eruptions, earthquakes or meteorite impacts) may not leave enough time for cell/organism adaptation, leading to mass extinctions (**Pope et al. 1998, Kerr 2013**), whereas chronic environmental conditions that originate from a progressive deviation from normal environmental fluctuations would first induce cell epigenetic homeostatic mechanisms. It is likely that epigenetic mechanisms of adaptation are trained by progressive environmental changes. However, epigenetic homeostatic mechanisms likely have a threshold of tolerance beyond which the final products of cell epigenetic responses, specific protein concentration, functionality and stability, would not be able to provide an effective response to the new environmental conditions. While the minimal needs for proteins can be easily met by blocking gene transcription, the optimal amount of protein and protein function cannot be attained in the case of an excessive demand, which exceeds the cell's epigenetic capabilities. In chronic stress conditions, some genes may be switched off by promoter methylation and others may be demethylated and chronically transcribed, "hypertranscribed". However, when "hypertranscribed" genes do not produce sufficient amounts of proteins to meet new cell needs, the genes are unable to provide an adequate response to the new environmental conditions. More generally, we can say that hypertranscribed genes are genes which are inadequate for the environmental context.

In these situations, gene duplications and/or gene editing ("genetic plasticity"), in search of more effective proteins, are two possible independent ways to achieve a new balance. Finally, CNV and SNP of hypertranscribed genes may allow cells to surgically explore a new fitness scenario, which increases their adaptability to stressful environmental conditions. In a changeable environment, like that on Earth, randomly emerging mechanisms able to produce these effects were likely positively selected. However, which mechanisms could make adaptation of hypertranscribed genes possible?

A first possible response to excessive demand for a specific protein is gene duplication. To synthesize eukaryotic proteins, first gene transcription, then splicing, and finally translation are sequentially induced. The "rate determining steps" in these "reactions" are likely the base pairing between codon and tRNA-anticodon bearing an amino acid during gene translation and the splicing events. In the case of hypertranscribed genes, an overflowing/excess of mature and "dis-spliced" (or badly-spliced) transcripts waiting for ribosomal protein synthesis is therefore expected. It is possible that, during stress conditions, the excess of multiple-sized transcripts from hypertranscribed genes together with stress-induced LINE-1 transposon transcription might favor the interaction of multiple-sized transcripts with LINE-1-encoded ORF2 multifunctional proteins. Finally, these associations may generate retropseudogene or fragmented gene insertions of hypertranscribed genes in the accessible chromatin regions of the hypertranscribed genes themselves. This hypothesis is supported by the observation that highly expressed housekeeping genes have high copy numbers of retropseudogenes (see

**Roy-Engel 2012**). As suggested above, in this scenario, retropseudogenes and gene fragments would represent the first step in gene duplication, just before the "viral" insertion of the promoters and new introns. The new gene copy would diverge from the original gene because of its fragmentation and/or as a consequence of the different promoter and intron insertions caused by new virus infections and/or the new transposon insertions, i.e. by cell environmental conditions.

Gene editing and generation of SNP in hypertranscribed/inadequate genes represent another way organisms can achieve a new cell balance in response to chronic stress conditions. This mechanism explores different gene point mutations to be tested in the new environmental context; those providing the highest level of cell fitness and survival capability will be inherited. What could the mechanism of surgical selection for editing an inadequate gene be?

**Figure 3A** shows the case in which a new biotic or abiotic pollutant chronically binds and interferes with a signal protein (**P**), affecting its binding with a second protein (**P1**) along its signal pathway. Both the protein (**P**) functional activity and its repression are inhibited by the pollutant and homeostatic epigenetic responses induce upregulation of gene transcription (and post-transcriptional protein expression and stabilization). An increase in protein synthesis is functional because it compensates for the lack of protein function induced by the binding with the pollutant (a negative feedback loop that reduces the pollutant-driven cell perturbation, but also favors the further binding with the pollutant). However, hypertranscription of the gene produces not only an increase in mRNA for protein synthesis, but also in the intronic transposons present (if any) in the hypertranscribed gene. In the case of sustained transcription, such as in chronic stress conditions, transposon RNAs from hypertranscribed intronic regions would be accumulated within the cells. Indeed, many stable lariat intronic RNAs derived from hundreds of different genes have been shown to be exported to the cytoplasm. They are short, mostly 100–500 nucleotides in length, and they are expressed at 10% or more of their cognate mRNA level (**Talhouarne and Gall 2018**). Importantly, stress conditions can also induce APOBEC upregulation, and I have already suggested that some transcribed retrotransposons can bind and guide these mutagenic enzymes. It is therefore possible that stable retrotransposon transcripts, following the DBR1-mediated lariat-debranching process, might guide APOBECs towards the genomic introns from which they were transcribed. When the chromatin of a gene is in an open state, such as during continuous gene (hyper)transcription, the specific intronic site would be readily accessible for editing.

This mechanism would foster the generation of mutations without DNA duplication on both gene alleles codifying for the protein involved in the binding with the pollutant, leaving the rest of the genome intact (**Fig.3B**). Although this mechanism could potentially occur in virtually all cells, it is likely that only long-lived cells, such as adult somatic stem cells or memory cells, could survive and respond to chronic environmental stressors. I have already suggested that the pollutant structure can bind to a specific tissue protein. Therefore, after chronic binding with pollutants, stem cells of specific tissues can be induced to randomly generate slightly different mutants (genetic plasticity) of hypertranscribed genes. Among the cell mutants expressing different protein variants (phenotypic plasticity), those bearing gene mutations which provide high cell fitness in the new organism environment would survive and expand, ultimately yielding a new gene version "inheritable" by tissue progeny (genetic assimilation). In line with this hypothesis, recent evidence has revealed intra-organismal and intra-tissue genetic mutations (genetic plasticity) in healthy individuals (see **Gottlieb et al. 2010**). Indeed, mosaicisms of both SNP (**Gottlieb et al. 2010**) and CNV (**Abyzov et al. 2012; O'Huallachain et al. 2012**) within and among normal tissues of the same individual have been observed. In addition, tissue specific translocations, pathological or non-pathological, can represent a third level of mosaicism and an expression of chromosomal plasticity.

As shown in **Figure 3B**, a protein mutant (**P**) able to restore binding with (**P1**) in the presence of a pollutant is a good solution for the survival of both tissue cells and the organism in which the cells reside. Interestingly, this hypothesized mechanism of gene editing mediated by APOBECs would produce different protein repertoires (phenotypic plasticity) and binding affinities for the pollutant, and eventually the mutated protein would somehow wrap around the pollutant, binding it with higher affinity than the original protein. This process strongly resembles the somatic hypermutation (SHM) of immunoglobulins (Igs) mediated by AID, the ancestor of the APOBEC family, during commitment to B cell memory (**McHeyzer-Williams et al. 2011, Salter et al. 2016, Senavirathne et al. 2015**). It is well known that generation of long-lived memory cells is characterized by new gene expressions mediated by epigenetic modifications that drive “naïve” cell differentiation into long-lived cells (**Gabrielli et al. 2016**). Therefore, these mechanisms may produce both epigenetic and genetic variations that yield a range of environmentally-induced phenotypes (phenotypic plasticity). Finally, environmental selection of the fittest phenotype would ultimately fix acquired gene versions into the genome, leading to genetic assimilation of environmentally-induced characteristics.

Is there any molecular evidence of this putative cell mechanism in human eukaryotic cells?

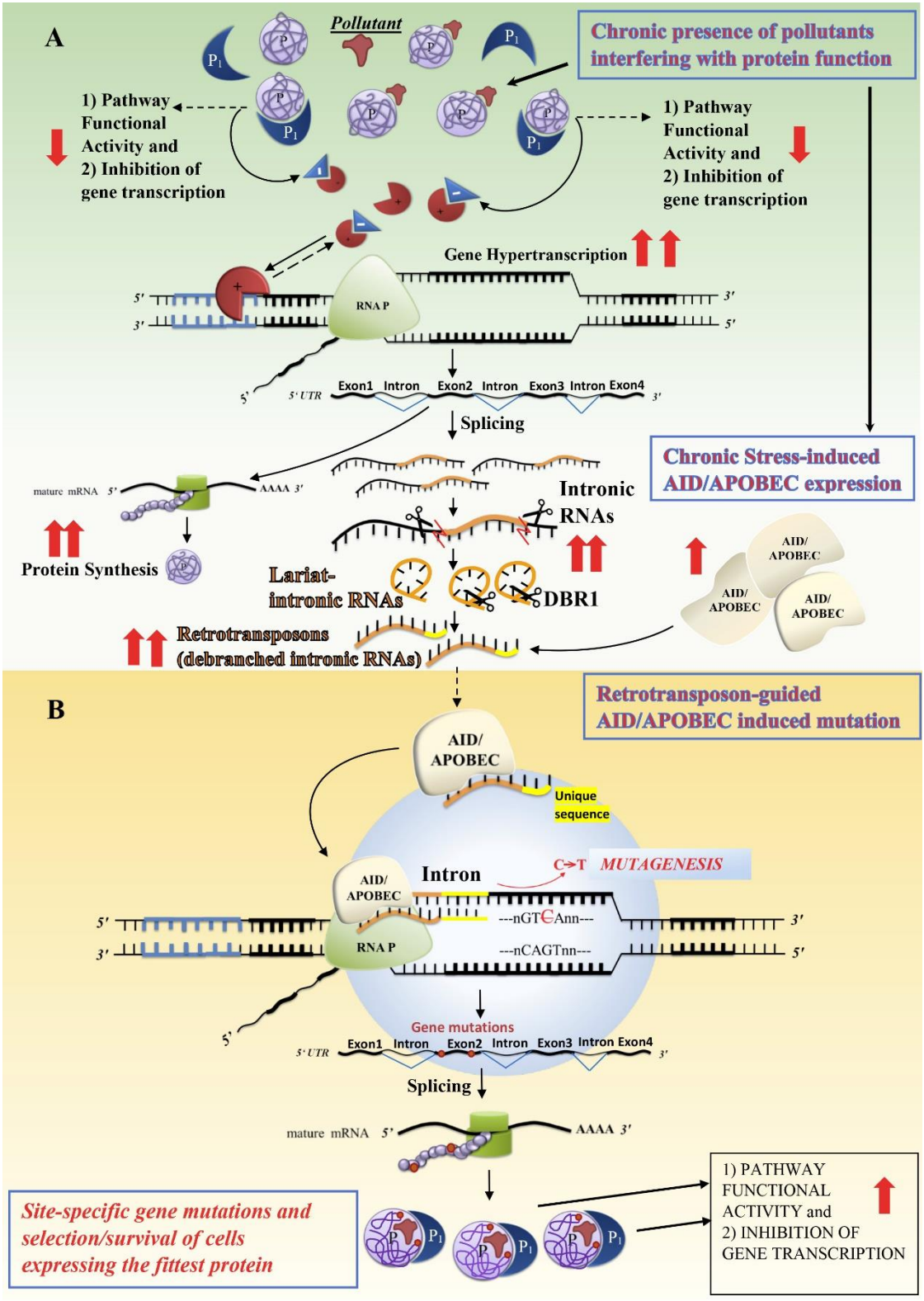


**Fig.3A,B. A *Putative cell mechanism of eukaryotic genome editing in response to chronic stress conditions.*** An example depicting chronic stress-induced site-specific mutagenesis mediated by APOBEC enzymes. (for reference, see the text; **abbreviations and symbols: RNA P**: RNA polymerase; **P**: protein; : transcription factor; : repressor, **DBR1**: lariat debranching enzyme).

***4.6. Learning from the immune system: the example of AID-mediated immunoglobulin recombination.*** As mentioned above, the mechanism model that I have proposed resembles SHM of Igs, a process which is usually coupled to heavy-chain class-switch recombination (CSR) of Ig locus, both of which are induced by AID (**McHeyzer-Williams et al. 2011**, **Zheng et al. 2015, Senavirathne et al. 2015**). SHM and CSR represent two pivotal processes in the immune system able to edit Ig genes in a way that adapts them to a new biotic pollutant (antigen) through targeted mutagenesis. During these processes, the Ig gene undergoes: 1) point mutations in the variable, antigen-binding coding sequences, which change antibody binding affinity for the antigen and 2) a B cell switch from IgM to IgG, IgE, or IgA, which changes the antibody effector function. These processes start when the binding affinity of the immunoglobulins (spontaneously produced by B cells) for an exogenous antigen is low and insufficient to eradicate the pathogen bearing it. The chronic presence of the pathogenic antigen induces chronic transcription of Ig genes (hypertranscription) and proliferation of B cells bearing low affinity antibodies, inadequate to counter the pathogen challenge/threat. Expansion of activated B cells reaches peripheral lymphoid organs where, promoted by antigen presenting cells, they encounter the biotic pollutant and undergo SHM and CSR processes. The mutations mediated by AID alter the specificity and affinity of the Ig protein for a "pathogenic pollutant" and generate a pool of B cell mutants (genetic plasticity). This model only partially explains the spectrum of somatic mutations observed in SHM during antigen-driven gene editing. In fact, the major mutations are represented not only by AID-mediated C-to-T (and correspondent guanosine-to-adenosine, G-to-A), but also by adenosine-to- guanosine (A-to-G) (**Steele et al. 2004**), pointing to the involvement of mutagenic mechanisms other than AID. In this regard, the adenosine deaminase enzymes acting on RNA (ADARs) have double stranded RNA binding domains and a C-terminal adenosine deaminase domain able to induce adenosine-to- inosine (A-to-I) RNA editing, finally read as A-to-G (**Steele et al. 2004, Zheng et al. 2017**). Recently Zheng and colleagues (**Zheng et al. 2017**) have shown that ADARs can edit both the RNA and DNA moieties of RNA-DNA hybrids, raising the possibility that ADARs, driven by specific double stranded RNA transposon transcripts, could effectively induce A-to-G DNA mutations. Indeed, analysis of the total mutation pattern indicates that the major transitions observed in SHM are A-to-G, C-to-T, and G-to-A, but not T-to-C (**Steele et al. 2004**), suggesting that AID-induced mutations (C-to-T) may involve both DNA strands (with a subsequent G-to-A mutations on the opposite DNA strand), while ADAR-mediated A-to-G mutations may involve only one DNA strand.

Regardless of the mechanism of somatic mutations, among the different B cells bearing different Ig mutants, only those bearing antibodies with a high affinity for the biotic pollutant proliferate and, finally, eradicate the pathogen. This ability enhances the organism's chances of survival and has therefore been inherited, although, in some cases, it may lead to the development of autoimmunity. This is an extraordinary mechanism of non-random, "on demand" mutation, in which environmental changes through a biotic "pollutant", drive the cells to edit specific exons of a specific eukaryotic gene, leaving the rest of the B cell genome intact. The chronic presence of cell-interacting pollutants would induce an intronic retrotransposon-mediated gene/protein plasticity. This mechanism provides the material for environmental selection of phenotypes, ultimately yielding new inheritable and "environmentally-independent" versions of Ig genes expressed on novel memory B cells, an example of "genetic assimilation" of acquired characteristics (**Waddington 1959**). Finally, the mechanism is able to adapt the gene structure, in such a way that it generates antibodies that, "wrapping" around the antigen, are more functional in their response to new environmental stressors. How does this molecular mechanism target AID to Ig locus for CSR and SHM?

Intriguingly, the main actor in these processes is AID, the ancestor member of the RNA-guided APOBEC family. SHM and CSR require complex interactions involving the selective recruitment of AID to transcriptionally active Ig regions. Subsequently, AID has been shown to translocate unidirectionally in concert with RNA polymerase

during transcription of Ig genes and to catalyzes cytidine deamination and C-to-T mutagenesis on the non-transcribed DNA strand of actively transcribed Ig genes (**Senavirathne et al. 2015**).
The molecular mechanisms that guide AID to repetitive switch (S) regions during CSR have been described by Zheng and colleagues (**Zheng et al. 2015**). CSR is a deletional-recombination process in repetitive DNA elements of S regions located upstream of each heavy-chain gene segment. AID deaminates cytosines within transcribed S regions, and the resulting point mutations activate the DNA double strand breaks required for CSR (**Zheng et al. 2015**). Importantly, the authors have demonstrated that RNA transcripts from intronic switch regions serve as molecular guides to target AID to S region DNA. Repetitive RNA transcripts derived from germline S intronic elements associate with AID and guide AID to the complementary S region DNA in a sequence-specific manner (**Zheng et al. 2015**). Moreover, the authors showed that DBR1, necessary for both debranching intronic lariats and retrotransposition (**Karst et al. 2000, Cheng and Menees 2004**), was also required for AID localization to S region DNA during CSR (**Zheng et al. 2015**), suggesting that stable intronic lariats, containing repetitive sequences of the S region, have to be processed by DBR1 before carrying out their functions.
Taken together, these data indicate that the RNA-mediated mechanism of targeting AID to ssDNA is mediated by intronic repetitive (viral) sequences generated by Ig hypertranscribed genes, which produce immunoglobulin proteins that are inadequate to respond to new environmental conditions, i.e. pathogens. A similar mechanism of RNA-targeting can be hypothesized for other mutagenic functions of APOBEC family members, as suggested in the putative mechanism model described above, although differences in the RNA-binding sites among APOBEC/AID systems are expected. Based on time of APOBEC/AID and retrotransposon expansion (see above, **Salter et al. 2016 and Roy-Engel 2012)**, the difference between the long repetitive sequences, such as those in the S region of Ig locus, which possess specific binding sites for AID protein and the retrotransposon (sub)families with specific affinity and binding sites for the different APOBECs, is in fact justifiable (**Hulme et al. 2007**, **Bach et al. 2008, Chiu et al. 2006, Zhang et al. 2010, Stenglein and Harris 2006, Feng et al. 2017**).

Finally, it can be hypothesized that, thanks to intronic transposons, the majority of cell proteins may be modified in a similar manner as immunoglobulins, and are thus able to plastically adapt their structures to a new “pollutant” interfering with their functions. In line with this hypothesis, one third of adaptive mutations affect virus-interacting human proteins, suggesting that they have occurred in response to viruses. These adaptive mutations may therefore represent one of the most dominant drivers of evolutionary changes in mammals (**Enard et al. 2016**). Hence, a mechanism developed to counter a biotic threat would be exapted and extended to abiotic pollutants for biological adaptive purposes. This hypothesis unveils a novel function of the “virus trained” immune system, a kind of “military” laboratory for the development of biomolecular technologies that are ultimately exapted for “civil” purposes. New biological technologies developed through viral training and viral domestication together with natural selection might be able to predispose life to counteract present challenges and to prevent future unpredictable biotic but also abiotic environmental threats.

***4.7. Retrotransposon-guided mutagenic enzymes: an ideal “Lamarckian” tool of evolvability and plasticity. The example of polyethylene metabolizing moth larvae.*** I would now like to make some observations that will help us to fully grasp some of the important consequences of the retrotransposon-guided mutagenic enzymes. The example described in **Figure 3** depicts a general signaling protein involved in pollutant binding; however, depending on the stereochemistry of both the pollutant and polymorphic protein domains (dependent on the genomic background), virtually all types of proteins with different functions may be involved (i.e. enzymes, transcription factors, receptors, hormones, onco-related proteins etc). Retrotransposon-guided APOBEC mutagenesis could be an inducible gene-targeting mechanism, which selectively edits hypertranscribed genes,

the ones that are inadequate to cope with emerging persistent and growing environmental stressors. Notably, these non-random genome editing mechanisms activated "on demand" in response to novel environmental pollutants are able to "choose" a few hyperfunctional genes in which to create novelties, i.e. gene mutations (genetic plasticity). Moreover, recent evidence has revealed that different intronic transcripts from the same locus are tightly and differently regulated in mammalian cells (**St Laurent et al. 2012**); therefore, retrotransposons transcribed from a specific intron would surgically direct a site specific mutagenesis in the intron and the adjacent exons, making it possible to modify specific protein domains.

The non-random genome editing mechanisms that I have described for humans would easily explain the incredible complexity of human protein structures and interactions. However, it is likely that similar mechanisms are present in other eukaryotic organisms, and altogether they might have provided a crucial contribution to complexity during "eukaryotic" evolution.

In line with these non-random DNA editing mechanisms, is the ability, developed by larvae of the wax moth, to metabolize polyethylene (**Bombelli et al. 2017**), a relatively recent man-made product. It is, in fact, possible that wax moth larvae, in novel environmental conditions with increasing amounts of plastic derivatives and the absence of wax, would have favored polyethylene (aspecific) binding to some proteins of the wax metabolic pathway. It is known that high concentrations of molecules, whether they be pollutants or non-pollutants, induce low affinity (aspecific) bindings to cellular structures, such as those shown in **Figure 3**. Consequent gene hypertranscription and protein upregulation would further favor the interaction with polyethylene. It is therefore possible that under conditions of steadily rising amounts of new molecules (e.g. polyethylene), "aspecific" bindings could be at the origin of the induction of some non-random genome editing responses. In the case of wax moth larvae, the plastic molecular interference may have pushed enzyme gene hypertranscription, ultimately inducing the wax enzyme gene to modify into a gene encoding a polyethylene metabolizing protein.

As shown in the **Figure 3**, through a "boomerang" mechanism, the transposon-guided APOBECs may edit genes codifying for the protein involved in the initial binding with the pollutant without DNA duplication. This means that mutations, originally occurring randomly during DNA duplication along the whole DNA molecule of viruses and prokaryotes, would surgically take place in the hypertranscribed genes of the vast eukaryotic genomes, leaving intact the rest of the genome. Finally, the eukaryotic non-random genome editing mechanisms, providing high eukaryotic DNA evolvability and plasticity to the changeable environmental challenges on Earth, have likely allowed the development of the extraordinary complexity of higher organisms, which need large genomes to store a huge amount of information. Importantly, through this mechanism, pollutants that do not directly bind to DNA, would be potentially mutagenic and able to induce DNA mutations in both alleles of hypertranscribed genes (only) of somatic tissues interacting with the pollutant. The chronic interaction between a pollutant and an exonic-derived protein domain would be the trigger for DNA editing induction; therefore, the tridimensional structures and arrangement of specific atoms in the space (i.e. their stereochemistry) of both tissue proteins (whose structures mainly depend on exon gene polymorphisms) and pollutants, would be the initial trigger of non-random DNA editing mechanisms. Notably, after mechanism induction, the intronic transposon polymorphisms can be crucial to guiding, or not guiding, APOBEC family members or other DNA mutagenic enzymes (for example ADARs or LINE-1-encoded ORF2p) to target sites, ultimately generating biodiverse responses and different interindividual gene-editing attempts. Moreover, mutations in tissue-specific genes will be expressed, as mutated proteins, only in specific tissue cells (mutated tissue); however, a tissue phenotypic variation may affect neighboring cells of other tissues in an organ, finally involving the whole organism. Within the different mutant cells, "aborted" proteins and proteins with different affinities (phenotypic plasticity) for the pollutant would be generated and cell bearing mutated proteins that

enhance cell fitness and cell survival in the presence of the pollutant (new cell environment) would be selected and expanded, generating a mutated tissue (a first level of selection). Within the organ, the mutated tissue interacts with neighboring cells of other tissues, only those restoring organ function in the presence of the pollutant (new organ environment) would allow organism survival (a second level of selection). Finally, among the different interindividual tissue/organ attempts, only those producing lasting fitness of the organ and thus of the organism in the new "polluted" ecosystem would be Darwinianly "inherited" (or "Waddingtonly" genetically assimilated). It is, in fact, likely that the most favorable mutations for the organisms would then survive long enough to be passed on, whereas mutations that are unfavorable or less favorable to the organisms (organisms bearing environmentally inadequate genomes) may produce rapid death or diseases (a third level of selection). The mechanisms through which some new somatic genes as well as their protein products produced by non-random genome editing mechanisms could be selected and inherited by the progeny will be discussed in detail in a future paper.

Notably, this non-random gene editing mechanism may also provide the rationale for Alu repeat enrichment in fragile sites of preneoplastic lesions, DNA regions with increased loss-of-heterozygosity frequency (**Tsantoulis et al. 2009**). In this regard, it is likely that in order to adapt cell tissues to new environmental conditions many aborted attempts and "pathological" by-products might be generated. Can non-random genome editing mechanisms unveil a link between pollution and chronic diseases? Indeed, transposons and mutagenic enzymes have been shown to be involved not only in species adaptation and evolution, but also in chronic diseases and, in particular, cancer progression. Notably, cancer cells are chronically "stressed" by organismal immune system, an "environmental" condition that might drive their adaptation/evolution. In this regard, it is interesting to explore the possibility of whether cancer cells could be exploited as model of eukaryotic cell evolution.

**5. Exploiting cancer as a model to study non-random genome editing mechanisms and evolution**

***5.1. A cell species that exploits non-random genome editing mechanisms to survive in chronic stress conditions: cancer***. Every "outside-in" signaling event from the environment to the cell yields an "inside-out" response from the cell to the environment and vice versa, leading to a fluctuating "epigenetic" balance or a search for a new "genetic" balance. Indeed, both epigenetic and genetic responses can be sequentially induced by environmental changes. An initial cause produces an effect which, through a boomerang reaction, influences the starting cause, creating negative or positive feedback loops that sometimes make it difficult to determine what the original cause (if there is one) and the cause-effect relationship of a phenomenon are. Epigenetic mechanisms are induced first, they are usually reversible and tend to reduce external changes around a fluctuating balance through "reactionary" negative feedback loops. Conversely, when a gene mutation is accidentally induced, for example, by a viral infection (environment), the mutation indirectly produces an environmental change (environmental plasticity), a process whereby a mutated cell actively alters its environment, a sort of cellular niche construction, that gives rise to "revolutionary" positive feedback mechanisms. Such positive feedback mechanisms can be readily observed in metabolic diseases, for example, the "environmental" toxicity of phenylalanine accumulation in phenylketonuria is a consequence of an enzymatic mutation and loss of function; however, this same phenomenon occurs in other biological contexts, including cancer. When a normal cell is transformed into a cancer cell, the environment suddenly becomes hostile, as immune cells and eventually drugs start to "threaten" cancer cell expansion (see **Zamai et al. 2007**). In some ways, cancers have a lot of functional similarities to viruses. Both can kill their host, produce genome selections, spread their deadly kiss and are challenged by immune system. Indeed, it has been suggested that cancer extracellular vesicle (exosomes) spreading could be involved in the process of metastasis (**Rajagopal and Harikumar 2018**). Moreover, it has been proposed that oncogenic nucleic acid materials and mutagenic

enzymes contained in extracellular vesicles (forms of eukaryotic horizontal genetic transmission) spread the tumor and its deadly activity to other tissues (**Rajagopal and Harikumar 2018, Stavrou and Ross 2015**), further highlighting virus-cancer similarities. In solid tumors, the disorganized three-dimensional structure creates cancer hypoxia and nutrient starvation, stress conditions that further drive cancer cell adaptation. However, to adapt to chronic and adverse environmental conditions, cancer cells likely activate the same smart non-random genome editing mechanisms, already described in normal cells, which promote cancer genome heterogeneity and ultimately cancer cell selection. There is in fact evidence that cancer cells use non-random DNA editing mechanisms, producing non-random clustered mutations similar to those generated in highly polymorphic human genes, such as HLA and KIR3DL1 (see **Norman et al. 2007**).

It is well known that cancer cells can exploit normally dormant embryonic pathways, acquiring embryonic and neotenic features (i.e., delaying maturity and retaining immature features) that foster developmental plasticity and face stressful environments. Indeed, cancer cells could represent new aggressive tissue species that rapidly evolve in a hostile "ecosystem", the organism. They behave similarly to an aggressive virus or a predator (like humans) that, though having a different level of organization, transforms or kills its "ecosystem", the cell or ecological niche, respectively. From this standpoint, it is not surprising that under selection pressure, cancer cells diverge with regard to gene copy numbers in subpopulations distributable in phylogenetic trees similar to those described for higher organisms (**Navin et al. 2011**). The pattern of cancer evolution is characterized by the sudden emergence of new tumor clones, a pattern defined as a 'punctuated clonal evolution' (**Navin et al. 2011**), borrowing the concept from the non-gradual species evolution observed in fossil records (**Gould and Eldredge 1993**). Moreover, saltational evolution has been suggested to play a role both under stress conditions, when the mutation rate is significantly increased (e.g. in stress-induced mutagenesis in bacteria), and in the evolution of tumor cells (**Katsnelson et al. 2019**). Indeed, tumor mutation rate, possibly due to the stress conditions above described, is substantially elevated compared to that of normal tissues and this may account for the increased probability of multimutational leaps in cancer cells (**Katsnelson et al. 2019**). Hence, cancer appears to be a good model to study, from which we can infer eukaryotic cell evolution.

Using cancer as a model for eukaryotic cell evolution allows us to verify previously hypothesized mechanisms of non-random genome editing. The development of drug resistant cell lines is a good example of cancer adaptation. In this regard, the mechanism of p53 non-random gene mutation induced by doxorubicin in the MCF-7 breast cancer cell line is currently under investigation. Chronic administration of increasing doses of doxorubicin induces doxorubicin-resistant MCF-7 cell lines with a reproducible identical C to T point mutation at the end (the last nucleotide) of intron 4 of the p53 gene (**Tsou et al. 2015**). The specific intron 4 mutation of p53 results in an abnormal splicing transcript with a 21 nucleotide deletion and in p53 loss (or gain) of function (**Tsou et al. 2015, Balcer-Kubiczek et al. 1995**). The mutation signature mediated by a consensus sequence of 4 (GTCA) nucleotides suggests the involvement of APOBEC3B (**Chan et al. 2015**), a well-known mutagenic enzyme associated with breast cancer and upregulated by doxorubicin (**Kanu et al. 2016**). Chronic DNA damage induced by doxorubicin likely upregulates the genome guardian p53 exon and intron transcription. It is therefore possible that the transcribed p53 retrotransposons, in particular the Alu element (with its unique sequence) present in the intron 4, could bind and surgically target APOBEC3B just upstream of the mutation site of the p53 gene, ultimately inducing the specific non-random point mutation and the alteration of a specific p53 protein domain.

***5.2. From gene hypertranscription to non-random repetitive element retrotransposition generating non-random novel regulatory gene networks: the example of melanism in the British peppered moth.*** **Figure 3B** shows how the constant presence of a protein-binding pollutant interaction could potentially induce hypertranscription and mutation of both protein gene alleles without cell proliferation. However, under deep

environmental changes (due, for example, to heavy industrial pollution), it is likely that several genes are induced to hypertranscription by different pollutants. Therefore, several intronic transposons from hypertranscribed genes could potentially generate multiple mutations (within a single generation in one genome), accounting for saltational evolution. Since the chromatin of hypertranscribed genes is in an open state and accessible even for transposition, a preferential exchange of retrotransposon "regulatory sequences" among the hypertranscribed genes would likely occur. Retrotransposition among hypertranscribed genes would generate new introns and/or new response elements in promoter and/or enhancer regions not only of genes, but also of retropseudogenes, possibly generating paralogous genes. Notably, this non-random retrotransposon exchange could produce novel regulatory gene networks, where a single transcription factor, originally associated with a specific gene, may link previously unrelated hypertranscribed genes through a common response element. In addition, non-random retrotransposition exchange may potentially involve both paternal and maternal homologous chromosomes bearing hypertranscribed genes, equipping the genes of both homologous chromosomes with the same regulatory response elements (introns, promoters, enhancers), a phenomenon that might precede chromosomal translocation homozygosity events (see later, and **Lin et al. 2009**). This new network mediated by a common transcription factor would be functional in accelerating and coordinating the cell response to a new specific environmental pressure, and therefore this mechanism might have been positively selected during evolution. Notably, these novel regulatory networks, although independent of gene mutations, would be "irreversibly memorized" in non-coding regions of the genome and therefore could generate inheritable and persistent ("genetically assimilated", to borrow a word from Waddington) alterations in the control of gene expression. These memorized changes in regulatory networks might be inherited by progeny, accounting for the well-known "epigenetic" inheritance. In this regard, it is noteworthy that the mutational event responsible for the melanism in the British peppered moth during industrial environmental changes is the insertion of a transposable element (**Van't Hof et al. 2016**), raising the possibility that environmental changes may have led to the generation of a novel regulatory gene network mediated by this transposon transposition. Notably, a pollutant-driven hypothesis of peppered moth melanism, as a selective mechanism in addition to differential bird predation by crypsis, has been proposed by Riley (**Riley 2013**). Since melanin has been shown to be able to chelate metal ions, an increase in melanization has been suggested to produce a localized accumulation of metals and potential cytoprotective effects in moths exposed to toxic levels of heavy metals (**Riley 2013**). In addition, it can be hypothesized that the non-random genome editing mechanisms induced by pollution could have selected both novel response elements for melanin gene transcription directly dependent on heavy metal presence and mutated melanin forms able to more strongly sequester heavy metals. The consequent addiction to heavy metals of the mutated melanin transcription and therefore of the peppered color of the moth would be consistent with the recent and rapid reversal of the process (the decline in the frequency of peppered forms) associated with a reduction in atmospheric pollution (**Cook 2013**).

I have already suggested that viral repetitive sequences inserted during viral infection or retrotransposition into intronic regions of a gene are expected to be similar to those inserted in promoter and enhancer regions of the same or other simultaneously hypertranscribed genes. Hence, such promoters and enhancers might also be targets of mutagenic enzymes driven by intronic retrotransposons of hypertranscribed genes. In this regard, members of the AID/APOBEC family have been shown either to efficiently deaminate methylated cytosines, leading to methylcytosine demethylation (**Ito et al. 2017**), or to induce C-to-T somatic mutation in non-coding promoter regions, generating a novel binding site for the oncogenic transcription factor, MYB (**Li et al. 2017**). The ability of AID/APOBEC family members to demethylate methylcytosine in regulatory (promoter or enhancer) regions might promote gene transcription. In this regard, in a model of breast cancer, it has been shown that APOBEC3B is able to increase expression of estrogen-responsive genes (**Periyasamy et al. 2015**),

suggesting that APOBEC3B might actually demethylate estrogen-responsive regulatory elements in breast cancer cells. The possibility of demethylating methylcytosine and editing C-to-T in the CG-rich islands of promoters and/or enhancers, promoting respectively either epigenetic and reversible ("environmentally-dependent") increased gene transcription or (similarly to retrotransposition) generation of "genome irreversible memorized" (or "genetically assimilated") novel responsive elements and regulatory networks that might be inherited by progeny (accounting for an "epigenetic" inheritance), points to additional roles for AID/APOBECs in transcriptional regulation and in new repetitive viral sequence generation. On the other hand, during chronic stress, APOBEC genes are hypertranscribed and therefore these same genes and even their regulatory regions can be intronic retrotransposon-mediated targets of their mutagenesis and/or of retrotransposition of new response elements, potentially generating new APOBEC members and new regulatory APOBEC gene networks. The chimeric APOBEC3A-3B deletion variant, which is linked to a higher risk of developing cancers, could be an example of a new APOBEC member (see **Caval e al. 2014**). The reciprocal potentiation of mutagenic enzymes and endogenous retroviral sequences allows the generation of new genes and responsive viral elements with new unique sequences without viral infection, making eukaryotic life stronger and stronger until the combinatorial possibilities of a finite DNA content have been exhausted (if this ever occurs). However, C-to-T somatic mutations in non-coding regulatory regions (**Li et al. 2017**) and new APOBEC gene variants (**Caval e al. 2014**) have been shown on cancer cells, raising the question of whether these phenomena are limited to pathological conditions or could be common aspects of eukaryotic cell development.

***5.3. Non-random chromosomal rearrangements and the involvement of non-random genome editing mechanisms in cancer cells following exposure to chemical and physical stressors.*** I have already suggested how gene hypertranscription and transposon transposition among hypertranscribed genes would disseminate common response elements into promoter, enhancer and intronic regions of hypertranscribed genes. However, what could be the mechanism leading to chromosomal translocation?

Gene hypertranscription can be induced by chronic stimulation mediated transcription factors, and it is well known that transcription factors also induce chromatin remodeling (**Li et al. 1997**), a step which precedes chromosomal translocations (**Roukos et al. 2013, Lin et al. 2009**). Using a model of prostate cell cancer, it has been shown that the ligand (DHT)-activated transcription factor, androgen receptor (AR) induces both specific gene transcriptions and chromosomal movements that create either mono-allelic or even bi-allelic spatial proximity among genes bearing the same intronic androgen response elements (AREs) (**Lin et al. 2009**). Notably, ARE repetitive sequences are one of the hormone response elements typically located within Alu repeats (**Babich et al. 1999, Polak and Domany 2006**). In this model, upon dihydrotestosterone (DHT) stimulation, AREs of different genes appear to move towards a common chromatin region, indicating that AR transcription factors bound to AREs may activate a nuclear myosin/actin-dependent mechanism (**Lin et al. 2009**) able to move them to a chromatin area of DHT-AR-ARE(Alu) complex accumulation. This mechanism calls to mind the chromosome movements in the mitotic spindle. Notably, the concomitant stimulation of DHT with genotoxic stressors produces both the proximity of ARE from distant chromosome regions and the recruitment of genotoxic stress-induced enzymes, AID and LINE-1-encoded ORF2 endonuclease (**Lin et al. 2009**). This association facilitates the generation of DNA double-stranded breaks and subsequent specific non-random translocations even on both homologous chromosomes, mediated by the non-homologous end joining (NHEJ) machinery (**Lin et al. 2009**), and possibly providing the rationale for Alu repeat enrichment in fragile chromosomal sites (**Tsantoulis et al. 2009, Fungtammasan et al. 2012**). Notably, the observed gene fusion events have been found to be present in 50-70% of prostate cancers in vivo (**Lin et al. 2009**), indicating that this model of tumor translocations mimics the in vivo events.
Interestingly, both chemical (Etoposide and Doxorubicin) and physical (50Gy of ionizing radiation) genotoxic stresses have been shown to reproducibly produce the same chromosomal translocations when combined with

DHT (**Lin et al., 2009**). To generate non-random chromosomal translocations, sequential events that produce proximity of intronic response elements (e.g. ARE) and specific targeting of genotoxic stress-induced enzymes are induced. In the prostate cancer model, the recruitment of LINE-1 ORF2 protein and AID acts through two independent mechanisms (**Lin et al. 2009**); therefore, the recruitment of AID, as in the case of CSR, might be mediated by its binding with intronic repetitive transcripts, while ORF2p may be mediated by its binding with intronic LINE-1 transcripts, both likely originating from hypertranscribed genes. Alternatively, AID and ORF2p might be independently driven to specific DNA regions through a direct binding with AR, as hypothesized for APOBEC3B and the estrogen receptor in a breast cancer model (**Periyasamy et al. 2015**). However, in the case of prostate cancer cells, AR should have specific binding sites for both AID and ORF2p (beside AREs), whose recruitments have been shown to be independent (**Lin et al. 2009**). Alternatively, an indirect AR binding with AID and ORF2p, mediated by different RNA repetitive "bridges", can be hypothesized. I have already suggested that transcription factors and anti-viral APOBEC members might evolve from two distinct viral nucleoproteins, both with an intrinsic ability to bind to repetitive nucleic acid sequences. Therefore, some viral sequences inserted into intronic regions might have been positively selected and "domesticated/educated" during evolution for their intrinsic functional ability to bind to both nucleoprotein-derived transcription factors and nucleoprotein-derived anti-viral enzymes, such as the mutagenic members of AID/APOBEC family or the ancient (ORF2) endonucleases. These types of intronic repetitive sequences would be able to create bridges between different RNA binding proteins and stress-induced mutagenic enzymes, depending on the intronic transposons of (hyper)transcribed genes. In this regard, it has been shown that AID translocates unidirectionally along with RNA polymerase during Ig gene transcription, but does not interact directly with RNA polymerase (**Senavirathne et al. 2015**), suggesting that a specific intronic repetitive sequence may mediate the interaction between AID and RNA polymerase during transcription. From this point of view, the intronic transposons would represent versatile modules, varying among different genes of the same individual and possibly varying among the same genes of different individuals, which would lead to individual responses.

Regardless of the way that the genotoxic stress-induced enzymes are recruited, non-random chromosomal translocations are induced by two independent mechanisms. On the one hand, the genotoxic stressors induce mutagenic enzyme expression and on the other hand, the transcription factors (in this case DHT-AR) induce specific gene hypertranscription, transposon expression and transposition among hypertranscribed genes. The target specificity in non-random chromosomal translocation would be assured by the presence of common response elements with a unique sequence, previously disseminated by transposition into promoter, enhancer and intronic regions among the hypertranscribed genes. These types of events can produce similar translocations on both paternal and maternal homologous chromosomes, a sort of "cell speciation" occurring without cell proliferative selection (cell cycle–independent) in a single-hit process.

***5.4. Solid cancer/leukemia/lymphoma as models of "cell speciation" and the example of mouse speciation in Seveso.*** Chromosomal translocations are usually pathogenic rearrangements often involved in cancer and, in particular, in leukemia and lymphoma development; however, they are also present in healthy individuals and in cord blood samples from healthy newborns (**Song et al. 2011**), suggesting that translocations are not themselves sufficient for the transformation of a normal cell into a malignant cell, which is likely also related to the cell environmental context, i.e. the organism in which the chromosomally rearranged cell lives. Indeed, there are several examples of non-pathogenic chromosomal imbalances with no phenotypic effects that are directly transmitted from parents to offspring (**Barber 2006**). In this regard, Robertsonian translocations (fusions between two acrocentric chromosomes), which can potentially lead to speciation (**Garagna et al. 1997**), are the most common chromosomal rearrangements in humans, occurring in approximately one in every thousand human newborns (**Song et al. 2016**). When the Robertsonian translocation is balanced, the

person carrying it (carrier) has a full genetic complement and is healthy. Nevertheless, Robertsonian translocation carriers usually have a reduced productive fitness due to meiotic segregation disturbances and interchromosomal effects, leading to a higher probability of genetically imbalanced gametes (**Song et al. 2016**). Indeed, sperm of Robertsonian translocation heterozygotes have a proportion of unbalanced chromosomal complements, leading to the increased risk of a chromosomally unbalanced fetus, infertility and miscarriage (**Song et al. 2016**); however, all sperm of Robertsonian translocation homozygotes are balanced, pointing to Robertsonian translocation homozygosity as a potential mechanism of human cell speciation (**Song et al. 2016**). Since intermediate translocation heterozygosity is self-sterilizing, a reproductive isolation mechanism known as hybrid incompatibility, homozygosity for a particular Robertsonian translocation chromosome, which theoretically can lead to the establishment of a new species, possibly occurs in a single-hit process. In this regard, I previously hypothesized that both paternal and maternal homologous chromosomes bearing hypertranscribed genes can non-randomly exchange retrotransposons, thus equipping genes of both homologous chromosomes with the same regulatory response elements (introns, promoters, enhancers). This mechanism may account for the direct generation of Robertsonian translocation homozygosity in the absence of a sterile heterozygote intermediate.

Finally, chromosomal translocations produced by non-random genome editing mechanisms would generate new persistent junctions that would reduce the distance among the genes bearing the same response elements, further accelerating the cell response to environmentally induced transcription factors. In the case in which both paternal and maternal homologous chromosome translocations occur in a single-hit process, rearrangement homozygosity could be a potential mechanism of “cell speciation”. Hence, chromosomal translocations may produce a new “species” of cell tissue that substitutes the old cell tissue species because it is more suitable to survive in the new organism environment. The new cell species can be either incompatible with the organism survival, as in the case of cancer/leukemia/lymphoma (cell-mediated diseases, some of which may be environmentally-induced through non-random genome editing mechanisms), or compatible and possibly advantageous to the organism, as likely occurred in the chromosomal speciation observed in mice after the Seveso disaster (**Garagna et al. 1997**). In the latter case, it can be hypothesized that non-random genome editing mechanisms may have produced both mice with higher resistance to dioxin toxicity and dioxin-induced chromosomal translocations directly in germline stem cells or indirectly transferred from somatic cells to those stem cells. This aspect and the possibility of non-random genome-editing-mediated diseases warrant further investigation and will be addressed in future perspective papers.

## 6. Conclusions and future perspectives

***6.1. Non-random genome editing mechanisms: a link between environmental changes and eukaryotic biological novelties.*** In the present work, I have explored several lines of evidence that support the existence of molecular mechanisms of non-random genome editing developed and exploited by human cells and, more generally, eukaryotic cells, to respond to new and chronic/stable environmental conditions.

When the cell environment is sufficiently stable, and cell reversible/epigenetic mechanisms are sufficient to respond to normal molecular fluctuations, the cell eukaryotic genome is adequate. However, in the face of new and chronic environmental changes (new pollutants/microorganisms and/or new molecule concentrations), the mechanisms of non-random genome editing would be induced to “surgically” produce several eukaryotic gene attempts (gene plasticity) and new regulatory sequences (“epigenetic” plasticity).

Finally, as the environmental changes become relatively stable, the best biological solutions would be selected and stably memorized in the DNA, accounting for both the genetic and the so called “epigenetic” inheritance. Non-random genome editing mechanisms may represent a link between environmental changes and

eukaryotic biological novelties, creating a dynamic world where biological plasticity and symbiosis are common consequences.

The existence of these genome editing mechanisms would establish the centrality of the environment, of which life is now a fundamental part, constructing specific niches/habitats (environmental plasticity). Indeed, the environment can influence biological structures at different levels, inducing both their variation/transformation and selection. However, only molecules (e.g. pollutants) that penetrate and reach cell compartments may generate cell environmental changes able to induce "molecular" DNA mutagenesis. Notably, tissue phenotypic variations induced by tissue-specific gene mutations create a cascade of events involving neighboring cells that sequentially affects organs and the fitness of the whole organism. Therefore, the outcome of the mutagenic responses in an organism does not depend only on the specific mutation but also on the response of the neighboring tissues (the environment around the mutated cells), which can be different in different organisms (e.g. some chromosome translocations can be either pathological or harmless). Similar to a cell signal, an environmental "outside-in" perturbation produced by exogenous "penetrating" molecules (pollutants) can induce an "inside-out" biological response that yields a new environmental condition at outer levels (niche construction/ environment plasticity). This process reoccurs cyclically, building up progressively higher levels of biological organization and complexity that, on the other hand, "eat" part of abiotic/inorganic environmental components.

***6.2. Non-random genome editing mechanisms reconcile gene-centered and "ecological" theories of evolution.***
The appearance of life as we know it, generated by the special environment on Earth, accelerated environmental changes (environmental plasticity), i.e. environmental evolution and the production of novel environmental heterogeneities through niche construction. This in turn drove biological structures to evolve and create novelties and higher levels of aggregation and complexity, which in turn, in an ongoing cycle, produced new environmental products. Notably, non-random genome editing mechanisms allow the storage of more information in the DNA database (large eukaryotic genomes) while still maintaining its flexibility and adaptability to environmental changes. Therefore, these mechanisms have likely allowed the development of higher organisms in the presence of rapidly adaptable microorganisms and, more generally, in Earth's changeable environment. Since evolutionary theories cannot disregard both the environmental context and the genetic mechanisms, the genome editing mechanisms may represent the missing link between ecological and gene-centered views of evolution. Indeed, non-random genome editing mechanisms may provide the molecular-genetic basis for the rapid biological plasticity described both in punctuated equilibrium by Gould and Eldredge (**Gould and Eldredge 1993**) and in "ecological" views of evolution (**Gilbert et al. 2015, Laland et al. 2014**) and hence may effectively reconcile gene-centered and "ecological" theories of evolution. It is for all of these reasons that I see the need to develop a new "hyper-modern" synthesis of evolutionary theory, in which Earth's changeable environment (Eco) (especially after the appearance of life) is a starting point that cyclically drives the biological evolution (Evo) and memorization (Memo) of the most adequate living forms, which ultimately aggregate (Poly) into increasingly complex organisms, holobionts.

Moreover, the existence of retrotransposon-guided proteins (mutagenic enzymes, multifunctional proteins, transcription factors etc) suggests that viral repetitive non-coding and apparently "junk" DNA elements, accounting for almost half of the human eukaryotic genome, may not be useless "selfish" or even parasitic elements, but rather, crucial elements of the sophisticated eukaryotic cell still involved in mediating response/adaptation to environmental challenges. Nevertheless, the majority of viral structures maintain their cell-killing and migrating features that exert strong selection pressure on their host organisms. How is it possible to reconcile virus cell-killing ability with virus domestication?

***6.3. Virus/virus-like particle evolution as migrating "organelles" of cells***

Some viral structures may also be seen as migrating "organelles" of cells that, going abroad to other biological "countries" (tissues of other organisms) undergo new ("stressful") conditions (e.g. new host immune pressure shaping virus diversity), and learn and acquire new skills (e.g. host-induced new viral immune evasion strategies and/or new structural or replication protein induced by host-driven domestication), which kill obsolete and no longer adequate biological "countries" unable to manage them (induction of cell and organism negative selection), while improving/modifying others that will be continuously tested facing changeable environmental challenges. In line with this hypothesis are the observations that genetic content can flow not only from virus to host cell, but also from host to viral genomes, as exemplified by host-derived: 1) stress-induced ERV reactivation (**Palmisano et al. 2012, De Vito et al. 2018**); 2) viral defence mechanisms; 3) viral capsid proteins and; 4) plasmids coding for viral replication proteins (**see Koonin et al. 2019, Kuprovic and Koonin 2017; Kuprovic et al. 2019, Kazlauskas et al. 2019**). These observations both justify the viral "escape hypothesis" (see **Kuprovic and Koonin 2017; Kuprovic et al. 2019)** and exemplify the "guns for hire" concept developed by Koonin and collaborators (**see Koonin et al. 2019)**. However, the back and forth horizontal transfers of genomic components between viral elements and cellular systems also reveals that: 1) an initial cause (e.g. viral infection) produces an effect which influences the starting cause through feedback loops (viral genome editing), making it difficult to determine what the original cause and the cause-effect relationship of a phenomenon are; 2) each migration in a new environment does not induce a mere and static genomic shuttling but also a genomic evolution/adaptation mediated by both random viral mutations and non-random cell-mediated genome editing mechanisms. These phenomena are expected to cyclically generate either locally and temporary inactive (non-interacting) elements or new parasites or new symbiotic elements. However, since symbiosis among biological elements is a more stable solution, a progressive aggregation and complexity (holobionts) is expected over time.

***6.4. A symbiotic reconciliation of the Darwinian and Lamarckian perspectives****.* The described mechanisms of genome editing perfectly match those whose existence I have "madly" hypothesized in the first part of this paper to justify a body of unexplainable "weird" evidence regarding eukaryotic species. Such mechanisms are activated "on demand" by novel environmental pressures, and they non-randomly "target" a small number of hyperfunctional and environmentally inadequate genes (codifying exons) or regulatory sequences (introns, promoters, enhancers, miRNA) in which they create novelties, i.e. "Lamarckianly" acquired non-random mutations induced by the (ab)use of gene transcription (as hypothesized for organs by Lamarck) that lead to phenotypic plasticity. The acquired phenotypic variants/novelties provide raw material for Darwinian natural selection, which will then determine the survival of the most effective version of genes/proteins and regulatory sequences in the new cell environment (first level of selection) and the most effective version of cell/tissue in the new organ/organism environment (second level of selection), hence suggesting multilevel environments and multilevel selections. In addition, it is likely that in new stable environmental conditions, eukaryotic organisms have developed mechanisms through which some somatic genes and regulatory sequences, proteins and cell adaptations produced by non-random genome editing mechanisms could be passed on and inherited by the progeny of a species (third level of selection); however, to transfer somatic gene and cell mutations to progeny, there must be cell "communication" mechanisms between somatic and germline cells. In this regard, spermatozoa from virtually all animal species have been shown to have a particular feature, i.e. the ability to spontaneously take up somatic extrachromosomal nucleic acid information present in extracellular vesicles (the gemmules of Darwinian pangenesis, eukaryotic "plasmid-like" forms of horizontal and vertical genetic transmission), which may be delivered to fertilized oocytes and embryos (**Spadafora 2017**). This possibility will be discussed in a future paper.

Genome evolution appears to be driven by environmental pressure and mediated by different evolvability mechanisms, from Darwinianly random in first and simple biological structures through error-prone nucleic acid replication (e.g. viroids) to those that are Lamarckianly orchestrated through selected gene hyper-transcription in (higher) organisms owning high-fidelity replication mechanisms. Altogether, these considerations "neotenically" renew/rejuvenate the old Lamarckian theory by the unveiling of new genome editing mechanisms and ultimately eliminate the conflict between the Lamarckian and Darwinian perspectives**.**

***6.3. Multilevel environments and multilevel selections***. The biological system seems to be formed by multilevel communicating nested environments in which a multistep selection process occurs sequentially. The environmental changes at each level of the biological hierarchy produce different sets of adaptations. I have already described that in the case of appearance of new environmental molecules (accidentally generated or migrated) interacting with the biological molecules present in a cell produce a cell imbalance. First of all, new genes/proteins and regulatory sequences (molecular levels of biological complexity) have to meet the needs of the cell in the new cell environment, and subsequently (bottom-up), among the new cell solutions, only those that meet the needs of the organ and organism in the new organ/organism environment will survive, while organisms bearing environmentally inadequate genomes will succumb. Among the adequate genetic solutions of the organism, only those that are transferable (if any) to progeny in the "species environment" will be inherited ("memorized" by the species). Finally, competition and selection can work at even higher levels of complexity, i.e. at the levels of intra- and inter-species communities and whole ecosystems ("megabionts"). On the other hand, the appearance (through an accidental generation or migration) of new viruses/cells/organisms (intermediate levels of biological complexity) interacting with biological entities in an ecosystem produces imbalances and/or new struggles for survival among biological entities at their same level of complexity (e.g. because of new intra-level biological interactions or new competitions for resources). In these cases, adaptation and selection start at intermediate levels of biological complexity and generate a flow of adaptations (like a fluid mosaic) both top-down to molecular level (e.g. new viral species bearing new biological molecules, genes/proteins) and bottom-up to ecosystem level (e.g. new organism species bearing new biological molecules and producing new "waste" products from their metabolism) through inter-level communication mechanisms between neighboring levels. Ultimately, multilevel adaptations lead to multilevel selection of both the most environmentally adequate "solutions" and symbiotic elements by exclusion of conflicting, death-producing biological entities. The degrees of freedom and possible solutions exponentially increase with each successive level of biological complexity; however, adaptation/selection at lower levels restricts the adaptive possibilities and actual degree of freedom at higher levels. Because lower (e.g. molecular) levels of biological complexity appear earlier than the emergence of higher forms of biological complexity, this means the latter is built from the "older" molecular levels that were selected in past (and different) environments on Earth. Among the older selections, there are some that never change (e.g. the genetic code, the number of cervical vertebrae in a giraffe etc), while other "epigenetic" adaptations may reverse. However, biological adaptations that are "fixed" in the past will bias subsequent biological adaptations restricting the solutions. For example, the RNA was likely the most stable molecule produced in the environmental conditions of the primordial soup and this aspect was likely determinant for the subsequent biological evolution (see later, on the origin of biomolecules), limiting the subsequent adaptive possibilities (degree of freedom). That is to say that the initial "imprinting" (determined by the initial environment) can be crucial/irreversible in determining some of the future developments, because the new biological "buildings" seem to partially come from the reuse and the redistribution of some old and stable "bricks" in a more "environmentally" efficient and stable manner (the future is partially built from some older and more stable "bricks").

In general, chronic environmental stress conditions in defined geographical areas (ecosystems) produce struggles for survival that rapidly induce the development of new biological tools and species (developmental plasticity or clades at different levels of complexity), which locally determine temporary periods of stability ("stasis").

The matryoshka-doll logic of multilevel selection theory has already proposed that natural selection operates at multiple levels of the biological hierarchy and that adaptations at lower levels influence higher levels (**Wilson and Wilson 2008**). Since in some cases, the biological traits "for the good of the group" may prevail against those for the good of the individual, the authors suggest that in determining a biological outcome (e.g. evolution of a new species), the "weight" of each level of selection has to be evaluated on a case-by-case basis (**Wilson and Wilson 2008**). It is possible that the level of organization of the new structures (abiotic or biotic molecules, virus, bacteria, eukaryotic cells, organisms, or others) that produces the initial biological imbalance/change in a specific environment/ecosystem, might be crucial in determining the "weight" of the different levels of selection. In a future paper, I will explore how and when biological traits could evolve altruistic behaviors ("for the good of the group"), exploiting genome editing, niche construction and multilevel selection processes.

***6.4. Biological systems as fractal systems with a nested "matryoshka doll" structure.*** Self-similar ("winning") structural and functional patterns across different levels of organization/complexity (self-similarity within a defined "scaling window ") are typical of fractal systems, which also exhibit a high level of (self-)organization and iterative pathways (**Losa 2009**), features that give living forms the ability to adapt to a changing environment, maximizing fitness. My general impression is that biological systems are fractal systems with a nested "matryoshka doll" structure. At different levels of organization/complexity, biological entities are formed from a sequential aggregation of slightly different basic modules ("monomers") that together form aggregates ("heteropolymers"), which in turn become basic modules for superior aggregations that may have patterns resembling those of lower levels although with higher levels of complexity (see Fig.1). For example, we can imagine a scenario in which different viruses become basic modules ("monomers") whose sequential aggregation leads to the formation of different cell types ("heteropolymers"). In turn, replication of different cells ("monomers") forms tissues ("homopolymers") that become basic modules for organ ("heteropolymers") formation and different organs ("monomers") become basic modules for organism ("heteropolymers") formation. Finally, different "homopolymers" of replicating species become basic modules for ecosystem ("heteropolymers") formation. Notably, a similar scheme/approach can be applied to atomic and subatomic elements and perhaps even celestial bodies, all distinguishable in species ("heteropolymers") formed by the aggregation of different "monomers". For example, basic modules of protons, electrons and neutrons aggregate into atoms, and atoms aggregate into molecules, following the rule of a progressive increase in matter stability during this aggregation. Can this rule be used to justify the selection of specific biomolecules on Earth?

***6.5. On the origin of biomolecules and virus-like structures.*** The starting point of the history of life can be traced to the particular environment that developed on Earth, namely the presence of huge amounts of water in which myriads of molecules could be synthesized (a primordial soup). In this regard, it has recently been hypothesized that nucleic acids, amino acids, and lipids were produced from a pair of simple compounds abundant on early Earth (**Patel et al. 2015**). These three classes of biomolecules may have given rise to aggregates able to produce a primordial form of life. Importantly, aggregation mediated by weak inter- and intra-molecular bonds stabilizes molecules and excludes them from further chemical reactions, creating a progressive selection of the most stable molecule aggregates. For example, during the hypothetical stage of an "RNA world", the formation of stem-loop structures of RNA through intra-molecular hydrogen-bonding

interactions (A-U and C-G) would generate more stable double helix RNA and non-enzymatic template-directed RNA ("heteropolymer") synthesis (see **Robertson and Joyce 2012**). The synthesis of a complementary oligonucleotide under the direction of a preexisting RNA sequence would likely foster phosphodiester bonds between nucleotides, in particular, between cytosine and guanosine because of their three stabilizing hydrogen bonds, possibly justifying the generation of both CG-rich viroid genomes, the first virus-like structures of the "RNA world" (**Durzyńska and Goździcka-Józefiak 2015, Trivedi et al. 2007**) and CG-rich promoter/enhancer regions, fossil forms of anciently inserted viral sequences. Moreover, it can be hypothesized that during the "RNA world" stage some amino acids were able to bind repetitive RNA sequences through weak molecular bonds (such as amino acid conserved regions of APOBEC or Cas proteins with viral repetitive transcripts) that may have fostered peptide bond formation (peptide synthesis) between amino acids coming in close proximity. This may have led to the selection of amino acids with a specific three-dimensional structure (levorotatory stereoisomers) and then primordial peptides (e.g. RNA-recognition motif domains, see **Kuprovic et al. 2019**) able both to bind repetitive RNA sequences and to increase molecular structural stability. Stable RNA-peptide aggregates ("heteropolymers") may have generated "naked" viruses (without envelopes), and subsequently, with affinity binding of lipids present in the primordial soup, different RNA-protein-lipid "viral" aggregates ("heteropolymers") were possibly generated, giving rise to enveloped viruses. The "biological bricks" were supplied by a passive diffusion from the primordial soup (primordial soup as a host) until competition for biological bricks occurred. Molecules capable of "stealing" amino acid (parasite-like mobile elements) and accelerating amino acid recruitment and proximity, such as tRNA-like structures involved in virus both replication and translation (see **Durzyńska and Goździcka-Józefiak 2015**), may have been selected. Aggregation of RNA elements coated with amino acids/peptides/proteins and/or able to recruit/steal amino acids from the primordial soup may have generated error-prone replication and "translation" systems (Random formation of first biological structures). Error-prone primordial forms of ribosome-like structures, generating myriads of different molecules (both RNAs and peptides) continuously tested and selected, were highly efficient to explore a new fitness/stability scenario, finally developing structures/aggregates with higher both stability and capability of reproduction. Of note, viral genes mainly code for proteins able to bind its own nucleic acid (RNA or DNA) code, this feature favors a positive feed-back loop, producing a "viral" structural stabilization because the "protective" protein RNA-coating ("RNA-peptide" structure) stabilize and protect RNA nucleic acid code. During the RNA world, among myriads of ribosome-like attempts, those able to reproduce both RNA and peptides with affinity for its own RNA and leading to stable RNA-peptide (i.e. naked virus-like) copies might have been positively selected. Indeed, this is not surprising if we think that spontaneous chemical reactions lead to chemical products that are more stable than their reactants, and biomolecules are chemical products originated from the primordial soup reactants.

Finally, as I have already hypothesized, cells may represent communities (heteropolymers) generated by sequential integrations of viral monomers composed of versatile operon modules and viral repetitive targeting sequences with regulatory functions, while organs and organisms can be thought of as aggregations (heteropolymers) of cell and organ versatile modules (monomers), respectively. Notably, all these versatile modules (except when they become selfish and predatory) can be seen as "biological bricks" that have been domesticated/educated to fulfill the functions of the higher (host) level. Finally, the primordial soup that initially supplied viral replication with its "biological bricks" (primordial soup as a host), was subsequently substituted by the advent of structures able to plan the "robbery" of the biological bricks from the environment, the cells, moving from primordial soup to cell as a viral host.

***6.8. Conclusive "philosophical" considerations.*** Of note, all living systems, catalyzing spontaneous chemical reactions, accelerate the formation of more stable products from reactants. Life may therefore represent a

sequential process of atom/molecule/cell/organisms aggregation and ”symbiosis” that occurs among (abiotic and biotic) entities engaged in persistent interactions, which usually arise from weak molecular bonds and lead to increasing matter stability and compaction. Indeed, biophysical data surprisingly indicate that atoms inside proteins, at interfaces of protein–protein complexes and at the protein-DNA interfaces are as closely packed as in crystalline solids (**Nadassy et al. 2001, Liang and Dill 2001**), suggesting that (biomolecular) matter inside living organisms may undergo a local compaction process and increase of density resembling those of celestial bodies in the Universe. From this point of view, life, as gravitation, works like a force of aggregation that progressively integrates the inorganic elements and biological entities interacting with it.

Finally, I would like to emphasize that my multidisciplinary approach, supported by a wealth of analytical data, is not a philosophical one. On the contrary, it is based on biological and physical mechanisms and is not intended to have any teleological implications. Importantly, the examples I have discussed are viewed as the result of chance and natural selection and not the consequence of something that was planned.

I am confident that the rational approach based on logical precepts adopted in the present paper will ultimately be appreciated by the scientific community. This work is a sort of message in a bottle that, sooner or later, I hope will go “viral” and inspire other scientists to make improvements in their respective scientific fields and to explore new areas of research. Hence, I would be grateful for any constructive criticism or corrections that might improve my personal and limited scientific view. Indeed, the mechanisms unveiled in the present work raise new issues and offer novel perspectives which could have repercussions in a wide range of scientific fields. Personally, I already have in my mind several scientific “stories” that I would like to tell in future works. In particular, I would like to explore several unexpected possible consequences stemming from the unveiled tools of evolvability, including explanations for several evolutionarily open issues and new strategies to prevent and counter chronic diseases.

If we can improve our understanding of where we come from and how we came to be, we can get a more complete picture of the present and perhaps even foresee the future. Otherwise we will only have an instantaneous and incomplete picture of the present.

**Declarations section**

**- Ethical Approval and Consent to participate**

Not applicable.

**- Consent for publication**

Not applicable.

**- Availability of supporting data**

Data generated, if any, are included in the article.

**- Competing interests**

The author declares that they have no competing interests.

**- Funding**

The work was supported by Italian grants "Progetti di Valorizzazione DiSB 2017 and 2018" of the University of Urbino Carlo Bo**.** There was a role of the funding body in editing the English.

**- Authors' contributions**

LZ is sole author and sole investigator. The author conceived of the article, wrote it and designed the figures. LZ read and approved the final manuscript.

**- Acknowledgements**

I am grateful to several friends who have dedicated time to listening to my “crazy” ideas. In alphabetical order: Luigi Bagella, Maria Balsamo, Federico Bastianelli, Gianni Braccioforte, Mauro Casalicchio, Angelo Clini, Pierandrea De Iaco, Paolo Del Grande, Alessandro Faragò, Luca Galluzzi, Guglielmo Giuliano, Armando Gregorini, Nadir Mario Maraldi, Alessandro Minelli, Rita Monticelli, Isabel Permanyer, Massimo Ranieri, Donata Ricci, Mariano Rocchi, Arturo Sala, Giorgia Santilli, Cinzia Sevignani, Carolina Simioni, Claudio Sorio and Guglielmo Vecchietti Massacci, my wife, Giuliana Benvenuti and my daughters, Chiara and Elena. I also wish to thank Dr. Gemma Nasoni for helping with the figure preparation and Prof. Timothy Bloom for his patience in editing the English. I would also like to apologize to those whose relevant contributions could not be cited because of my limits.

**- Author information**

**Affiliations**

Loris Zamai, Department of Biomolecular Sciences, via Ca’ le Suore 2, University of Urbino Carlo Bo, 61029 Urbino and INFN-Gran Sasso National Laboratory, Assergi, L'Aquila, Italy.

**Corresponding author**

Correspondence to Loris Zamai: loris.zamai@uniurb.it

**References**

**Abyzov A, Mariani J, Palejev D, Zhang Y, Haney MS, Tomasini L, Ferrandino AF, Rosenberg Belmaker LA, Szekely A, Wilson M, Kocabas A, Calixto NE, Grigorenko EL, Huttner A, Chawarska K, Weissman S, Urban AE, Gerstein M, Vaccarino FM. Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. Nature. 2012;492(7429):438-42. doi: 10.1038/nature11629.**

**Agrawal L, Lu X, Qingwen J, VanHorn-Ali Z, Nicolescu IV, McDermott DH, Murphy PM, Alkhatib G. Role for CCR5Delta32 protein in resistance to R5, R5X4, and X4 human immunodeficiency virus type 1 in primary CD4+ cells. J Virol. 2004;78(5):2277-87.**

**Ajon M, Fröls S, van Wolferen M, Stoecker K, Teichmann D, Driessen AJ, Grogan DW, Albers SV, Schleper C. UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili. Mol Microbiol. 2011;82(4):807-17. doi: 10.1111/j.1365-2958.2011.07861.x.**

**Arjen E. van't Hof, Pascal Campagne, Daniel J. Rigden, Carl J. Yung, Jessica Lingley, Michael A. Quail, Neil Hall, Alistair C. Darby, Ilik J. Saccheri. The industrial melanism mutation in British peppered moths is a transposable element. Nature 534, 102–105. 2016**

**Babich V, Aksenov N, Alexeenko V, Oei SL, Buchlow G, Tomilin N. Association of some potential hormone response elements in human genes with the Alu family repeats. Gene. 1999;239(2):341-9.**

**Bach D, Peddi S, Mangeat B, Lakkaraju A, Strub K, Trono D. Characterization of APOBEC3G binding to 7SL RNA. Retrovirology. 2008;5:54. doi: 10.1186/1742-4690-5-54.**

**Balcer-Kubiczek EK, Yin J, Lin K, Harrison GH, Abraham JM, Meltzer SJ. p53 mutational status and survival of human breast cancer MCF-7 cell variants after exposure to X rays or fission neutrons. Radiat Res 1995;142(3):256-62.**

**Barber JC. Directly transmitted unbalanced chromosome abnormalities and euchromatic variants. J Med Genet. 2005;42(8):609-29.**

**Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, Cullen BR. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci U S A. 2006;103(23):8780-5.**

**Bombelli P, Howe CJ, Bertocchini F. Polyethylene bio-degradation by caterpillars of the wax moth Galleria mellonella. Curr Biol. 2017;27(8):R292-3. doi: 10.1016/j.cub.2017.02.060.**

**Caval V, Suspène R, Shapira M, Vartanian JP, Wain-Hobson S. A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3'UTR enhances chromosomal DNA damage. Nat Commun. 2014;5:5129. doi: 10.1038/ncomms6129.**

**Chan K, Gordenin DA. Clusters of Multiple Mutations: Incidence and Molecular Mechanisms. Annu Rev Genet. 2015;49:243-67. doi: 10.1146/annurev-genet-112414-054714.**

**Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, Kim J, Kwiatkowski DJ, Fargo DC, Mieczkowski PA, Getz G, Gordenin DA. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet 2015;47(9):1067-72. doi:10.1038/ng.3378.**

**Cheng Z, Menees TM. RNA branching and debranching in the yeast retrovirus-like element Ty1. Science 2004, 303:240–243.**

**Chiu YL, Witkowska HE, Hall SC, Santiago M, Soros VB, Esnault C, Heidmann T, Greene WC. High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. Proc Natl Acad Sci U S A. 2006;103(42):15588-93.**

**Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science. 2016;351(6277):1083-7. doi: 10.1126/science.aad5497.**

**Cook, LM. The rise and fall of the Carbonaria form of the peppered moth. Q Rev Biol. 2003;78(4):399-417. doi:10.1086/378925.**

**De Vito A, Lazzaro M, Palmisano I, Cittaro D, Riba M, Lazarevic D, Bannai M, Gabellini D, Schiaffino MV. Amino acid deprivation triggers a novel GCN2-independent response leading to the transcriptional reactivation of non-native DNA sequences. PLoS One. 2018;13(7):e0200783. doi:10.1371/journal.pone.0200783.**

**Dimitriu T, Ashby B, Westra ER. Transposition: a CRISPR way to get around. Curr Biol. 2019;29:R866-9. doi: 10.1016/j.cub.2019.08.010**

**Durzyńska J, Goździcka-Józefiak A. Viruses and cells intertwined since the dawn of evolution. Virol J. 2015;12:169. doi: 10.1186/s12985-015-0400-7.**

**Enard D, Cai L, Gwennap C, Petrov DA. Viruses are a dominant driver of protein adaptation in mammals. Elife. 2016;5. pii: e12469. doi: 10.7554/eLife.12469.**

**Feng Y, Goubran MH, Follack TB, Chelico L. Deamination-independent restriction of LINE-1 retrotransposition by APOBEC3H. Sci Rep. 2017;7(1):10881. doi: 10.1038/s41598-017-11344-4.**

**Feschotte C. The contribution of transposable elements to the evolution of regulatory networks. Nat Rev Genet. 2008; 9(5): 397–405.**

**Fitzgerald DM, Hastings PJ, Rosenberg SM. Stress-Induced Mutagenesis: Implications in Cancer and Drug Resistance. Annu Rev Cancer Biol. 2017; 1:119-140.**

**Fungtammasan A, Walsh E, Chiaromonte F, Eckert KA, Makova KD. A genome-wide analysis of common fragile sites: what features determine chromosomal instability in the human genome? Genome Res. 2012;22(6):993-1005. doi: 10.1101/gr.134395.111.**

**Gabrielli S, Ortolani C, Del Zotto G, Luchetti F, Canonico B, Buccella F, Artico M, Papa S, Zamai L. The Memories of NK Cells: Innate-Adaptive Immune Intrinsic Crosstalk. J Immunol Res. 2016; 2016:1376595. doi: 10.1155/2016/1376595.**

**Galhardo RS, Hastings PJ, Rosenberg SM. Mutation as a stress response and the regulation of evolvability. Crit Rev Biochem Mol Biol. 2007;42(5):399-435.**

**Garagna S, Zuccotti M, Redi CA, Capanna E. Trapping speciation. Nature. 1997;390(6657):241-2.**

**Gilbert SF, Bosch TC, Ledón-Rettig C. Eco-Evo-Devo: developmental symbiosis and developmental plasticity as evolutionary agents. Nat Rev Genet. 2015;16(10):611-22. doi: 10.1038/nrg3982.**

**Gottlieb B, Beitel LK, Alvarado C, Trifiro MA. Selection and mutation in the ‘‘new’’ genetics: an emerging hypothesis. Hum Genet 2010;127(5):491–501**

**Gould SJ, Eldredge N. Punctuated equilibria comes of age. Nature. 1993; 366:223–227.**

**Haerter JO, Sneppen K. Spatial structure and Lamarckian adaptation explain extreme genetic diversity at CRISPR locus. MBio. 2012;3(4):e00126-12. doi: 10.1128/mBio.00126-12.**

**He F, Vestergaard G, Peng W, She Q, Peng X. CRISPR-Cas type I-A Cascade complex couples viral infection surveillance to host transcriptional regulation in the dependence of Csa3b. Nucleic Acids Res. 2017;45(4):1902-1913. doi: 10.1093/nar/gkw1265.**

**Hesselberth JR. Lives that introns lead after splicing. WIREs RNA 2013, 4:677–691.**

**Holmes EC. What does virus evolution tell us about virus origins? J Virol. 2011;85:5247–51.**

**Horn AV, Klawitter S, Held U, Berger A, Vasudevan AA, Bock A, Hofmann H, Hanschmann KM, Trösemeier JH, Flory E, Jabulowsky RA, Han JS, Löwer J, Löwer R, Münk C, Schumann GG. Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity. Nucleic Acids Res. 2014;42(1):396-416. doi: 10.1093/nar/gkt898.**

**Hulme AE, Bogerd HP, Cullen BR, Moran JV. Selective inhibition of Alu retrotransposition by APOBEC3G. Gene. 2007;390(1-2):199-205.**

**Imamichi H, Dewar RL, Adelsberger JW, Rehm CA, O'Doherty U, Paxinos EE, Fauci AS, Lane HC. Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. Proc Natl Acad Sci U S A. 2016;113(31):8783-8. doi: 10.1073/pnas.1609057113.**

**Ito F, Fu Y, Kao SA, Yang H, Chen XS. Family-Wide Comparative Analysis of Cytidine and Methylcytidine Deamination by Eleven Human APOBEC Proteins. J Mol Biol. 2017;429(12):1787-1799**

**Jangam D, Feschotte C, Betrán E. Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. Trends Genet. 2017;33(11):817-831. doi: 10.1016/j.tig.2017.07.011.**

**Jeffs AR, Benjes SM, Smith TL, Sowerby SJ, Morris CM. The BCR gene recombines preferentially with Alu elements in complex BCR-ABL translocations of chronic myeloid leukaemia. Hum Mol Genet. 1998;7(5):767-76.**

**Kanu N, Cerone MA, Goh G, Zalmas LP, Bartkova J, Dietzen M, McGranahan N, Rogers R, Law EK, Gromova I, Kschischo M, Walton MI, Rossanese OW, Bartek J, Harris RS, Venkatesan S, Swanton C. DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer. Genome Biol. 2016;17(1):185. doi: 10.1186/s13059-016-1042-9.**

**Karst SM, Rutz ML, Menees TM. The yeast retrotransposons Ty1 and Ty3 require the RNA Lariat debranching enzyme, Dbr1p, for efficient accumulation of reverse transcripts. Biochem Biophys Res Commun. 2000, 268:112–117.**

**Katsnelson MI, Wolf YI, Koonin EV. On the feasibility of saltational evolution. Proc Natl Acad Sci U S A. 2019;116(42):21068-21075. doi: 10.1073/pnas.1909031116..**

**Kazlauskas D, Varsani A, Koonin EV, Krupovic M. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. 2019 Nat Commun. 2019;10(1):3425. doi: 10.1038/s41467-019-11433-0.**

**Kerr RA. Mega-eruptions drove the mother of mass extinctions. Science. 2013;342(6165):1424. doi: 10.1126/science.342.6165.1424.**

**Kinomoto M, Kanno T, Shimura M, Ishizaka Y, Kojima A, Kurata T, Sata T, Tokunaga K.All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. Nucleic Acids Res. 2007;35(9):2955-64.**

**Klompe SE, Vo PLH, Halpin-Healy TS, Sternberg SH. Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. Nature. 2019;571(7764):219-225. doi: 10.1038/s41586-019-1323-z.**

**Koito A, Ikeda T. Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases. Front Microbiol. 2013;4:28. doi: 10.3389/fmicb.2013.00028. eCollection 2013.**

**Koonin EV, Kuprovic M. Evolution of adaptive immunity from transposable elements combined with innate immune systems. Nat Rev Genet. 2015;16(3):184-92. doi: 10.1038/nrg3859.**

**Koonin EV, Makarova KS, Wolf YI, Krupovic M. Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. Nat Rev Genet. 2019 Oct 14. doi: 10.1038/s41576-019-0172-9.**

**Koonin EV, Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol. 2017;37:67-78. doi: 10.1016/j.mib.2017.05.008.**

**Koonin EV, Wolf YI. Is evolution Darwinian or/and Lamarckian? Biol Direct. 2009;4:42. doi: 10.1186/1745-6150-4-42.**

**Kuprovic M, Dolja VV, Koonin EV. Origin of viruses: primordial replicators recruiting capsids from hosts. Nat Rev Microbiol. 2019;17(7):449-458. doi: 10.1038/s41579-019-0205-6.**

**Kuprovic M, Koonin EV. Multiple origins of viral capsid proteins from cellular ancestors Proc Natl Acad Sci U S A. 2017;114(12):E2401-E2410. doi: 10.1073/pnas.1621061114.**

**Laland K, Uller T, Feldman M, Sterelny K, Müller GB, Moczek A, Jablonka E, Odling-Smee J, Wray GA, Hoekstra HE, Futuyma DJ, Lenski RE, Mackay TF, Schluter D, Strassmann JE. Does evolutionary theory need a rethink? Nature. 2014;514(7521):161-4. doi: 10.1038/514161a.**

**Lee Y, Rio DC. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem. 2015;84:291-323. doi: 10.1146/annurev-biochem-060614-034316.**

**Li Q, Wrange O, Eriksson P. The role of chromatin in transcriptional regulation. Int J Biochem Cell Biol 1997;29(5):731-42.**

**Li WY, Reddy R, Henning D, Epstein P, Busch H. Nucleotide sequence of 7 S RNA. Homology to Alu DNA and La 4.5 S RNA. J Biol Chem. 1982;257(9):5136-42.**

**Li Z, Abraham BJ, Berezovskaya A, Farah N, Liu Y, Leon T, Fielding A, Tan SH, Sanda T, Weintraub AS, Li B, Shen S, Zhang J, Mansour MR, Young RA, Look AT. APOBEC signature mutation generates an oncogenic enhancer that drives LMO1 expression in T-ALL. Leukemia. 2017;31(10):2057-2064.**

**Liang J, Dill KA. Are proteins well-packed?. Biophys J. 2001;81(2):751–766. doi:10.1016/S0006-3495(01)75739-6**

**Lin C, Yang L, Tanasa B, Hutt K, Ju BG, Ohgi K, Zhang J, Rose DW, Fu XD, Glass CK, Rosenfeld MG. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell. 2009;139(6):1069-83. doi: 10.1016/j.cell.2009.11.030.**

Losa GA. The fractal geometry of life. Riv Biol. 2009;102(1):29-59.

**Lozynskyi R, Lozynska M. The study of the mechanisms of the different phenotypical manifestations in patients with reciprocal translocations. Biophotonics and New Therapy Frontiers. 2006; 6191, 386-397. doi: 10.1117/12.663011.**

**McClintock B. The significance of responses of the genome to challenge. Science. 1984; 226(4676):792-801.**

**McHeyzer-Williams M, Okitsu S, Wang N, McHeyzer-Williams L. Molecular programming of B cell memory. Nat Rev Immunol. 2011 Dec 9;12(1):24-34. doi: 10.1038/nri3128.**

**Medstrand P, van de Lagemaat LN, Mager DL. Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome Res. 2002;12(10):1483-95.**

**Miller WJ1, McDonald JF, Pinsker W. Molecular domestication of mobile elements. Genetica. 1997;100(1-3):261-70.**

**Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? Trends Genet. 2007;23(4):183-91.**

**Mourier T, Nielsen LP, Hansen AJ, Willerslev E. Transposable elements in cancer as a by-product of stress-induced evolvability. Front Genet. 2014;5:156. doi: 10.3389/fgene.2014.00156. eCollection 2014.**

**Muckenfuss H, Hamdorf M, Held U, Perkovic M, Löwer J, Cichutek K, Flory E, Schumann GG, Münk C. APOBEC3 proteins inhibit human LINE-1 retrotransposition. J Biol Chem. 2006;281(31):22161-72.**

**Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, Muthuswamy L, Krasnitz A, McCombie WR, Hicks J, Wigler M. Tumour evolution inferred by single-cell sequencing. Nature 2011;472(7341):90-4. doi: 10.1038/nature09807.**

**Nadassy K, Tomás-Oliveira I, Alberts I, Janin J, Wodak SJ. Standard atomic volumes in double-stranded DNA and packing in protein--DNA interfaces. Nucleic Acids Res. 2001;29(16):3362–3376. doi:10.1093/nar/29.16.3362**

**Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell. 2014;156(5):935-49.**

**Norman PJ, Abi-Rached L, Gendzekhadze K, Korbel D, Gleimer M, Rowley D, Bruno D, Carrington CV, Chandanayingyong D, Chang YH, Crespí C, Saruhan-Direskeneli G, Fraser PA, Hameed K, Kamkamidze G, Koram KA, Layrisse Z, Matamoros N, Milà J, Park MH, Pitchappan RM, Ramdath DD, Shiau MY, Stephens HA, Struik S, Verity DH, Vaughan RW, Tyan D, Davis RW, Riley EM, Ronaghi M, Parham P. Unusual selection on the KIR3DL1/S1 natural killer cell receptor in Africans. Nat Genet. 2007;39(9):1092-9.**

**O'Huallachain M, Karczewski KJ, Weissman SM, Urban AE, Snyder MP. Extensive genetic variation in somatic human tissues. Proc Natl Acad Sci USA. 2012;109(44):18018–23.**

**Palmisano I, Della Chiara G, D'Ambrosio RL, Huichalaf C, Brambilla P, Corbetta S, Riba M, Piccirillo R, Valente S, Casari G, Mai A, Martinelli Boneschi F, Gabellini D, Poli G, Schiaffino MV. Amino acid starvation induces reactivation of silenced transgenes and latent HIV-1 provirus via down-regulation of histone deacetylase 4 (HDAC4). Proc Natl Acad Sci U S A. 2012;109(34):E2284-93. doi: 10.1073/pnas.1202174109.**

**Patel BH, Percivalle C, Ritson DJ, Duffy CD, Sutherland JD. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. Nat Chem. 2015;7(4):301-7. doi: 10.1038/nchem.2202.**

**Periyasamy M, Patel H, Lai CF, Nguyen VTM, Nevedomskaya E, Harrod A, Russell R, Remenyi J, Ochocka AM, Thomas RS, Fuller-Pace F, Győrffy B, Caldas C, Navaratnam N, Carroll JS, Zwart W, Coombes RC, Magnani L, Buluwela L, Ali S. APOBEC3B-Mediated Cytidine Deamination Is Required for Estrogen Receptor Action in Breast Cancer. Cell Rep. 2015;13(1):108-121. doi: 10.1016/j.celrep.2015.08.066.**

**Piacentini L, Fanti L, Specchia V, Bozzetti MP, Berloco M, Palumbo G, Pimpinelli S. Transposons, environmental changes, and heritable induced phenotypic variability. Chromosoma. 2014;123(4):345-54. doi: 10.1007/s00412-014-0464-y.**

**Pinto Y, Gabay O, Arbiza L, Sams AJ, Keinan A, Levanon EY. Clustered mutations in hominid genome evolution are consistent with APOBEC3G enzymatic activity. Genome Res. 2016;26(5):579-87. doi: 10.1101/gr.199240.115.**

**Platt RN, Vandewege MW, Ray DA. Mammalian transposable elements and their impacts on genome evolution. Chromosome Res. 2018;26(1-2):25-43. doi: 10.1007/s10577-017-9570-z.**

**Polak P, Domany E. Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. BMC Genomics. 2006;7:133. doi: 10.1186/1471-2164-7-133**

**Pope KO, D'Hondt SL, Marshall CR. Meteorite impact and the mass extinction of species at the Cretaceous/Tertiary boundary. Proc Natl Acad Sci U S A. 1998;95(19):11028-9.**

**Rajagopal C, Harikumar KB. The Origin and Functions of Exosomes in Cancer. Front Oncol 2018;8:66. doi: 10.3389/fonc.2018.00066.**

**Ratner HK, Sampson TR, Weiss DS. I can see CRISPR now, even when phage are gone: a view on alternative CRISPR-Cas functions from the prokaryotic envelope. Curr Opin Infect Dis. 2015;28(3):267-74. doi: 10.1097/QCO.0000000000000154.**

**Riley, PA. A proposed selective mechanism based on metal chelation in industrial melanic moths. Biological Journal of the Linnean Society. 2013;109 (2):298–301. doi:10.1111/bij.1206**

**Robertson MP, Joyce GF. The origins of the RNA world. Cold Spring Harb Perspect Biol. 2012;4(5). pii: a003608. doi: 10.1101/cshperspect.a003608.**

**Roukos V, Voss TC, Schmidt CK, Lee S, Wangsa D, Misteli T. Spatial dynamics of chromosome translocations in living cells. Science 2013;341(6146):660-4. doi: 10.1126/science.1237150.**

**Roy-Engel AM. LINEs, SINEs and other retroelements: do birds of a feather flock together? Front Biosci (Landmark Ed). 2012; 17: 1345–1361.**

**Salter JD, Bennett RP, Smith HC. The APOBEC Protein Family: United by Structure, Divergent in Function. Trends Biochem Sci. 2016;41(7):578-594. doi: 10.1016/j.tibs.2016.05.001.**

**Scott WG, Klug A. Ribozymes: structure and mechanism in RNA catalysis. Trends Biochem Sci. 1996;21(6):220-4.**

**Seed KD, Lazinski DW, Calderwood SB, Camilli A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature. 2013;494(7438):489-91. doi: 10.1038/nature11927.**

**Senavirathne G, Bertram JG, Jaszczur M, Chaurasiya KR, Pham P, Mak CH, Goodman MF, Rueda D. Activation-induced deoxycytidine deaminase (AID) co-transcriptional scanning at single-molecule resolution. Nat Commun. 2015;6:10209. doi: 10.1038/ncomms10209.**

**Song J, Ii X, Sun L, Xu S, Liu N, Yao Y, Liu Z, Wang W, Rong H, Wang B. A family with Robertsonian translocation: a potential mechanism of speciation in humans. Mol Cytogenet. 2016; 9: 48. doi: 10.1186/s13039-016-0255-7**

**Song J, Mercer D, Hu X, Liu H, Li MM. Common leukemia- and lymphoma-associated genetic aberrations in healthy individuals. J Mol Diagn. 2011;13(2):213-9. doi: 10.1016/j.jmoldx.2010.10.009.**

**Spadafora C. Sperm-Mediated Transgenerational Inheritance. Front Microbiol 2017;8:2401. doi:10.3389/fmicb.2017.02401.**

**St Laurent G, Shtokalo D, Tackett MR, Yang Z, Eremina T, Wahlestedt C, Urcuqui-Inchima S, Seilheimer B, McCaffrey TA, Kapranov P. Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. BMC Genomics. 2012;13:504. doi: 10.1186/1471-2164-13-504.**

**Stavrou S, Ross SR. APOBEC3 Proteins in Viral Immunity. J Immunol. 2015;195(10):4565-70. doi: 10.4049/jimmunol.1501504.**

**Steele EJ, Franklin A, Blanden RV. Genesis of the strand-biased signature in somatic hypermutation of rearranged immunoglobulin variable genes. Immunol Cell Biol. 2004;82(2):209-18. doi: 10.1046/j.0818-9641.2004.01224.x**

**Stenglein MD, Harris RS. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. J. Biol. Chem. 2006; 281:16837–16841.**

**Strecker J, Ladha A, Gardner Z, Schmid-Burgk JL, Makarova KS, Koonin EV, Zhang F. RNA-guided DNA insertion with CRISPR-associated transposases. Science. 2019;365(6448):48-53. doi:10.1126/science.aax9181.**

**Talhouarne GJS, Gall JG. Lariat intronic RNAs in the cytoplasm of vertebrate cells. Proc Natl Acad Sci U S A. 2018;115(34):E7970-E7977. doi: 10.1073/pnas.1808816115.**

**Trivedi S, Gunwant S, Swati P. Analysis of similarities between viroid, prokaryote and eukaryote genomes to revisit theories of origin of viroids. J Cell & Mol Biol. 2007;6(1): 9–18.**

**Tsantoulis PK, Kotsinas A, Sfikakis PP, Evangelou K, Sideridou M, Levy B, Mo L, Kittas C, Wu XR, Papavassiliou AG, Gorgoulis VG. Oncogene-induced replication stress preferentially targets common fragile sites in preneoplastic lesions. A genome-wide study. Oncogene. 2008;27(23):3256-64.**

**Tsou SH, Chen TM, Hsiao HT, Chen YH. A critical dose of doxorubicin is required to alter the gene expression profiles in MCF-7 cells acquiring multidrug resistance. PLoS One 2015;10(1):e0116747. doi: 10.1371/journal.pone.0116747.**

**Van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ.The industrial melanism mutation in British peppered moths is a transposable element. Nature. 2016;534(7605):102-5. doi: 10.1038/nature17951.**

**Volff JN. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays. 2006; 28:913–22.**

**Waddington CH. Canalization of development and genetic assimilation of acquired characters. Nature. 1959; 183(4676):1654-55.**

**Westra ER, Buckling A, Fineran PC. CRISPR-Cas systems: beyond adaptive immunity. Nat. Rev. Microbiol. 2014;12, 317–326.**

**Wildschutte JH, Williams ZH, Montesion M, Subramanian RP, Kidd JM, Coffin JM. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. Proc Natl Acad Sci U S A. 2016; 113(16):E2326-34.**

**Wilson, David Sloan; Wilson, E. O. (2008). "Evolution 'for the good of the group' ". American Scientist. 96 (5): 380–389. doi:10.1511/2008.74.1 (https://doi.org/10.1511%2F2008.74.1).**

**Wissing S, Montano M, Garcia-Perez JL, Moran JV, Greene WC. Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. J Biol Chem. 2011;286(42):36427-37. doi: 10.1074/jbc.M111.251058.**

**Wright AV, Nuñez JK, Doudna JA. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell. 2016;164(1-2):29-44. doi: 10.1016/j.cell.2015.12.035.**

**Wright BE. A biochemical mechanism for nonrandom mutations and evolution. J Bacteriol. 2000;182(11):2993-3001.**

**Zamai L, Ponti C, Mirandola P, Gobbi G, Papa S, Galeotti L, Cocco L, Vitale M. NK cells and cancer. J Immunol. 2007;178(7):4011-6.**

**Zhang W, Du J, Yu K, Wang T, Yong X, Yu XF. Association of potent human antiviral cytidine deaminases with 7SL RNA and viral RNP in HIV-1 virions. J Virol. 2010;84(24):12903-13. doi: 10.1128/JVI.01632-10.**

**Zhang Y, Romanish MT, Mager DL. Distributions of transposable elements reveal hazardous zones in mammalian introns. PLoS Comput Biol. 2011;7(5):e1002046. doi: 10.1371/journal.pcbi.1002046.**

**Zheng S, Vuong BQ, Vaidyanathan B, Lin JY, Huang FT, Chaudhuri J. Non-coding RNA generated following lariat debranching mediates targeting of AID to DNA. Cell 2015; 161:762-773.**

**Zheng Y, Lorenzo C, Beal PA. DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA. Nucleic Acids Res 2017;45(6):3369-3377. doi: 10.1093/nar/gkx050.**