Genome evolution is driven by gene expression-generated biophysical constraints through RNA-directed genetic variation: A hypothesis

Didier Auboeuf

The biogenesis of RNAs and proteins is a threat to the cell. Indeed, the act of transcription and nascent RNAs challenge DNA stability. Both RNAs and nascent proteins can also initiate the formation of toxic aggregates because of their physicochemical properties. In reviewing the literature, I show that co-transcriptional and co-translational biophysical constraints can trigger DNA instability that in turn increases the likelihood that sequences that alleviate the constraints emerge over evolutionary time. These directed genetic variations rely on the biogenesis of small RNAs that are transcribed directly from challenged DNA regions or processed from the transcripts that directly or indirectly generate constraints or aggregates. These small RNAs can then target the genomic regions from which they initially originate and increase the local mutation rate of the targeted loci. This mechanism is based on molecular pathways involved in anti-parasite genome defence systems, and implies that gene expression-related biophysical constraints represent a driving force of genome evolution.

Keywords:
- DNA instability; evolution; gene expression; RNA; RNA processing; translation

Introduction

The modern synthesis theory of evolution postulates that the evolution of living organisms results from random mutations that are selected or filtered by natural or sexual selection, or that are under neutral evolution when they do not affect gene product functions. “Random mutations” means that mutations occur at random locations in the genomic DNA or that mutations occur randomly with respect to the biological effects they may drive. However, increasing evidence indicates that the DNA location of mutations is not random. For example, the sequencing of an increasing number of human genomes from parent-offspring trios indicates that de novo mutations in the germline are not randomly distributed across the genome [1–5]. It seems instead that the DNA location of mutations is associated with local genomic factors including local sequence context (e.g. CpG dinucleotides), DNA biochemical (e.g. methylation) and structural features (e.g. non-B-DNA structures), or chromatin marks and topology [1–5]. This is likely because these genomic factors impact on different mutational processes such as DNA damage or editing, or processes involved in DNA repair [6–9]. This interplay raises the possibility that some cellular processes could impact on the mutational rate of genomic loci by impacting on local genomic factors.

In this context, increasing evidence indicates that RNA molecules target specific genomic loci through several mechanisms (e.g. base-pairing to complementary nascent RNAs or to complementary DNA strands), and can impact on the local genomic factors described above. RNAs can direct chromatin or DNA biochemical and structural modifications, and can guide to targeted loci, enzymes involved in DNA sequence modifications (e.g. DNA endonucleases and editing enzymes) [10–13]. RNAs are also involved in DNA repair and recombination, and they direct complex DNA rearrangements [14–16]. These
observations have led several authors to propose that RNAs are involved in “re-writing” the genome [14–19]. Consequently, some biological processes could result in the biogenesis of RNAs that target complementary genomic loci and that modulate the mutation rate of the targeted-loci. If such a process exists, where might these RNAs originate, and could they increase the local mutational rate of targeted genomic locations in a non-random way with respect to the biological effects they potentially drive?

While it has long been established that transcription is coupled to translation in prokaryotic cells, one of the most exciting recent discoveries in eukaryotic cells, is that the different steps of the gene expression process are physically coupled. Indeed, transcription is coupled to RNA processing, and mRNA metabolism is coupled to protein metabolism [20, 21]. In this article, I show that the physical proximity or the tight interplay between different steps of gene expression provides the molecular framework for a potential feedback from the gene expression process back to DNA mutation process. More specifically, since the biogenesis of any gene product (i.e. RNA or protein) in the crowded interior of a cell can threaten cellular and genomic integrity, I will propose that the cell might overcome these problems through molecular pathways that facilitate changes to genomic sequences that produce “toxic molecules” (Box 1). Since parasitic nucleic acids are also a threat to the integrity of the cell and its genome, the same molecular pathways, which rely on small RNAs to mutate or remove exogenous toxic nucleic acids, are presumably also at work to modify or remove toxic endogenous nucleic acid sequences.

Co-transcriptional biophysical constraints shape gene architecture

The interplay between co-transcriptional biophysical constraints, DNA instability, and RNA processing drives DNA sequence evolution

It is now clearly established that transcription generates topological and biophysical constraints on DNA with consequences on genome stability. For example, negative and positive supercoils are generated behind and in front, respectively of the transcribing RNA polymerases. This then results in the formation of non-B DNA structures and transcription or replication roadblocks that can induce DNA instability (e.g. DNA breaks) [22, 23]. In addition, the newly synthesized RNA molecules can re-hybridize to the template DNA strand, to create co-transcriptional R-loops (composed of an RNA:DNA hybrid and a displaced single-stranded DNA). These R-loops are a major source of DNA instability [23, 24]. Transcription can therefore challenge DNA integrity (e.g. by inducing DNA breaks) both in an RNA-dependent and -independent manner, through co-transcriptional biophysical constraints.

As well as generating genome instability, co-transcriptional constraints trigger the co-transcriptional processing of nascent RNAs (e.g. splicing, 3’-end RNA processing) in eukaryotic cells. For example, chromatin compactness and transcriptional roadblocks impacts on co-transcriptional splicing or trigger transcription termination, RNA cleavage and 3’-end RNA processing [20, 25–28]. The formation of R-loops has also been involved in 3’-end RNA processing [29, 30]. It is believed that co-transcriptional physical constraints induce pausing of RNA polymerase, which increases the time window required for the co-transcriptional recruitment of RNA processing factors on the nascent RNAs [20, 26, 31, 32].

While co-transcriptional physical constraints impact on co-transcriptional RNA processing, increasing evidence indicates that co-transcriptional RNA processing itself alleviates these constraints, and protects DNA from transcriptional-mediated damages. Indeed, inhibition of different RNA processing steps, including splicing, leads to genome instability [33–36]. Several mechanisms could explain the protective effect of co-transcriptional RNA processing on DNA integrity. First, RNA-binding proteins involved in RNA processing may coat the nascent RNA and prevent it from hybridizing back to the DNA template. Supporting this model, depletion of several splicing factors lead to R-loop formation and genome instability [33–39]. Second, co-transcriptional RNA processing could favor the removal of newly synthesized transcripts from chromatin since splicing is coupled to exon junction complex recruitment, which contributes to co-transcriptional RNA packaging and RNA export. Supporting this model, depletion of factors involved in the coupling between RNA processing, RNA packaging and export, results in DNA instability [40, 41].

Box 1

The concept of toxicity used throughout this manuscript must be understood in a general sense and covers several notions. For example, transcription is toxic to DNA because it can cause DNA breaks. RNAs also have some degree of toxicity to DNA since they can re-hybridize to DNA and generate R-loops that are genotoxic. RNAs produced from some genomic repeated elements, like retrotransposons, are toxic to genomes because they can invade it and alter their functioning by interfering with the activity of genes. RNAs are also potentially toxic because of their ability to form molecular aggregates that are toxic to the cell as it is now clearly established in a variety of pathologies. Similarly, protein biogenesis can lead to the formation of toxic molecular aggregates that can be linked, for example, to nascent protein misfolding. The concept of toxicity used in this manuscript relies therefore on the notions that (i) a genome is used by a cell to produce molecules, RNAs, some of which being used to produce proteins; (ii) these biogenesis processes are potentially “dangerous” for the cell because the act of production of these molecules or these molecules themselves can create deleterious biophysical constraints within the crowded intracellular environment.
prokaryotes may have a similar protective role to eukaryotic co-transcriptional RNA processing, by taking nascent RNA off the DNA template [42].

Because of the interplay between co-transcriptional biophysical constraints, DNA instability and co-transcriptional RNA processing, I propose that co-transcriptional biophysical constraints within a transcriptional unit increase DNA damage locally, that is, in the vicinity of the physical constraints (Fig. 1). As long as the constraints persist, the resulting DNA instability increases the probability that mutations occur at the challenged locus. The high local mutational rate may lead to the emergence of RNA processing sites over evolutionary time because they alleviate transcription-mediated genotoxicity and thus increase the DNA stability of their host gene.

Are small RNAs involved in directing genetic variations in response to transcription-induced genome instability?

The mutational process in Fig. 1 that ultimately overcomes co-transcriptional DNA instability likely involves small RNAs. Indeed, small noncoding RNAs are produced in the vicinity of double-stranded DNA break (DSB) sites [43–46]. These small RNAs can induce chromatin modifications to assist DSB repair or drive the local recruitment of proteins involved in DNA repair [43–47]. Recent work also suggested that RNAs can be used as templates for homologous recombination in bacteria, yeast, and human. Indeed, RNAs can anneal to complementary DNA sequences and serve as templates for DNA repair by reverse transcription [48–51]. RNAs can also direct endonucleases or DNA-editing enzymes, like the activation-induced cytidine deaminase, to genomic loci, which results in increased DNA instability and mutation there [12, 13, 16].

Collectively, these observations raise the possibility that transcription-mediated DNA damage induces the biogenesis of small RNAs (either by transcription or nascent RNA cleavage) that subsequently modulates the local mutational rate. It would be interesting to look, in the future, at whether co-transcriptional biophysical constraints induce specific types of small RNAs, DNA damage and mutations that are more likely to generate splicing or polyadenylation sites or some RNA processing regulatory sequences. For example, since co-transcriptional R-loops are made of a displaced single-stranded DNA that is the preferential substrate for DNA editing enzymes, it would be interesting to test whether the mutations mediated by DNA-editing enzymes increase the generation of RNA (regulatory) processing sites. Even if this is not the case, the important take home message from Fig. 1, is that local DNA instability that is induced by co-transcriptional constraints persists until sequences (e.g. RNA processing sites) that alleviate the co-transcriptional physical constraints emerge over evolutionary time.

Understanding genomic features in light of co-transcriptional physical constraints driving DNA sequence evolution

The mutational process hypothesis outlined in Fig. 1 implies that biophysical parameters that impact co-transcriptional events (and that therefore more-or-less directly impact DNA stability) can be modified during evolution not because of randomly located mutations. Rather, mutations preferentially occur where co-transcriptional parameters induce genomic instability. In prokaryotes, this model implies that transcriptional elongation rate directly contributes to codon-use evolution. Indeed, a fast transcriptional elongation rate would create genetic instability when the nascent RNA is not efficiently translated. Therefore, RNA-mediated genotoxicity would favor the emergence of codons that synchronize the kinetic parameters of translation and transcription [42, 52].
A consequence in eukaryotes of the mutational process in Fig. 1 is the emergence, over evolutionary time, of a large number of cryptic and alternative RNA processing sites. Indeed, if co-transcriptional constraints trigger both DNA instability and RNA processing, which in turn alleviates these constraints, transcription-mediated genome instability would constantly favor the emergence of new RNA processing sites. This conclusion leads on to a novel interpretation of the huge number of available RNA processing sites (splicing and polyadenylation sites) present in most eukaryotic genes. It is often assumed that alternative RNA processing sites are generated by random mutations and are then selected during evolution depending on the cellular functions of the gene products they allow the synthesis [53]. However, it is conceivable that RNA processing sites are generated over evolutionary time because of the DNA instability triggered by co-transcriptional biophysical constraints. Neo-formed RNA processing sites would be passed down the generations because they contribute to the genetic stability of the transcribed loci they are embedded in. Neo-formed RNA processing sites within a gene would next be filtered depending on their potential impact on the cellular function of the gene products (see below).

The interplay between co-transcriptional constraints, DNA instability and co-transcriptional RNA processing could also contribute to explain the evolutionary success of some mobile elements, like the Alu elements in primates. Alu elements, belonging to a class of retroelements termed SINEs (short interspersed elements), contribute to 11% of the human genome. Despite the fact that the expansion of these elements increases the size of transcribed genomic regions and is a threat to the hosting genome [54–56], Alus may advantage the hosting genome by reducing transcription-mediated genomic instability. Indeed, it is now well established that Alu elements provide both polyadenylation and splicing sites [57–60]. This means that the genotoxicity of Alu elements might be counterbalanced by the ability of these elements to spread RNA processing sites within their host genomes. It is interesting also to note that Alu sequences are co-transcriptionally wrapped up by RNA processing factors and favor nascent RNA folding, which may collectively help to take Alu-containing nascent RNAs off chromatin [61, 62]. Although it is known that T-rich sequences are important for Alu element insertion [59], we do not know yet the rules that could direct the insertion of Alu elements into specific DNA locations. Could Alu elements be preferentially inserted where co-transcriptional biophysical constraints are creating genomic instability? Alu elements could take advantage of the single-stranded DNA in co-transcriptional R-loops. T-rich sequences that induce RNA polymerase pausing [63] could increase the likelihood of Alu insertion, especially if they are downstream of GC-rich sequences that increase the likelihood of co-transcriptional R-loops [64]. Insertion of Alu elements within these unstable DNA regions would bring pseudo-RNA-processing sites which would bring RNA processing-mediated genome stability. The evolutionary success of Alu elements might therefore be due to their ability to disseminate alternative (or pseudo-) RNA processing sites within their host genomes, as these elements simultaneously stabilize transcribed-genomic loci and increase the molecular diversity generated from these loci [53].

Mutations generated by the transcription-mediated mutational process could next be filtered during evolution based on the gene products’ functions. A mutation may alleviate a co-transcriptional constraint but disturbs the function of the gene product and give rise to deleterious biological consequences. However, the flexibility of RNA processing pathways [65] could “buffer” the potential deleterious effects of newly generated RNA processing sites on gene product functions. Indeed, newly generated and weak alternative RNA processing sites could be used in cases of “emergency,” when RNA polymerase gets stuck in a locus. The strength of RNA processing sites in a given locus therefore likely relies on the evolutionary-controlled equilibrium between the sites’ effects on the stability of the locus and on the functions of its gene products. Having focus above on co-transcriptional biophysical constraints, I will next address the potential interplay between biophysical constraints occurring during translation and the evolution of coding gene sequences.

Co-translational biophysical constraints shape coding sequences

Formation of RNA and protein aggregates is a threat to the cell homeostasis

Before describing how co-translational biophysical constraints might shape coding sequences over evolutionary time, it is important to first highlight that one of the main pitfalls of the gene expression process is the formation of toxic RNA and protein aggregates. Aggregation can result from the increase of the local concentration of proteins and RNAs since their physicochemical properties make them prone to form aggregates. For example, protein aggregates can be seeded by increased local protein concentration, which might be critical at the protein production site and because many proteins contain aggregation-prone intrinsically disordered regions [66, 67]. Protein aggregates can also be initiated by protein unfolding during translation, as peptides emerging from ribosomes can form “spurious” contacts with peptides from the same nascent polypeptide [66, 67]. Likewise, the physicochemical properties of RNAs, their ability to interact with each other through base pairing, and their ability to interact with RNA-binding proteins that contain aggregation-prone intrinsically disordered regions, make the RNAs prone to form aggregates [68, 69]. Therefore, the main question is not why proteins and RNAs form aggregates, but is rather what the mechanisms preventing the formation of aggregates are in the crowded interior of the cells?

One straightforward mechanism is RNA cleavage. For example, it has been shown that co-translational unfolding of nascent proteins can induce co-translational mRNA cleavage. This has been observed during...
the endoplasmic reticulum stress response where an mRNA can be co-translationally cleaved if it is in the process of producing an unfolded nascent protein that is translocating into the endoplasmic reticulum [70]. A similar process occurs widely, free in the cytoplasm, where nascent protein unfolding can induce co-translationally cleaved and thus translation arrest [71, 72]. In addition, several translation-associated processes, including nonsense-mRNA mediated decay, “no-go” decay and the nonstop decay pathways, induce mRNA cleavage if a specific step of the protein synthesis process (e.g. translation termination) is inefficient [73, 74]. It has also been recently demonstrated that synonymous codons affect the kinetics of translation elongation, which impacts co-translationally mRNA cleavage [74–76]. Therefore, mRNAs can be co-translationally cleaved when their translation is inefficient or results in the synthesis of nascent proteins that initiate aggregate formation.

While RNA cleavage is often associated with RNA degradation, there is increasing evidence that cleaved RNAs can give rise to small functional RNAs. There are indeed numerous examples of mature coding and noncoding RNA molecules being cleaved by endoribonucleases and giving rise to small RNAs that regulate diverse biological processes [77–82]. An emerging concept is that cleavage-derived small RNAs are involved in feedback loops, and allow cells to “fight” potentially toxic RNAs. This is illustrated by the piRNA pathway [83–87]. piRNAs were originally identified as small RNAs that are cut out of transcribed retrotransposons. The cleavage of retrotransposon RNAs decreases their ability to invade their host’s genome [88, 89]. In addition, the retrotransposon-derived piRNAs are loaded onto proteins of the Argonaute family, which directs the cleavage of any transcripts that contain retrotransposon-complementary sequences [88–91]. But these small RNAs can also target the genome regions that produce their precursors and induce targeted transcriptional gene silencing [83, 85, 87, 88]. Interestingly, the piRNA pathway is not restricted to retrotransposon-derived RNAs as (i) piRNA-like molecules can also be derived from pseudogenes, 5'- or 3'-UTRs and even coding mRNA sequences [91–95]; (ii) piRNA-like molecules have also been shown to post-transcriptionally regulate mRNAs [91, 94]; (iii) piRNA-like molecules have been shown to target genomic regions that do not contain retrotransposons [96–100]. The piRNA pathway illustrates how toxic RNAs (e.g. retrotransposon RNAs) can be cleaved and trigger the biogenesis of small RNAs that next contribute to targeted-RNA cleavage (i.e. post-transcriptional gene silencing, PTGS) or transcriptional gene silencing (TGS). Both pathways can be described as feedback loops since they inhibit the production or accumulation of the toxic RNAs (Fig. 2).

The question now to be addressed is whether small RNAs deriving from potentially toxic precursor RNAs impact on the precursors’ DNA sequences as part of a cellular process that “fights” against genome-generated toxic RNAs? This possibility is supported by the direct and indirect roles of small RNAs in chromatin and DNA biochemical modifications, DNA repair, and DNA sequence modifications and recombination at targeted loci, as described in the previous part.

In summary, (i) co-translational events can trigger RNA cleavage; (ii) cleaved RNAs can be further processed into small RNAs; and (iii) small RNAs can impact on DNA stability. The next section describes how these molecular pathways could work together to provide a molecular framework to link co-translational biophysical constraints to directed-mutations within coding sequences.

**One sequence can impact on DNA, RNA, and protein features**

Protein coding sequences are clearly shaped by functional constraints depending on the amino acid chain sequence. However, there is now clear evidence that the biophysical processes of protein synthesis and unfolding also contribute to shape coding sequences, as even synonymous sites appear to be under evolutionary constraints [101–105]. It is believed that synonymous sites that are neutral at the
amino acid chain level, are not neutral in terms of quantitative and qualitative parameters of biophysical processes like protein synthesis and folding. The preference for specific synonymous codons depends on their effects on RNA secondary structures, and on the fact that they determine the nature of the anti-codons (tRNAs) to be used during translation. Both RNA secondary structure and codon usage impact on translation kinetics and protein folding [66, 103–105]. Therefore, coding sequences could be selected over evolutionary time, based not only on the encoded amino acids, but also based on their impact on co-translational biophysical processes. Could these co-translational biophysical processes drive genetic variations?

I propose that co-translational biophysical constraints that cause nascent protein mis-folding and aggregate formation, trigger co-translational mRNA cleavage. Cleaved mRNAs could then initiate the biogenesis of small RNAs that next target the loci they originate from and increase the local mutational rate of the targeted loci. A special class of nucleic acid sequences, namely G-rich tracts, illustrates how such a molecular framework could be straightforward.

G-rich DNA or RNA strands can form G-quadruplexes that are topologically polymorphic secondary structures. These structures that are mutation hotspots impact several processes at the DNA, RNA and protein level (Fig. 3, left panel) [106–115]. Because of the G-quadruplex features, RNA molecules containing G-quadruplexes may allow a feedback from translation to DNA mutational processes. Indeed, if a co-translational event is inefficient or altered (e.g., if a nascent peptide is misfolded and initiates aggregation), there is a probability that translationally repressed mRNAs will be co-translationally cleaved in the vicinity of structures, like G-quadruplexes, that reduce the motion of ribosomes and that can act as translational roadblocks (Fig. 3, right panel) [106, 111–115]. These RNA structures may help the recruitment of endoribonucleases and, alternatively, the translationally repressed mRNAs might be cleaved anywhere and trimmed by exoribonucleases that can be blocked at stable RNA secondary structures, like G-quadruplexes [116–119]. Therefore, co-translational biophysical constraints may result in the production of G-quadruplex-containing RNA fragments and mRNA-derived small RNAs. Accordingly, G-quadruplexes are involved in regulating the biogenesis of small RNAs, such as piRNAs and they are present in several kinds of cleavage-derived small RNAs [116, 119, 120]. Therefore, RNA secondary structures like G-quadruplexes might play a role in the biogenesis of small RNAs from precursor RNAs by impacting on RNA cleavage, by protecting RNA fragments from exoribonucleases, or by initiating the biogenesis of small RNAs.

G-quadruplex-containing small RNAs may next target the genomic loci they originate from, by base pairing with nascent RNAs or with strands of opened DNA. Indeed, it has been shown that G-quadruplex-containing RNAs can form stable RNA:DNA hybrids, that is, where G-quadruplexes are made of half RNA and half DNA [109, 121, 122]. These RNA:DNA hybrids form R-loops, which can lead to DNA instability. G-quadruplex containing RNAs may also direct enzymes like DNA-editing enzymes (e.g., Activation-induced cytidine deaminase) to targeted loci. Indeed, it has been shown that after transcription and splicing, the lariats produced from immunoglobulin gene introns that contain repeated sequences (i.e. the switch regions) are de-branched and used for the biogenesis of G-quadruplex-containing small RNAs. These small RNAs can be bound by Activation-induced cytidine deaminase and guide the enzyme to the genomic intronic switch-regions in a sequence-specific manner [13]. The interaction between the G-quadruplex-containing RNAs and one of the DNA switch-region strands may lead to the formation of R-loop structures, within which the single-stranded DNA is the preferential substrate of the Activation-induced cytidine deaminase's enzymatic activity [13]. Deaminated DNA next engages the base excision and mismatch...
RNAs then trigger mutations of the genomic locus that encodes the B protein. This triggers the formation of aggregates and the biogenesis of small RNAs. These small protein B and helps its folding (left panel). If a mutation modifies the affinity of the protein A interactions occur co-translationally. For example, protein A interacts with the nascent protein B and helps its folding (left panel). If a mutation modifies the affinity of the protein A (A', right panel) for its nascent protein B partner and induces its co-translational misfolding, this triggers the formation of aggregates and the biogenesis of small RNAs. These small RNAs then trigger mutations of the genomic locus that encodes the B protein.

Figure 4. A: Heat shock proteins (HSP) bind to nascent polypeptides and help the nascent protein fold (left panel). HSPs are believed to buffer genetic alteration by permitting some variation of protein sequence. In the absence of HSPs (right panel), a larger number of nascent protein misfold, which results in increased mutational load. B: Some protein–protein interactions occur co-translationally. For example, protein A interacts with the nascent protein B and helps its folding (left panel). If a mutation modifies the affinity of the protein A (A'; right panel) for its nascent protein B partner and induces its co-translational misfolding, this triggers the formation of aggregates and the biogenesis of small RNAs. These small RNAs then trigger mutations of the genomic locus that encodes the B protein.

Repair machineries to generate double-stranded DNA breaks, which creates genetic variability within the immunoglobulin loci [13, 123]. This mechanism also occurs outside the immunoglobulin loci ("off-targets") in DNA regions that can form G-quadruplex structures [13, 123]. In conclusion, RNAs containing G-quadruplexes can direct genetic variability.

Other specific DNA and RNA sequences and structures likely play a similar role to G-quadruplexes. Of particular interest are short tandem repeats like trinucleotide repeats that are involved in many genetic diseases. These sequences generate biophysical constraints during DNA replication and transcription and are highly mutagenic [124–126]. They also contribute to the formation of structured RNAs and can induce ribosome stalling during the elongation phase of translation [127–130]. Remarkably, mRNAs containing trinucleotide repeats can be cleaved and initiate the biogenesis of small functional RNAs [131–133]. Finally, RNAs containing trinucleotide repeats can form DNA:RNA hybrids or triplexes in trans [134]. In conclusion, some sequences (e.g. G-quadruplexes, trinucleotide repeats) have features that could allow a straightforward feedback from translation to directed-genetic variation.

Do co-translational physical constraints drive DNA sequence co-evolution?

In the molecular pathway in Figs. 2 and 3, biophysical parameters impacting on co-translational events can be modified during evolution not because of randomly located mutations but because co-translational events trigger co-translational cleavage of mRNA and the biogenesis of small RNAs that next increase the local mutational rate of targeted loci. A consequence of this model is that protein chaperones that help protein folding during translation should “buffer” this mutational process (as the proteins involved in RNA processing do, see part 1). Recent evidence has shown that protein chaperones, like heat shock proteins (HSPs) that help protein folding during translation, couple protein and RNA homeostasis, and are involved in piRNA biogenesis pathways, impact on genome evolution [135–140]. It has been suggested that on the one hand HSP chaperones buffer mutations, as they allow some protein sequence variations by helping protein folding, and on the other hand, the HSP knockdown induces the apparition of de novo mutations [139–146]. An interesting possibility is that the absence of HSP chaperones increases co-translational aggregate formation and results in the production of mutagenic small RNAs (Fig. 4A).

Consequently, proteins interacting with a nascent polypeptide could also contribute to directed-mutations by impacting co-translational protein folding. Indeed, in contrast to what is often believed, many events affecting proteins occur during translation, which includes protein-protein interactions [147–149]. If a mutation affects a protein A that interacts with a nascent protein B and alters its folding during translation, this could trigger the mutational process described above and lead to mutation of the protein B encoding gene (Fig. 4B). Therefore, if biophysical constraints trigger mutations that alleviate the initiating constraints, these mutations can in turn create other constraints anywhere within interacting networks, which will trigger mutations in other genes. Consequently, the interplay between biophysical constraints and mutational processes could explain the evolution of protein interaction networks.

Evolution of coding sequences may not just be fuelled by random mutations. Mutational processes may also be triggered by co-translational biophysical parameters that can feedback on DNA sequences through the biogenesis of structured and mutagenic small RNAs.

A widespread driving force shaped in a species-specific manner

Genome defence systems and RNA-mediated genome evolution are the two faces of the same coin

If gene expression-generated biophysical constraints drive genetic variations, this process is likely to be ubiquitous. First, co-transcriptional and co-translational biophysical constraints rely on the physicochemical properties of nucleic acids and proteins. Second, transcription and translation are universal. Finally, this concept
Figure 5. RNAs produced from exogenous or endogenous (e.g. retrotransposons) parasites are toxic. Some cellular RNAs associated with aggregate formation could be directed to the pool of toxic RNAs. This pool of RNAs could trigger genetic variation.

relied on the basic notion (that could apply to any cell) that some expressed-genomic sequences are toxic (Box 1) and challenge the integrity of the cell and its genome. Since every living organism has evolved specific molecular pathways to “fight” parasitic nucleic acids, which relies on RNA-guided immunity, nucleic acid cleavage or editing, it would be expected that each organism uses the same (or similar) molecular pathways to fight both parasitic nucleic acids and endogenous toxic sequences [83–87, 150–152].

The porous nature of the frontier between “self” cellular nucleic acids and endogenous or exogenous parasitic “non-self” nucleic acids supports the notion of an interplay between genome defense systems and RNA-mediated genome evolution. Indeed, it was believed that cellular RNA decay and small RNA biogenesis pathways were distinct pathways, the first one being involved in “self” RNA degradation and the second one being used to fight “non-self” parasite RNAs. However, there is now considerable evidence that both pathways are tightly connected [83–87, 153, 154]. For example, and as already mentioned, piRNAs that allow the cell to fight against retrotransposon invasion are produced from retrotransposons, coding genes, and pseudogenes. In addition, cellular RNAs are massively edited, as parasite RNAs are, and can be used to generate RNAs that activate the immune response [155]. The L1-ribonucleoprotein particle, which is responsible for the genomic insertion of retrotransposon-derived RNAs, can also create processed pseudogenes when it allows genomic integration of mRNA sequences (self RNAs) after reverse transcription [156, 157]. Collectively, these observations suggest that cellular “self” RNAs can at some points get entangled in the biological pathways normally used to fight parasite “non-self” nucleic acids (Fig. 5). Extrapolating from the cross-talk between anti-parasitic nucleic acid pathways and cellular RNA metabolism pathways would explain how and why some cellular RNAs could become “mutagenic.”

The genome defense systems against parasitic nucleic acids and the RNA-mediated genetic variations of endogenous toxic genomic sequences could actually be two faces of the same coin as these processes both just remove or modify toxic sequences. As a consequence, the evolutionary trajectory of a genome would directly depend on the parasitic nucleic acids it met. This means that although biophysical constraints-mediated genome evolution could be a widespread driving force, the precise molecular pathways that drive genetic variations could be specific to each organism depending on the precise genome defense systems it has. While cells may take advantage of molecular pathways involved in fighting parasitic nucleic acids to modify their own genome, it is interesting to underline that the eukaryotic gene expression process has recently been proposed to be shaped by the “combat” against parasitic nucleic acids [158].

Are RNA-directed genetic variations triggered by co-transcriptional and co-translational biophysical constraints interconnected?

In the first section, I proposed that co-transcriptional biophysical constraints shape gene architecture. In this context, it is interesting to underline that genome-wide waves of transcription occur during the development and differentiation of male germ cells. These genome-wide waves of transcription are the consequences of the genome-wide epigenetic “reprogramming” occurring during male germ cell development and differentiation [159–162]. As a consequence, male germ cells produce the most complex set of coding and noncoding transcripts and alternative splicing variants [159–162]. Therefore, male germ cells may experience extensive transcription-mediated genomic instability that could explain the large-scale apoptosis of immature sperm cells [4, 163, 164]. Another consequence of transcription of genome-wide waves in male germ cells is the expression of a wide variety of retrotransposon-derived RNAs, which results in the activation of the piRNA pathway [159–162]. As described in section 1, the expression of retrotransposon-derived RNAs (e.g. Alu) and the genomic insertion of these elements could, at some point, help male germ cells to alleviate local co-transcriptional constraints and therefore increase the survival of the germ cell with this particular kind of de novo mutations.

However, de novo mutations could generate deleterious gene products. Mattick et al. recently described a molecular pathway by which de novo mutation filtering could be performed during male germ cell development. In this scenario spermatogonia die if they have mutations that do not pass molecular “quality control” [159]. This filtering of de novo mutations during gametogenesis likely reduces transmission of deleterious genetic variants to the next generation. If de novo mutations, resulting from insertion of Alu elements within an intronic locus pass the spermatogenic “quality control,” then this could result in the Alu element exonization being weakly recognized as an alternative exon [53]. The buffering activity of the splicing process would first decrease the likelihood of generating deleterious gene products (i.e. weak splice sites are more often missed). However, if co-transcriptional physical constraints “push” toward the acquisition of stronger RNA processing sites, this would lead to the increase in the inclusion rate of new exons (e.g. Alu exons) during splicing. Meanwhile, the newly included exons would create constraints during translation (e.g. nascent protein mis-folding). These co-translational constraints would, in turn, trigger mutations in the corresponding coding sequences through G-quadruplex-containing small RNAs deriving from Alu sequences. Indeed, retrotransposons
are prone to form G-quadruplex structures and they may have contributed to
the spread of G-quadruplex structures within genomes [165, 166]. Therefore, some retrotransposons (e.g. Alu) may not only spread RNA processing sites (see first part) but also G-quadruplexes that could help retrotransposon-derived exons to rapidly evolve as coding exons through the mutational process relying on co-translational biophysical constraints. Therefore, Alu-derived exons could evolve to encode peptides that do not create deleterious constraints during translation. One general prediction of this model is that biophysical parameters impacting on co-translational events evolve together with biophysical parameters impacting on co-transcriptional events. This would imply the existence of a relationship between codon optimization, translation, and transcription, as was recently suggested [167–169].

In addition to favoring germ cell survival, by decreasing transcription-mediated genomic instability, some de novo mutations could provide germ cell growth advantage. Massive production of undifferentiated spermatogonia and their large-scale apoptosis is thought to reflect a spermatogonial selection process (“selfish spermatogonial selection”) [4, 163, 164]. Recent evidence indicates that many evolutionarily new genes are specifically expressed first in the testis. This has led to the view of the testis as a “nursery” for new gene products and the view that genes can emerge dependent on testis-specific function (the “out of testis” hypothesis) [160, 170, 171]. Therefore, not only do de novo mutations occur during gametogenesis but some of them might be filtered and eventually selected during this process.

**Box 2**

This box aims at describing some experimental settings that could help test the proposed hypothesis, focusing first on prokaryotic cells and next on eukaryotic cells. Since translation is coupled to transcription in prokaryotic cells, modifying optimal toward non-optimal codons in highly transcribed genes, is expected to increase local genomic instability. This would favor the re-emergence of optimal codons that would “synchronize” the dynamic of transcription and translation, therefore decrease genomic instability. Likewise, it may be possible to engineer coding sequences leading to the biogenesis of proteins having a high probability to misfold when emerging from ribosomes. Co-translational (therefore co-transcriptional) misfolding of nascent proteins is expected to increase local genomic instability, which would end up when new sequences alleviating protein misfolding emerge. Since RNA processing is coupled to transcription in eukaryotic cells, it might be possible to insert within gene bodies (e.g. in introns) DNA sequences that create constraints during transcription (e.g. R-loop prone sequences). These sequences would increase the local DNA instability, which would favor the emergence of RNA processing sites. It might also be possible to engineer an eukaryotic cellular model in which the folding of a protein can be challenged during translation in a controlled manner. A prediction resulting from the proposed hypothesis is that small RNAs corresponding to pieces of the parent mRNAs should be detectable, as should an increase of the mutational rate of the corresponding locus. While the experiment settings described above are expected to be associated with genetic variations at targeted loci, the characterization of the underlying mutational processes involved in prokaryotic and eukaryotic cells would allow to decipher whether genetic variations are random or can be driven by dedicated processes. The interplay between antiparasite genome defence systems and RNA-directed genetic variations could be addressed in unicellular organisms by exposing them to different stressful environments, when their genome defence systems are either active or inactive. One prediction is that mutation-mediated organismal adaptation would be strongly impaired in cell without an efficient antiparasite genome defence system.

**Conclusion and outlook**

If gene expression-generated biophysical constraints direct genome evolution, then organismal evolution may not just be fuelled by random mutations. First, some de novo mutations would preferentially occur in an RNA-directed manner in genomic regions that generate constraints or some kind of toxicity when they are expressed (Box 1). Second, de novo mutations would not randomly occur with respect to the biological effects they may drive. Indeed, small RNA-directed genetic variations would start because genomic sequences are toxic and it would end when the sequences are no longer toxic. Different experimental designs could be developed to test the proposed hypothesis (Box 2). Since the same gene expression-generated constraints exist in genomes of different individuals from the same species, mutations in certain genomic locations could recur frequently amongst individuals. This is in contrast with random mutations that have a low probability of occurring several times at the same location. The high frequency of these constraint-derived mutations in a population would increase their penetrance. RNA-directed mutation process may also explain why some mutations are more frequent than others and have been recurrently generated during evolution [172–175]. It is also interesting to note that many disease-associated mutations often affect RNA processing sites [176].

What could be the link between gene expression-generated biophysical constraints directing genome evolution and phenotype? First, it cannot be excluded that the cellular micro-environment can impact on biophysical parameters (e.g. co-translational protein folding) which could trigger a mutational pathway that would end up alleviating the environment-mediated constraints. Therefore, modifications of the cellular environment could trigger a mutational process that increases the likelihood of generating genetic variants impacting gene products involved in the environment-mediated constraints. Related to this phenomenon, I recently proposed that mutations in cancer cells might be directed and adapted to the tumoral micro-environment. This would help explain why most (if not all) anticancer
therapies failed because of tumor cell resistance [177].

Second, the concept of gene expression-generated biophysical constraints directing genome evolution implies that (i) the mutational process-driving force is still operating until the constraints are alleviated; and (ii) any new mutation generated in response to a constraint can create new constraints. This network could help explain some kinds of protein co-evolution [178] and therefore the “coordinated” evolution of genes involved in the same genetic circuit driving complex phenotypes.

Acknowledgment

I thank Cyril Bourgeois, Franck Mortreux, Marina Touillaud, Julian Venables, and Sophie Pantalacci for a critical reading of the manuscript.

The authors have declared no conflict of interest.

References

1. Goldmann JM, Wong WS, Pinelli M, Farrah T, et al. 2016. Parent-of-origin-specific signatures of de novo mutations. Nat Genet 48: 935–9.

2. Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, et al. 2016. Timing, rates and spectra of human germline mutation. Nat Genet 48: 126–33.

3. Franciolli LC, Polak PP, Koren A, Meneloua A, et al. 2015. Genome-wide patterns and properties of de novo mutations in humans. Nat Genet 47: 822–6.

4. Arheim N, Calabrese P. 2016. Germline stem cell competition, mutation hot spots, genetic disorders, and older fathers. Annu Rev Genomics Hum Genet 17: 219–43.

5. Campbell CD, Eichler EE. 2013. Properties and rates of germline mutations in humans. Trends Genet 29: 575–84.

6. Makova KD, Hardison RC. 2015. The effects of chromatin organization on variation in mutation rates in the genome. Nat Rev Genet 16: 213–23.

7. Burman B, Zhang ZZ, Pegoraro G, Lieb JD, et al. 2015. Histone modifications predispose genome regions to breakage and translocation. Genes Dev 29: 1393–402.

8. Reijns MA, Kemp H, Ding J, de Proos SM, et al. 2015. Lagging-strand replication shapes the mutational landscape of the genome. Nature 518: 502–6.

9. Wang G, Vasquez KM. 2014. Impact of alternative DNA structures on DNA damage, DNA repair, and chromatin stability. DNA Repair 19: 143–51.

10. Holoch D, Moazed D. 2015. RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet 16: 71–84.

11. Nowacki M, Haye JE, Fang W, Vijayan V, et al. 2010. RNA-mediated epigenetic regulation of DNA copy number. Proc Natl Acad Sci USA 107: 3516–21.

12. Sabin LR, Delas MJ, Hannon GJ. 2013. Dogma derailed: the many influences of RNA on the genome. Mol Cell 49: 783–94.

13. Zheng S, Vuong BQ, Vaidyanathan B, Lin JY, et al. 2015. Non-coding RNA generated following lariat debranching mediates targeting of AID to DNA. Cell 161: 762–73.

14. Khanduja JS, Calvo IA, Joen RI, Hill IT, et al. 2016. Nuclear non-coding RNAs and genome stability. Mol Cell 60: 245–53.

15. Keshin H, Shen Y, Huang F, Patel M, et al. 2014. Transcript-RNA-templated DNA recombination and repair. Nature 515: 436–9.

16. Fang W, Landweber LF. 2013. RNA-mediated gene rearrangement: hypotheses and evidence. BioEssays 35: 84–7.

17. Shapiro JA. 2016. The basic concept of the read-write genome: mini-review on cell-mediated DNA modification. Bio Systems 140: 35–7.

18. Herbert A. 2004. The four Rs of RNA-directed evolution. Nat Genet 36: 19–25.

19. Mattick JS, Mehler MF. 2008. RNA editing, DNA recoding and the evolution of human cognition. Trends Neurosci 31: 227–33.

20. Bentley DL. 2014. Coupling mRNA processing with transcription in time and space. Nat Rev Genet 15: 163–75.

21. Walters RW, Parker R. 2015. Coupling of Ribostasis and Proteostasis: Hsp70 Proteins in mRNA Metabolism. Trends Biochem Sci 40: 552–9.

22. Kim N, Jinks-Robertson S. 2012. Transcription as a source of genome instability. Nat Rev Genet 13: 204–14.

23. Gaillard H, Aguilera A. 2016. Transcription as a Threat to Genome Integrity. Annu Rev Biochem 85: 291–317.

24. Hamperl S, Starte CA. 2014. The combination of co-transcriptional RNA/DNA hybrid structures to DNA damage and genome instability. DNA Repair 19: 84–94.

25. Jimeno-Gonzalez S, Soberon G. 2013. Properties of de novo mutations in humans. Nat Genet 45: 1393–402.

26. Defective histone supply causes changes in RNA polymerase II elongation rate and cotranscriptional pre-mRNA splicing. Proc Natl Acad Sci USA 112: 14840–5.

27. Henriques, Catter M, Stirling PC. 2014. Stable pausing by RNA polymerase on the genome. Nature 511: 362–5.

28. Gowrishankar J, Harinarayanan R. 2004. Why is transcription coupled to translation in bacteria? Mol Microbiol 54: 598–603.

29. Francia S, Michelin F, Saxena A, Tang D, et al. 2012. Site-specific Dicer and Drosha RNA processes control the DNA-damage response. Nature 488: 231–5.

30. Wei W, Ba Z, Gao M, Wu Y, et al. 2012. A role for small RNAs in DNA double-strand break repair. Cell 149: 101–12.

31. d’Adda di Fagagna F. 2014. A direct role for small non-coding RNAs in DNA damage response. Trends Cell Biol 24: 171–8.

32. Yang YQ, Qi Y. 2015. RNA-directed repair of DNA double-strand breaks. DNA Repair 32: 82–5.

33. Chowdhury D, Choi YE, Braut M. 2013. Charity begins at home: non-coding RNA functions in DNA repair. Nat Rev Mol Cell Biol 14: 181–9.

34. Storici F, Bebenek K, Kunzel TA, Gordenin DA, et al. 2007. RNA-templated DNA repair. Nature 447: 336–41.

35. Rao C, Bielli P, Pagliarini V, Sette C. 2015. The interplay between DNA damage response and RNA processing: the unexpected role of splicing factors as gatekeepers of genome stability. Front Genet 6: 35.

36. Li X, Niu T, Manley JL. 2007. The RNA binding protein RNP51 alleviates ASF/SF2 depletion-induced genomic instability. RNA 13: 2108–15.

37. Naro C, Bielli P, Pagliarini V, Sette C. 2016. The splicing-associated protein Nrl1 suppresses homologous recombination-dependent R-loop formation in fission yeast. Nucleic Acids Res 44: 1703–17.

38. Tankiwa M, Sanjiv K, Helleday T, Herr P, et al. 2016. The spliceosome U2 snRNP factors promote genome stability through distinct mechanisms; transcription of repair factors and R-loop processing. Oncogene 6: 6280.

39. Dominguez-Sanchez MS, Barroso S, Gomez-Gonzalez B, Luna R, et al. 2011. Genome instability and transcription elongation impairment in human cells depleted of TTH/TREX. PLoS Genet 7: e1002386.

40. Bhatia V, Barroso SI, Garcia-Rubio ML, Tumini E, et al. 2014. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. Nature 511: 362–5.

41. Gowrishankar J, Harinarayanan R. 2004. Why is transcription coupled to translation in bacteria? Mol Microbiol 54: 598–603.

42. Francis S, Michelin F, Saxena A, Tang D, et al. 2012. Site-specific Dicer and DROSHA RNA processes control the DNA-damage response. Nature 488: 231–5.

43. Wei W, Ba Z, Gao M, Wu Y, et al. 2012. A role for small RNAs in DNA double-strand break repair. Cell 149: 101–12.

44. d’Adda di Fagagna F. 2014. A direct role for small non-coding RNAs in DNA damage response. Trends Cell Biol 24: 171–8.

45. Yang YQ, Qi Y. 2015. RNA-directed repair of DNA double-strand breaks. DNA Repair 32: 82–5.

46. Chowdhury D, Choi YE, Braut M. 2013. Charity begins at home: non-coding RNA functions in DNA repair. Nat Rev Mol Cell Biol 14: 181–9.

47. Storici F, Bebenek K, Kunzel TA, Gordenin DA, et al. 2007. RNA-templated DNA repair. Nature 447: 336–41.

48. Rao C, Bielli P, Pagliarini V, Sette C. 2015. The interplay between DNA damage response and RNA processing: the unexpected role of splicing factors as gatekeepers of genome stability. Front Genet 6: 35.

49. Li X, Niu T, Manley JL. 2007. The RNA binding protein RNP51 alleviates ASF/SF2 depletion-induced genomic instability. RNA 13: 2108–15.

50. Naro C, Bielli P, Pagliarini V, Sette C. 2016. The splicing-associated protein Nrl1 suppresses homologous recombination-dependent R-loop formation in fission yeast. Nucleic Acids Res 44: 1703–17.
50. Keskin H, Meers C, Storici F. 2016. Transcription RNA supports precise repair of its own DNA gene. RNA Biol 13: 157–65.

51. Frasch M. 2013. Non-coding RNA: sequence-specific guide for chromatin modification and DNA damage signaling. Front Genet 6: 320.

52. Proshkin S, Rahmouni AR, Mironov A, Nucler E. 2010. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science 328: 504–8.

53. Keren H, Lev-Maor G, Ast G. 2010. Allosteric switching and evolution: diversification, exon definition and function. Nat Rev Genet 11: 345–55.

54. Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. Nat Rev Genet 10: 691–703.

55. Ule J. 2013. Alu elements: at the crossroads between disease and evolution. Biochem Soc Trans 41: 1532–5.

56. Kaep K, Speek M. 2011. Intronic retroelemental folding toward different RNA polymerase II. Mobile Genet Elements 2: 154–7.

57. Daniel C, Behm M, Ohman M. 2015. The role of Alu elements in the cis-regulation of RNA processing. Cell Mol Life Sci 72: 4063–76.

58. Schmitz J, Brosius J. 2011. Exonization of transposed elements: A challenge and opportunity for evolution. Biochimie 93: 1925–38.

59. Deininger P. 2011. Alu elements: know the SINES. Genome Biol 12: 236.

60. Doucet AJ, Wilusz JE, Miyoshi T, Liu Y et al. 2015. A 3’ poly(A) tract is required for LINE-1 retrotransposition. Mol Cell 60: 728–41.

61. Attig J, Ruiz de Los Mozos I, Haberman N, Wang Z et al. 2016. Splicing repression allows the gradual emergence of new Alu-exons in primate evolution. eLife 5: e19545.

62. Liang D, Wilusz JE. 2014. Short intronic repeat sequences facilitate circular RNA production. Genes Dev 28: 2233–47.

63. Chi A, Dirac-Sveistrup AB, Hartmann H et al. 2012. DBDRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. Nature 484: 386–9.

64. Anderson SK, Proudfoot NJ. 2014. A double-edged sword: R-loops as threats to genome integrity and powerful regulators of gene expression. RNA Biol 11: 386–91.

65. Papasaikas P, Valcarcel J. 2016. The spliceosome: the ultimate RNA chaperone and sculptor. Trends Biochem Sci 41: 33–45.

66. Buhr F, Jha S, Thommen M, Mittelstaedt J et al. 2016. Synonymous codons direct cotranslational folding of different protein conformations. Mol Cell 61: 341–51.

67. Pallares I, Ventura S. 2016. Understanding and predicting protein misfolding and aggregation: insights from proteomics. Trends Cell Biol 26: 2570–81.

68. Ramaswami M, Taylor JP, Parker R. 2013. Altered ribostasis: RNA-protein granules in degenerative disorders. Cell 154: 727–36.

69. Jarvelin AI, Noerenberg M, Davis I, Castello A et al. 2015. The new (dis)order in degenerative diseases: Piwi Interacting RNA pathway in the germline. Genome Res 25: 368–80.

70. Pantoano L, Jodar M, Bak M, Ballesca JL et al. 2015. The small RNA content of human sperm reveals pseudogene-derived piRNAs complementary to protein-coding genes. RNA 21: 1085–95.

71. Williams Z, Proshkin S, Mihailovic A, Lin C et al. 2015. Discovery and characterization of piRNAs in the human fetal ovary. Cell Rep 13: 854–63.

72. Zhang P, Kang YJ, Gou LT, Wang J et al. 2016. MIWI and piR-B23 mediated cleavage of messenger RNAs in mouse testes. Cell Res 25: 193–207.

73. Yamich J, Heo SJ, Dhabhi J, Martin DI et al. 2015. PiRNA-like small RNAs mark extended 3’UTRs present in germ and somatic cells. BMC Genomics 16: 462.

74. Fu A, Jacobs DI, Zhu Y. 2014. Epigenome-wide analysis of piRNAs in gene-specific DNA methylation. RNA Biol 11: 1301–12.

75. Zhang X, Liu P, Liu J et al. 2016. IL-4 inhibits the biogenesis of an epigenetically suppressive PIWI-interacting RNA to upregulate CD1a Molecules on monocytes/dendritic cells. J Immunol 196: 1591–603.

76. Yuan H, Wu QL, Sun CY, AI LS et al. 2015. PiRNA-823 contributes to tumorigenesis by regulating de novo DNA methylation and angiogenesis in multiple myeloma. Leukemia 29: 198–206.

77. Eskin V, Magliocco S, Formica D, Gianfrancesco F. 2011. PiR_05520 belongs to Piwi-associated RNAs regulates expression of the human melanin receptor 1A gene. PLoS ONE 6: e22727.

78. Watanabe T, Tomazay S, Mitsuya K, Tokoty Y et al. 2011. Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. Science 332: 848–52.

79. Wilke CO, Drummond DA. 2010. Signatures of protein biophysics in coding sequence evolution. Curr Opin Struct Biol 20: 385–9.

80. Cal P, Papp B, Lercher MJ. 2006. An integrated view of protein evolution. Nat Rev Genet 7: 337–48.

81. Bali V, Bebek Z. 2015. Decoding mechanisms by which silent codon changes influence protein biosynthesis and function. Int J Biochem Cell Biol 64: 58–67.

82. Shankar R, Wilke CO, Seo S et al. 2013. RNA secondary structures: a key to biological regulation and complexity. Nucleic Acids Res 41: 2073–94.

83. Rhodes D, Lipp HS. 2015. G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res 43: 8627–37.

84. Carle CM, Zaher HS, Chalker DL. 2016. A parallel G quadruplex-binding protein regulates the boundaries of DNA elimination events in Tetratrichomonas thermophila. PLoS Genet 12: e1005842.

85. Agarwala P, Pandey S, Maiti S. 2015. The tale of RNA G-quadruplexes. Org Biomol Chem 13: 5570–8.

86. Simonis R, Fratta P, Neidle S, Parkinson GN et al. 2015. G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. FEBS Lett 589: 1653–60.

87. Kowk CK, Marsico G, Sahakyan AB, Chambers VS et al. 2016. RNA-Seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat Methods 13: 841–4.

References

Bioessays 39, 10, 1700069, © 2017 The Authors. BioEssays Published by Wiley Periodicals, Inc. 1700069 (11 of 13)
110. Bolduc FC, Garant JM, Allard FE, Perreault JP. 2016. Irregular G-quadruplexes found in the untranslated regions of human mRNAs influence translation. J Biol Chem 291: 21751–60.

111. Endoh T, Sugimoto N. 2016. Mechanical insights into ribosomal progression over-coming RNA G-quadruplex from periodical translation suppression in cells. Sci Rep 6: 22719.

112. Olsthoorn RC. 2014. G-quadruplexes within pri-mRNA: the missing link in pri-mRNA disease? Nucleic Acids Res 42: 9327–33.

113. Bashkirov VI, Vourekas A. 2016. Pathogenic mechanism in Huntington’s disease involves small CAG-repeats with neurotrophic activity. PLoS Genet 8: e1002481.

114. Marti E, Estvill X. 2013. Non-small coding RNAs add complexity to the RNA pathogenic mechanisms in trinucleotide repeat expansion disorders. Front Mol Neurosci 6: 45.

115. Costantino L, Koshland D. 2015. The Yin and Yang of R-loop biology. Curr Opin Cell Biol 34: 39–45.

116. Costantino L, Koshland D. 2015. The Yin and Yang of R-loop biology. Curr Opin Cell Biol 34: 39–45.

117. Costantino L, Koshland D. 2015. The Yin and Yang of R-loop biology. Curr Opin Cell Biol 34: 39–45.

118. Costantino L, Koshland D. 2015. The Yin and Yang of R-loop biology. Curr Opin Cell Biol 34: 39–45.

119. Costantino L, Koshland D. 2015. The Yin and Yang of R-loop biology. Curr Opin Cell Biol 34: 39–45.

120. Costantino L, Koshland D. 2015. The Yin and Yang of R-loop biology. Curr Opin Cell Biol 34: 39–45.

121. Costantino L, Koshland D. 2015. The Yin and Yang of R-loop biology. Curr Opin Cell Biol 34: 39–45.

122. Costantino L, Koshland D. 2015. The Yin and Yang of R-loop biology. Curr Opin Cell Biol 34: 39–45.

123. Chandra V, Bortnick A, Murce C. 2015. AID targeting: old mysteries and new challenges. Trends Immunol 36: 257–55.

124. Schmidt MH, Pearson CE. 2016. Disease-associated repeat instability and mismatch repair. DNA Repair 38: 117–26.

125. Budworth H, McMurray CT. 2013. Bidirectional transcription of trinucleotide repeats: roles for exclusion repair. DNA Repair 12: 672–84.

126. Brahmachari SK, Meera G, Sarkar PS, Balagurumoothy P, et al. 1995. Simple repeat elements in mitochondria: a mouse genome: structure and functional significance. Electrophoresis 16: 1705–14.

127. Green KM, Linsalata AE, Todd PK. 2016. RAN translation—what makes it run? Brain Res 1647: 30–42.

128. Blaszczzyk L, Lypniwski W, Kliśsek A. 2017. Structures of RNA repeats associated with neurological diseases. Wiley Interdiscip Revs: RNA 8: e1412.

129. Wolf AS, Grayhack EJ. 2015. Asc1, homolog of human RACK1, prevents frameshifting in yeast by ribosomes stalled at CGA codon repeats. RNA 21: 1427–37.

130. Letzring DP, Wolf AS, Brue CE, Grayhack EJ. 2013. Translation of CGA codon repeats in yeast involves quality control components and ribosomal protein L1. RNA 19: 1208–17.

131. Nalavade R, Griesche N, Ryan DP, Hildbrand S, et al. 2013. Mechanisms of RNA-induced toxicity in CAG repeat disorders. Cell Death Dis 4: 6752.

132. Bajzer-Coronel M, Porta S, Kagerbauer B, Mateu-Bustamante T, et al. 2012. A pathogenic mechanism in Huntington’s disease involves small CAG-repeated RNA with neurotrophic activity. PLoS Genet 8: e1002481.

133. Martí E, Estvill X. 2013. Non-small coding RNAs add complexity to the RNA pathogenic mechanisms in trinucleotide repeat expansion disorders. Front Mol Neurosci 6: 45.

134. Colak D, Zaninovic N, Cohen MS, Rosenwaks Z, et al. 2014. Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. Science 343: 1002–5.

135. Jarosz D. 2016. The Human HSP70 family of chaperones: where do we stand? Cell Stress Chaperones 21: 379–404.

136. Pennisi R, Ascenzio P, di Massa A. 2015. Hsp90: mRNA and DNA repair. Biomolecules 5: 2589–618.

137. Ichiyangi T, Ichiyangi K, Ogawa A, Kuramochi-Miyagawa S, et al. 2014. HSP90alpha plays an important role in p53 RNA biogenesis and retrotransposition repression in mouse. Nucleic Acids Res 42: 11903–11.

138. Garagarza V, Yin H, Weiner MM, Wang J. 2011. Drosophila Piwi functions in Hsp90-mediated suppression of phenotypic variation. Nat Genet 43: 153–8.

139. Specchia V, Piacentini L, Tritto P, Fantl V, et al. 2010. Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. Nature 463: 662–5.

140. Jarosz DF, Lindquist S. 2010. Hsp90 and environmental stress transform the adaptive value of natural genetic variation. Science 330: 1820–4.

141. Piacentini L, Fantl V, Specchia V, Bozetti MP, et al. 2014. Transposons, environmental changes, and heritable induced phenotypic variability. Chromosoma 123: 345–54.

142. Paaby AB, Rockman MV. 2014. Cryptic genetic variation: evolution’s hidden substrate. Nat Rev Genet 15: 247–58.

143. Fares MA. 2015. The origins and evolutionary and functional significance of R-loop formation. RNA 21: 147.

144. Hildebrand A003772. 2013. Translation of CGA codon repeats in yeast by ribosomes stalled at CGA codon repeats. RNA 21: 1427–37.

145. Rehwinkel J. 2015. RNA degradation in antiviral immunity and autoimmunity. Trends Immunol 36: 179–88.

146. Cecere G, Grishok A. 2014. A nuclear perspective on RNAi pathways in metazoans. Biochim Biophys Acta 1839: 223–33.

147. Savva YA, Rezaei A, St Laurent G, Reenan RA. 2016. Reprogramming, circular reasoning and self versus non-self: one-stop shopping with RNA editing. Front Genet 7: 100.

148. Roberts TG, Morrice CL. 2013. Not so pseudo anymore: pseudogenes as therapeutic targets. Pharmacogenomics 14: 2023–34.

149. Mandal PK, Ewing AD, Hancks DC, Kazazian HH, Jr. 2013. Enrichment of processed pseudogene transcripts in 1L-rRNA-ribonuclease particles. Hum Mol Genet 22: 3733–40.

150. Madhani HD. 2013. The frustrated gene: origins of eukaryotic gene expression. Cell 153: 744–53.

151. Werner A, Piatek MJ, Mattick JS. 2015. Transpositional shuffling and quality control in male germ cells to enhance evolution of complex organisms. Ann NY Acad Sci 1314: 158–63.

152. Soumillon M, neculae A, Weier M, Brwand D et al. 2013. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep 3: 2179–90.

153. Rathke C, Baarends WM, Awe S, Renkawitz-Pohl R. 2014. Chromatin dynamics during spermiogenesis. Biochim Biophys Acta 1839: 155–68.

154. Lin X, Han M, Cheng L, Chen J et al. 2016. Expression dynamics, relationships, and transcriptional regulations of diverse transcripts in mouse spermatogenic cells. RNA 22: 1011–24.

155. Maher GJ, McGowan SJ, Giannoulatos E, Verrilli C, et al. 2016. Visualizing the origins of selfish de novo mutations in individual seminiferous tubules of human testes. Proc Natl Acad Sci USA 113: 2464–9.

156. Goriely A, McGrath JJ, Hultman CM, Wilkie AO, et al. 2013. “Selfish spermatozoal gonial selection”: a novel mechanism for the association between advanced paternal age and neurodevelopmental disorders. Am J Psychiatry 170: 599–608.

157. Keijovsny E, Lexa M. 2014. Quadruplex-forming DNA sequences spread by retro-transposons may serve as genome regulators. Mob Genet Elements 4: e28084.

158. Keijovsny E, Tokan V, Lexa M. 2015. Transposable elements and G-quadruplexes. Chromosome Res 23: 615–23.

159. Trotta E. 2013. Selection on codon bias in yeast: a transcriptomic hypothesis. Nucleic Acids Res 41: 9382–95.

160. Zhou Z, Deng Y, Zhou M, Li L et al. 2016. Codon usage is an important determinant of gene expression levels largely through its...
effects on transcription. Proc Natl Acad Sci USA 113: E6117–E25.

169. Newman ZR, Young JM, Ingolia NT, Barton GM. 2016. Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. Proc Natl Acad Sci USA 113: E1362–71.

170. Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. Genome Res 20: 1313–26.

171. Carelli FN, Hayakawa T, Go Y, Imai H, et al. 2016. The life history of retrocopies illuminates the evolution of new mammalian genes. Genome Res 26: 301–14.

172. Lind PA, Farr AD, Rainey PB. 2015. Experimental evolution reveals hidden diversity in evolutionary pathways. eLife 4: e07074.

173. Lek M, Karcewski KJ, Minikel EV, Samocha KE, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536: 285–91.

174. Storz JF. 2016. Causes of molecular convergence and parallelism in protein evolution. Nat Rev Genet 17: 239–50.

175. Livnat A. 2013. Interaction-based evolution: how natural selection and nonrandom mutation work together. Biol Direct 8: 24.

176. Manning KS, Cooper TA. 2017. The roles of RNA processing in translating genotype to phenotype. Nat Rev Mol Cell Biol 18: 102–14.

177. Auboeuf D. 2016. Putative RNA-directed adaptive mutations in cancer evolution. Transcription 7: 164–187

178. Pazos F, Valencia A. 2008. Protein co-evolution, co-adaptation and interactions. EMBO J 27: 2648–55.