Noncoding RNAs: Persistent Viral Agents as Modular Tools for Cellular Needs

Günther Witzany

Telos-Philosophische Praxis, A-5111-Bürmoos, Austria

It appears that all the detailed steps of evolution stored in DNA that are read, transcribed, and translated in every developmental and growth process of each individual cell depend on RNA-mediated processes, in most cases interconnected with other RNAs and their associated protein complexes and functions in a strict hierarchy of temporal and spatial steps. Life could not function without the key agents of DNA replication, namely mRNA, tRNA, and rRNA. Not only rRNA, but also tRNA and the processing of the primary transcript into the pre-mRNA and the mature mRNA are clearly descended from retro-"elements" with obvious retroviral ancestry. They seem to be remnants of viral infection events that did not kill their host but transferred phenotypic competences to their host and changed both the genetic identity of the host organism and the identity of the former infectious viral swarms. In this respect, noncoding RNAs may represent a great variety of modular tools for cellular needs that are derived from persistent nonlytic viral settlers.

Key words: **noncoding RNAs; regulatory networks; addiction modules; persistent viral life style; nucleic acid language**

Introduction

Current knowledge indicates that DNA is a stable information storage medium. As proposed in an article by Vetsigian, Woese, and Goldenfeld, it serves as an "evolutionary protocol" for evolutionary novelties and selected properties. A wide variety of small RNAs regulate key cellular processes of replication as well as genetic arrangements, rearrangements, recombination and repair, and even inventions. In most cases they act after being transcribed from the stable DNA storage medium into a pretranscriptional or transcriptional modus, with the advantage of being active prior to all translated proteins. Elements such as micro-RNAs, small nuclear and small nucleolar RNAs, tRNA, and rRNA as well as the assemblies of the spliceosome and ribosome are vital to all life processes. In this respect I remember a quote from the Austrian philosopher Ludwig Wittgenstein: "The meaning of a word is its use within a language and to understand a sentence means to understand a language. To understand a language means to be the master of a technique."

What are the "masters" of the technique used to edit nucleotide sequences of the genetic text according to combinatorial (syntactic), context-sensitive (pragmatic), and content-specific (semantic) rules, according to Charles Morris the obligate and nonreducible levels of rules that are inherent to any kind of language or language-like codes?

Why Editing Needs Editors

For a long time the textbook conviction held sway that the genetic content arrangements of DNA sequences are evolutionary results of random mutations and their selection. Then it was noticed that DNA regions which code for proteins are decreasing with organismic

Address for correspondence: Günther Witzany, Telos-Philosophische Praxis, Vogelsangstraße 18c, A-5111-Bürmoos, Austria. Voice/fax: 0043 62746805. witzany@sbg.at

Natural Genetic Engineering and Natural Genome Editing: Ann. N.Y. Acad. Sci. 1178: 244–267 (2009).
doi: 10.1111/j.1749-6632.2009.04989.x

complexity (in humans 1.5%) and former "junk DNA" plays increasing important roles in gene regulation. Additionally, it was assumed that this "junk DNA" is a kind of selfish genetic element, with an inherent tendency to replicate themselves, in contrast with its helpful and symbiotic functions within host genomes.

For many decades it was common practice to speak about the "genetic code" with its inherent language-like features. Long before, and even after, Manfred Eigen's suggestion that the nucleic acid sequences are comparable to and function like any real language coherent with a (molecular) syntax, linguistic, and communicative vocabulary was commonly used in genetics, cell biology, and molecular biology: genetic code, code without commas, misreading of the genetic code, coding, genetic storage medium DNA, genetic information, genetic alphabet, genetic expression, messenger RNA, cell-to-cell communication, immune response, transcription, translation, nucleic acid language, amino acid language, recognition sequences, recognition sites, protein coding sequences, repeat sequences, etc.

In contrast with the evolutionary paradigm of random assemblies of nucleic acids that constitute the genetic text we do not know any real-life languages or codes which emerged as a randomly derived mixture of characters.

If Manfred Eigen's suggestion is still valid, and the description of nucleic acid sequence order in terms of linguistics (molecular syntax) makes sense for the future, we should look at the current scientific knowledge of "language" and "communication."

Every language is based on signs, whether they are signals or symbols. In humans and other animals they are transported auditively, visually, or tactilely. In nonhuman living beings they are transported by small molecules in crystallized, fluid, and gaseous form. Additionally these signs can be combined coherently with combinatorial rules (syntax). Signs are not generated and used by themselves, but in real-life languages by living beings. These sign-generating and sign-using agents live *in vivo* in continued changing interactions and environmental circumstances. This is the context (pragmatics) in which a living being is interwoven. This context determines the meaning (semantics) of the signs in messages that are used to communicate and to coordinate single as well as group behavior.

Therefore, we can understand that the same sentence, or the same syntactic sequence order, of any language or code can have different, and in extreme cases, opposite meanings and therefore transport different messages. The important consequence of this fact is that it is not possible to extract the meaning of an informational content solely out of the syntactic structure, but someone has to identify the context within which the living being uses this syntactic structure.

The primary agents are not the sequences of signs, not the rules which determine sequence orders, but the living agents. Without living agents there are no signs, no semiotic rules, no signalling, and no communication. Paradoxically, without signs, semiotic rules, signalling, and communication, no living agents could coordinate growth and development.

If we assume the genetic code to function language-like, knowing that no language which has been observed functions by itself, then we have to postulate living agents that are competent to use signs coherent with syntactic, pragmatic, and semantic rules. Adapted to the genetic code, this means that there must be living agents competent in generation and integration of meaningful nucleotide sequences, and meaningful nucleotide sequences are not a randomly derived mixture of nucleotides.

Natural genome editing from a biocommunicative perspective means competent agent-driven generation and integration of meaningful nucleotide sequences into preexisting genomic content arrangements and the ability to (re-)combine and (re-)regulate them according to the context-dependent (i.e., adaptational) purposes of the host organism.

DNA: Information Storage Medium and Life Habitat of Endogenous Viruses

There is increasing evidence that all cellular life is colonized by exogenous and/or endogenous viruses in a nonlytic but persistent lifestyle. A persistent lifestyle in cellular life-forms most often seems to derive from an equilibrium status reached by at least two competing genetic settlers and the immune function of the host that keeps them in balance. Persistent settlement of host genomes means that if we postulate agent-driven genetic text editing then we have to look at their *in vivo* life strategies to understand their habits and the situational contexts that determine their content arrangements. Then we can reconstruct nucleic acid sequences that function as a code, not as a statistically random-like mixture of nucleotides, but as informational content in a syntactic order that is coherent with the whole sequence space generated by agents that are linguistically competent in nucleic acid language, that is, the genetic code. As in every language each character, word, and sentence together with starts, stops, commas, and spaces in-between has content and a text-formatting function and is generated by competent agents.

If we imagine that humans and one of the simplest animals, *Caenorhabditis elegans*, share a nearly equal number of genes (ca. 20,000) it becomes obvious that the elements that create the enormous diversity are not the protein-coding genes but their higher order regulatory network that is processed by the mobile genetic elements, such as transposons and retroposons and the noncoding RNAs[1] (Juergen Brosius created the appropriate image of "genes floating in a sea of retroposons"). If we consider the important role of the highly structured and ordered regulatory network of noncoding RNAs as not being randomly derived, one of the most favorable models with explanatory power is the virus-first thesis. This means that the evolution of the noncoding RNA world is the result of persistent viral life strategies.

The whole range of mobile genetic agents that are competent to edit the genetic code/nucleic acid language not only edit, but also regulate key cellular processes of replication as well as transcription, translation, recombination, repair, and even inventions via a wide variety of small RNAs. In this respect, DNA is not only an information-storing archive but a life habitat for linguistically-competent RNA agents, most of them seemingly of viral or subviral descent. To understand their natural genome editing competence we have to look not only at their linguistic competence in editing and regulating correct nucleotide sequences but at their communicative competence too, that is, how they interact with each other, how they compete within host organisms, how they symbiotically interact with host organisms to ward off competing parasites, and what life strategies they share. Persistent infection lifestyles that do not harm hosts and symbiotic, cooperating viral swarms may be more successful in evolutionary terms for integrating advantageous phenotypes into host organisms than "selfish" agents.

A Viral Progenitor of the Eukaryotic Nucleus?

If we look at persistent viral life strategies we should first look at the role of the eukaryotic nucleus. The eukaryotic cell most probably evolved by a symbiogenetic integration event of former free-living bacteria. This integration, however, cannot explain the progenitor of the eukaryotic nucleus because its key features could not have derived from prokaryotes.[2,3] The eukaryotic nucleus has numerous key features, proteins, and RNAs that are not found in any prok, aryote. Interestingly, these key features are present in certain prokaryote viruses.[4–6] These viruses use linear chromosomes, telomere repeats, multiple membranes, histone-packaged chromosomes with marking effect for self/nonself identification, and nuclear pores.

No single virus encompasses all of these key features, but every key feature of the eukaryotic nucleus is present in some large dsDNA viruses. This requires consideration of a process in which different viral competences were integrated into a single dsDNA virus that was the progenitor of the eukaryotic nucleus. Alternatively, a large dsDNA virus functioned as a simple eukaryotic nucleus and later integrated other viral competences. On examination of the key features of several candidates for this integration, the focus is primarily on prokaryotic, eukaryotic, and archaeal phages.

Prokaryotic phages such as cyanophages have double-stranded DNA, DNA and RNA polymerases similar to eukaryotes. Eubacterial phages possess linear double-stranded DNA, telomeres, DNA and RNA polymerases, chromatin, and internal membranes. Archaeal phages with linear double-stranded DNA have telomere repeats similar to eukaryotes. They also possess chromatin and an internal lipid tendency to nonlytic, persistent, (and often mixed) infections.[4]

Other DNA viruses share similar features that are characteristic of the eukaryotic nucleus but are not found in prokaryotes. An example is the vaccinia virus (poxvirus).[7] These viruses have a membrane-bound segregation of transcription and translation, multiple membranes, and their DNA synthesis combines membrane loss and a cell-cycle-dependent restoration as well as an actin/tubulin-bound transport system[4,8] and, interestingly, nuclear pores.[9] Cytoplasmic DNA viruses (African swine fever virus) have chromatin and linear chromosomes with telomeres. PhycoDNA viruses have mRNA capping, introns, and diverse DNA replication proteins. Torque Teno Virus (TTV) 1-4 have linear double-stranded DNA genomes with a molecular basis for the evolution of eukaryotic chromatin; they also have capsids which integrate internal and external lipid proteins.[4]

In addition, all these viruses have the capability for self and nonself identification. All viruses mark their genomes, RNAs, and proteins by different kinds of chemical modifications, for example, methylation. This marking allows the differentiation between self and nonself. Nonself may be other viruses, the host genome, or host-related transcripts.[4]

Evolutionary Roles of Viruses as Natural Genome Editors

To understand the evolutionary emergence of the eukaryotic nucleus it could be useful to reconstruct the natural genome-editing competences of viruses.[10] Recent research in microbiology, based on comparative genomics and phylogenetic analyses, has demonstrated that life must be viewed from the perspective of the crucial role played by viruses.[4,11–15]

This contradicts former concepts that focused on viruses in the framework of (i) escape theories, that is, viruses are intact or deformed genetic parasites that escaped from cellular life, or that viruses (ii) evolved from cellular ancestors or (iii) that they are not living beings because they cannot live without cellular life. From these perspectives, viruses could not play crucial roles in the evolution of cellular life. Interestingly, phylogenetic analyses do not support the former concept of RNA- and DNA-viruses descending from cellular life. These analyses also show that DNA and RNA viruses most probably did not have a common ancestor but evolved independently. Viruses probably have to be placed at the very beginning of life, long before cellular life evolved.[4]

Persistent Viral Life Strategies are Beneficial for Their Hosts

Acute viruses that exhibit lytic action induce disease and even death. In contrast, a persistent lifestyle implies compatible interactions with the host, either by being integrated into the host genome or within its cell plasma.[16] The result is nondestructive symbiosis during most life stages of the host. The persistent lifestyle allows the virus to transmit complex viral

phenotypes to the host organism. This process, which changes both the genetic identity of the host and the identity of its persistent settler, enables the host to broaden its evolutionary adaptational potential and may promote the formation of new species.[4]

The persistent lifestyle of viruses is typically tissue specific, that is, host tissues are colonized by different nonlytic viruses which integrate themselves into the host cytoplasm, for example, as plasmids or into the host genomes, and co-evolve with them. A common habit of persistent viral settlers is that during host cell replication they function in a tissue-specific, replication-cycle dependent manner. Interestingly, micro-RNAs in eukaryotic cells have similar tissue-specific or developmental expression patterns.[17] Micro-RNAs play important roles in Dicer- and Risc-mediated mRNA degradation or mRNA translation inhibition.[18] This implies an RNAi immune function. Because micro-RNAs act on mRNAs, not on proteins, they are probably encoded by persistent nuclear DNA viruses.[19] We will look at these competences later.

Persistent Status through Addiction Modules

The persistent status is the result of multiple colonization events into a host. This neutralizes former antagonistic and incompatible features of competing viral agents without harming the host.[20–22] Most of the endogenous or exogenous inhabitants inherent to bacteria, protozoa, plants, animals, and fungi are a complementary mix of formerly antagonistic viral features. They can still be identified today as toxin/antitoxin, restriction/modification-, insertion/deletion modules, that is, complementary counterpart regulatory functions that do not harm the host.[4,23,24] As symbiotic neutralization and counterpart regulation, they represent new phenotypic features. One feature is regulated exactly by the antagonist according to developmental stages in the cell cycle, replication, tissue growth, or similar developmental contexts. Should this suppressor function become unbalanced, then the normally downregulated part may become lytic with potentially lethal consequences, as documented for symbiodinium and its major role in coral bleaching.[25]

Retroviral Competences in a Persistent Symbiotic Lifestyle: Endogenization

Endogenous retroviral competences in the persistent status are often characterized by features expressed only in the strict time window of a developmental process, such as axis formation, trophectoplast formation, or the S phase of the cell cycle. In these highly specialized contexts they are replicated through signalling, which blocks the suppression of the replication process. After the function is fulfilled, a signal once again initiates suppressor function. Retroelements, with their (i) higher-order regulatory functions, (ii) capability for genetic creativity, and (iii) capacity for innovation of new regulatory patterns and combinations descended from retroviruses, which can easily be identified by their three essential parts *gag*, *pol,* and *env*.[26–28] Most endogenous retroviruses have been degraded into formerly connected domains, but they can still be recognized by retroposons or one of these three genes,[21,29–31] which means their formerly connected genomic content may be used by host organisms as single or networking modular tools for a variety of new regulatory functions. The *gag* gene encodes structural proteins, *pol* encodes enzymes such as reverse transcriptase, protease, ribonuclease and integrase functions, and *env* encodes envelope proteins, surface and transmembrane proteins and proteins causing host cell fusion and immunosuppression.[22]

Reconstructing the Highways that also Play Important Roles in Persistence

Interestingly we also find small DNA viruses as genetically stabilized and co-evolved

persistent viruses that do not trigger and are not part of an immune response of the host organism. Their active role is regulated and depends on the cell cycle of the host in which they are transcribed, replicated, and silenced again during a certain phase of the cell cycle. This means their persistent status is changed into an active role only during a strictly defined phase of the cell cycle in which they are needed for a specialized function and are still competent, most likely being adapted especially for this function. After fullfilling this function they disappear again. To ensure this function and its fine-tuned regulation, they need a highly conserved regulative viral protein domain with characteristic host interactions to use (manipulate) the host replication for their special needs. Their behavioral pattern is adapted to the host and circumvents the acute lytic phase.[32]

We can find similar reproduction patterns in RNA viruses. Replication of retroviruses in eukaryotes depends on successful entry into the membrane of host cells.[33] Some retroviruses circumvent this active entry of the cellular membrane in that they wait until the start of the cell-cycle phase-dependent dissolution of the cell membrane during replication. Once the direct path to the nucleus is free, retroviral RNA is transported to the eukaryotic nucleus and integrates into the host genome. Through transcription of the host genome a complete viral RNA genome is processed and transported out of the nucleus through the cytoplasm to the cell membrane. Here the reproduced viral genomes assemble and are encapsulated by viral *gag* encoded structural proteins and leave the host cell. Retroviral life integrates dozens of retroviral competences such as replication, transcription, translation, repair, trafficking to and from the nucleus, splicing, alternative splicing, and 3′ end processing. Transport to the nucleus and afterwards from the nucleus to the membrane again gives an overview about agent-driven evolution of eukaryotic cells if we think about the eukaryotic nucleus being an ancient dsDNA virus.

In contrast to these retroviral and often lytic replication cycles, the overwhelming majority are nonlytic but persistent retroviruses. They infect the host organism and integrate their *gag, pol, env* functional parts into the host genome and adapt to the replication cycle of the host organism without leaving the host cells. In becoming part of the identity of the host genotype they change the genome formation and transfer a phenotype to the host that noninfected host genomes do not possess. We are still at the beginning of imagining how the persistent lifestyle of viruses plays a role in the evolution, development, and genomic regulatory ratio of eukaryotes.[4]

This advantageous behavior of not leaving the host genome again, but reaching a persistent status within the host genome (most probably by becoming part of an addiction module, i.e., a competing genetic settler that creates an equilibrium status balanced by the host immune system), can be better understood if we look at the patterns of retroviral movement during infection events.

Patterns of Retroviral Movement: The Kinesin/Dynein Addiction Module

It is the prokaryotic viruses that tend to use pores during cell destruction and exit. Their entry pores are their tail plates, which can also be toxins. In eukaryotes, the most relevant pores are on the mitochondria, associated with apoptosis. However, eukaryotic DNA viruses (adenovirus) do often bind to nuclear pores and these are clearly also associated with microtubules that transport virus to the pore. This pore-building ability plays an evolutionary role in all tubular structures that connect multicellular tissues of eukaryotic organisms.

The alternative way of most eukaryotic viruses, including retroviruses, to pass through the cell membrane of the host is by endocytosis followed by receptor binding.[34] A well-known behavioral motif then occurs: the release of its RNA or DNA into the host cell and its immediate spread into smaller parts, such as reverse transcriptases and preintegration

complexes. This is an important behavioral motif in relation to noncoding RNA transcript processing into small noncoding RNA species, as discussed later.

These smaller parts move on the "highway" of the actin-based cytoskeleton and its microtubules in the direction of the nucleus[35] by using the kinesin-motor protein superfamily.[36] After reaching the nuclear membrane these viral parts pass through the membrane through nuclear pores by using importin proteins (which consists of two subunits with complementary functions: an indicator for an ancient addiction module itself) which bind to a special recognition sequence. Afterwards, integrase, which is also produced and used by DNA viruses for the same purpose, integrates these viral parts into the genome. The integration process is not random but involves strict coherence to the syntax of the nucleotide sequences of the host.

If the host cell replicates, these RNA viral parts are transcribed into DNA. As DNA sequences they pass through the membrane of the nucleus again to move towards the cell membrane. Again they use the "highway" of microtubules, but unlike earlier, they use dynein as the motor protein.[36] Interestingly, the change in direction of the kinesin/dynein transporter proteins depends on suppression of antagonists of the dynein or kinesin. This could be an indicator that both are part of an addiction module of former competing genetic parasites. Prior to becoming an internal part of cellular transport it could have been an external system for movement in and out of cells. And indeed there is a connection between these two motor proteins and retroviral (*gag*)-parts.[28,37]

When they reach the cell membrane the viral parts form a patch, via protein assembly on internal membranes, that is divided from the cell by exocytosis and produces its own capsule in which the RNA genome matures. In other contexts, similar proteins are involved in cytokinesis,[38] and most interestingly, in neural and immunological synaptic communication.[39,40] Other retroviruses build their capsule, not at the plasma membrane, but at the centrioles.[30] Both sites of capsid building and transport depend on intact *env* and *gag* codings and the recycling of membrane parts.[37] The *gag* parts bind the kinesin motor proteins that are needed for mictrotubulin transport. The exact transport of the retroviral RNA through the cytoplasm depends on interactions with numerous host proteins that build the so-called RNA-Transport-Granulat (RTG).[37] This RTG is also present in nearly all cells, such as fibroblasts, T-cells, and epithelial cells.[41,42]

Retroviral RNA editing for processing of extracellular viral parts is a very complex process: retroviral RNA editing functions in a similar way to cellular mRNA processing but is much more regulated by cis- and transactive mechanisms, which seem to be a former retroviral competence. Export of retroviral RNA out of the nucleus requires several RNA helicases, such as RNA helicase A, DEAD box proteins, DDX1 and DDX2.[43–46]

After retroviral RNA processing in the nucleus by (i) alternative splicing, (ii) 3′end processing, and (iii) RNA transport from the nucleus through the cytoplasm on the microtubule highways, they reach the membrane where they assemble. Transport depends on intact micro-tubules.[47] One part of the retroviral RNA is not spliced but is translated into structural proteins to form the capsules.[48–51] The addiction module of the antagonistic motor proteins kinesin/dynein (see note above), which drive retroviral RNA transport, plays an important role in the cell division of eukaryotic cells.[52,53] Different kinesin proteins regulate the movement (Kip2p) and also the direction/orientation of this movement (Kip3p).[54]

Interestingly, mitotic spindle processing without kinesin/dynein transcription does not function as well as positioning of the two poles and the segregation of the chromosomes in the anaphase.[55–58] Both motor protein families have genetic similarities with archaea and eubacteria, indicating their important roles in prokaryotes as well.[59–61] Motor proteins and cytoskeleton interactions are very

specific and also interconnect with the Golgi apparatus.[62,63]

Agents of Natural Genome Editing: Genetic Settlers in a Comfortable DNA Habitat

Recent research shows extensive dynamic DNA remodelling by mobile agents, such as transposons, retroposons, and noncoding RNAs, which are able to cause a wide variety of DNA arrangements, rearrangements, and recombinations.[64–67] Some authors refer to these as agents of genomic creativity,[21] mobile or regulatory elements[68–70] or entities,[71] while others refer to transposable elements,[72] noncoding RNA populations,[73] endogenous mutators,[74] and still others to mobile DNA species or genetic parasites.[4,75] Together, these agents enable complex organisms to integrate several temporal steps and a great variety of coordinated signalling processes in eukaryotic cell replication, fix them in a conserved DNA storage medium, and if necessary, resolve conservation, change, rearrange, or newly construct the whole genomic content and sequence order.[76]

The DNA information storage medium is and has to be edited. I predict a future discussion on how to refer to these editing agents, for example, as interactions of more or less chemical molecules or as "non-random genetic change operators" (Shapiro 2007, personal communication) or as natural genome editors.

From a bio-communicative perspective—which investigates combinatorial (syntactic), content-specific (semantic), and contextual (pragmatic) rules of genetic text processing—it is important to note that there can be no editing without competent agents that edit, that is, an editor or most likely a swarm of editors.[77] For example, the spliceosome or even the ribosome works as an integrated network of several small nuclear RNAs and their associated proteins.[65] They are clearly authorities in competently acting upon the molecular syntax of the DNA language.

Life could not function without the key agents of DNA replication, namely mRNA, tRNA, and rRNA. Not only rRNA, but also tRNA and the processing of the primary transcript into the pre-mRNA and the mature mRNA, are clearly descended from retro-"elements"[78–82] with obvious retroviral ancestry.

It is now possible to appreciate how sophisticatedly the competent agents act in the case of endogenous retroviral swarms, that have reached a persistent and nonlytic lifestyle. We also know that all related retro-elements share a common genome-editing capacity similar to transposable elements. Nonetheless, it remains difficult to reconstruct how all these DNA-encoded RNA agents reached a persistent status in hundreds, thousands, and tens of thousands of elements. We only know that they act in a precisely coordinated manner that would be impossible without competent signalling. This includes a strict capacity for self/nonself identification, which is a major asset of RNAs in general and of small nucleolar RNAs in particular.[83]

Persistent endogenous agents that are competent in natural genome editing apparently prefer a special kind of habitat characterized as noncoding DNA sectors. They use a syntax mainly consisting of direct or inverted repeats. They colonized DNA genomes by inserting their sites beside coding elements; then they use these coding elements for different needs. So we have to look at sequence orders that consist of noncoding elements, such as repeats and coding elements. In the human genome, only 3% of coding regions remain. The remaining 97% serve as a habitat for persistent viral operators that orchestrate a highly sophisticated division of labor. From these regions they can actively regulate coding sequences because they are able to change specific DNA content throughout the genome. All eukaryotic DNA replication processes share a cut-and-paste process in which noncoding elements, that is,

introns, are spliced out; later the remaining exons that code for proteins are combined into a coherent protein-coding content ready for translation.

As opposed to persistent endogenous agents of natural genome editing in eukaryotes, we find persistent exogenous agents in prokaryotes that are competent in natural genome editing of the prokaryotic gene pool. This process has long been visualized as horizontal gene transfer and is now recognized as occurring via plasmids, phages, and transposons, all with viral ancestors.[84]

It is difficult to perceive mere molecules or molecule buildings as being "competent" to process the sophisticated DNA language. It is less difficult to think of viruses as these subject-like agents.

Large Noncoding RNAs: Competent Regulators of Gene Expression

Noncoding RNAs that function in gene regulation coordinate and organize various actions, such as chromatin modification and epigenetic memory, transcriptional regulation, control of alternative splicing, RNA modification and RNA editing, control of mRNA turnover, control of translation, and signal transduction.[85]

In contrast to former opinions about the expression levels of genomes it is now increasingly clear that most eukaryotic genomes are highly expressed and a great abundance of noncoding RNAs with regulatory functions are transcribed. Most of these noncoding RNAs are alternatively spliced and divided into smaller RNAs that are integral parts of ribonucleoprotein (RNP) complexes. They regulate nearly all aspects of gene regulation. Small RNA species include micro-RNAs, small interfering RNAs, small nuclear RNAs, and small nucleolar and transfer RNAs.[86]

Although recent research has tried to evaluate the enormous regulatory networks of small RNAs, the role of thousands of longer transcripts is not yet clear. We know that they play important roles in histone modification, methylation, that is, epigenetic control of developmental processes such as the mammalian HOX clusters,[85] and also transcriptional interference, promoter inactivation, and effects on enzymatic pathways. Interestingly, these large noncoding RNAs are found as interlacing and overlapping sense and antisense transcripts derived from introns or intergenic regions, which means that they stem from the preferred life habitat of persistent viral settlers. That they may be of viral descent is indicated by their developmental stage- and tissue-specific expression patterns, which are typical habits of persistent viral settlers. Similar to their smaller relatives they are involved in the formation of ribonucleoprotein (RNP) complexes.

Micro-RNAs and Their Associated Proteins are RNPs

Small noncoding RNAs also share a special competence for epigenetic regulation of gene expression[87] and are derived from repetitive genomic sequences.[88] Repetitive genomic sequences indicate descent from retroviral infection events.[89] The capacity for epigenetic regulation of gene expression includes the "recognition" (identification) of specific sequences in other nucleic acids and is common to RNAs,[83] especially small nuclear RNAs and tRNAs that (i) identify splice junctions in both pre-mRNAs and codons, and (ii) process both the subunits of the spliceosome and the ribosome.[17] This implicates their capacity for self/nonself distinction as well as for identifying the molecular syntax. If one of these does not function, that is, error or damage occurs, the regulation or structural features of small noncoding RNAs do not function.

Maybe a key feature of noncoding RNAs is that they share an analog/digital language-competence,[90] such that in their secondary molecular structure they can act as molecular adaptors to the protein world, whereas their nucleotide word order seems to be digitally structured information. Their ability to edit

the molecular syntax of genetic texts according to different needs is exemplified by recent research on the functions of learning and memory in mammalian neuronal networks.[91,92] This is possible in that RNA editing alters transcripts from loci encoding proteins involved in neural cell identity, resulting in DNA recoding.[93]

The action of endogenous RNA species such as micro-RNAs, small interfering RNAs, and piwi RNAs in animals and plants is a mediating process in that they guide the binding of protein complexes to specific nucleic acid sequences.[94] There action starts both as a regulatory process at the transcriptional level (e.g., by endogenous siRNAs)[95] when the action potential is activated out of the "evolutionary protocols" fixed in the DNA storage medium, or at the post-transcriptional level in that they stabilize messenger RNA and translation into proteins.[96] This means that small noncoding RNAs do not solely mediate the transfer of genetic information from DNA to protein but also act as sequence-specific regulators in the expression of other RNA transcripts and, interestingly, in silencing specific transposons.[18,87]

Control patterns include mRNA degradation (siRNAs,), translational repression (miRNAs), heterochromatin formation, and transposon control (piwiRNAs).[87] Endogenous miRNAs and siRNAs share biogenesis and can perform interchangeable functions.[97]

They cannot be distinguished by their chemical composition or their action. But they differ in their production pathways: (i) miRNAs derive from genetic sequences that are different to known genes, siRNAS derive from mRNAs, transposons, viruses, (ii) miRNAs are processed out of transcripts that can form RNA hairpins, siRNAs are processed from long bimolecular RNA duplexes, (iii) miRNAs are always conserved in related organisms, endogenous siRNAs are rarely conserved, (iv) miRNAs are produced from genes that are specialists in silencing of different genes, siRNAs are typically auto silencing, such as viruses, transposons, and repeats of centromeres.[18] SiRNAs are expressed from extended double-stranded regions of long inverted repeats and can inhibit the expression of nearly any target gene in response to double-stranded RNA and have a very efficient and ancient immune function against genetic parasites.[98,99] This functions by identifying foreign RNA sequences and inhibiting their replication. The ancient RNAi immune function is based on self/nonself identification competence.[98] Many of these elements are retroposons or transposons and are encoded in the repetitive sequences of the genome[99] and therefore are clearly of viral origin.

Micro-RNAs are single-stranded RNAs of 19–25 nucleotides in length and are generated from endogenous hairpin transcripts of 70 nucleotide precursor miRNAs.[100] The transcription of this pre-miRNA is processed by RNA polymerases pol II and pol III. Whereas pol II produces the messenger RNA, small nucleolar and small nuclear RNAs of the spliceosome pol III produce shorter noncoding RNAs, such as tRNAs, some rRNAs, and a nuclear RNA that is part of the spliceosome.[18] They control not only developmental timing, hematopoiesis, organogenesis, apoptosis, and cell proliferation but also fat metabolism in flies, neuronal patterning in nematodes, and control of leaf and flower development in plants.[18] Most of them are processed out of introns!

It is predicted that every metazoan cell type at each developmental stage has a distinct miRNA expression profile.[18] The most characteristic differences of miRNAs acting in plants and animals are found in the stem loop.

Micro-RNAs, as well as small interfering RNAs, seem to have descended from transposable elements with an inherent regulatory ratio on gene regulation that is fulfilled by a variety of small interfering RNAs or micro-RNAs that act in a coordinated manner in that they share a division of labor in hierarchical steps of suppression and amplification. This is indicated in transposable elements that encode both siRNAs and miRNAs.[101] They can be found in intronic regions and build stem loop structures (hairpins) as a common feature of active RNA species, such as ribozymes. The

defense mechanism of host genomes against transposable element invaders through siRNA evolved into miRNAs with a new regulatory complexity and a new phenotype. First evolving as an immune function, it was later co-opted as a tool for complex regulatory pathways for host gene expression.[101] This co-option of identical competences for different purposes seems to be a common evolutionary pattern for regulatory controls that can be flexibly altered and rearranged to cause phenotypic variation without altering basic components.[102]

Micro-RNAs are acknowledged as key regulators of gene expression. This means that dysfunctions of these regulatory mechanisms may lead to dysregulated genes with a cascade of disease-causing consequences.[103] This means that the smallest RNAs are the basis of gene regulation, although in their original function they had an immune function against viruses and similar agents. Later on their function was adapted for eukaryotic gene expression. Sometimes these new functions overlap with old features when virus infection is acute[104] and the balanced (persistent) status is disturbed. The eukaryotic signalling pathways that act during gene expression or DNA replication and the role of transport from the cytoplasm through nuclear pores into the nucleoplasm, genetic expression, and retransport into the cytoplasm and the role of membranes, show that viruses had, and still have, manipulatory abilities in host genomes and host cells.[105] In particular, the ability of persistent invaders to self-splice out before export to the cytoplasm—an ability derived from the division of transcription and translation through nuclear membranes with pores—produces an analog protein-coding data set ready for translation and determines the identity of the host without harming its reproductive cycle.

RNPs and Their Functions

The complex network of all RNA interactions and the great variety of functions as preconditions of eukaryotic complexity include gene-gene interactions as well as the integration and regulation of most of the gene activities on different levels, such as chromatin structure, DNA-methylation, transcription, RNA splicing, RNA translation, RNA stability, and RNA signalling. This means that most of the functions of gene control that we currently know[17,66,67,73] descended from retro-elements. One important feature is that noncoding RNAs function as RNPs and not as naked RNAs.[106]

Some of them are small nuclear and small nucleolar RNAs (see below). Interestingly, the nucleolus plays an important role with similar performance to other endogenous retroelements during a small time window in the cell cycle. Nucleoli appear during the interphase of mammalian cells and are the locus of ribosome processing. Nucleoli also control regulation of the cell cycle.[107] In general, noncoding RNAs play important roles as intermediate products in the function and structure of the host genome, for example, in the reproductive cycle of host cells from the start to the end of the S-phase. This is a common feature of persistent endogenous retroviruses that adapted to the host genome as an addiction module. They now represent a phenotypic function of the host organism as neutralized (balanced) former competing viral agents. If one of these agents is damaged or its regulation becomes unstable the counterpart may still become virulent with potential disease-causing consequences for the host organism.[89]

Regulatory functions are restricted to specific phases of the cell cycle. One antagonist is suppressed and therefore the other can be produced at increasing rates according to the phase of the cell cycle, for example, production of tubulin and/or nanotubulin structures[108] or production of proteins with a variety of functions in cell division. At the end of this specific phase of the cell cycle the suppression of the antagonist ceases, mediated by specific signalling processes, and the antagonist then suppresses the increased production of its counterpart. This is the precondition for terminating its function in this phase of the cell cycle.

These interactions represent a kind of universal module with key regulatory function. Noncoding RNAs are involved in nearly all of these key functions of cellular activities in all domains of life (bacteria, eukarya, archaea).

Small Nuclear and Small Nucleolar RNPs

Most of these functions are fulfilled by noncoding RNAs, which act as binding partners to ensure the correct position of the nucleic acid target molecule for its enzymatic functions. Normally this process works as Crick-Watson base-pairing which includes a lot of proteins. These interconnected networks between noncoding RNAs and proteins are termed the RNPs.[106]

Some of these noncoding RNPs are small nuclear (sn) and small nucleolar (sno) RNPs. They are competent in multiple functions, many of them for intracellular transport and motility. In human cells they are encoded within introns. Both act in a complementary fashion: snRNPs retrieve the snoRNAs from the introns. On the other hand, snoRNAs are required by the snRNPs for posttranscriptional modifications. Their interdependency seems to be typical for viral derived addiction modules. Both snRNPs and snoRNPs depend on stable but inactive pre-RNPs, which only can mature if they are located far from their later active position.[106]

In most vertebrates, the snoRNAs come from introns of pre-RNA transcripts. These snoRNAs are integrated into a complex structure by endonucleases, exonucleases, and helicases.[109,110] Interestingly, in the yeast *Saccharomyces cerevisiae,* most snoRNAs are not encoded in introns.[83] This could be an indicator of an earlier evolutionary phase in which snoRNAs had not yet become endogenous by an infection event by retro-agents.

Most of the snoRNAs of vertebrates that are encoded in their introns are transcribed by polymerase II and produced after splicing through exonucleolytic trimming. They are highly lineage specific and form a separate family of mobile genetic elements.[111]

Genes coding for snoRNAs switch on host genes by retroposition. They also play important roles in excision and integration of specific sequences. They are as important as Alu, SINEs, and LINEs.[111] Interestingly, the best-known mobile genetic elements are integrated into snoRetroposons. The insertion is characterized by their individually different target site duplication-repeats, for example, full-length LINEs, that is, ALUs in primates. snoReverse Transcriptases (snoRTs) are also inserted into other mobile elements as well as into DNA transposons.[111] Retro-elements seem to be compatible with each other and are integrated into small nucleolar reverse transcriptase sequences, each of them with its individual target-site duplication repeat.[111] This could be an indicator of identification processes and identity modules. Many of the snoRNA genes are retroposons with retroviral ancestry.[111] snoRTs that are part of the introns of host genes can be used as RNAs with novel functions. After a retroposition event they can silence the snoRNA copy of their parents and lead to a new function of the new snoRNA.[111] This seems to be a common feature of endogenous retroposons and transposons in that these genetic invaders are driving forces of genomic creativity: the same competences in different contexts may lead to new functions and phenotypes.

Transport of small nucleolar RNPs is characerized by both nuclear and cytoplasmic cycles. Cajal bodies play important roles in intranuclear transport as centers for RNP assembly, transport, modification, and editing. Their production is a complex highly coordinated and conserved process. When the small nuclear RNPs arrive in the nucleoplasm they spread over the whole interchromatin space. Newly generated RNPs are assembled in the Cajal bodies before they are integrated into the nucleolar subdomains fibrallin and interchromatin clusters. This seems to be an indicator that they are processed and modified within

the Cajal bodies and specialized (informed) for specific functions.[106] In addition to their function as *de novo* assembling factors of RNPs, multicompetent Cajal bodies play an important role in modification and recycling processes of U4/U6 snRNP complexes that are remnants of splicing processes.[106] Cajal bodies are nuclear organelles with high motility and important roles in splicing, ribosome processing, and transcription.[112] It has been observed that Cajal bodies move from one end of the nucleus to the other, assemble with other Cajal bodies, or divide themselves into smaller ones.[113] These endogenous retro-elements are still important agents in natural genetic engineering and natural genome editing.

Important Roles of Small Nucleolar RNPs in Eukaryotic Genome Editing

Small nucleolar RNAs seem to derive from a very ancient group of endogenized RNA viruses with a wide variety of functions such as 2′-O-methylation, pseudouridylation of many classes of RNAs, rRNA processing, and synthesis of telomeric DNA.[114]

Today we know two families of snoRNAs: the C/D and the H/ACA RNAs. They are produced in the nucleolus and have a variety of functions within the nucleolus but also function out of the nucleolus as a recreation center for special substances. Eukaryotic cells have several dozen species of snRNAs and 200 known snoRNAs (C/D and H/ACA RNAs). These RNAs are one of the most diverse transacting RNAs currently known. They are available not only in eukaryotes but also in archaea. These snoRNPs (C/D H/ACA) share important functions such as: protein translation, mRNA splicing, genome stability, ribosome function, and modifications in snRNAs of eukaryotes, tRNAs in archaea, and neuronal mRNAs in mammals. One kind of H/ACA RNA, telomerase RNA, is needed for telomere production.[115]

Most of the known snoRNPs guide modifications of other ncRNAs. Both motifs are simple and very ancient and are part of the telomerase RNA.[116,117] Depending on their function, C/D and H/ACA finally orientate to nucleoli, Cajal bodies, or telomeres.[106]

We know that one H/ACA RNP functions in vertebrate telomere synthesis. Telomerase RNA and its protein partner, telomerase reverse transcriptase, are both target orientated and strictly regulated. Telomerase enzymes are active in telomere elongation only in the S-phase of the cell cycle. Interestingly, telomerase does not elongate telomeres continuously but at different times with interruptions in between cellular subcycles, especially in the shorter ones. The 3′ region of the endogenous human telomerase RNA possesses all the structural features of H/ACA boxes of snoRNAs. Only in the S-phase of the cell cycle do both telomerase RNA and telomerase reverse transcriptase move toward telomeres, otherwise they are in different (waiting?) positions.[118] The accumulation of human telomerase RNA in nucleoplasmic Cajal bodies in certain cancer cells occurs only during the S-phase of the cell cycle when telomerases are generated.[119] This means that the accumulation process is cell cycle dependent and time limited. Telomerase transport in the S-phase of the cell cycle is advantageous because telomerase activity in eukaryotes is limited to chromosome replication and suppresses destructive functions of telomerases at nontelomere elongation locations such as chromsome repair and repair of double-strand breaks.[113]

Cajal bodies in particular play important roles in the generation and function of telomerase RNPs. Human telomerase RNA has an H/ACA motif that it shares with small Cajal body RNAs and a wide variety of snoRNAs. The snoRNA family regulates modifications and cleavage of ribosomal RNAs in the nucleolus, and the small Cajal body RNA family regulates modifications of small nuclear RNAs within the Cajal body.[120]

Besides the nucleolus, the Cajal bodies are the best investigated nucleoplasmatic organelles. They are enriched by spliceosomal and nucleolar RNPs. Especially in the

interphase, many nucleolar intermediates emerge and play important roles in transport and modification of a variety of cellular RNAs.[121] C/D and H/ACA boxes guide the posttranscriptional pol-II specific spliceosomal snRNAs via pseudouridylation, which is characteristic of posttranscriptional RNA modifications in eukarya and archaea and generally plays an important role in the correct functioning of cellular RNAs. These guide RNAs share a Cajal body-specific localization signal, the CAB Box.[122] It is suggested that pseudouridylation guide RNPs play important roles in the processing of rRNA and in function-control/regulation of telomerases in eukaryotes. Ribosomal RNAs play important roles and therefore are DNA encoded in hundreds of transcription units. These units are organized in large tandem arrays. In active synthesis these rRNA loci form nucleoli with important roles in evolution and development. These rRNA loci are habitats for mobile elements like retroposons.[123] Pseudouridylation pockets also seem to direct snoRNAs to the nucleolus, which means that they have identity functions.[124]

In yeast, telomerase is not associated with H/ACA snoRNPs but with SM snRNPs, which demonstrates that telomerase is generated and regulated in different organisms in different ways.[106]

The Enigma of tRNA

Parallel to the ribosomal rRNAs, tRNA is one of the most important RNAs and is essential for the replication of both RNA genomes and DNA genomes. It seems to be a very ancient competence because it functions in a wide variety of contexts in quite different ways. Investigations of the variety of functions of tRNA show that it seems to be an addiction module-like association of different predecessors dating to the RNA world. The 3′ end of the tRNA structure could be the start of an ancient replicase. Replicase is the CCA-adding enzyme that is necessary to complete nucleotides that were lost because of incorrect starts. This CCA-adding activity has been the first telomerase function with its function of completing lost nucleic acid end sequences during replication. If it is really as old as suggested it must be present in all three domains of life: archaea, bacteria, eukarya. These CCA-adding enzymes are found in all three domains of life, are part of the same nucleotidyltransferase superfamily, and are very similar in their function.[79,80]

The genomic tag hypothesis is another part of the puzzle over the role of noncoding RNA abilities. It is suggested that the one half of tRNA evolved to mark single-stranded RNA genomes for replication in the early RNA world. The second half of the tRNA evolved separately as a primer of templated protein synthesis and started the RNP world. Both parts derived from independent RNA agents, which together built a kind of addiction module. Then this module became involved in translation from RNA template into protein. That tRNA plays important roles in the replication of single-stranded RNA viruses of bacteria, plants, and mammals, replication of duplex DNA plasmids of fungal mitochondria, retroviral replication, and also replication of present chromosomal telomeres, is less noticed nowadays.[79,80]

Ancient and Prominent: Reverse Transcriptases

Last but not least let's have a look at one of the most prominent and ancient natural genome editing agents: reverse transcriptases, such as telomerases that function in telomere maintenance are noncoding retro-elements. This indicates that their ancestry is of retroviral origin.[118] The vast majority of retro-elements use tRNAs or RNAs with strong secondary structures to process reverse transcription. Interestingly, there is a similarity between the function of tRNA in the production of proteins from RNA information, and reverse transcriptase, an enzyme that is crucial for turning RNA into DNA.[91]

Retroposition—a billions of years-old process—still plays important roles in building the structure and function of genomes in a continuing interaction between host genomes and the colonizing life strategies of mobile genetic invaders, an everlasting evolution-driving process of rearrangement, renovation, and innovation.[125]

Copying from RNA into DNA generally involves reverse transcriptases. Recent research has demonstrated that overlapping epigenetic marking in eukaryotic cells is an important evolutionary feature to silence the expression of mobility of these mobile elements.[72] Mobile elements can silence single genes as well as larger chromosomal regions and, therefore, play an important role in the evolution of diversity. They share the ability to recombine, rearrange, repair, and insert into genomic content with other retro-elements.[126,127] They influence neighboring genes through alternative splicing and are active agents as enhancers and promoters or act by polyadenylation patterns.[72]

Reverse transcriptases play key roles in mobile elements, such as transposons and retroposons, both of viral origin. One type of retroposon has direct repeats at its ends (LTR), whereas others do not (non-LTRs). Interestingly, the number of retroposons increases with every transposition (transposition duplication) so that they can expand genomes: LINE-1 is 20% of the human genome. In contrast, transposons contain a code for the transposase protein. This enzyme identifies the terminal inverted repeats that flank mobile elements, excises them and integrates itself instead. The gap at the donor site is repaired in a cut-and-paste transposition or is filled up with a copy of the transposon by a gap repair technique.[72] Transposons can also integrate themselves in phages and plasmids and are transferred with them into other cells.[84] This is evidence for self/nonself differentiation.

In contrast to nonmobile telomeres and centromeres, mobile sequences, such as transposons and retroposons, [128] and noncoding repetitive elements, such as LTRs, SINEs, and LINEs, enable far-reaching DNA rearrangement and reorganization.[64,129,130] Together, they play a decisive role in the evolution of new genomic structures.[130–132] The repetitive sequences are highly species-specific and are more suitable for determining the identity of species than the coding sequences.[4]

This does not mean that only mobile sequences represent ancient genetic settlers. The similarity of telomeres and centromeres—nonmobile repeat elements—in descent and their relatively poor loci of inverted repeats or retro-elements could indicate an ancient immune function that protects both from massive invasions by genetic parasites.[133,134]

Major Roles of Reverse Transcriptases in Natural Genome Editing

In addition, reverse transcriptases play key roles in altering genomic structures[135] and therefore, in evolutionary processes facilitated by natural genome editing. Reverse transcriptases are used to generate (i) copies of mRNAs that they need for integration into a genome, and (ii) copies of non-mRNAs, such as small nucleolar RNAs, one of the largest classes of noncoding RNAs,[109] which, like DNA copies, are SINEs. SINEs can initiate new genes that code for small RNAs with regulatory abilities in existing genes.

One further key feature of reverse transcriptases is that they are a primer for retroposons such as LTRs (copia, gypsy, Ty1, IAPs, HERVs). Non-LTRs (Het-A/TART, SINEs, LINEs) act like telomerases in several arthropods and plants. Moreover, reverse transcriptases are encoded and used by open reading frames (ORF), ORF1 (an RNA-binding and shuttling protein), ORF2 (endonuclease, reverse transcriptase activities), as well as ALUs (manipulation of LINE-1 function for mobilization), group II self-splicing introns and snoRNAs (type 1–3 retroposons), all of which have important regulatory functions.[106,111,113,136] Reverse transcriptases are also found in retroviruses of mammals and birds, in the hepadnavirus of mammals and

birds, and the caulimovirus of plants, in LTR retroposons of animals, plants, fungi, and protozoa, in non-LTR retroposons of animals, plants, fungi, and protozoa, in group II introns of bacteria, fungi, plant mitochondria, chloroplasts, and plastids, in mitochondrial plasmids of *Neurospora* mitochondria, and in multiple single-stranded DNAs.[4,137] Although many researches believe that non-LTRs evolved before retroviruses, there is recent evidence that viruses are their ancestors.[142]

Reverse transcriptases (DNA polymerases) together with RNA-dependent RNA polymerases replicate positive-strand RNA viruses, double-stranded RNA viruses, negative-strand RNA viruses, and retroviruses. RNA-dependent RNA polymerases produce dsRNAs in that they copy single-stranded RNA templates into dsRNAs. The RNA-dependent RNA polymerases are initiated by two or more complementary micro-RNA sites[72] that could indicate an addiction module because of its counterpart regulation. These polymerases are also involved in the coupling of heterochromatin for the production of siRNAs.[138] The RNAi system is competent in posttranscriptional gene silencing and is, therefore, a crucial instrument in keeping the balance between the need for expression and the need for silencing.[139]

As mentioned above, ORFs also code for reverse transcriptase. Many organisms have ORFs that code for proteins with sequences very similar to retroviral reverse transcriptases.[140,141] RNA-dependent DNA polymerase (reverse transcriptase) is related to RNA-dependent RNA polymerase. Rooting these lines of descent in RNA-dependent RNA polymerases yields two groups: (i) group 1 contains LTR retroposons, RNA viruses, DNA viruses; (ii) group 2 contains non-LTR retroposons, bacterial and other organelle parts.[73]

The telomerase function is cell-cycle regulated. It functions only when its suppression is removed. Once the telomerase function in telomere replication is fulfilled, a signal initiates its suppression again. A disturbed signalling process may lead to uncontrolled cell replication. Telomerase has to be transported to telomere repeats for its elongation during the S-phase of the cell cycle. The delivery agents are, again, Cajal bodies, small nucleolus-like organelles that are competent in (i) splicing, (ii) ribosome production, and (iii) transcription.[112,119]

Natural Genome Editors: Communal Interacting Agents with Persistent Status

In contrast to DNA with its stable features and enormous information storage potential, RNA is involved in the active parts of copying and coding processes, as demonstrated in new sequence generation, replicative processes, gene invention, and higher order regulations in all key processes of life. Prior to the evolution of cellular life it is proposed that very simple structured RNA (pre-RNAs) started by growing through base-pairing mechanisms without coding features. The selection of an RNA population, a direct product of error prone unedited RNA replication, is known as the quasispecies theory.

Growth by base-pairing mechanisms is different to the growth of primitive secondary structures of single-stranded RNAs, which can stabilize and replicate themselves as hairpin/stem loop structures with inherent coding capabilities. When coding began, catalytic functions connected with syntactic rule-ordered information enabled these simple molecules to act as semiotic agents: they became capable of generating nucleic acid sequences with a functional meaning that had to be recognized, identified, and interpreted correctly in the situational context, with a combinatorial pattern of the base pairs that differed from the diverse features inherent in nonself agents of the same structure, that is, self/nonself identification.[142] Biotic competences differ from abiotic interactions, because in contrast to the latter, biotic competences

which became active may even fail. These abilities are still evident in the t-loop structure of tRNAs, a variety of ribozymes, and self-enforcing RNAi loops, which couple heterochromatin assembly to siRNA production.[138]

The high density of early RNA life led to competing situations in which it was an advantage to escape into DNA informational storage and protein-based cellular life as outlined in another article.[143] Protein and DNA invention were a prerequisite for the evolution of divergence and the variety of life because all evolutionary inventions could be stored as evolutionary protocols[77] in this stable storage medium. The information content of the human genome is comparable to an archive of 5000 books with 300 pages each.

It appears that all the detailed steps of evolution stored in DNA that are read, transcribed, and translated in every developmental and growth process of each individual cell depend on RNA-mediated processes, in most cases interconnected with other RNAs and their associated protein complexes and functions in a strict hierarchy of temporal and spatial steps. It is clear that this regulatory order could not evolve by chance or that it represents solely a randomly derived mixture of nucleotides, but that it is composed of individual functions and integration into one developmental target, in strict coherence to the syntax of the nucleic acid language.

Today we are beginning to realize the degree of abundance and variety of RNA species with their different, sometimes complementary and competing roles in all key processes of life. RNAs play complementary roles in information processing and regulation.[144] In most cases they have an inheritable status, being integrated in the genome of organisms, and are termed endogenous. In other cases they are ancient individuals living in the cytoplasm of cells as persistent nonlytic parasites similar to DNA settlers, with important endosymbiotic roles. Their relation to viruses is close and some virologists consider an evolutionary tree of RNA species and RNA viruses. Interestingly, some DNA viruses have other features (linear chromosome, telomere ends, intron-like structures) that indicate a different origin with ancient roots comparable to RNA viruses. These features connect the evolutionary roots of archaea and eukaryotes, because ancient dsDNA viruses have similar features in viruses of archaea and the eukaryotic nucleus. A special feature is the RNA proofreading and repair ability of RNA polymerases, which would be the precondition of an RNA genome in the early RNA world because of the relatively unstable RNA structures.[145–147] On the other hand, this instability is a necessary precondition for the high productivity of different RNA sequences with the rapidly adapting features necessary in the early RNA world, that is, a large variety of different RNA identities.

In particular, the mode of replication of eukaryotes shows a wide variety of hierachical ordered processes that each depend on signalling processes to indicate the successful termination of the preceding process. A temporal order exists, with time windows processed by different process design connected with the cell cycle, along with transport systems of signals and complex messages, agents, co-agents and helpers, such as ancillary proteins and a network of interwoven regulatory elements which suppress or amplify the start and stops of semioses, production, regulation, and the whole toolbox of natural genetic engineering. We now know that all these processes involve RNAs, which become active via transcription out of the DNA storage medium prior to translation into proteins.

During translation from digital DNA storage into the analog code of protein language it is interesting that DNA information contains multiple RNA- and protein meanings, that is, from the same genetic data set it is possible to transcribe multiple RNA species or translate a variety of proteins according to the higher order regulations inherent in epigenetic control and/or transcriptional, pre- or posttranscriptional modification targets. Additionally, in eukaryotes we find some noncoding RNA species, such as small nuclear, small nucleolar, small

Cajal body, and small interfering RNAs. There are some indications that noncoding DNA plays a role in the *de novo* generation of genes.[148]

Nearly all genetic and genome editing processes involve RNAs, which in most cases function as a network. Most of the functions are performed, conducted, and regulated by noncoding RNAs, which are encoded in intronic DNA in most cases with repetitive syntax. Most of their functions are active only in an intermediate stage of RNA processing, which is regulated by strict starts and stops that are signal mediated. Error in these regulations is interconnected with organismic diseases.

Noncoding RNAs are similar to all kinds of mobile genetic elements such as LTRs, non-LTRs, SINEs, LINEs, snoRNAs, and snRNAs, all of persistent viral origin. They are integrated via addiction modules, that is, an agent/antagonist relationship between competing genetic settlers that are neutralized and balanced by their antagonism and the immune system of the host. This is an advantage for both the host, which attains a new genetic phenotype that noninfecte relatives do not possess, and the viral settlers, which both survive and co-evolve within a new genomic habitat.

Transfer RNA (tRNA) and ribosomal RNA (rRNAs), with their key functions in protein translation, represent such addiction modules, containing several subunits, each of them necessary for function. The whole cell cycle of eukaryotes is regulated by these noncoding RNAs.

Obviously DNA functions as both a relatively stable information storage medium, an evolutionary protocol to fix advantageous innovations, and as a comfortable habitat for persistent genetic settlers. If all the RNA capabilities derive from viruses or similar agents that compete in the available global pool of organismal genomes—viruses and their relatives are ten times more abundant than cellular genomes—then only those that give their hosts an advantageous genomic identity that is able to ward off an abundance of competing genetic settlers will survive. They must be able to build addiction modules (genetic and genomic innovations) together with the host immune system, each of them a unique culture-dependent habitat. Only then will the survival of both the genetic settlers and their host populations be likely.

Conclusion

The biocommunicative approach investigates both communication processes within and among cells, tissues, organs and organisms, as sign-mediated interactions, and nucleotide sequences as code, that is, language-like text, which follows in parallel three kinds of rules: combinatorial (syntactic), context-sensitive (pragmatic), and content-specific (semantic). From the bio-communicative perspective editing genetic text sequences requires—similar to signalling codes between cells, tissues and organs—biotic agents that are competent in correct sign use according to combinatorial, context-sensitive, and content-specific rules. Otherwise, neither communication processes nor nucleotide sequence generation or recombination can function. Even if the process of following these rules is very conservative, variations, alterations, and change may occur. The agents rule following identities may fail, and nucleic acid sequence grammar can be damaged or deformed and may become syntactically incorrect.

At the genomic level we can identify a remarkable process of change from a mechanistic view to the perspective of nonmechanistic genetic content processing. If we look at the current knowledge of hierarchical and temporal order of single steps and substeps in replication and transcription processes there must be natural genome editing agents that are competent both in generation of meaningful nucleotide sequences and in the use of these sequences according to different needs, such as integration, modification, recombination, and extraction into preexisting genetic texts. As we will see, in this respect DNA is not only an information storing archive but a life habitat for nucleic

acid language competent RNA agents of viral or subviral descent. These agents are competent in almost error-free editing of nucleotide sequences according to combinatorial, context-sensitive, and content-specific rules. They even generate nucleotide sequences *de novo*. They are also able to generate new rules of use for nucleic acid sequence modules by rearrangements in the higher order regulatory network of noncoding domains. Thus ancient sequences of the DNA storage medium may be used as modular tools in a wide variety of different contexts for new functions, made possible through the different meaning of syntactically identical sequences.

Competent agents in nucleic acid language are not solus ipse agents but are competent as mutual or parasitic "swarms" or "clouds," most of them RNA-based communities that share these competences. Their competence is a communal one, each of them being capable of self and nonself identification. The interactive competence of a community enables each individual to be competent. If we look at interacting communities, such as ribosomes and spliceosomes (each containing subunits without which they cannot function), we see their communal competence. If we look at the hierarchical processes of gene expression, transcription, RNA processing, mRNA and tRNA transport for translation we can also see communally acting agents. From the virus-first perspective they are now mutually interacting but may derive from formerly competing agents. mRNA and tRNA maturation in eukaryotes in particular also seem to reflect communal processing.

Formerly competing agents have reached an equilibrium status balanced by the immune response of the infected host to achieve a persistent lifestyle in the host genome, for example, toxin/antitoxin-modules. The number of communal interacting agents represented by ribosomes, spliceosomes, or even the consortium that cooperates as the adaptive immune system[139] ranges from a few to hundreds and thousands. In the latter case communal agents interact in DNA rearrangements with enormous consequences for many protein-based products that play important roles in immune functions.

This view could change the construction of research projects, that is, shifting the focus from mutational (random) changes of nucleotide sequences to investigating nucleotide sequences from the perspective of viral-derived sequences that now play important roles in the regulation of cellular functions. Their status within one of many addiction modules can be changed by nonbeneficial circumstances for the cell (e.g., stress) and they may become lytic again, resulting in a wide variety of diseases.

Acknowledgments

I am grateful to Luis P. Villarreal, Director of the Center of Virus Research at the University of California–Irvine for critical comments and helpful suggestions.

Conflicts of Interest

The author declares no conflicts of interest.

References

1. Claverie, J.M. 2005. Fewer genes, more noncoding RNA. *Science* **309:** 1529–1530.
2. Bell, P.J.L. 2001. Viral eukaryogenesis: Was the ancestor of the nucleus a complex DNA virus? *J. Mol. Evol.* **53:** 251–256.
3. Bell, P.J.L. 2006. Sex and the eukaryotioc cell cycle is consistent with a viral ancestry for the eukaryotic nucleus. *J. Theor. Biol.* doi: 10.1016/j.jtbi.2006.05.015.
4. Villarreal, L.P. 2005. *Viruses and the Evolution of Life*. ASM Press. Washington.
5. Forterre, P. 2006. The origin of viruses and their possible roles in major evolutionary transitions. *Vir. Res.* **117:** 5–16.
6. Forterre, P. 2006. Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: A hypothesis for the origin of cellular domain. *Proc. Natl. Acad. Sci. USA* **103:** 3669–3674.
7. Takemura, M. 2001. Poxviruses and the origin of the eukaryotic nucleus. *J. Mol. Evol.* **52:** 419–425.
8. Van Lent, J.W.M. & C. Schmitt-Keichinger. 2006. Viral movement proteins induce tubule formation in plant and insect cells. In *Cell-Cell Channels*. F. Baluska, D. Volkmann & P. Barlow, Eds.: 1–13.

Landesbioscience-Springer Verlag, Berlin, Heidelberg, New York.
9. Bapteste, E., R.L. Charlebois, *et al*. 2005. The two tempos of nuclear pore complex evolution: Highly adapting proteins in an ancient frozen structure. *Gen. Biol.* **6:** R85.
10. Witzany, G. 2007. The agents of genomic creativity. In *Biosemiotics in Transdisciplinary Contexts. Proceedings of the Gathering 6 in Biosemiotics*. G. Witzany, Ed.: 95–103. Umweb. Helsinki.
11. Forterre, P. 2001. Genomics and early cellular evolution. The origin of the DNA world. *Comptes rendus de l'Académie des sciences. Série 3, Sciences de la vie* **324:** 1067– 1076.
12. Forterre, P. 2002. The origin of DNA genomes and DNA replication proteins. *Curr. Opin. Microbiol.* **5:** 525–532.
13. Forterre, P. 2005. The two ages of the RNA world, and the transition to the DNA world: A story of viruses and cells. *Biochimie* **87:** 793–803.
14. Koonin E.V., T.G. Senkevich & V.V. Dolja, 2006. The ancient virus world and evolution of cells. *Biol. Dir.* **1:** 29.
15. Tran, E., J. Brown & E.S. Maxwell. 2004. Evolutionary origins of the RNA-guided nucleotide modification complexes: From the primitive translation apparatus? *Tren. Biochem. Sci.* **29:** 343–350.
16. Gorinsek, B., F. Gubensek & D. Kordis. 2004. Evolutionary genomics of chromovirus in eukaryotes. *Mol. Biol. Evol.* **21:** 781–798.
17. Mattick, J.S. 2003. Challenging the dogma: The hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* **25:** 930–939.
18. Bartel, D.P. 2004. MicroRNAs: Genomics, biogenesis, mechanism and function. *Cell* **116:** 281–297.
19. Cullen, B.R. 2006. Viruses and microRNAs. *Nat. Gen.* **38:** 25–30.
20. Ryan, F.P. 2004. Human endogenous retroviruses in health and disease: a symbiotic perspective. *J. Roy. Soc. Med.* **97:** 560–565.
21. Ryan, F.P. 2006. Genomic creativity and natural selection: A modern synthesis. *Biol. J. Linn. Soc.* **88:** 655–672.
22. Ryan, F.P. 2007. Viruses as symbionts. *Symbiosis* **44:** 11–21.
23. Gerdes, K. 2000. Toxin-antitoxin modules may regulate synthesis of macromolecules during nutritional stress. *J. Bac.* **182:** 561–572.
24. Makarova, K.S., N.V. Grishin & E.V. Koonin. 2006. The HicAB cassette, a putative novel, RNA-targeting toxin-antitoxin system in archaea and bacteria. *Bioinform* **22:** 2581–2584.
25. Witzany, G. 2007. *The Logos of the Bios 2. Bio-Communication*. Umweb. Helsinki.
26. Rashkova, S., S.E. Karam, R. Kellum *et al*. 2002. Gag proteins of the two Drosophila telomeric retrotransposons are targeted to chromosome ends. *J. Cell Biol.* **159:** 397–402.
27. Peterson-Burch, B.D. & D.F. Voytas. 2002. Genes of the Pseudoviridae (Ty1/copia retrotransposons). *Mol. Biol. Evol.* **19:** 1832–1845.
28. Tang, Y., U. Winkler, E.O. Fredd, *et al*. 1999. Cellular motor protein KIF-4 associates with retroviral gag. *J. Virol.* **73:** 10508–10513.
29. Gao, X., E.R. Havecker, P.V. Baranov, *et al*. 2003. Translational recoding signals between gag and pol in diverse LTR retrotransposons. *RNA* **9:** 1422–1430.
30. Sfakianos, J.N. & E. Hunter. 2003. M-PMV capsid transport is mediated by env/gag interactions at the pericentriolar recycling endosome. *Traffic* **4:** 671–680.
31. Ivanyi-Nagy, R. & J.L. Darlix. 2008. Ancient retrotransposons as possible remnants of the primitive RNP World. In *Retrotransposition, Diversity and the Brain*. F.H. Gage & Y. Christen, Eds. Springer. Berlin/Heidelberg, doi: 10.1007/978-3-540-74966-0_4.
32. Shadan, F.F. & L.P. Villarreal. 1996. The evolution of small DNA viruses of eukaryotes: Past and present considerations. *Vir. Gen.* **11:** 239–257.
33. Nisole, S. & A. Saib. 2004. Early Steps of retrovirus replicative cycle. *Retrovir.* **1:** 9.
34. Greber, U.F. 2002. Signalling in viral entry. *Cell Mol. Life Sci.* **59:** 608–626.
35. Greber, U.F. 2005. Viral trafficking violations in axons: The herpesvirus case. *Proc. Natl. Acad. Sci. USA* **102:** 5639–5640.
36. Vale, R.D. 2003. The molecular motor toolbox for intracellular transport. *Cell* **112:** 467–480.
37. Cochrane, A.W., M.T. McNally & A.J. Mouland. 2006. The retrovirus RNA trafficking granule: From birth to maturity. *Retrovir.* **3:** 18.
38. Carlton, J.G. & J. Martin-Serrano. 2007. Parallels between cytokinesis and retroviral budding: A role for the ESCRT machinery. *Science* **316:** 1908–1912.
39. Dustin, M.L. & D.R. Colman. 2002. Neural and immunological synaptic relations. *Science* **298:** 785.
40. Piguet, V. & Q. Sattenaut. 2004. Dangerous liaisons at the virological synapse. *J. Clin. Invest.* **114:** 605.
41. Shav-Tal, Y. & R.H. Singer. 2005. RNA localisation. *J. Cell Sci.* **118:** 4077–4081.
42. Mingle, L.A., N.N. Okuhama, J. Shi, *et al*. 2005. Localization of all seven messenger RNAs for the actin-polymerazitation nucleator arp2/3 complex in the protrusions of fibroblasts. *J. Cell Sci.* **118:** 2425–2433.
43. Yedavalli, V.S., C. Neuveut, Y.H. Chi, *et al*. 2004. Requirement of DDX3 dead box RNA helicase for

HIV – I Rev-RRE export function. *Cell* **119:** 381–392.
44. Reddy, T.R., W. Xu & F. Wong-Staal. 2000. General effect of Sam68 on rev/Rex regulated expression of complex retroviruses. *Oncogene* **19:** 4071–4074.
45. Li, J., H. Tang, T.M. Mullen, *et al*. 1999. A role for RNA helicase A in posttranscriptional regulation of HIV type 1. *Proc. Natl. Acad. Sci. USA* **96:** 709–714.
46. Tang, P.Z., C.H. Tsai-Morris & M.L. Dufau. 1999. A novel gonadotropin-regulated testicular RNA helicase. A new member of the dead-box family. *J. Biol. Chem.* **274:** 37932–37940.
47. Basyuk, E., T. Galli, M. Mougel, *et al*. 2003. Retroviral genomic RNAs are transported to the plasma membrane by endosomal vesicles. *Develop. Cell* **5:** 161–174.
48. Butsch, M. & K. Boris-Lawrie. 2002. Destiny of unspliced retroviral RNA: Ribosome and/or virion? *J. Virol.* **76:** 3089–3094.
49. Kaye, J.F. & A.M.Lever. 1999. Human immunodeficiency virus types I and 2 differ in the predominant mechanism used for selection of genomic RNA for encapsidation. *J. Virol.* **73:** 3023–3031.
50. Poon, D.T., E.N. Chertova & D.E. Ott. 2002. Human immunodeficiency virus type 1 preferentially encapsidates genomic RNAs that encode Pr55(Gag): Functional linkage between translation and RNA packaging. *Virol.* **293:** 368–378.
51. Butsch, M. & K. Boris-Lawrie. 2000. Translation is not required to generate virion precursor RNA in human immunodeficiency virus type 1-infected T-cells. *J. Virol.* **74:** 11531–11537.
52. Cottingham, F.R. & M.A. Hoyt. 1997. Mitotic spindle positioning in *Saccharomyces cerevisiae* is accomplished by antagonistically acting microtubule motor proteins. *J. Cell. Biol.* **138:** 1041–1053.
53. Terada, S. & N. Hirokawa. 2000. Moving on to the cargo problem of microtubule-dependent motors in neurons. *Curr. Opin. Neurobiol.* **10:** 566–573.
54. Miller, R.K., K.K. Heller, L. Frisen, *et al*. 1998. The kinesin-related proteins Kip2p and Kip3p, function differently in nuclear migration in yeast. *Mol. Biol. Cell.* **9:** 2051–2068.
55. Hoyt, M.A., L. He, K.K. Loo, *et al*. 1992. Two *Saccharomyces cerevisieae* kinesin-related gene products required for mitotic spindle assembly. *J. Cell. Biol.* **118:** 109–120.
56. Saunders, W.S., D. Koshland, D. Eshel, *et al*. 1995. *Saccharomyces cerevisieae* kinesin- and dynein-related proteins required for anaphase chromosome segregation. *J. Cell. Biol.* **128:** 617–624.
57. Hoyt, M.A., A.A. Hyman & M. Bähler. 1997. Motor proteins of the eukaryotic cytoskeleton. *Proc. Natl. Acad. Sci. USA* **94:** 12747–12748.
58. Haraguchi, K., T. Hayashi, T. Jimbo, *et al*. 2006. Role of the kinesin 2 family protein, KIF3, during mitosis. *J. Biol. Chem.* **281:** 4094–4099.
59. Defeu-Soufo, J.H. & P.L.Graumann. 2004. Dynamic movement of actin-like proteins within bacterial cells. *EMBO Rep.* **5:** 789–794.
60. Graumann, P.L. 2004. Cytoskeletal elements in bacteria. *Curr. Opin. Microbiol.* **7:** 565–571.
61. Graumann, P.L. & H.J. Defeu-Soufo. 2004. An intracellular actin motor in bacteria? *Bioessays* **26:** 1209–1216.
62. Preuss, M.L., D.R. Kovar, Y.R.J. Lee, *et al*. 2004. A plant-specific kinesin binds to actin microfilaments and interacts with cortical microtubules in cotton fibers. *Plant Physiol.* **136:** 3945–3955.
63. Lu, L., Y.R.J. Lee, R. Pan, *et al*. 2005. An internal motor kinesin is associated with the Golgi apparatus and plays a role in trichome morphogenesis in Arabidopsis. *Mol. Biol. Cell.* **16:** 811–823.
64. Shapiro, J.A. 2002. Genome organization and reorganization in evolution. *Ann. N. Y. Acad. Sci.* **981:** 111–134.
65. Vaughn, M.W. & R. Martienssen. 2005. It's a small RNA world, after all. *Science* **309:** 1525–1526.
66. Mattick, J.S. 2001. Non-coding RNAs: The architects of eukaryotic complexity. *EMBO Rep.* **2:** 986–991.
67. Mattick, J.S. 2006. The underworld of RNA. *Nat. Gen.* **38:** 393.
68. Eickbush, T. 1999. Mobile introns: Retrohoming by complete reverse splicing. *Curr. Biol.* **9:** 11–14.
69. Brosius, J. 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* **238:** 115–134.
70. Brosius, J. 2003. The contribution of RNAs and retroposition to evolutionary novelties. *Genet.* **118:** 99–115.
71. Daubin, V. & H. Ochman, 2004. Start-up entities in the origin of new genes. *Curr. Opin.Gen. Dev.* **14:** 616–619.
72. Slotkin, R.K. & R. Martienssen. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8:** 272–285.
73. Mattick, J. 2007. A new paradigm for developmental biology. *J. Exp. Biol.* **210:** 1526–1547.
74. Kidwell, M.G. & D.R. Lisch. 2001. Perspective: Transposable elements, parasitic DNA and genome evolution. *Evol.* **55:** 1–24.
75. Nakamura, T.M. & T.R. Cech. 1998. Reversing time: The origin of telomerase. *Cell* **92:** 587–590.
76. Shapiro, J.A. 2006. Genome informatics: The role of DNA in cellular computations. *Biol. Theory* **1:** 288–301.
77. Vetsigian, K., C. Woese & N. Goldenfeld. 2006. Collective evolution and the genetic code. *Proc. Natl. Acad. Sci. USA* **1003:** 10996–10701.

78. Rao, A.L.N., T.W. Dreher, L.E. Marsh, *et al*. 1989. Telomeric function of the tRNA-like structure of brome mosaic virus RNA. *Proc. Natl. Acad. Sci. USA* **86:** 5335–5339.
79. Maizels, A. & A.M. Weiner. 1993. The genomic tag hypothesis: Modern viruses as molecular fossils of ancient stretegies for genomic replication. In *The RNA World*. R.F. Gesteland & J.F. Atkins, Eds.: 577–602. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.
80. Maizels, N., A.M. Weiner, D. Yue, *et al*. 1999. New evidence for the genomic tag hypothesis: Archaeal CCA-adding enzymes and tRNA substrates. *Biol. Bull.* **196:** 331–334.
81. Flavell, A.J. 1995. Retroelements, reverse transcriptase and evolution. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* **110:** 3–15.
82. Eickbush, T.H. & D.G. Eickbush. 2007. Finely orchestrated movements: Evolution of the ribosomal RNA genes. *Genet.* **175:** 477–485.
83. Filipowicz, W. 2000. Imprinted expression of small nucleolar RNAs in brain: Time for RNomics. *Proc. Natl. Acad. Sci. USA* **97:** 14035–140357.
84. Frost, L.S., R. Leplae, A.O. Summers, *et al*. 2005. Mobile genetic elements: The agents of open source evolution. *Nat. Rev. Microbiol.* **3:** 722–732.
85. Amaral, P.P., M.E. Dinger, T.R. Mercer, *et al*. 2008. The eukaryotic genome as an RNA machine. *Science* **319:** 1787–1789.
86. Yazgan, O. & J.E. Krebs. 2007. Noncoding but nonexpendable: Transcriptional regulation by large noncoding RNA in eukaryotes. *Biochem. Cell. Biol.* **85:** 484–496.
87. Chu, C.Y. & T.M. Rana. 2007. Small RNAs: Regulators and guardians of the genome. *J. Cell. Physiol.* **213:** 412–419.
88. Farazi, T.A., S.A. Juranek & T. Tuschl. 2008. The growing catalogue of small RNAs. Exiqon. The 2008 Collection Booklet 35–69.
89. Witzany, G. 2006. Natural genome editing competences of viruses. *Acta Biotheor.* **54:** 235–253.
90. St.Laurent III, G. & C. Wahlestedt. 2007. Noncoding RNAs: Couplers of analog and digital information in nervous system function. *Trends Neurosci*. 30: 612–631.
91. Presutti, C., J. Rosati, S. Vincenti, *et al*. 2006. Non coding RNA and brain. *BMC Neurosci.* **7:** S5.
92. Mehler, M.F. & J.S. Mattick. 2007. Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiol. Rev.* **87:** 799–823.
93. Mattick, J. & M.F. Mehler. 2008. RNA editing, DNA recoding and the evolution of human cognition. *Trends Neurosci.* **31:** 227–233.
94. Ambros, V. & X. Chen. 2007. The regulation of genes and genomes by small RNAs. *Develop.* **134:** 1635–1641.
95. Watanabe, T., Y. Totoki, A. Toyoda, *et al*. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453:** 539–543.
96. Chen, K. & N. Rajewsky. 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Gen.* **8:** 93–103.
97. Doench, J.G., C.P. Petersen & P.A. Sharp. 2003. siRNAs can function as miRNAs. *Gen. Develop.* **17:** 438–442.
98. Fire, A. 2005. Nucleic acid structure and intracellular immunity: Some recent ideas from the world of RNAi. *Quart. Rev. Biophys.* **38:** 303–309.
99. Sontheimer, E.J. & R.W. Carthew. 2005. Silence from within: Endogenous siRNAs and miRNAs. *Cell* **122:** 9–12.
100. Kim, V.N. 2005. MicroRNA biogenesis: Coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* **6:** 376–385.
101. Piriyapongsa, J. & I.K. Jordan. 2008. Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA* **14:** 1–8.
102. Mattick, J. & Gagen. 2001. The evolution of controlled multitasked gene networks: The role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.* **18:** 1611–1630.
103. Calin, G.A. & C.M. Croce. 2006. MicroRNA signatures in human cancers. *Nat. Rev. Can.* **6:** 857–866.
104. Quellet, D.L., M.P. Perron, L.A. Gobeil, *et al*. 2006. Micro RNAs in gene regulation: When the smallest govern it all. *J. Biomed. Biotech.* **69616:** 1–20.
105. Miller, S. & J. Krijnse-Locker. 2008. Modification of intracellular membrane structures for virus replication. *Nat. Rev. Microbiol.* **6:** 363–374.
106. Matera, A.G., R.M. Terns & M.P. Terns. 2007. Non-coding RNAs: Lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Mol. Cell. Biol.* **8:** 209–220.
107. Hiscox, J.A. 2002. The nucleolus – a gateway to viral infection. *Arch. Virol.* **147:** 1077–1089.
108. Rustom, A., R. Saffrich, I. Markovic, *et al*. 2004. Nanotubular highways for intercellular organelle transport. *Science* **303:** 1007–1010.
109. Zemann, A., A. Beckke, M. Kiefmann, *et al*. 2006. Evolution of small nucleolar RNAs in nematodes. *Nuc. Ac. Res.* **34:** 2676–2685.
110. Richard, P., A.M. Kiss, X. Darzacq, *et al*. 2006. Co-transcriptional recognition of human intronic box H/ACA snoRNAs occurs in a splicing-independent manner. *Mol. Cell. Biol.* **26:** 2540–2549.

111. Weber, M.J. 2006. Mammalian small nucleolar RNAs are mobile genetic elements. *PLoS Genet.* **2:** e205.
112. Platini, M., I.Goldberg, A.I. Lamond, *et al*. 2002. Cajal Body dynamics and association with chromatin are ATP dependent. *Nat. Cell. Biol.* **4:** 502–508.
113. Tomlinson, R.L., T.A. Ziegler, T. Supakorndej, *et al*. 2006. Cell cycle-regulated trafficking of human telomerase to telomeres. *Mol. Biol. Cell.* **17:** 955–965.
114. Kiss, T. 2002. Small nucleolar RNAs: An abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109:** 145–148.
115. Kiss, A.M., B.E. Jády, X. Darzacq, *et al*. 2002. A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. *Nuc. Ac. Res.* **30:** 4643–4649.
116. Clouet d'Orval, B., M.L. Bortolin, C. Gaspin, *et al*. 2001. Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp. *Nuc. Ac. Res.* **29:** 4518–4529.
117. Wang, C. & U.T. Meier. 2004. Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J.* **23:** 1857–1867.
118. Witzany, G. 2008. The viral origins of telomeres, telomerases and their important role in eukaryogenesis and genome maintenance. *Biosem.* **2:** 191–206.
119. Jady, B.E., E. Bertrand & T. Kiss. 2004. Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *J. Cell. Biol.* **164:** 647–652.
120. Fu, D. & K. Collins. 2006. Human telomerase and Cajal body ribonucleoproteins share a unique specificity of Sm protein association. *Gen. Dev.* **20:** 531–536.
121. Darzacq, X., B.E. Jady, C. Verheggen, *et al*. 2002. Cajal body-specific small nuclear RNAs: A novel class of 2′-O-methylation and pseudouridylation guide RNAs. *EMBO J.* **21:** 2746–2756.
122. Richard, P., X. Darzacq, E. Bertrand, *et al*. 2003. A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *EMBO J.* **22:** 4283–4293.
123. Eickbush, T.H. & D.G. Eickbush. 2007. Finely orchestrated movements: Evolution of the ribosomal RNA genes. *Genet.* **175:** 477–485.
124. Narayanan, A., A. Lukowiak, B.E. Jády, *et al*. 1999. Nucleolar localization signals of box H/ACA small nucleolar RNAs. *EMBO J.* **18:** 5120–5130.
125. Brosius, J. 2005. Echoes from the past – are we still in an RNP world? *Cyt. Gen. Res.* **110:** 1–4.
126. Teng, S.C., B. Kim & A. Gabriel. 1996. Retrotransposon reverse transcriptase-mediated repair of chromosomal breaks. *Nature (London)* **383:** 641–644.
127. Coffin, J.M., A.H. Hughes & H.E. Varmus. 1997. *Retroviruses*. Cold Spring Harbor Laboratory Press. New York.
128. Volff, J.N. 2006. Turning junk into gold: Domestication of transposable elements and the creation of new genes in eukaryotes. *BioEssays* **28:** 913–922.
129. Sternberg, R. 2002. On the roles of repetitive DNA elements in the context of a unified genomic-epigenetic system. *Ann. N. Y. Acad. Sci.* **981:** 154–188.
130. Shapiro, J.A. & R. Sternberg. 2005. Why repetitive DNA is essential to genome function. *Biol. Rev.* **80:** 1–24.
131. Shabalina, S.A. & N.A. Spiridonov. 2004. The mammalian transcriptome and the function of noncoding DNA sequences. *Gen. Biol.* **5:** 105e.
132. Sternberg, R. & J.A. Shapiro. 2005. How repeated retroelements format genome function. *Cyt. Gen. Res.* **110:** 108–116.
133. Nosek, J., P. Kosa & L. Tomaska. 2006. On the origin of telomeres: A glimpse at the pre-telomerase world. *Bioessays* **28:** 182–190.
134. Villasante, A., J.P. Abad, R. Planello, *et al*. 2007. *Drosophila* telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. *Gen. Res.* **17:** 1909–1918.
135. Haoudi, A. & J.M. Mason. 2000. Reverse trannscriptase can stabilize or destabilize the genome. *Genome* **43:** 949–956.
136. Batzer, M.A. & D.L. Deininger. 2002. ALU repeats and human genomic diversity. *Nat. Rev. Gen.* **3:** 370–380.
137. Temin, H.M. 1985. Reverse transcription in the eukaryotic genome: Retroviruses, pararetroviruses, retrotransposons and retrotranscripts. *Mol. Biol. Evol.* **2:** 455–468.
138. Sugiyama, T., H. Cam, A. Verdel, *et al*. 2005. RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc. Natl. Acad. Sci.* **102:** 152–157.
139. Grewal, S.I. & S.C. Elgin. 2007. Transcription and RNA interference in the formation of heterochromatin. *Nature* **447:** 399–406.
140. Xiong, Y. & T.H. Eickbush. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9:** 3353–3362.
141. Mesnard, J.M. & G. Lebeurier. 1991. How do viral reverse transcriptases recognize their RNA genome? *FEBS Lett.* **287:** 1–4.
142. Villarreal, L.P. 2009. *Origin of Group Identity: Viruses, Addiction and Cooperation*. Springer Verlag, New York.

143. Witzany, G. 2008. Bio-communication of bacteria and their evolutionary roots in natural genome editing competences of viruses. *Open Evol. J.* **2:** 44–54.
144. Kenzelmann, M., K. Rippe & J.S. Mattick. 2006. RNA: Networks & imaging. *Mol. Sys. Biol.* **2:** 44.
145. Jeffares, D.C., A.M. Poole & D. Penny. 1998. Relics from the RNA world. *J. Mol. Evol.* **46:** 18–36.
146. Poole, A.M., D.C. Jeffares & D. Penny. 1998. The path from the RNA world. *J. Mol. Evol.* **46:** 1–17.
147. Poole, A.M. & D.T. Logan. 2005. Modern mRNA proofreading and repair: Proof that the last universal common ancestor (LUCA) had an RNA genome? *Mol. Biol. Evol.* **22:** 1444–1455.
148. Levine, M.T., C.D. Jones, A.D. Kern, *et al*. 2006. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc. Natl. Acad. Sci. USA* **103:** 9935–9939.