

Reverse Transcription of Corona and other RNA viruses

Karin Moelling*

Department of Medical Microbiology, University Zürich, Zürich, Switzerland

Abstract

The human genome is full of endogenized DNA copies from retroviruses. Not only retroviruses but almost all RNA viruses can leave DNA footprints in mammalian genomes. The viral DNA sequences from RNA viruses often cover single structural viral genes not the full length viral genomes and are unable to produce infectious virus particles. Molecular clocks allow to date endogenization events back to about 90 Million years, but they are still ongoing. Endogenous sequences often maintain an open reading frame suggesting that they are selected for some function and supply some evolutionary benefits for the host. Indeed, the sequences protect against superinfection by related viruses and against disease. The viral DNA copies require a reverse transcriptase for DNA synthesis from the viral RNA. This enzyme is available from mobile genetic elements such as retrotransposons or LINES which were shown to reverse transcribe the RNA from Bornavirus, Ebola-, Marburg Disease and other viruses. Recently, Corona viral RNA has been described as DNA copy in the human genome. Again, only some viral sequences were detected and no full length genomes or infectious virus. RNA vaccines could potentially also contribute to some DNA copies even though this has not been shown. The presence of such endogenized viral genes and their expression is of benefit for the host and protective against superinfection or disease. This would include DNA copies from Corona viral RNA vaccines. A potential genotoxic effect due to integration and possible gene disruption cannot be excluded but is estimated to be rare. The risk of gene toxicity is certainly lower for an RNA vaccine than for a Corona viral infection. Therefore a vaccination is beneficial and we could even profit from protein expression of endogenized viral DNA as a genetic vaccine.

Keywords: RNA viruses • Integration • Endogenization • Complementary DNA • Reverse transcriptase • Retrotransposon

Introduction

Reverse transcription

The coronavirus disease 2019 (COVID-19) was first reported in December 2019 in Wuhan, China. Retroviruses replicate their RNA genome by retrotranscribing it into a DNA copy as an obligatory step, which integrates into host chromosomal DNA as DNA proviruses. Then progeny virus is produced by RNA transcription, protein synthesis, particle assembly and release by budding out of the host cells without lysis [1]. The name reverse transcriptase (RT) indicates that a "normal" mechanism seemed to be reversed. Now this mechanism is looked at as the original or major one. If the RNA world preceded the DNA world a copy mechanism from RNA to DNA was required possibly initially even without a protein enzyme [2]. The next surprise came with the sequencing of the human genome, the detection of retroviral sequences in the mammalian genomes [3]. Almost 50% of the human genome consists of retrovirus related sequences.

Nucleic acids that replicate by reverse transcription are summarized as the retroelements. They are wide spread and occur in higher plants, higher animals, fungi, insects, and bacteria. Retroelements are classified into the real retroviruses, para retroviruses with DNA genomes but RT-dependent replication, eukaryotic retroelements or retrotransposons, and bacterial RT-retroelements [4].

Retroviruses occur in two versions, as infectious particles replicating via integrated DNA copies in somatic cells but also as endogenized sequences in germ cells with subsequent inheritance a rare event. Human endogenous retroviruses HERVs and various shorter retroviruses like derivatives

populate the human genome [3]. Only some of them, about 8% to 10% of the human and mouse genomes, contain full length retroviruses, others are defective viral genomes with various sized deletions. Numerous genetic elements consist of solo retroviral promoters, the long terminal repeats, LTRs, which can regulate viral and neighboring cellular genes.

Para retroviruses replicate via an RT from pregenomic RNAs but carry DNA inside their particles [1]. There are numerous mobile genetic elements or transposable elements, which can move within the cell by a cut and paste or copy and paste mechanism as transposons or retrotransposons, respectively. The first kind leads to transposition of cellular genetic elements to other locations within the genome; the latter one requires an RT as intermediate to copy a mRNA into DNA, which leads to gene duplication upon integration. These movements are locked within the cell. The RT for these events originates from LINE-1 or LINE elements, long interspersed nuclear element, which encode two proteins, an endonuclease and an RT [5].

LINE elements

LINE elements make up 17 to 21% in the human genome, 850.000 copies, 7000 bases long coding for an endonuclease and RT. Most importantly, only about 100 LINES are active and have the potential for mobilization. Their RT can also reverse transcribe cellular mRNAs or RNA from the SINE elements. SINEs are smaller elements, less autonomous, representing about 13% presence of the genome, and borrow the RT from the LINES. In addition there are the LTR retrotransposons amounting in mammals to about 20% of the genome. These integrated fossil retroviral elements play major roles, since they contribute to evolution, defense, and adaptation to environmental stress, genomic diversity and sometimes to

*Corresponding Author: Karin Moelling, Department of Medical Microbiology, University Zürich, Zürich, Switzerland, Tel: +41 79 800 8337; E-mail: moelling@molgen.mpg.de

Copyright: © 2021 Moelling K. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received date: 20 May, 2021; Accepted date: 03 June 2021, Published date: 10 June, 2021

cancer [6,7].

Endogenous retroviruses

Integration of retroviruses is an obligatory step during replication and normally occurs in somatic cells. It involves the RT and the viral integrase for integration. Rarely integration occurs into the chromosome of germ cells leading to endogenization which includes inheritance to future generations. The viral sequences then become part of the genome of the offspring of the host and the viral sequence will be inherited in a Mendelian fashion. So far mainly the simple animal retroviruses succeeded in becoming endogenously inherited viruses. For that the virus needs to enter germ cells, which may require viral receptors. Recently such receptors for HIV have been identified on germ cells suggesting that HIV may also integrate one day and become endogenized [8]. Endogenous lentivirus has not been detected but simple retroviruses only, possibly because lentiviruses may be younger.

Syncytin and phenix

Several proteins encoded by HERVs have been reported to play important roles in host physiology for example formation of a placenta, which allows mammals to grow embryos in their placenta. This happened by the infection of mammals with the Human endogenous retrovirus HERV-W, which integrated and expressed a protein syncytin into syncytioblasts in the placental trophoblast layer between two lipid layers. It is closely related to the HIV protein gp41, the transmembrane protein with immune suppressive function in HIV infected people and AIDS disease [9]. It caused immunosuppression in female vertebrates and thereby allowed the growth of embryos in the placenta and abrogated the need of a mother to lay eggs. The immunosuppressive effect on the mother prevents immune rejection of the embryo. Such a protective retroviral infection occurred independently many times during evolution in different mammalian species. It is documented in 18 different species [10,11]. This happened throughout 85 Mio years [12].

The relationship of syncytin to HIV is reminiscent today of the effect of HIV on cells in culture by the fusogenic envelope protein gp41, which induces syncytia, cell fusion leading to giant cells. This is a simple indicator for the presence of HIV in cell culture [12].

Close to 50% of the human genome consists of retrovirus like sequences [3]. The sequences indeed originate from real viral infections millions of years ago. This was demonstrated by deriving a consensus sequence of a dozen of mutated endogenous retroviral sequences. The consensus sequence was then synthesized, transfected into cells, where virus was produced and detectable by electron microscopy, where it clearly resembles today's retroviruses. It was designated as Phenix retrovirus which entered the mammalian genome about 5 Mio years ago [13].

Some of the endogenized retroviruses in the mammalian genomes may have become endogenized between 30 and up to 93 Mio years ago [11].

The integration of the HERV-K in the human genome was analyzed to learn about an endogenous HERV and its effect on neighboring gene expression. The 450.000 HERV elements make up 8% of the human genome [3]. HERVs and other regulatory elements (RE) can influence genome regulatory elements such as enhancers, promoters, and splice and polyadenylation sites. The integrated viral promoters can contribute to the transcriptomes which make up 48.0000 transcripts per cell. They could be in sense or antisense orientation to the neighboring genes and might thereby upregulate or downregulate their expression by sense or antisense effects.

Reverse oriented proviruses may even be beneficial by protecting with antisense RNA mechanism from newly infecting retroviruses [14].

Indeed the HERV-K retrovirus family members contribute substantially to the cellular transcriptome. Expression of their viral RNA has been shown to modulate host gene expression. The HERV-K (HML-10) (a human mouse like insert-10) was studied in detail as a model. It invaded the human genome about 35 Mio years ago and is enriched within the introns. The

retroviral promoters LTRs exhibit activity and the HML-10 express an LTR-primed RNA transcript active in the opposite direction of the neighboring gene, coding for the pro-apoptotic Death Associated Protein (DAP3), and thereby suppressed apoptosis. This resulted in unreduced cellular proliferation which is a characteristic of cancer cells. However, in this case the weak expression of the viral transcripts did not allow concluding how important the event was in general and its possible role in cancer [15]. It might be possible that integrated retroviruses 35 Mio years old within the genomes have become harmless and were selected not to do any harm. Others may have been eliminated. This raises the question about the effect of newly endogenized sequences. Such a process can be followed in real time in Koalas.

Koalas

Most importantly this process of endogenization can be actively ongoing today and is not an event specific or exclusive to ancient viral fossils. There is an exciting example of such an ongoing process of endogenization within the Australian Koalas. They were under threat of extinction and were put into custody for recovery. There they got infected by a retrovirus related to Gibbon Ape Leukemia virus, a cancer virus. The virus is in Koalas an intermediate involved in both, vertical and horizontal transmission, endogenization as well as infection. It allows to study the process and consequences of retroviral endogenization in action. For about 100 years the process of virus endogenization has been ongoing in Koalas and thereby protects against superinfection. Many Koalas died of the leukemic effect of the retrovirus, but some of the Koalas became immune during their life time. This case demonstrates that a retrovirus can become endogenous and thereby protective against viral superinfection [8,16,17]. The human genome is full of retroviral elements which lead to host protection against secondary infections of closely related other viruses. Koalas teach us that endogenization is still an actively ongoing process today.

RT for Borna, Ebola and other Viruses

Endogenous viral DNA copies were thought to be molecular viral fossils in the mammalian genomes and were assumed to be restricted to endogenized retroviruses and related sequences. However, besides retroviruses there are other viruses which entered the mammalian genetic material. These comprise RNA viruses and DNA viruses. Thus, not only retroviruses but also non retroviral viruses have entered the genome.

More recently elements from ancient Borna viruses have been detected in genomes of several mammalian species, including humans. This was proven by a comprehensive study with a sequence comparison of various viral species using an informatics approach which was performed by Belyi and colleagues [18,19]. 5.666 viral genes from all known non-retroviral families with single stranded RNA genomes were matched against the germline of 48 vertebrate species to determine if such viruses could contribute to the vertebrate genetic heritage. Eighty integrated genomic DNA sequences found in vertebrate species originated from RNA viruses from 40 Mio years ago. Two families of circulating virus families dominated such as Borna viruses and Filoviruses (which comprise Ebola and Marburg disease viruses). Both types cause lethal neurological diseases or hemorrhagic fever. Borna viruses are negative single stranded RNA viruses, which establish non-cytopathic persistent infections in the nuclei of host cells in many vertebrate species. Only some of the Borna virus genes were detected, those for the major nucleocapsid protein, N (p40), the matrix protein M, the RNA-dependent RNA-polymerase and some glycoproteins. The size of the endogenous nucleoproteins is about 366 amino acids. Endogenization was generally limited to one or very few copies of a related viral gene per species, suggesting that later integrations failed or provided little advantage to the host. This appears to be different for the endogenous retroviruses which are numerous.

Borna virus sequences tend to be found in animals and humans which are resistant to Borna virus infections or do not develop severe disease

symptoms, while horses lacking the sequences can fall seriously ill. This pathology leads to the conclusion that Borna viral sequences serve as restriction factors against Borna viral infections. They protect against de novo infections and disease. Such a defense mechanism is known also in bacteria, where it is called "superinfection exclusion", and means "no entry", indicating immunity of the host [17]. Importantly, the often long reading frames for the integrated viral sequences are consistent with the notion, that their products provide some important biological advantage to the species. The reverse may also be true, the viruses could also benefit for their own survival which is provided by the natural reservoirs for their persistence and transmission. However today they are not replication competent.

Borna virus sequences tend to be found in animals and humans which are resistant to Borna virus infections or do not develop severe disease symptoms, while horses lacking the sequences can fall seriously ill. This pathology leads to the conclusion that Borna viral sequences serve as restriction factors against Borna viral infections. They protect against de novo infections and disease. Such a defense mechanism is known also in bacteria, where it is called "superinfection exclusion", and means "no entry", indicating immunity of the host [17]. Importantly, the often long reading frames for the integrated viral sequences are consistent with the notion, that their products provide some important biological advantage to the species. The reverse may also be true, the viruses could also benefit for their own survival which is provided by the natural reservoirs for their persistence and transmission. However today they are not replication competent. The detected virus sequences belong to viruses with extremely high mortality rates in some susceptible species reaching up to 80% in horses if they develop Borna viral disease and up to 90% in humans infected with Ebola virus.

Endogenized Ebolavirus sequences also comprised only some of their genes, mainly the genes for the major viral nucleocapsid protein NP and a polymerase cofactor VP35 [20]. Substitutions within the endogenous sequences lead to reduced pathogenicity and infection. The mutated virus still develops antibodies that render the host resistant to subsequent virus challenge.

The mutation rate known to be high for RNA viruses is severely reduced in integrated DNA copies and was estimated to be lower by four orders of magnitude. Borna viruses replicate in the nucleus while Filoviruses in the cytoplasm, yet in both cases DNA copies can be produced. LINE expression occurs inside the nucleus, so that its RT and nuclear RNA genomes have a higher chance to interact!

Thus, endogenous sequences of the Borna and Filoviruses viruses can lead to prevention of subsequent superinfection by the respective viruses. The endogenized genes cause a natural resistance to the pathogenic effect of the virus, whereby the expression of one gene is sufficient. Only under special conditions resistance can be overcome by high viral dose infection as shown for FV-1, an endogenous retroviral sequence in mice [21].

Other RNA virus Sequences Integrated as DNAs

There are other RNA viruses known such as a positive strand RNA virus, Dicistro virus, an Israeli acute paralysis virus of bees, for which viral sequences have been described integrated into the germline of bees from different hives [22]. Similar observations have been made in mice with endogenous retroviral sequences related to a capsid gene, the FV-1 locus, which confers resistance to infection by some retroviruses [23]. Other RNA virus related elements originate from Filoviruses other than Ebola and Borna viruses such as Nyamanini virus or Tamara bat virus. Endogenous filovirus elements were detected in several mammalian species.

In humans the hepatitis C virus, an RNA virus, has been shown to be integrated into the DNA of infected patients. The viral sequences were present in 4 out of 51 patients [24]. In this and other cases it needs to

be determined whether germline integrations may be simply accidents or provide the host with some important selectable advantage [18]. One answer comes from some more studies and the fact that the endogenous sequences are highly conserved. A detailed study by Katzourakis and colleagues also focused on non-retroviral integrations but other endogenous viral elements in animal genomes [11]. The authors concluded that members of all known virus genome types and replication strategies were detected in animal genomic sequences, which shows the evolutionary history of viruses and even their ancient origin. Some genes remained functionally active for many millions of years, suggesting that the captured viral sequences contributed to host evolution and must have had a selective advantage. They identified seven families including double stranded RNA viruses (such as Reoviridae), positive sense RNA viruses (Flaviviridae), as well as both segmented (Orthomyxoviridae, Bunyaviridae) and non-segmented families of negative sense RNA viruses (Borna-, Filo and Rhabdoviridae (such as rabies)) [11].

Also here lethal virus infections can apparently be overcome by integration of viral genomes as protection against superinfection [11,18]. Viral products from integrated viral genes correlate with resistance to the disease and may thereby provide some biological advantage to the species. Not only the host may have an advantage and protection against superinfection but also the virus could benefit as a somehow more resistant species, which may provide natural reservoirs for their persistence and transmission.

The timeline for integration was determined and a phylogenetic tree for the integration events constructed. Borna-, Filo-, Circo (containing circular single-stranded DNA) and Parvo virus (containing single-stranded linear DNA) can integrate in the genomes of primates, rodents, carnivora, bats, and other species [11]. Dependo viruses, single stranded DNA viruses, are also found in several animal genomes. Dependo virus like elements depend on Adeno or herpes viruses for replication. Many of the virus families replicate in the host cell nucleus. All of them entered the germlines between 30 and 93 Mio years ago [11].

LCMV

Another integration event of an RNA virus has been described in mice for the Lymphocytic Choriomeningitis virus (LCMV), which is an Arenavirus, characterized by a segmented ambisense RNA. It contains an RNA which is in part of positive and in part of negative polarity, transcription and translation are coupled (Nguyen and Haenni, 2003). In order to integrate in mice genomes, LCMV gets support from murine intracisternal A-type particles (IAP). These are defective endogenous retroviruses in rodents, which are enveloped particles, contain LTRs, gag-pol genes expressing an RT, and are budding from the Endoplasmic Reticulum. One can even replace the IAP by HIV. This strengthens the notion on the role of an RT for complementary cDNA formation [25]. After acute infection of mice with LCMV, cDNA sequences were detected in the spleen of mice, which are a natural host and permissive for replication of this virus leading to a persistent viral infection. LCMV sequences were detected as cDNA due to interaction between the IAPs and the infecting exogenous RNA from LCMV in permissive animal species [25]. Integrations with LCMV were estimated as 1 in 10^3 to 10^5 cells [25]. Integrated sequences are small and cover subgenomic regions and are unable to produce infectious virus. Only a small fraction of the cells express viral antigens which stimulate immune reactions or autoimmunity. The LCMV effects were interpreted as a general mechanism of integrated DNAs, like a "naturally produced DNA vaccine" [25]. Thus, also in this study the positive protective effect of endogenized viral sequences is emphasized.

DNA vaccines

DNA vaccines have been shown previously in several studies; among them were analyses with injected plasmid DNAs. One of them was coding for HIV gene products as vaccine in HIV patients [26], another one coded for a cytokine, Interleukin IL-12, acting as immunotherapy against malignant melanoma in cancer patients [27], and tuberculosis in mice [28] with

some efficacy. In humans the treated patients showed no side effects or autoimmune reactions. Integrated DNA was not major concern, because nuclear DNA might induce DNA antibodies and cause Sjögren Syndrome or other diseases [27]. The technique of using plasmid DNA as vaccines is presently followed by about ten companies against Corona virus. These genetic vaccines are using plasmid DNA instead of RNA (WHO, link). Two of them are in Korea (Genexine, GeneOne Life Sciences), furthermore Rottapharm Biotech, a Canadian Company, Entos Pharmaceuticals which produces DNA to enter gut cells not for injection, the Australian based Technovalia which developed a gene gun instead of needle injections.

The vaccine DNA has to enter the nucleus of the cell to produce a mRNA. This is one step more than required for RNA vaccines. Till now plasmid DNA vaccines have only been approved in veterinary medicine [20]. Whether an endogenization event is involved, is unknown but a concern.

Integration is a Genotoxic Event

Retroviral integration into the human genome is a natural replication step. It normally happens in somatic not in germ cells and viral genes are therefore not inherited to further generations. However, analysis of the human genome has indicated that there are many retroviral sequences integrated and inherited, many of them mutated or deleted to short fragments but the number is high, almost 50% [3]. Even though such integrated genetic viral information can have a protective effect on the host, there is the concern that integration events into the genomic DNA could be a genotoxic event. It may lead to the interruption of functional genes or activate regulatory mechanisms which can alter gene expression. In particular, retroviruses harbor the promoter sequences LTRs at their termini, which are strong promoters normally activating viral genes but potentially also neighboring genes by downstream promotion with potential mutagenic effects. Retroviruses were developed for gene therapy initially as integrating retroviral vectors supplying defective or missing genes into a host. Meanwhile, non-integrating viral vectors are used and applied by ex vivo treatment, while in vivo gene therapy especially with retroviral vectors is avoided and forbidden, since germ cells could be affected. This will avoid infection of germ cells and the transmission to future generations [29,30].

The many retroviruses or retroviral-like elements present in our genome entered by an integration event, similarly active transposons or retrotransposons enter the genomes by the cut and paste or copy and paste mechanisms yet all of these are naturally occurring events and not man made. The consequences of mobile DNA are often unknown but are also of concern as origin of human diseases [7].

Furthermore, viral as well as non-retroviral RNA viruses can also act as mutagens as a consequence of cDNA integration. Integration of cDNA into the host chromosome can always be a genotoxic event and may cause harm to the host DNA by interrupting a gene. Even though the human genome consists to about only 2% of exons coding for human genes, the residual 98% non-coding sequences might still contribute to some unexpected effect [3]. It is important to note, that only retroviral integration events include the activity of retroviral promoters, LTRs. Without such a promoter activation of neighboring genes is much weaker. LTRs are retroviral specific and other RNA viruses will lack such activating effects.

Integration of viral sequences can be considered as an important contribution to immune protection against superinfection and prevention of disease, to gene modifications, genetic complexity, to development, to evolution and others. Viruses can be considered primarily as drivers of evolution especially by integration. Pathogenicity may be a side effect [4,31].

Corona virus integrated DNA

In Dec 2020 an Archive preprint from R Jaenisch and colleagues was published in which they describe the reverse transcription of Corona virus RNA which can lead to integration of Corona viral DNA into the

human genome [32,33]. In primary cells from patients Corona viral DNA sequences were described as fused to cellular DNA sequences. Such "chimeric" transcripts were interpreted as consistent with integration. Cells can be spiked with excess RT either by LINE elements or HIV infection. The resulting overexpression of the RT can lead to increased integration frequencies [33]. Without overexpression of the RT integration was much reduced. Chimeric DNA mapped mainly to the nucleocapsid gene of the Corona viral genomes, which leads to the most abundant RNA. Cytokine exposure as a potential mechanism in patients could activate an RT and increase the effect of DNA integration.

The authors discuss that the long term expression of Corona viral RNA in patients might be due to expression of integrated Corona viral sequences after infection [34]. These patients however, do not shed replication competent virus, but produce viral RNA sequences detectable by PCR. The expression is mainly restricted to sequences coding for the structural protein N, one of the major viral transcripts. Expression of complete viruses is excluded. Thus, there is no fear of production of infectious viruses.

Similar mechanisms may occur with Zika, Dengue or influenza viruses as stated by the authors. Expression of the N protein could enhance the adaptive immunity against the virus.

This is similar to the data mentioned above [11,18,19]. The authors speculate that detection of viral RNA by PCR may lead to false interpretation, inhibitors may be misinterpreted as useless and patients are assumed to produce virus instead of recovery. They discuss their result as explanation for understanding some puzzling results on long term expression of positive Corona viral RNA. The follow-up paper clarifies some of the mentioned results. The possibility of experimental artifacts during sample preparation is discussed [33].

Longterm expression of viral RNA sequences has recently been described in immune suppressed patients including HIV patients, who are unable to clear the virus. This is a risk for the development of Corona viral mutants within the host [32].

The result was apparently a surprise even though similar phenomena have been known [11,18,19]. What remains unclear is the quantification of the effect and thereby its biological relevance.

Among the many Corona viral vaccines under development several are also based on DNA. Nothing is known about integration events on DNA virus vector vaccines some of which are based on adeno viruses from Chimpanzees, Ad5 and Ad26, the modified DNA Vaccinia Ankara Virus (MVA), which was an early vaccine against pox viruses, previously used for gene therapy against HIV. All DNA viruses may be able to integrate at non-specific sites except for Adeno-Associated virus which enters at a specific location and was therefore developed for gene therapy in humans [1].

Genomic analyses of integrated complementary DNA from RNA virus sequences in the human genome allowed the determination of the timepoint when this happened and data allowed to construct a phylogenetic tree with the integration events dating back to round 90 Mio years [11]. There is no information mentioned on ancient Corona viral sequences in the human genome. If this is a relevant occurrence one would expect to find Corona sequences in the human genome. This is especially expected in light of the recent evidence on the so called "Russian Flue" which turned recently out to be a Corona pandemic from 1889 till 1900 [35].

Four seasonal corona viruses are known every winter in the Northern hemisphere. The viruses cause 15 to 30% of the flu symptoms in winter seasons. The names are: NL63 (Netherland) isolated from a child preferentially affecting children, HKU1 Hongkong virus isolated 2005 from mice, 229E a lab isolate numbered by lab series, isolated in 1965, and most importantly OC43, an organ culture isolate, from cattle, the only one among the four we know of as a pandemic virus. It started in 1889 in St Petersburg and was distributed worldwide to Moskow, Berlin, Paris, London, USA and to far East. The virus may have undergone some reassortment and became

unexpectedly complex and dangerous. It came in three waves or more between 1889 until 1900 it showed many small outbreaks. About 1.5 Mio people were estimated to have died of a world population then of about 1.5 billion.

The disease was presumably mis named Russian Flue since more recently it was attributed to a Corona pandemic due to sensitivity deficiencies in nose and throat which have so far been correlated with Corona viruses only, not with Influenza. Van Ranst in Belgium and Lone Simonson in Roskilde in Denmark identified Corona viruses as cause of the "Russian Flue" which became one of the present seasonal viruses, OC43 [35,36]. Sequence analysis has not indicated integrated Corona viral DNA sequences in known data bases and was also not mentioned in any of the sequence analyses quoted here.

Target-primed reverse transcription

The procedure for integration is assumed to be predominantly mediated by LINE elements with their RT. Borna viral mRNA is reverse transcribed and integrated into the host genome. This was attributed to non-LTR retrotransposons which encode and express functional RTs. The mechanisms of endogenization are supposed to involve the RT coded for by LINEs via target primed reverse transcription from viral RNA. For that the RT transfers the invading viral RNA to a breakpoint of cellular DNA which is cleaved by the endonuclease coded for by LINE. One strand of the DNA is extended at its 3'-end by primer extension by the RT, copying the viral RNA. This DNA is then the matrix for a second DNA strand, which is copied by displacing or degrading the initial RNA. Then repair occurs [37] (Figure 1). Alternatively, recombination events may occur between retrotransposons and exogenous RNA viruses [38].

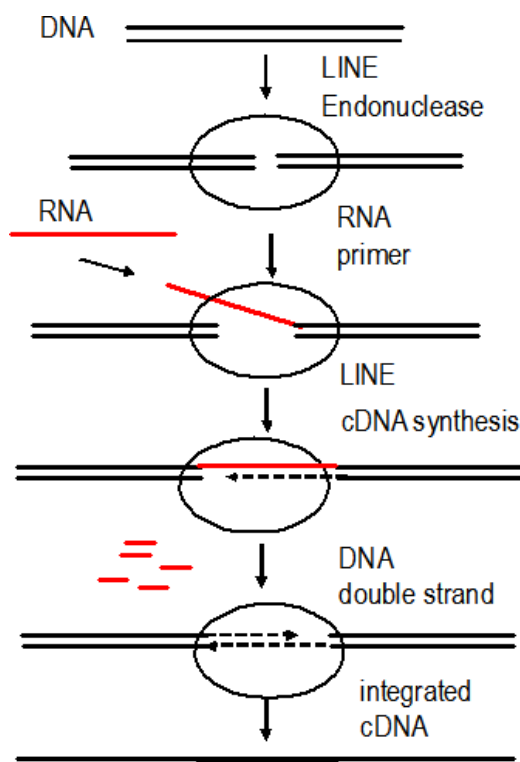


Figure 1. Model for target-primed reverse transcription of a viral RNA into a complementary cDNA and its integration into the host DNA.

Consequences-Danger or Risk?

The findings summarized here establish that not only retroviruses but genetic material derived from all known viral genome types and replication

strategies can enter the animal germ line. Thus there is an extensive gene flow from viruses to animal genomes. Animal genomes are therefore documents or archives of viral host interactions. This is greatly broadening the scope of paleovirological studies and indicating a more significant evolutionary role for gene flow from virus to animal genomes than has previously been recognized [11].

Endogenization of cDNA from RNA viruses is a natural event and has been known for decades. It can be dated back by molecular clocks to integration events many millions of years. Integration often includes only some viral genes, and often only a few copies. The genes maintained open reading frames throughout their history. This allows the conclusion, that they serve some function for the host and some evolutionary advantage. Also for the virus is integrated stable DNA and a safe harbor, even though most endogenous viruses are truncated and do not produce infectious viruses. In the cases known and described no virus production of infectious virus occurred.

The presence of Corona viral RNA sequences present in genomes as cDNA is not a surprising result. Also as in other cases, only some sequences are integrated and no infectious virus is produced. Surprisingly, no data indicate the presence of Corona viral cDNA in any of the comprehensive genomic sequencing studies. This would be important information to assess the significance of their genomic integration.

In all mentioned cases integrated viral cDNA sequences were of beneficial effects they protect against other viruses and even viral diseases.

Protection against superinfection by the endogenized viral sequences can be attributed to the protein expression of intact or faulty nucleocapsid proteins as endogenous gene products which could block viral replication or induce assembly of non-infectious viral particles. The defective viruses represent asymptomatic reservoirs. Such particles could become good immunogens providing immune protection in the host. Possibly protection could be partly due to the antiviral activity of interferon response to the RNA. It is interesting to note that these viral sequences have not developed stop codons over the past 40 million years suggesting that expression is essential for maintaining resistance of the host against these viruses. Even though these studies documented mainly ancient relatives of some RNA viruses having left their DNA copies in the germline of vertebrate hosts, this process is ongoing till today as documented with Koalas and by recent reports, indicating that such events might still occur today [39-45].

Conclusion

If one generalizes these observations to Corona virus infections or RNA vaccines, then endogenization of Corona virus sequences could also be beneficial. Their expression could even enhance vaccination effects or prevent disease! Whether this plays a role in resistance of some people, is unknown.

The RNA vaccine comprises the sequence for the spike protein, the RNA is injected into muscle cells, it is short lived, has no promoter sequence and is rather unstable. Even presumably low levels of Corona viral RNA present in the RNA vaccines has to be evaluated and compared with a viral infection. The virus infection will produce many orders of magnitude more RNA with the potential of integration as cDNA than an immunization with RNA.

To what extent an RNA vaccine against Corona viruses could be copied into a cDNA and integrated has not been discussed or described.

Integration can be a genotoxic event. This could be a concern. However also this is a natural consequence of a Corona viral infection. It should be much reduced in the vaccine.

Finally, integrated cDNA spike sequence of Corona viruses and the Corona viral RNA vaccine could both be of benefit and could represent in both cases free additional DNA vaccine! We do not need to worry!

Conflict of Interest

We have no conflict of interests to disclose and the manuscript has been read and approved by all named authors.

References

- Flint JS, Enquist LW, Racaniello VR and Skalka AM. *Principles of Virology*. DC: ASM Press, 2015.
- Wilson, DS and Szostak JW. "In vitro selection of functional nucleic acids." *Annu. Rev. Biochem* 68(1999):611–647.
- Lander, ES, Linton LM, Birren B, Nusbaum C and Zody MC, et al. "Initial sequencing and analysis of the human genome." *Nature* 409 (2001):860–921.
- Moelling, K. "Half a century of reverse transcriptase happy birthday!" *Genome Biology* 22(2021):31.
- Balestrieri, E, Matteucci C, Cipriani C and Ricceri L, et al. "Endogenous retroviruses activity as a molecular signature of neurodevelopmental disorders." *Int. J. Mol. Sci.*20(2019): 6050.
- Feschotte, C, and Pritham EJ. "DNA transposons and the evolution of eukaryotic genomes." *Annu. Rev. Genet* 41 (2007):331–368.
- Kazazian, HH and Moran JV. "Mobile DNA in Health and Disease." *N Engl J Med* 377(2017):361-370.
- Colson, P, Ravaux I, Tamalet C and Baptiste E, et al. "HIV infection en route to endogenization: Two cases." *Clin. Microbiol. Infect* 20 (2014):1280–1288.
- Denner, J, and Young PR. "Koala retroviruses: characterization and impact on the life of koalas." *Retrovirology* 10(2013):108.
- Blaise, S, de Parseval N, Bénit L and Heidmann T. "Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution." *Proc Natl Acad Sci USA* 100 (2003):13013-8.
- Katzourakis, A and Gifford RJ. "Endogenous viral elements in animal genomes." *PLOS genetics* (2010): 6, 11.
- Mi, S, Lee X, Veldman GM and Racie L, et al. "Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis." *Nature* 403(2000):785–9.
- Dewannieux, M, Harper F, Richaud A and Ribet D, et al. "Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements." *Genome Res* 16(2006), 1548–1556.
- Mack, M, Bender K and Schneider PM. "Detection of retroviral antisense transcripts and promoter activity of the HERV-K(C4) insertion in the MHC class III region." *Immunogenetics* 56(2004):321–32.
- Broecker, F, Horton R, Heinrich J and Schweiger M R, et al. "The intron-enriched HERV-K(HML-10) family suppresses apoptosis, an indicator of malignant transformation." *Mob. DNA* 7(2016):25.
- Broecker, F and Moelling K. "Evolution of immune systems from viruses and transposable elements." *Front Microbiol* (2019).
- Broecker, F and Moelling K. "What viruses tell us about evolution and immunology: Beyond Darwin?" *Ann. New York Acad Sci* (2019).
- Belyi, VA, Levine AJ and Skalka AM. "Unexpected inheritance: Multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes." *PLoS Pathog* 6(2010).
- Belyi, VA, Levine AJ and Skalka AM. "Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the parvoviridae and circoviridae are more than 40 to 50 million years old." *J. Virol* 84 (2010):12458–12462.
- Leung, DW, Ginder ND, Fulton DB and Basler CF, et al. "Structure of the Ebola VP35 interferon inhibitory domain." *Proc Natl Acad Sci USA* 106 (2009):411–416.
- Boone, LR, Innes CL, Glover PL and Linney E. "Development and characterization of an Fv-1-sensitive retrovirus-packaging system: Single-hit titration kinetics observed in restrictive cells." *J Virol* 63(1989):2592–2597.
- Maori, E, Lavi S, Mozes-Koch R and Peretz Y, et al. "Isolation and characterization of Israeli acute paralysis virus, a dicistrovirus affecting honeybees in Israel: evidence for diversity due to intra and inter species recombination." *J Gen Virol* 88(2007): 3428–3438.
- Bishop, KN, Bock M, Towers G and Stoye JP. "Identification of the regions of Fv1 necessary for murine leukemia virus restriction." *J Virol* 75 (2001):5182– 5188.
- Zemer R, Cohen YK, Naftaly T and Klein A. "Presence of hepatitis C virus DNA sequences in the DNA of infected patients." *Eur J Clin Invest* 38 (2008):845–848.
- Klenerman, P, Hengartner H and Zinkernagel RM. "A non-retroviral RNA virus persists in DNA form." *Nature* 390(1997):298-301.
- Weber, R, Bossart W, Cone R and Moelling K. "Phase I Clinical Trial with HIV-1 gp160 Plasmid Vaccine in HIV-1 Infected Asymptomatic Subjects." *Eur J Clin Microbiol Infect Dis* 20(2001):800–803.
- Heinzerling, L, Burg G, Dummer R and Oberholzer PA, et al. "Intratumoral injection of DNA encoding human interleukin 12 into patients with metastatic melanoma: Clinical efficiency." *Hum Gene Ther* 16(2005):35–48.
- Lowrie, DB, Tacson RE, Bonato VLD and Faccioli LH, et al. "Tuberculosis DNA vaccine Sequella." *Nature* 400(1999):269-271.
- Jadzary, SD and Poh CL. "Recent advances in delivery of veterinary DNA vaccines against avian pathogens." *Vet. Res* 50(2018):78.
- Moelling K. "Viruses more friends than foes". *World Scientific Press*, 2020.
- Moelling, K. "Within-host and between-host evolution in SARS-CoV-2 new variant's source." *Viruses* 13(2021):751.
- Zhang, L, Richards A, Khalil A and Haiting Ma, et al. "SARS-CoV-2RNA reverse transcribed and integrated into the human genome." *bioRxiv preprint* (2020).
- Zhang, L, Richards A, Barrasa I and Young R A, et al. "Reverse transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed in patient-derived tissues." *PNAS* 118 (2021).
- Yuan B. "Recurrence of positive SARS-CoV-2 viral RNA in recovered COVID-19 patients during medical isolation observation." *Sci Rep* 10 (2020):11887.
- Vijgen, L, Keyaerts E, Moës E and Wollants E, et al. "Complete genomic sequence of human coronavirus OC43: Molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event." *J Virol* 79 (2005):1595–1604.
- Surprising discovery: Coronavirus has previously brought down the world briefing on 15 December 2020.
- Eickbusch, TH. "Repair by retrotransposition." *Nature genetics* 31 (2002):1-3.
- Geuking, MB, Weber J, Dewannieux M and Heidmann T, et al. "Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration." *Science* 323 (2009):393–396.
- Horie, M, Honda T, Suzuki Y and Daito T, et al. "Endogenous non-retroviral RNA virus elements in mammalian genomes." *Nature* 463 (2010):84–87.
- Cohen, J. "Further evidence supports controversial claim that SARS-CoV-2 genes can integrate with human DNA." *Science* (2021).
- Moelling, K and Broecker F. "The reverse transcriptase-RNase H: From viruses to antiviral defense." *Ann. N. Y. Acad. Sci.* 1341 (2015):126–135.
- Nguyen, M and Haenni AL. "Expression strategies of ambisense viruses." *Virus Res* 93(2003):141-50.
- Palardini M, Cervasi B, Reyes-Aviles E and Ortiz A M, et al. "Low levels of SIV infection in sooty mangabey central memory CD4 T cells are associated with limited CCR5 expression." *Nature Medicine* (2011).
- Tarlinton R E, Meers J and Young P R. "Retroviral invasion of the koala genome." *Nature* 442(2006):79–81.
- Draft landscape and tracker of COVID-19 candidate vaccines briefing on 1 June 2021.