ISSN 0026-8933, Molecular Biology, 2017, Vol. 51, No. 2, pp. 205–215. © Pleiades Publishing, Inc., 2017.
Original Russian Text © V.M. Blinov, V.V. Zverev, G.S. Krasnov, F.P. Filatov, A.V. Shargunov, 2017, published in Molekulyarnaya Biologiya, 2017, Vol. 51, No. 2, pp. 240–250.

REVIEWS

UDC 578.264.9

Viral Component of the Human Genome

V. M. Blinov[a], V. V. Zverev[a], G. S. Krasnov[a, b, c], F. P. Filatov[a, d, *], and A. V. Shargunov[a]

[a] *Mechnikov Research Institute of Vaccines and Sera, Moscow, 105064 Russia*
[b] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 111911 Russia*
[c] *Orekhovich Research Institute of Biomedical Chemistry, Moscow, 119121 Russia*
[d] *Gamaleya Research Center of Epidemiology and Microbiology, Moscow, 123098 Russia*
**e-mail: ffelix001@gmail.com*

Received December 27, 2015; in final form, April 27, 2016

Abstract—Relationships between viruses and their human host are traditionally described from the point of view taking into consideration hosts as victims of viral aggression, which results in infectious diseases. However, these relations are in fact two-sided and involve modifications of both the virus and host genomes. Mutations that accumulate in the populations of viruses and hosts may provide them advantages such as the ability to overcome defense barriers of host cells or to create more efficient barriers to deal with the attack of the viral agent. One of the most common ways of reinforcing anti-viral barriers is the horizontal transfer of viral genes into the host genome. Within the host genome, these genes may be modified and extensively expressed to compete with viral copies and inhibit the synthesis of their products or modulate their functions in other ways. This review summarizes the available data on the horizontal gene transfer between viral and human genomes and discusses related problems.

Keywords: virus, horizontal gene transfer, cell receptor
DOI: 10.1134/S0026893317020066

INTRODUCTION

The relationships between viruses and their hosts are in fact more complex and diverse than is generally perceived. If their interactions are transient and limited to the single virus host paradigm, it largely determines the diagnosis of an infection, and the virus is considered an absolute parasite; this is the approach currently practiced by clinicians. However, on a larger time scale, transient interactions of this kind represent only a specific case of a more general role of viruses, since they are the basis of the evolutionary progression of the whole biological system where viruses and their hosts are constantly adapting to each other, either gaining certain advantages or suffering considerable losses. This is a continuous, ongoing process with a varying rate. Both sides act on the population level and exhibit different extents of aggression and plasticity. For this reason, it can be extremely difficult to determine the causes and the nature of a pathology (especially chronic) in a given individual because they are often modified by virological events that occurred many generations ago. Therefore, it seems fairly reasonable to consider a potential viral origin (in a broad sense) for almost any human disease with unclear etiology, even though it appears noninfectious, especially taking into account that, in particular, this origin implies the possibility of horizontal gene transfer, a phenomenon that is most efficiently mediated by viruses and largely determines the evolutionary progress as the consequence of the total sum of numerous elementary interspecies interactions [1].

The initiation of a viral infection depends on the presence of specific receptors on host cells, i.e., on the host sensitivity to a particular virus. The host's ability to resist infection or to develop a mutual relationship with the virus determines a favorable outcome, while weaker elements can be eliminated from the host population by the lethal course of infection. The distribution of genes that encode these receptors in the human genome is shown in Fig. 1 [2].

The parallel evolution of viruses and their host's sensitivity on the immune response level can lead to a decrease in the virus's pathogenicity [3]. For instance, this process underlaid the spreading of the low-virulence poxvirus variant, *alastrim*, in the years that preceded the eradication of smallpox [4–6].

Viruses recognize specific receptors on the surface of host cells; otherwise, if the expression of these receptors is inhibited, they become noninfectious. For instance, this happens if the human gene that encodes the CCR5 receptor to HIV is damaged by a 32-bp deletion [7]. In some cases, a virus receptor can be blocked by a protein, the gene of which was previously acquired by the host from the same virus via horizontal transfer. On the molecular level, a mutualistic solution has the form of a latent infection; a viral agent can per-

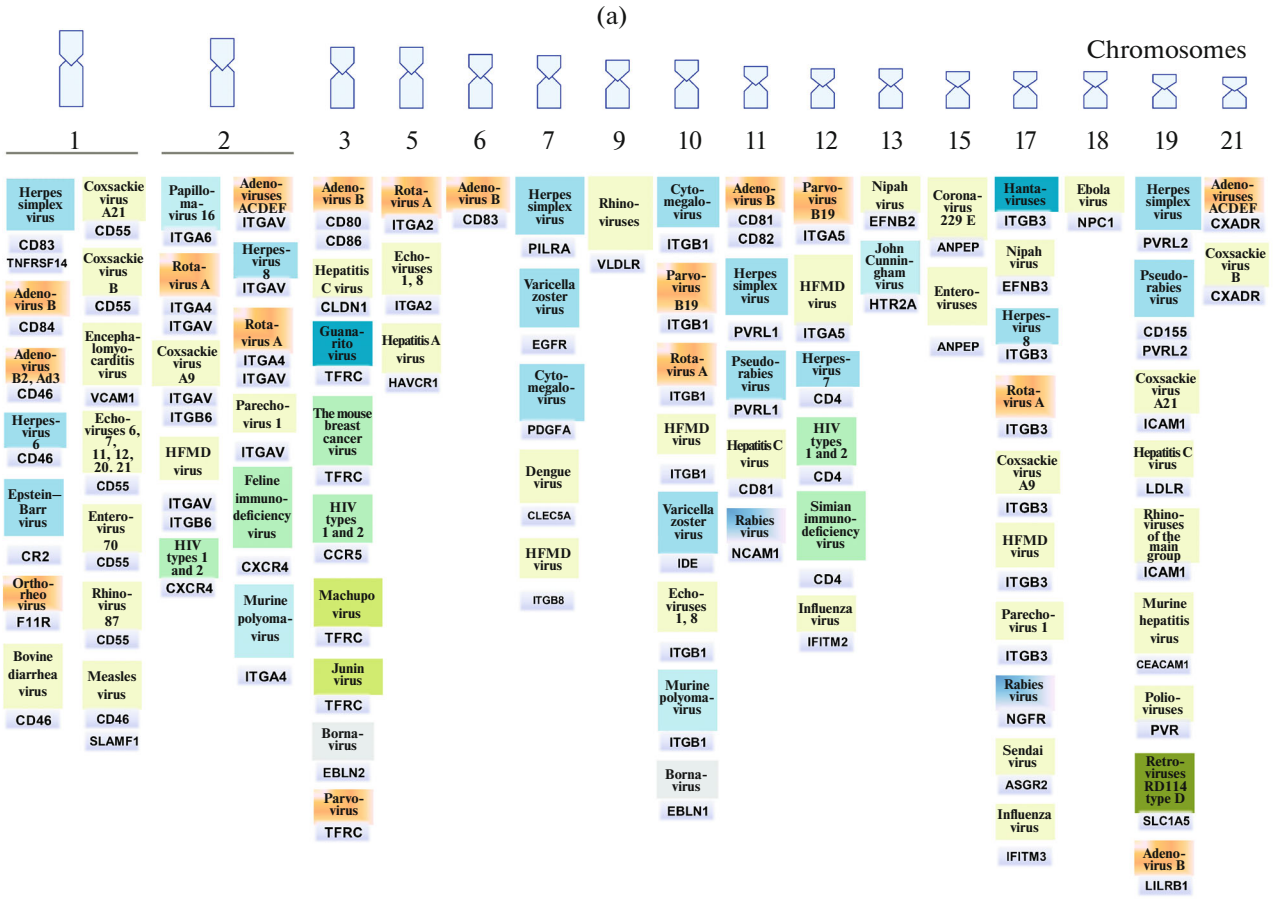

Fig. 1. (a) Distribution of viral receptor genes in human chromosomes; (b) viruses shown in the same colors as the corresponding receptors are ordered according to the Baltimore classification. In the second row, ss and ds stay for viruses with single- and double-stranded genomes, respectively; for single-stranded genomes, positive (+) or negative (–) sense is indicated.

sist for a long time in a host's cells, which protect it from external factors, while the host organism can make use of the reactivation of the virus, the expression of certain viral genes, or the production of latent RNAs for its own benefit; at the same time, sporadic reactivation and release from the host body enable the virus to maintain the level of genomic variation sufficient for its evolutionary promotion.

Horizontal transfer involves fragments of genetic information that vary strongly in size, in particular depending on the buffer genome capacity of each participant. In the human genome, this capacity is determined by the portion of chromosomal DNA, which does not contain species-specific protein-encoding sequences and, thus, can basically make a place for novel information that will be modified to reach a new balance. If we consider full-size genes, the essential sequences occupy ~10% of the human genome, while only 2.5% of the genome are gene exons.

The reverse process, i.e., the acquisition of host genes or shorter sequences by viruses, is also possible, although viral genomes obviously have a lower absolute capacity for storing the acquired material. However, there still are certain provisions; for instance, it was shown that up to 80% of genes in the herpesvirus [8–11] and adenovirus [12–15] genomes can be removed and substituted with foreign DNA without losing virus viability.

In some cases, it is difficult to identify the direction of the initial horizontal transfer (i.e., to determine whether a gene was transferred from the virus to the host or the other way around), because these genes start to perform important functions in both the virus and the host.

Considerable interest is drawn to viral tRNA-like (clover leaf, or L-form) structures present in the human genome and in some viral genomes, such as alphavirues and endogenous retroviruses. These structures can participate in the stabilization of viral RNA, as well as in viral replication and translation; all of these functions are determined by the folding of these structures [16, 17].

In human DNA, viral insertions can be present as full-size genome sequences, but also as smaller

(b)

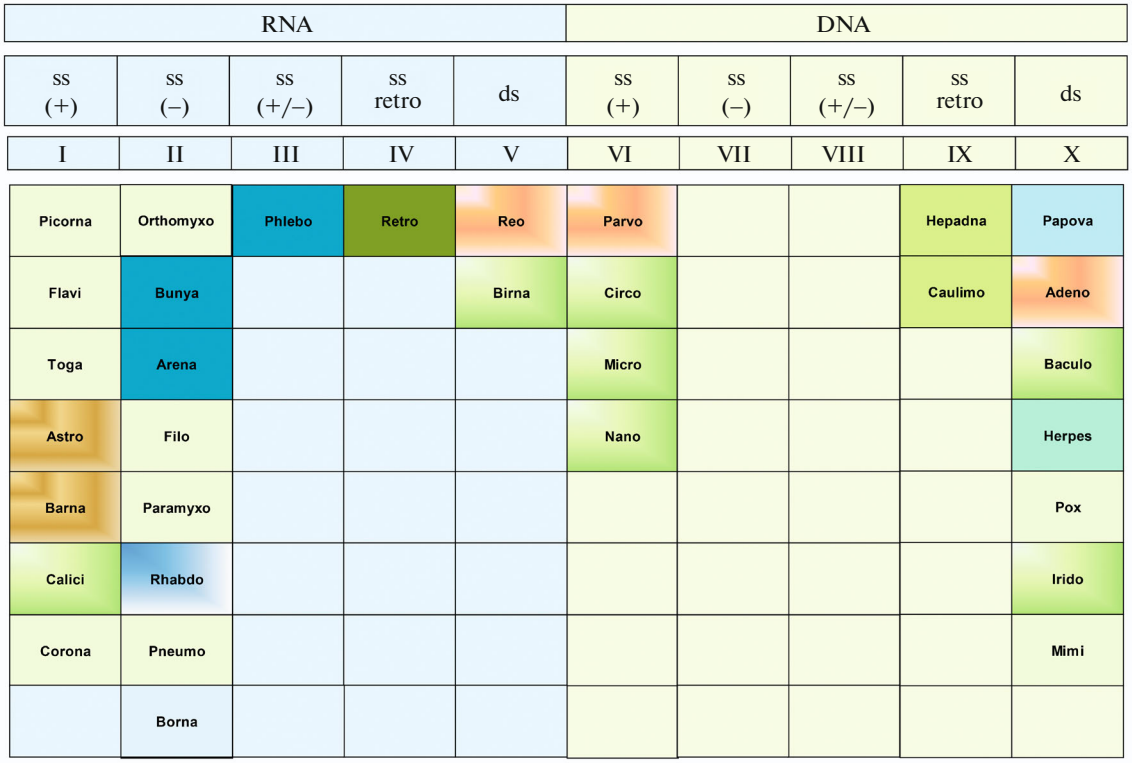

Fig. 1. (Contd.)

genome segments, individual viral genes or their clusters, and short sequence fragments. The genetic material of all known virus types using all possible replication strategies can reach animal germ cells and be transferred to subsequent generations, which determines the evolutionary role of the gene flow from viruses to animals [1].

INSERTION DONORS

RNA viruses

Retroviruses have certainly left the most extensive and frequent evolutionary ancient viral traces in host DNA: in the human genome, sequences of human endogenous retroviruses (HERV) amount to ~8% of its total size [18–20] and are derived from at least 504 phylogenetically different sources [20]. This fact is related to specific characteristics of the retrovirus replication machinery, which transcribes the genetic information carried by the viral DNA into DNA that is subsequently incorporated into the host genome. These incorporated viral sequences can be maintained in the host genome for a long time, either in the initial form, or with some modifications, and can be inherited. Modifications are determined by the activity of a number of different factors, including mobile elements of the host genome and transposons, which makes it usually very difficult to identify the source of a given insertion. Further on, viral genes integrated into host chromosomes can act as alleles that modify the host phenotype and sometimes provide a considerable selective advantage. For example, it is believed that this phenomenon contributed to the evolution of viviparous placental mammals; their genomes carry retroviral insertions that encode syncytins, proteins that serve to form the syncytiotrophoblast layer of the placenta and to ensure the immunological tolerance of the mother towards the embryo. Importantly, not all of the mammals (and not only mammals) possess a well-developed placenta, but those who do also have syncytins derived from surface glycoproteins of different retroviruses, the insertion of which occurred at different moments of mammalian evolution. In retroviruses, these surface glycoproteins contain immunosuppressing domains [2, 21–26], and it is these domains that are used by the novel hosts. Relic retroviral sequences (HERV-K) can be found in the genomes of human ancestors, Old World primates, nearly to the moment of their separation from the New World primates [27]. Human chromosomal DNA contains 30–50 HERV-K copies, some of which contain genes that exhibit low levels of expression in normal testicular and placental tissue [28]. At the same time, endogenous retroviruses and retrotransposons can induce carcinogenesis in somatic cells [29, 30].

Other RNA viruses. In fact, the scope of the described phenomena is not limited to retroviruses as such, since the ubiquity of retroviral elements in animal genomes, their activity in germline cells [31], along with the fact that viral replication depends significantly on RNA expression, allow retroviruses to contribute in different ways to the insertion of nonretroviral genes into animal germline cells.

The genomic integration of nonretroviral genes can be mediated by nonhomologous recombination with chromosomal DNA [32] or by interaction with retroelements of the host cells [33, 34]. It has been shown that retrotransposons can help the host borrow sequences from different RNA and DNA viruses. For example, recombination between the RNA of the lymphocytic choriomeningitis virus and a murine IAP retrotransposon results in reverse transcription of the RNA and in its integration into the host genome [35]. In some species, such insertions were shown to provide advantages; for example, bees that have acquired the gene that encode structural proteins of dicistrovirus become resistant to this agent, which causes acute paralysis in wild-type individuals [36]. This phenomenon is very common in the kingdoms of plants and fungi [37, 38].

In human, genomic incorporation of nonretroviral sequences has been described for such RNA viruses as Ebola [21, 22, 39, 40] and Marburg viruses (family *Filoviridae*), agents of the Born disease (family *Bornaviridae*) [23, 37, 38, 41], and polioviruses (order *Picornavirales*) [32]. The first two families belong to the order *Mononegavirales*, which also includes paramyxoviruses and rabdoviruses. These viruses have different virion structures, but share a common trait: their genome is composed of a single-stranded negative-sense RNA molecule up to 19 kb long. Filoviruses cause extremely severe acute infections in humans (with 60–90% lethality), and bornaviruses causes an equally severe disease in horses (with up to 90% lethality). Bornaviruses have also been detected in humans; in patients with severe mental disorders, such as schizophrenia [42], as well as in individuals without any pronounced clinical presentation in the cells of which these viruses persisted over long periods of time [43].

Fragments of filo- and bornavirus genomes are inserted into the host genome via interaction with LINEs [44], the most common mobile elements in higher eukaryotes, which bear a reverse transcriptase gene. The most frequent findings in the host genomes are inserted fragments containing bornavirus genes *N* and *L*, which encode nucleocapsid protein N and RNA-dependent RNA polymerase (p190), respectively. It is difficult to say when these fragments were inserted into the host genomes; presumably, this event occurred about 40 MA ago. Its initiation and rate varied among different hosts and for different bornaviruses species. At present, in many animal species, bornavirus-derived genes have evolved into homologous own genes, *EBLN1* and *EBLN2* [44]. These acquisitions provided an important selective advantage, enabling the host to resist devastating bornavius epizootics. It was shown that animals that possess *EBLN* genes are resistant to species-specific bornavirus infections, or the course of the disease is less severe in them [37]. The molecular basis of this resistance is the excessive synthesis of a protein N analog, which inhibits its p190 polymerase and, thus, decreases the virus yield. The initial functions of inserted genes can undergo gradual modification with time; this phenomenon is referred to as *exaptation*.

Filovirus genes appropriated by vertebrates are those that encode NP nucleoprotein and structural VP35 protein, which inhibits interferon production in the host. Cellular analogs of these genes have been found in bats, bandicoot, wallaby, and other animals of the area. The borrowed viral genes can be partially transcribed, and the resulting truncated N-terminal NP fragments (the full-size sequence of the host gene apparently is not expressed [44]) compete with the corresponding viral component, inhibiting the replication of the virus [45].

Retroviruses are not the only group for which nucleotide sequences can be fixed in eukaryotic genomes. In the 1970s, V.M. Zhdanov, an outstanding Soviet virologist, hypothesized that it should be possible for other RNA viruses [46]. In a later work, L.Yu. Frolova et al. showed that tRNA-encoding sequences in the human genome homologous to tRNA-like elements encoded by LTRs of endogenous retroviruses, can act as targets for alphaviruses [47].

Presumably, virus-specific revertase is not at all necessary for a viral nucleic acid to insert into a host genome [22]. Some features of this process are similar in very different viruses and are worth a detailed analysis. For example, the stable incorporation of viral genes into the host genome typical for retroviruses was also described in filoviruses (although it is less frequent), while in many other viruses, this feature is less prominent or unknown at all.

The Ebola virus genome has an interesting feature, which probably does not explain its ability to integrate into the host genome (most probably occurring by homologous recombination), but is nevertheless worth mentioning. It is a short (25–27 amino acids) immunosuppressing fragment p15E, which exhibits a high level of homology in Ebola and retroviruses [22]. Importantly, the function of this domain is activated as a result of the incorporation of an additional adenine, which results in a reading frame shift. This adenine insertion in p15E serves as a marker of pathogenicity of both retroviruses and the Ebola virus when they infect a new host.

The human genome contains a number of immunosuppressing fragments: they are expressed within syncytin genes in plancetal cells [23]. Insertions of nonretrovirus genes can basically have occurred as

result of interactions between the gene source and a retrovirus (most commonly, a LINE retrotransposon).

Among the proteins, the genes of which were borrowed from a virus, those that perform a primarily protective function are the most likely to be fixed in their new environment. For example, Fv1 is a protein similar to CA protein of the murine leukemia virus and competing with it; it binds the viral capsid and blocks reverse transcription, providing insensitivity to infection. In human and other animals, there are also protective proteins encoded by genetic elements of viral origin incorporated in the host genome. In particular, TRIM5 (and its analogs present in some primates) can inhibit the proliferation of some retroviruses in largely the same way as Fv1 and, at the same time, affects the proinflammatory transcription factors NF-κB and AP-1, which control the expression of genes involved in immune response, apoptosis, and cell cycle [48].

DNA Viruses

DNA retroviruses. Elements derived from genomes of DNA retroviruses (e.g., the duck hepatitis B virus) are found significantly less frequently in host DNA than those originating from RNA retroviruses [1, 44, 49]), even though they possess revertase, which could be expected to enable efficient incorporation of viral sequences into the host genome.

Other DNA viruses. Host genomes also bear traces of encounters with DNA viruses. Most commonly, these are members of the large parvovirus family (*Parvoviridae*): dependoparvoviruses (adenoassociated agents that can replicate only in the presence of a helper adenovirus or herpesvirus), nonpathogenic in humans. Dependoparvovirus genes have been found in the DNA of pigs, cattle, rats, mice, and other animals [1, 50]. Probably (although it has not been proven), these insertions serve to protect against parvovirus infections. Papillomaviruses were also shown to integrate into human genome [51].

Apart from the above agents, it was also found that the pig genome contains relic copies of the circovirus genome (noninfectious for humans). Genes of nano- and geminiviruses were found in plant genomes [1, 50]. Human DNA was also found to contain genes and larger genome elements of herpesviruses, including the Epstein–Barr virus (human herpesvirus 4, HHV4), human herpesvirus 6 (HHV6), and other members of this superfamily [52].

Finally, at least 34% of the human genome is composed of fairly large virus-like sequences: the so-called selfish DNA, the origin of which is unclear, while the only observed type of activity is autoreplication. The most active group are transposons of the LINE class (long interspersed elements); in human, they harbor approximately one in thousand genetic mutations [53–55]. On the whole, virus-like components of the human genome account for nearly a half of the chromosomal DNA, and some of them play an important role in the host organism, but hardly anything is known about the origin and functions of the others. Antiviral host response can transform an acute infection into the chronic or even the latent form (as it happens with herpesviruses), and a reservoir of viruses is thus conserved in their natural host, who will remain their target in case of reactivation.

It should be underlined that, for many important reasons, the body of relevant data available from the existing publications is far from being complete [1]. First of all, not all of the known viruses have been studied as potential sequence donors for the host genome, and not all the potential host species have their genomes sequenced. Secondly, an insertion of a viral sequence may represent a temporary outcome of a single infection event and will not be maintained in subsequent host generations. Moreover, the host species may also be eliminated from the evolution (as a result of extinction), which means that the acquired insertions will only be conserved if they are no longer limited to the extinct species and have become specific for a more general taxonomic branch, such as genus, order, or higher. Next, the relatively recent insertions may be insufficiently widespread to be identified and thus evade observation. Finally, the ability to incorporate parts of the viral genome into the chromosomal DNA of host germline cells can vary strongly among different taxonomic groups of viruses, i.e., orders, families, genera, and even species

If insertions of viral sequences remain functionally active in the host cell genome, they can give rise to either proteins that function in a new environment or untranslated RNAs of different sizes. If these insertions are inactive, they can merely witness a history of close and evolutionary long-term interactions between the virus and the host.

The characteristic trait of human herpesviruses is that, in their typical latent state, they can persist and replicate in the form of an episome in the direct vicinity of the host genomic DNA. These viruses have long coexisted with their hosts and the hosts' phylogenetic ancestors, and their genomes carry full-size genes that in turn were captured at some moment from the host and can often be expressed in their new environment [52, 56]. The nucleotide sequences of these genes, the encoded amino acids, and even the functions of the resulting proteins do not correspond strictly to their cell counterparts, but the range of their functions in the virus certainly suggests a relationship between the viral and the host genes. Some herpesviruses (such as HHV8 associated with Kaposi's sarcoma) have acquired genes of serpentines (G-protein-coupled receptors; GPCR), which they employ at the early stage of lytic infection of lymphoid cells and of sarcoma itself; moreover, the inhibition of viral DNA synthesis does not affect the functioning of these

genes. Their function is not quite clear, since they act mainly at the early stages of infection: *GPCR* transcripts, which have mainly bicistronic structure, protect the 5'-region of the coding sequence of another HHV8 gene, *K14*; while monocistronic transcripts analyzed in model experiments did not exhibit such properties. This probably indicates that the translation of *GPCR* transcripts may be reinitiated and suggests the need for further analysis of all the functions of GPCR itself in the pathogenesis of Kaposi's sarcoma.

The product of the GPCR-encoding gene *BILF1* captured by HHV4 acts as a specific inhibitor of class I major histocompatibility complex. The reverse process can also be imagined easily, as well as subsequent GPCR modifications as a result of the repeated gene capture with the acquisition of new functions. In fact, no such data are currently available, but the research has only begun very recently.

A recent study detected the first endogenous herpesvirus (genus *Roseolovirus*) in the genome of the *Philippine tarsier* (order Primates), while insertions of nearly full-size HHV8- and HHV4-like genomes were found in the DNA of other primates: aye-aye, lemur, and chimpanzee [57].

In this context, it is also worth mentioning that more data exist on the integration of herpesviruses into host genomes [58], as well as on the incorporation of herpesvirus 6 DNA (HHV6; also of the genus *Roseolovirus*) into the telomeric zones of human chromosomes. The significance of these insertions (which can be transmitted vertically if occur in germline cells) for human diseases or for the functioning of the immune system is currently absolutely unclear [56, 59, 60].

Similar data were obtained for viruses of other taxonomic groups. A systematic analysis showed that sequences derived from a wide range of animal viruses other than retroviruses are present as endogenous elements in mammalian, avian, and insect genomes. These elements of animal genomes represent the full spectrum of viral replication strategies [1]; moreover, the larger the sample of animal genomes, the wider the diversity of endogenous viral elements. Obviously, the more ancient these elements are, the smaller the number of host species needed to detect them. To identify more recent viral insertions, a much wider sample on the order and genus level is required. However, as we have pointed out above, the currently available data are far from complete.

The diversity of the known virus isolated as represented in virus gene/protein banks is but a small portion of the total virus diversity. In view of their likely ancient origin, members of the many virus families may be much more widely distributed among their mammalian hosts than we currently imagine, both as separate entities and as genome fragments in the host DNA. This is also reflected in the virus phylogeny, which was constructed using endogenous virus insertions along with exogenous viruses; close exogenous relations frequently are either not identified or were only described in the past decade [61–63]. The recently discovered relationship between filoviruses and marsupials has suddenly specified this infraclass of mammals as a potential filovirus reservoir. The presence of viral insertions may become an important factors for evaluation of findings obtained using metagenomic approaches [48, 64, 65].

All of these data have been discussed in several comprehensive reviews [1, 2, 66]. The capture of certain fragments of a viral genome by the host may be a random event; however, should the captured genes reach germline cells and prove useful to the host, they can become fixed in subsequent generations.

It may prove interesting to consider information that concerns shorter insertions of viral origin. They can be expected to exist in much higher numbers, since the probabilities of their incorporation, as well as of transposition or multiplication, seem higher than for larger insertions. Fragments of 20—30 nt long would have the same size as the biologically active RNAs, such as those involved in RNA interference.

Numerous studies, e.g., those reviewed in [67], indicate that such molecules can directly participate in the regulation of mammalian viruses. For instance, liver-specific miR-122 suppresses the replication of hepatitis C virus [68], while several human microRNAs, such as miR-199a-3p, miR-210, and miR-125a-5p, inhibit hepatitis B virus [69]. Other human microRNAs are targeted against the influenza virus (miR-323, miR-491, and miR-654), the vesicular stomatitis virus (miR-27 and miR-93), and against HIV (miR-28, miR29a, miR-125b, miR-150, miR-223, and miR-382) [70]. There are miRNAs against HHV4 and HHV8 herpesviruses (miR-17/92 and miR-106b/25) [71–73], Coxsackie virus (miR-342-5p) [74], and human papilloma virus [74]. It is also known that some viral miRNAs can circulate in human. For instance, miRNA-UL112-1 is encoded by the human cytomegalovirus (HHV5; genus *Cytomegalovirus*, family *Herpesviridae*), and its target is mRNA of the pre-early viral protein IE1 [75]; miRNA of HHV4 is targeted against BART LMP mRNA [76, 77], while miRNAs of herpes simplex viruses (HHV1 and HHV2), miR-H2, miR-H3, and miR-H4, are directed against mRNA of viral proteins ICP0 and ICP34.5 [78]. Similar miRNA–mRNA pairs were detected in experimental models of a retroviral infection [79]. It was shown that endogenous small interfering RNAs (siRNAs) can regulate the activity of some endogenous retroviruses [80].

However, it is currently hardly possible to identify the targets of microRNAs based on their sequences only, since they do not need to be strictly complementary. In addition to the CLIP-seq and PAR-CLIP techniques that employ immunoprecipitation of RNA–protein complexes with subsequent sequencing [81], it may prove helpful to analyze the data on the coexpression of miRNAs and their putative gene tar-

gets [82]. Two other groups of molecules participating in RNA interference are siRNAs and so-called piwiRNAs (or piRNAs). The latter are 24–29 nt long; their activity is mediated by a specific mechanism involving enzymes of the PIWI family, and their targets are usually transposons, retrotransposons, and endogenous retroviruses [83]. The 20–30-nt long fragments that form siRNAs and piRNAs must be almost fully complementary to their target sequences (with mismatches of no more than 2 nt). These RNAs are mainly targeted at exons, which facilitates the in silico search for putative small insertions in the recipient DNA. The functional activity of siRNAs and piRNAs should be determined experimentally. The detection of nucleotide fragments of the specific size in the *Escherichia coli* genome invited the hypothesis that bacteria possess an antiviral immune defense system [84], which later became the basis for the development of the revolutionary gene editing tools, CRISPR Cas9/Cpf1 [85, 86].

We attempted to identify 20- to 30-nt-long homology stretches (which we provisionally refer to as *hits*) in the human genome and in the genomes of human adenoviruses and herpesviruses and showed that their number was significantly higher than in extended samples that included either viruses that do not directly infect human, such as bacteriophages, or artificially generated sequences of the same as the herpesvirus genome [87].

Later, we showed that, even in a given group of viruses (HHVs), the portion of the viral exome that corresponds to hits, i.e., short 20–28-nt long sequences homologous between the virus and the host, is specific to each virus type and, as the first approximation, can be related to the destructive effect of the viral infection [88].

We have also proposed a hypothesis that reactivated HHV1 could exhibit oncolytic properties in vivo [88], because the body of *RL1* (which encodes ICP34.5 protein) contains a large number of virus/human homologous sequences (hits); in artificial HHV1-based oncolytic constructs, this gene is switched off first. Obviously, our hypothesis requires experimental verification.

Homo sapiens AS A RECIPIENT OF VIRAL INSERTIONS

The Fig. 2 schematically shows the load of human DNA with viral relics. We define the known viral insertions and their derivatives, including transposons, as relics, since we are currently able to identify only those that were incorporated into the host genome long time ago.

Human translated genes constitute approximately 2% of the total information volume of genomic DNA, whereas as virus relics and virus-like structures (with some reservations, in addition to endogenous retroviruses, these may be assumed to include retrotransposons and DNA transposons) amount to nearly half of the human genome.

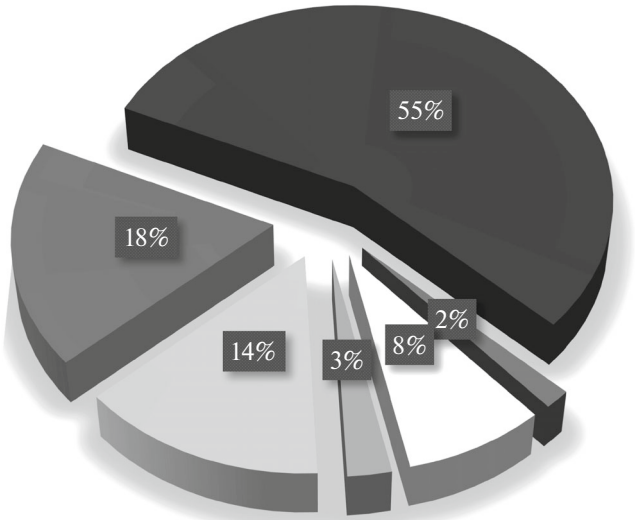

Fig. 2. Portions of the human genome occupied by protein-coding genes (2%) and relic viral sequences: endogenous retroviruses (8%), DNA transposons (3%), SINEs (14%), and LINEs (18%). Dark area of the human genome (55%) also contains other viral sequences (see text), genes whose activity is limited to transcription, and structures whose function is currently unknown.

This information suggests that there might be a need to reconsider the notion of a virus. From the traditional medical point of view, a virus is a parasite that infects a sensitive cell and ultimately causes its destruction. The virus (its genome) can also infiltrate the host DNA, occupying the free space or adding to its size. By replicating in the host genome, the virus ensures its genetic diversity and at the same time acquires a safe depot protecting it from external factors. This viral activity can be reasonably classified as parasitic.

However, from this viewpoint, we overlook the potential evolutionary advantages that the host population may gain as result of a viral infection: viruses are the most efficient vectors, which might transfer novel genetic information. For this reason, we propose to define a virus as an information carrier indistinguishable from a specific one and possessing an autonomous autoreplication program that employs the addressee's reading, synthesizing, and metabolic machinery for its own realization. This definition underlines the major evolutionary function of viruses as carriers of genetic information and is free from moral judgment, which is foreign to nature. Moreover, this definition has a more general character than the one specifying viruses a parasites.

Certainly, these considerations do not imply that *Homo sapiens* does not need to struggle with destructive viral infections. Moreover, the fact that, in nature,

the horizontal transfer of viral genes can render the human host resistant to devastating viral infections suggests the possibility of an utterly new approach to this struggle based on the thorough understanding of the molecular biology of interactions between the virus and the cell. For instance, if it is known that overexpression of a certain viral protein inhibits the infection cycle, it would be reasonable to add to the host–virus system a gene construct that could ensure efficient expression of the suppressor protein. More attempts are currently undertaken to employ short DNA or RNA sequences for antiviral therapy using the RNA interference technique (more precisely, to silence the expression of viral genes) [89–91].

As for the short regions of homology to herpesvirus or adenovirus DNA (which can basically be viral insertions), their distribution in the human genome is specific for every virus species, and on the whole, as we showed previously [87], it can hardly be random.

CONCLUSIONS

It was not the main objective of our brief review to list all of the available data on the question. The body of these data is rapidly growing, which, however, does not seem to change the existing consensus view on the problem. Obviously, mutual insertions of genomic fragments in virus–host pairs occur regularly, if not very frequently. Their exact frequency remains to be determined both on the population level and on the level of single individuals. The objective is not as much to determine the number of already incorporated foreign genes but rather to find out how this frequency depends on the type of the virus and its host. It is also important to determine the consequences of each insertion, either transient or fixed, can have for the human host within the lifetime of the hosts or of their ancestors. It should be evaluated how often these insertions reach the host's germline cells and what key factors govern these events. There are currently no answers to these questions, but these issues are addressed in different areas of study, which sometimes produces unexpected results. For instance, the Haldane and Waddington problem about the number of generations were required to obtain recombinant inbred strains was initially only solved by the authors for the cases of two and three genes [92]. Recently, for the general case Samal and Martin proposed an approach based on a statistical formalism rarely used in areas other than physics [93]; surprisingly, it provided an exact solution to the problem involving any given number of genes. Although the above example lies in the area of population genetics, while the number of studies on our topic of interest is insufficient to allow population-level generalizations, it nevertheless confirms the need to continue the efforts and emphasizes the productivity of breaking traditional thinking patterns, even in such conservative fields as medical diagnostics and therapy. The exchange of genetic information between living organisms is a complex albeit an infrequent phenomenon and generates a considerable uncertainty when we attempt to provide a comprehensive explanation of the causes and nature of a given pathological condition. The dynamic equilibrium between the human genome sensu stricto and the integrated viral sequences that perform protective and regulatory functions in the host organism represents a much deeper relationship than the organism's interaction with the internal microbiota. Currently, these relationships constitute the subject of separate omics-type research. They are the consequence of a multisided (not just a two-sided) encounter, where one side is represented by a viral population, which is heterogeneous in each individual case and consists of different proportions that are both fully featured and defective, as well as mutant particles, while the other side is a multicellular organism that developed as a result of macro- and microevolution. Viruses have enriched the host genome in functional virus relics, the amounts and diversity of which greatly exceed the total of the host's own genes. On the other hand, the multicellular organism has developed a system of antiviral defense and continues to develop it while employing the full range of available molecular mechanisms. Predicting the consequences of this encounter and the results of their targeted modification for the benefit of the infected macroorganism is a problem that can only be solved after considerable advances in the techniques of analyzing and modeling the underlying mechanisms.

REFERENCES

1. Katzourakis A., Gifford R.J. 2010. Endogenous viral elements in animal genomes. *PLoS Genet*. **6**, e1001191.
2. Blinov V.M., Gaisler V., Krasnov G.S., Shargunov A.V., Shurdov M.A., Zverev V.V. 2014. Cellular analogs of viral proteins. *Zh. Mikrobiol. Epidemiol. Immunobiol*. **2**, 101–113.
3. Esteban D.J., Hutchinson A.P. 2011. Genes in the terminal regions of orthopoxvirus genomes experience adaptive molecular evolution. *BMC Genomics*. **12**, 261.
4. Blinov V.M., Shchelkunov S.N., Sandakhchiev L.S. 1993. A probable molecular factor responsible for generalization of variola virus infection. *Dokl. Akad. Nauk*. **328**, 108–111.
5. Shchelkunov S.N., Marennikova S.S., Blinov V.M., Rezenchuk S.M., Tetmenin A.V., Chzhikov V.E., Gutorov V.V., Safronov P.F., Kurmanov R.K., Sandakchiev L.S. 1993. Entire coding sequence of the variola virus. *Dokl. Akad. Nauk*. **328**, 629–632.
6. Shchelkunov S.N., Resenchuk S.M., Totmenin A.V., Blinov V.M., Marennikova S.S., Sandakhchiev L.S. 1993. Comparison of the genetic maps of variola and vaccinia viruses. *FEBS Lett*. **327**, 321–324.
7. Samson M., Libert F., Doranz B.J., Rucker J., Liesnard C., Farber C.M., Saragosti S., Lapoumeroulie C., Cognaux J., Forceille C., Muyldermans G., Verhofstede C., Burtonboy G., Georges M., Imai T., et al. 1996. Resistance to HIV-1 infection in Caucasian indi-

viduals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature*. **382**, 722–725.

8. Knipe D.M., Ruyechan W.T., Roizman B., Halliburton I.W. 1978. Molecular genetics of herpes simplex virus: Demonstration of regions of obligatory and non-obligatory identity within diploid regions of the genome by sequence replacement and insertion. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 3896–900.
9. Longnecker R., Roizman B. 1987. Clustering of genes dispensable for growth in culture in the S component of the HSV-1 genome. *Science*. **236**, 573–576.
10. Filatov F. 1980. DNA of herpes group viruses. *Usp. Sovrem. Biol.* **89,** 90–103.
11. Baines J.D., Roizman B. 1991. The open reading frames UL3, UL4, UL10, and UL16 are dispensable for the replication of herpes simplex virus 1 in cell culture. *J. Virol.* **65**, 938–944.
12. Volpers C., Kochanek S. 2004. Adenoviral vectors for gene transfer and therapy. *J. Gene Med.* **6** (Suppl. 1), S164–S171.
13. Danthinne X., Imperiale M.J. 2000. Production of first generation adenovirus vectors: A review. *Gene Ther.* **7**, 1707–1714.
14. Goncalves M.A., de Vries A.A. 2006. Adenovirus: From foe to friend. *Rev. Med. Virol.* **16**, 167–186.
15. Crystal R.G. 2014. Adenovirus: The first effective in vivo gene delivery vector. *Hum. Gene Ther.* **25**, 3–11.
16. Hammond J.A., Rambo R.P., Filbin M.E., Kieft J.S. 2009. Comparison and functional implications of the 3D architectures of viral tRNA-like structures. *RNA.* **15**, 294–307.
17. Ganser L.R., Mustoe A.M., Al-Hashimi H.M. 2014. An RNA tertiary switch by modifying how helices are tethered. *Genome Biol.* **15**, 425.
18. Belshaw R., Pereira V., Katzourakis A., Talbot G., Paces J., Burt A., Tristem M. 2004. Long-term reinfection of the human genome by endogenous retroviruses. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4894–4899.
19. Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., Funke R., Gage D., Harris K., Heaford A., Howland J., et al. 2001. Initial sequencing and analysis of the human genome. *Nature*. **409**, 860–921.
20. Garazha A., Ivanova A., Suntsova M., Malakhova G., Roumiantsev S., Zhavoronkov A., Buzdin A. 2015. New bioinformatic tool for quick identification of functionally relevant endogenous retroviral inserts in human genome. *Cell Cycle.* **14**, 1476–1484.
21. Volchkov V.E., Blinov V.M., Netesov S.V. 1992. The envelope glycoprotein of Ebola virus contains an immunosuppressive-like domain similar to oncogenic retroviruses. *FEBS Lett.* **305**, 181–184.
22. Blinov V.M. 1999. "Dormant" immunosuppressive domain in filoviruses. *Proceedings of XI International Congress of Virology, Sydney, August 9–13.* No. 5147.
23. Bukreyev A., Volchkov V.E., Blinov V.M., Netesov S.V. 1993. The GP-protein of Marburg virus contains the region similar to the 'immunosuppressive domain' of oncogenic retrovirus P15E proteins. *FEBS Lett.* **323**, 183–187.
24. Bear G. 2004. When genes go walkabout. *Proc. Am. Philos. Soc.* **148**, 324–331.
25. Lynch V.J., Leclerc R.D., May G., Wagner G.P. 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat. Genet.* **43**, 1154–1159.
26. Villarreal L.P. 2004. Can viruses make us human? *Proc. Am. Philos. Soc.* **148**, 296–323.
27. Sverdlov E.D. 2000. Retroviruses and primate evolution. *Bioessays.* **22**, 161–171.
28. Tonjes R.R., Lower R., Boller K., Denner J., Hasenmaier B., Kirsch H., Konig H., Korbmacher C., Limbach C., Lugert R., Phelps R.C., Scherer J., Thelen K., Lower J., Kurth R. 1996. HERV-K: The biologically most active human endogenous retrovirus family. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **13** (Suppl. 1), S261–S267.
29. Rowe W.P., Pincus T. 1972. Quantitative studies of naturally occurring murine leukemia virus infection of AKR mice. *J. Exp. Med.* **135**, 429–436.
30. Suntsova M., Garazha A., Ivanova A., Kaminsky D., Zhavoronkov A., Buzdin A. 2015. Molecular functions of human endogenous retroviruses in health and disease. *Cell. Mol. Life Sci.* **72**, 3653–3675.
31. Giordano R., Magnano A.R., Zaccagnini G., Pittoggi C., Moscufo N., Lorenzini R., Spadafora C. 2000. Reverse transcriptase activity in mature spermatozoa of mouse. *J. Cell Biol.* **148**, 1107–1113.
32. Gmyl A., Belousov E., Maslova S., Khitrina E., Chetverin A., Agol V. 1999. Nonreplicative RNA recombination in poliovirus. *J. Virology*. **73**, 8958–8965.
33. Jern P., Coffin J.M. 2008. Effects of retroviruses on host genome function. *Annu. Rev. Genet.* **42,** 709–732.
34. Geuking M.B., Weber J., Dewannieux M., Gorelik E., Heidmann T., Hengartner H., Zinkernagel R.M., Hangartner L. 2009. Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. *Science.* **323**, 393–396.
35. Barbot W., Dupressoir A., Lazar V., Heidmann T. 2002. Epigenetic regulation of an IAP retrotransposon in the aging mouse: Progressive demethylation and de-silencing of the element by its repetitive induction. *Nucleic Acids Res.* **30**, 2365–2373.
36. Maori E., Lavi S., Mozes-Koch R., Gantman Y., Peretz Y., Edelbaum O., Tanne E., Sela I. 2007. 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**, 3428–3438.
37. Belyi V.A., Levine A.J., Skalka A.M. 2010. Unexpected inheritance: Multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. *PLoS Pathog.* **6**, e1001030.
38. Horie M., Honda T., Suzuki Y., Kobayashi Y., Daito T., Oshida T., Ikuta K., Jern P., Gojobori T., Coffin J.M., Tomonaga K. 2010. Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature*. **463**, 84–87.
39. Narayanan R. 2014. Ebola-associated genes in the human genome: Implications for novel targets. *MOJ Proteomics Bioinform.* **1** (5), 00032. doi 10.15406/mojpb.2014.01.00032

40. Cardenas W.B., Loo Y.M., Gale M., Jr., Hartman A.L., Kimberlin C.R., Martinez-Sobrido L., Saphire E.O., Basler C.F. 2006. Ebola virus VP35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling. *J. Virol.* **80**, 5168–5178.
41. Feschotte C. 2010. Virology: Bornavirus enters the genome. *Nature*. **463**, 39–40.
42. Nunes S.O., Itano E.N., Amarante M.K., Reiche E.M., Miranda H.C., de Oliveira C.E., Matsuo T., Vargas H.O., Watanabe M.A. 2008. RNA from Borna disease virus in patients with schizophrenia, schizoaffective patients, and in their biological relatives. *J. Clin. Lab. Anal.* **22**, 314–320.
43. Bode L., Ludwig H. 2003. Borna disease virus infection, a human mental-health risk. *Clin. Microbiol. Rev.* **16**, 534–545.
44. Horie M., Tomonaga K. 2011. Non-retroviral fossils in vertebrate genomes. *Viruses*. **3**, 1836–1848.
45. Watanabe S., Noda T., Kawaoka Y. 2006. Functional mapping of the nucleoprotein of Ebola virus. *J. Virol.* **80**, 3743–3751.
46. Zhdanov V.M. 1975. Integration of viral genomes. *Nature*. **256**, 471–473.
47. Frolova L.Y., Tsivkovskii R.Y., Sivolobova G.F., Oparina N.Y., Serpinsky O.I., Blinov V.M., Tatkov S.I., Kisselev L.L. 1999. Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA*. **5**, 1014–1020.
48. Belshaw R., Dawson A.L., Woolven-Allen J., Redding J., Burt A., Tristem M. 2005. Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): Implications for present-day activity. *J. Virol.* **79**, 12507–12514.
49. Gilbert C., Feschotte C. 2010. Genomic fossils calibrate the long-term evolution of hepadnaviruses. *PLoS Biol.* **8**, e1000495.
50. Kapoor A., Simmonds P., Lipkin W.I. 2010. Discovery and characterization of mammalian endogenous parvoviruses. *J. Virol.* **84**, 12628–12635.
51. Jeon S., Allen-Hoffmann B.L., Lambert P.F. 1995. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *J. Virol.* **69**, 2989–2997.
52. Holzerlandt R., Orengo C., Kellam P., Alba M.M. 2002. Identification of new herpesvirus gene homologs in the human genome. *Genome Res.* **12**, 1739–1748.
53. Deininger P.L., Batzer M.A. 2002. Mammalian retroelements. *Genome Res.* **12**, 1455–1465.
54. Wessler S.R. 1998. Transposable elements and the evolution of gene expression. *Symp. Soc. Exp. Biol.* **51,** 115–122.
55. Kidwell M.G., Lisch D. 1997. Transposable elements as sources of variation in animals and plants. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 7704–7711.
56. Arbuckle J.H., Medveczky M.M., Luka J., Hadley S.H., Luegmayr A., Ablashi D., Lund T.C., Tolar J., De Meirleir K., Montoya J.G., Komaroff A.L., Ambros P.F., Medveczky P.G. 2010. The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes in vivo and in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5563–5568.
57. Aswad A., Katzourakis A. 2014. The first endogenous herpesvirus, identified in the tarsier genome, and novel sequences from primate rhadinoviruses and lymphocryptoviruses. *PLoS Genet.* **10**, e1004332.
58. Morissette G., Flamand L. 2010. Herpesviruses and chromosomal integration. *J. Virol.* **84**, 12100–12109.
59. Pellett P.E., Ablashi D.V., Ambros P.F., Agut H., Caserta M.T., Descamps V., Flamand L., Gautheret-Dejean A., Hall C.B., Kamble R.T., Kuehl U., Lassner D., Lautenschlager I., Loomis K.S., Luppi M., et al. 2012. Chromosomally integrated human herpesvirus 6: Questions and answers. *Rev. Med. Virol.* **22**, 144–155.
60. Arbuckle J.H., Pantry S.N., Medveczky M.M., Prichett J., Loomis K.S., Ablashi D., Medveczky P.G. 2013. Mapping the telomere integrated genome of human herpesvirus 6A and 6B. *Virology*. **442**, 3–11.
61. Presti R.M., Zhao G., Beatty W.L., Mihindukulasuriya K.A., da Rosa A.P., Popov V.L., Tesh R.B., Virgin H.W., Wang D. 2009. Quaranfil, Johnston Atoll, and Lake Chad viruses are novel members of the family *Orthomyxoviridae*. *J. Virol.* **83**, 11599–11606.
62. Attoui H., MohdJaafar F., Belhouchet M., Tao S., Chen B., Liang G., Tesh R.B., de Micco P., de Lamballerie X. 2006. Liao ning virus, a new Chinese seadornavirus that replicates in transformed and embryonic mammalian cells. *J. Gen. Virol.* **87**, 199–208.
63. Li L., Kapoor A., Slikas B., Bamidele O.S., Wang C., Shaukat S., Masroor M.A., Wilson M.L., Ndjango J.B., Peeters M., Gross-Camp N.D., Muller M.N., Hahn B.H., Wolfe N.D., Triki H., et al. 2010. Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. *J. Virol.* **84**, 1674–1682.
64. Victoria J.G., Kapoor A., Dupuis K., Schnurr D.P., Delwart E.L. 2008. Rapid identification of known and new RNA viruses from animal tissues. *PLoS Pathog.* **4**, e1000163.
65. Li L., Victoria J.G., Wang C., Jones M., Fellers G.M., Kunz T.H., Delwart E. 2010. Bat guano virome: Predominance of dietary viruses from insects and plants plus novel mammalian viruses. *J. Virol.* **84**, 6955–6965.
66. Buzdin A., Khodosevich K., Mamedov I., Vinogradova T., Lebedev Y., Hunsmann G., Sverdlov E. 2002. A technique for genome-wide identification of differences in the interspersed repeats integrations between closely related genomes and its application to detection of human-specific integrations of HERV-K LTRs. *Genomics*. **79** (3), 413–422.
67. Russo A., Potenza N. 2011. Antiviral effects of human microRNAs and conservation of their target sites. *FEBS Lett.* **585**, 2551–2555.
68. Jopling C.L., Yi M., Lancaster A.M., Lemon S.M., Sarnow P. 2005. Modulation of hepatitis C virus RNA abundance by a liver-specific microRNA. *Science*. **309**, 1577–1581.
69. Liu W.H., Yeh S.H., Chen P.J. 2011. Role of microRNAs in hepatitis B virus replication and pathogenesis. *Biochim. Biophys Acta*. **1809**, 678–685.
70. Houzet L., Jeang K.T. 2011. MicroRNAs and human retroviruses. *Biochim. Biophys Acta*. **1809**, 686–693.

71. Skalsky R.L., Corcoran D.L., Gottwein E., Frank C.L., Kang D., Hafner M., Nusbaum J.D., Feederle R., Delecluse H.J., Luftig M.A., Tuschl T., Ohler U., Cullen B.R. 2012. The viral and cellular microRNA targetome in lymphoblastoid cell lines. *PLoS Pathog*. **8**, e1002484.

72. Riley K.J., Rabinowitz G.S., Yario T.A., Luna J.M., Darnell R.B., Steitz J.A. 2012. EBV and human microRNAs co-target oncogenic and apoptotic viral and human genes during latency. *EMBO J*. **31**, 2207–2221.

73. Kang J.G., Majerciak V., Uldrick T.S., Wang X., Kruhlak M., Yarchoan R., Zheng Z.M. 2011. Kaposi's sarcoma-associated herpesviral IL-6 and human IL-6 open reading frames contain miRNA binding sites and are subject to cellular miRNA regulation. *J. Pathol*. **225**, 378–389.

74. Zheng Z.M., Wang X. 2011. Regulation of cellular miRNA expression by human papillomaviruses. *Biochim. Biophys. Acta*. **1809**, 668–677.

75. Murphy E., Vanicek J., Robins H., Shenk T., Levine A.J. 2008. Suppression of immediate-early viral gene expression by herpesvirus-coded microRNAs: Implications for latency. *Proc. Natl. Acad. Sci. U. S. A*. **105**, 5453–5458.

76. Lo A.K., To K.F., Lo K.W., Lung R.W., Hui J.W., Liao G., Hayward S.D. 2007. Modulation of LMP1 protein expression by EBV-encoded microRNAs. *Proc. Natl. Acad. Sci. U. S. A*. **104**, 16164–16169.

77. Lung R.W., Tong J.H., Sung Y.M., Leung P.S., Ng D.C., Chau S.L., Chan A.W., Ng E.K., Lo K.W., To K.F. 2009. Modulation of LMP2A expression by a newly identified Epstein–Barr virus-encoded microRNA miR-BART22. *Neoplasia*. **11**, 1174–1184.

78. Jurak I., Griffiths A., Coen D.M. 2011. Mammalian alphaherpesvirus miRNAs. *Biochim. Biophys. Acta*. **1809**, 641–653.

79. Klase Z., Houzet L., Jeang K.T. 2011. Replication competent HIV-1 viruses that express intragenomic microRNA reveal discrete RNA-interference mechanisms that affect viral replication. *Cell Biosci*. **1**, 38.

80. Golden D.E., Gerbasi V.R., Sontheimer E.J. 2008. An inside job for siRNAs. *Mol. Cell*. **31**, 309–312.

81. Hafner M., Lianoglou S., Tuschl T., Betel D. 2012. Genome-wide identification of miRNA targets by PAR-CLIP. *Methods*. **58**, 94–105.

82. Krasnov G.S., Dmitriev A.A., Melnikova N.V., Zaretsky A.R., Nasedkina T.V., Zasedatelev A.S., Senchenko V.N., Kudryavtseva A.V. 2016. CrossHub: A tool for multi-way analysis of The Cancer Genome Atlas (TCGA) in the context of gene expression regulation mechanisms. *Nucleic Acids Res*. **44**, e62.

83. Haasnoot J., Westerhout E.M., Berkhout B. 2007. RNA interference against viruses: Strike and counterstrike. *Nat. Biotechnol*. **25**, 1435–14543.

84. Nakata A., Amemura M., Makino K. 1989. Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *J. Bacteriol*. **171**, 3553–3556.

85. Cong L., Ran F.A., Cox D., Lin S., Barretto R., Habib N., Hsu P.D., Wu X., Jiang W., Marraffini L.A., Zhang F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*. **339**, 819–823.

86. Zetsche B., Gootenberg J.S., Abudayyeh O.O., Slaymaker I.M., Makarova K.S., Essletzbichler P., Volz S.E., Joung J., van der Oost J., Regev A., Koonin E.V., Zhang F. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*. **163**, 759–771.

87. Zabolotneva A., Tkachev V., Filatov F., Buzdin A. 2010. How many antiviral small interfering RNAs may be encoded by the mammalian genomes? *Biol. Direct*. **5**, 62.

88. Filatov F., Shargunov A. 2015. Short nucleotide sequences in herpesviral genomes identical to the human DNA. *J. Theor. Biol*. **372**, 12–21.

89. Wasi S. 2003. RNA interference: The next genetics revolution? In: *Horizon Symposia; Understanding the RNAissance.* Nature Pub. Group, pp. 1–4.

90. Mansoori B., Sandoghchian Shotorbani S., Baradaran B. 2014. RNA interference and its role in cancer therapy. *Adv. Pharm. Bull*. **4**, 313–321.

91. Bender E. 2014. The Second Coming of RNAi. *Scientist*. **28**, 52-57.

92. Haldane J.B., Waddington C.H. 1931. Inbreeding and linkage. *Genetics*. **16**, 357–374.

93. Samal A., Martin O.C. 2015. Statistical physics methods provide the exact solution to a long-standing problem of genetics. *Phys. Rev. Lett*. **114**, 238101.

Translated by D. Timchenko