

ERVs and LINEs—along novel lines of thinking

Peer Terborg

A major part of the genomes of organisms is made up of what scientists now call transposable and transposed elements (TEs). The most complex TEs are endogenous retroviruses (ERVs) and long interspersed nuclear elements (LINEs). Approximately 8% of the human genome is made of ERVs and 17% of LINEs. A growing number of investigations define these elements as important structural and regulatory elements of the genome and they are increasingly appreciated as a major driving force of evolution.¹ The mainstream opinion still interprets these genetic elements as the remnants of ancient invasions of RNA viruses, although, like protein coding genes, more and more functions are attributed to them. Previously, I referred to these elements as variation-inducing genetic elements (VIGEs),^{2,3,4} since they appear to be particularly

good at generating novel genetic contexts and regulatory environments. In this short perspective, some unexpected novel functions of ERVs and LINEs will be highlighted.

LINEs

The first class of TEs (or VIGEs), which has recently gained a lot of attention, is LINEs. Although current philosophers of nature believe that LINEs—like ERVs—have their origin in RNA viruses, which invaded the genomes in ancient times, this view is untenable, knowing that presently there are no RNA viruses resembling LINEs. LINEs have a unique genetic make-up, and the only reason to perceive them as RNA virus remnants is that they have a reverse transcriptase enzyme resembling that of ERVs. Still, the actual origin of LINEs is completely unknown. LINE1, the only transposable element active in the human genome, is a complex genetic element with two open reading frames: ORF1 and ORF2. The protein coded by ORF2 provides essential enzymatic activities for the reverse transcription, as well as for integration of a newly transposed copy of LINE1. LINE1 propagates through a copy-paste

mechanism, thereby leaving identical copies on different positions in the genome. The exact role of ORF1 is unclear. It specifies a protein with protein-binding properties, but it can also function as a nucleic acid chaperone.⁵ Why do organisms contain such extremely elaborate mechanisms to induce variation in their genomes?

From immunology, we know that T and B cells also have mechanisms to produce variation in their DNA sequences to rapidly increase the specificity of their intruder-recognition systems (T-cell receptor and immunoglobulin rearrangements). The VIGE hypothesis holds that LINEs are a tool to induce or deliver variation. But where? Could they be involved in learning processes in the brain? Here, billions of differentiated neurons require continuous plasticity to operate in neuronal networks.⁶

One of the most unexpected novel functions of LINE1 is constructing a layer of fine-tuning in the neural networks in the brain. The mammalian brain is an extremely complex organ made up of a thousand different types of neurons that perform a variety of functions. In 2015, an Australian team of researchers revealed that the DNA of hippocampal and cortical neurons is

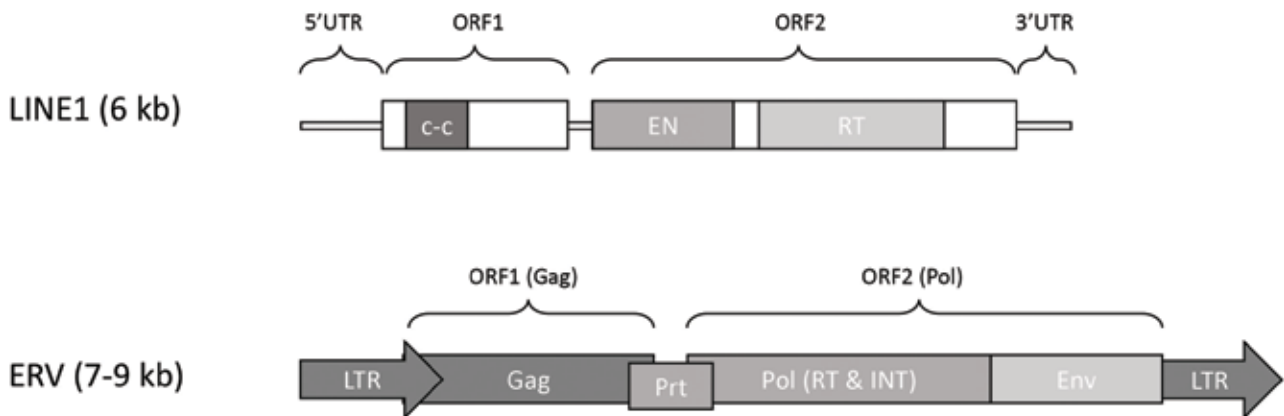


Figure 1. Genetic organization of the two 'evolutionarily' unrelated transposable elements, LINE1 and ERV. LINE1 is about 6 kilobases (kb) long, while ERVs vary between 7 and 9 kb. UTR = untranslated region; ORF = open reading frame; c-c = RNA & protein-binding protein; EN = endonuclease; RT = reverse transcriptase; LTR = long terminal repeat; Prt = protease; Pol = polymerase, which codes for a reverse transcriptase (RT); and an integrase (INT). Some ERVs also contain an Env gene, which codes for an envelope protein.

distinct, due to LINE1 mobilizations and retrotranspositions, which contribute to cell mosaicism. Most neurons in the brain have alterations to their DNA that make each neuron genetically unique. The researchers suggested that LINEs were potentially involved in building and fine-tuning neuronal networks.⁷ Similarly, a group of neuroscientists of the Salk Institute in California, US, showed that LINE1 in healthy neurons does not just insert DNA but also removes it.⁸ The researchers described how LINE1 deletes whole genes and consequently causes disparity between neurons, since such variations may affect the expression of genes critical to the developing brain.

The findings may explain what makes our thoughts and sensations so unique, and why identical twins can be so different. Because the brain is composed of billions of differentiated neurons—not one is identical—the genome requires a mechanism to induce such variation, just like the immune system requires variation-inducing mechanisms to generate millions of distinct antibodies and T-cell receptors. The enormous amount of variation required cannot be coded into the genome, as it would exceed the dimensions of the cell. If the information needed to produce the different neurons had to be recorded in the genome, it would be too large to function as a data processing system. Here, LINE1 functions to generate variation in the neurons and they clearly provide a genomic mechanism to increase processing power. That this variation-inducing mechanism does not act randomly is evident from the fact that no cancers of hippocampal cells are known to medical science.

Another recently identified function of LINEs is the formation of eukaryotic ‘operons’. In microbiology, an operon is understood as the functional unit of the DNA of prokaryotes (bacteria), which consists of several collinear

genes that are expressed together and code for proteins with related functions (such as an integrated metabolic pathway). In the genomes of higher organisms, the eukaryotes, collinear operons are uncommon. Nevertheless, many proteins must be expressed together in cooperating networks. A 2016 study suggested that interactions between distant DNA regions make it possible for different genes to be expressed together.⁹ Hence, LINEs may function to bring together co-expressed genes operating in functional biological networks, comparable with bacterial genes expressing together in operons. These higher, three-dimensional genomic structures, which regulate the accessibility of the genes through chromatin changes, may form through LINE-RNA interactions via Hoogsteen base pairing.¹⁰ We must come to consider the genome of eukaryotes as spatial networks of interactive elements to form regulatory platforms for clustered gene expression.

Although several studies had identified LINE1 as an essential factor for murine preimplantation development, the details of ‘how and what’ were unknown.¹¹ In 2017, a study published in *Nature Genetics* demonstrated that LINE1 activity regulates the chromatin dynamics and is essential for normal embryonic development in mice.¹² The report demonstrated that appropriate genome-wide LINE1 chromatin activation/silencing is required for early embryonic development. Embryos with activated LINE1 had greater chromatin accessibility and a larger nuclear volume, whereas embryos with repressed LINE1 had less chromatin accessibility. Here, the LINE1 system appears to function as a generic mechanism for gene regulation during immediate early embryogenesis. Thus, when normal epigenetic control over gene expression is not yet in place,

LINEs regulate the accessibility of the genes by modifying the chromatin.

ERVs

Earlier, in a series of papers, I argued that the origin of RNA viruses can be understood as genetically modified ERVs which acquired virulence genes and thus became disease-causing agents.²⁻⁴ The ‘VIGE-first hypothesis’ holds that RNA viruses have their origin in ERVs, and that ERVs were created for/with a purpose. ERVs are made of two genes, *gag* and *pol*, which are also found in all modern RNA viruses. This fact is also the most vital argument for why endogenous retroviruses are always interpreted as remnants of ancient genomic invasions of RNA viruses. The *pol* gene encodes a large protein with four distinct enzymatic activities: a protease, a reverse transcriptase, an RNase, and an integrase. To produce the individual proteins, the protease, which is synthesized first, proteolytically releases the other three enzymes from the precursor sequence. The transcribed, full-length ERV RNA then functions as a template for reverse transcriptase, the enzyme that catalyzes the synthesis of a double-stranded RNA-DNA hybrid.

Next, the RNase enzyme removes the RNA part, and the remaining single-stranded DNA forms a circular molecule. This circular single-stranded DNA serves as a template for the synthesis of a second DNA strand. The double-stranded DNA copy can now be put back in the genome with the help of the integrase enzyme. The position where this happens is determined by repetitive DNA sequences flanking the ERV element and/or by the sequence specificity of the endonuclease (integrase).^{13,14} Alternatively, the RNA molecule can be packed in a capsule consisting of three proteins, which are specified by the *gag* gene, and the whole thing looks very much like a virus. Why this packaging is necessary

is unclear, but it may prevent the RNA molecule from docking to the wrong places in the cell. On the other hand, the protein-coated viral-like particles may contain biologically active molecules which have to be protected and/or delivered to the right places. In other words, we are dealing with a subcellular transport system.

In 2018, two publications addressing this possibility appeared simultaneously in *Cell*.^{15,16} Neurons use a virus-like construct to pass on messenger RNAs that code for the building blocks of that virus-like construction. These building blocks are known as activity-regulated cytoskeleton-associated protein (ARC). Although the ARC protein was for a long time suspected to be involved in learning and memory processes, nobody knew how or why. ARC is homologous to the *gag* proteins, which are found in all RNA viruses and ERVs. Although ARC is required for synaptic plasticity and cognition, and mutations in this gene are linked to autism and schizophrenia, its biological function is largely undefined.

The publications in *Cell* now shed some light on this matter. Jason Shepherd and colleagues from the University of Utah, USA, transferred the ARC gene into bacteria, and observed that ARC proteins self-assemble into capsids which look very much like virus coats.¹⁵ The researchers concluded that that neuronal ARC gene encodes a repurposed retrotransposon *gag* protein that packages intercellular RNA to mediate intercellular communication in the nervous system. Purified ARC capsids are taken up and transfer ARC mRNA into the cytoplasm of neurons. Apparently, the neurons need ARC in such large amounts that they require a special delivery system. Furthermore, these results show that ARC exhibits molecular properties similar to those of retroviral *gag* proteins. Of course, the authors spun an evolutionary story

around their findings, claiming that ARC is derived from a vertebrate lineage of Ty3/gypsy retrotransposons. In a comment on a Dutch media site, Shepherd admitted: “Other neuroscientists would have laughed at me if I had claimed something like that before.” His response identifies the junk DNA hypothesis of the Darwinian paradigm as a science stopper, and shows that questioning junk DNA still induces scoff and laughter from a scientific community blinded by the erroneous idea that our genome is made of viruses.

In the same issue of *Cell*, a research group from the University of Massachusetts further disclosed another function of ARC proteins.¹⁶ They discovered that the motor neurons of fruit flies control muscles by releasing extracellular vesicles which are packed with ARC capsids. Here too, the ARC protein forms capsid-like structures. They bind *dAr1* mRNA in neurons and they are uploaded into extracellular vesicles that are transferred from motor neurons to muscles. The more active the neurons are the more capsids are delivered. These results point to a trans-synaptic mRNA transport mechanism involving retrovirus-like capsids and extracellular vesicles. The paper also reports how cultured genetically modified mouse neurons, which do not express the ARC gene, integrated ARC capsids and started to use the delivered ARC mRNAs. Again, we see a sophisticated delivery system at work, not viruses. The researchers asked whether this form of transport may also play a role in the delivery of additional mRNAs and proteins, and perhaps may promote the spread of Alzheimer’s and other neurological disorders.¹⁶

Considering these novel facts, we are compelled to also ask whether the ERV system itself is some sort of common delivery mechanism, since ERV-like vesicles readily

leave and enter cells of the placenta. Unfortunately, nobody is really interested in studying this fascinating possibility. Still, it has recently been reported that ERVs can act as DNA regulatory elements^{17,18} as long non-coding RNAs,^{19,20} and as triggers for the innate immune system.²⁰ ERVs in the human genome are able to bind ‘signal transducer and activator of transcription’ 1 (STAT1), an effector of the interferon (IFN) pathway involved in immune responses. The enrichment of ERVs in IFN-regulated genes suggest that they play an active role as regulators of essential immune system genes.²¹

Conclusions

In biology, everything is regulated and controlled. Although we have only recently started to study truly the functions of TEs we have already found that they accomplish many crucial functions in regulating gene expression, differentiation, and development. About 10 years ago, I started to name TEs after their functions in the genome: variation-inducing genetic elements (VIGEs).^{2,3} In the light of current knowledge, this term still seems to be appropriate, although their functions now go far beyond inducing variation.

New studies keep providing unexpected functions for TEs, indicating they are an integral part of originally designed genomes, which we should refer to as *baranomes*.²² This is clear from the DNA-nucleosome binding rules that their sequences tightly follow²³ and by increasing evidence that the activity of TEs is tightly controlled by (epi) genetic mechanisms and specific RNA molecules.²⁴ It is outlandish to claim this intricate genetic system came about by an ancient invasion of RNA viruses. In my opinion, the mainstream opinion still has the order of events upside down: the genomes of the eukaryotes are not built of

remnants of RNA viruses. Rather, RNA viruses have their origin in the genome, and to be precise in ERVs. Life was created good; RNA-viruses and the diseases they induce first appeared after the Fall.

References

1. Mobile DNA elements in primate and human evolution, Xing, J., Witherspoon, D.J., Ray, D.A., Batzer, M.A., and Jorde, L.B., *Am. J. Phys. Anthropol. Suppl* **45**:2–19, 2007.
2. Terborg, P., The design of life: part 3—an introduction to variation-inducing genetic elements, *J. Creation* **23**(1):99–106, 2009.
3. Terborg, P., The design of life: part 4—variation inducing genetic elements and their function, *J. Creation* **23**(1):107–114, 2009.
4. Terborg, P., The ‘VIGE-first hypothesis’—how easy it is to swap cause and effect, *J. Creation* **27**(3):105–112, 2013.
5. Martin, S.L., Li J., and Weisz, J.A., Deletion analysis defines distinct functional domains for protein-protein and nucleic acid interactions in the ORF1 protein of mouse LINE-1, *J. Mol. Biol.* **304**(1):11–20, 2000.
6. Erwin, J.A., Marchetto, M.C., and Gage, F.H., Mobile DNA elements in the generation of diversity and complexity in the brain, *Nat. Rev. Neurosci.* **15**:497–506, 2014.
7. Upton, K.R., Gerhardt, D.J., Jesuadian, J.S., Richardson, S.R., Sánchez-Luque, F.J., Bodea, G.O., Ewing, A.D., Salvador-Palomeque, C., van der Knaap, M.S., Brennan, P.M., Vanderver, A., and Faulkner, G.J., Ubiquitous L1 mosaicism in hippocampal neurons, *Cell* **161**(2):228–39, 2015.
8. Erwin, J.A., Paquola, A.C., Singer, T., Gallina, I., Novotny, M., Quayle, C., Bedrosian, T.A., Alves, F.I., Butcher, C.R., Herdy, J.R., Sarkar, A., Lasken, R.S., Muotri, A.R., and Gage, F.H., L1-associated genomic regions are deleted in somatic cells of the healthy human brain, *Nature Neurosci.* **19**:1583–1591, 2016.
9. Corradin, O., Cohen, A.J., Luppino, J.M., Bayles, I.M., Schumacher, F.R., and Scacheri, P.C., Modeling disease risk through analysis of physical interactions between genetic variants within chromatin regulatory circuitry, *Nat Genet.* **48**(11):1313–1320, 2016.
10. Hoogsteen base pairing occurs under physiological conditions to yield a wider DNA groove, which exactly fits an RNA molecule. My own unpublished research demonstrates that, due to their genetic make-up, LINE1 sequences more easily form Hoogsteen base pairs and may thus contribute to the 3D spatial distribution of the DNA in the nucleus. In addition, single stranded LINE sequences easily hybridize trans-chromosomally and thus may further substantiate a spatial DNA network.
11. Beraldi, R., Pittoggi, C., Sciamanna, I., Mattei, E., and Spadafora, C., Expression of LINE-1 retroposons is essential for murine preimplantation development, *Mol. Reprod. Dev.* **73**:279–872, 2006.
12. Jachowicz, J.W., Bing, X., Pontabry, J., Bošković, A., Rando, O.J., and Torres-Padilla, M.E., LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo, *Nature Genetics* **49**:1502–1510, 2017.
13. Barabaugh, P.J., Post-transcriptional regulation of transposition by Ty retrotransposons of *Saccharomyces cerevisiae*, *J. Biol. Chem.* **270**:10361–10264, 1995.
14. Wilke, C.M., Maimer, E., and Adams, J., The population biology and evolutionary significance of Ty elements in *Saccharomyces cerevisiae*, *J. Genetics* **86**:155–173, 1992.
15. Pastuzyn, E.D., Day, C.E., Kearns, R.B., Kyrke-Smith, M., Taibi, A.V., McCormick, J., Yoder, N., Belnap, D.M., Erlendsson, S., Morado, D.R., Briggs, J.A.G., Feschotte, C., and Shepherd, J.D., The neuronal gene ARC encodes a repurposed retrotransposon gag protein that mediates intercellular RNA transfer, *Cell* **172**(1–2):275–288, 2018.
16. Ashley, J., Cordy, B., Lucia, D., Fradkin, L.G., Budnik, V., and Thomson, T., Retrovirus-like gag protein ARC1 binds RNA and traffics across synaptic boutons, *Cell* **172**(1–2):262–274, 2018.
17. Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L., Embryonic stem cell potency fluctuates with endogenous retrovirus activity, *Nature* **487**(7405):57–63, 2012.
18. Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., Nix, D.A., Peterson, C.M., Tapscott, S.J., Carrell, D.T., and Cairns, B.R., Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons, *Nat. Genet.* **49**(6):925–934, 2017.
19. Wang, J., Xie, G., Singh, M., Ghanbarian, A.T., Raskó, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N.V., Schumann, G.G., Chen, W., Lorincz, M.C., Ivics, Z., Hurst, L.D., and Izsvák, Z., Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells, *Nature* **516**(7531):405–409, 2014.
20. Durruthy-Durruthy, J., Sebastiano, V., Wossidlo, M., Cepeda, D., Cui, J., Grow, E.J., Davila, J., Mall, M., Wong, W.H., Wysocka, J., Au, K.F., and Reijo Pera, R.A., The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming, *Nat. Genet.* **48**(1):44–52, 2016.
21. Grow, E.J., Flynn, R.A., Chavez, S.L., Bayless, N.L., Wossidlo, M., Wesche, D.J., Martin, L., Ware, C.B., Blish, C.A., Chang, H.Y., Pera, R.A., and Wysocka, J., Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells, *Nature* **522**(7555):221–225, 2015.
22. Terborg, P., The design of life: part 2—Baranomes, *J. Creation* **22**(3):68–76, 2008.
23. Huda, A., Mariño-Ramírez, L., Landsman, D., and Jordan, I.K., Repetitive DNA elements, nucleosome binding and human gene expression, *Gene* **436**(1–2):12–22, 2009.
24. Di Giacomo, M., Comazzetto, S., Saini, H., De Fazio, S., Carrieri, C., Morgan, M., Vasiliauskaitė, L., Benes, V., Enright, A.J., and O’Carroll, D., Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis, *Mol. Cell* **50**(4):601–608, 2013. Ti tem qui alia dollora ipiti volere volut