nature reviews molecular cell biology

https://doi.org/10.1038/s41580-023-00674-z

Review article

Check for updates

Activation of human endogenous retroviruses and its physiological consequences

Nicholas Dopkins & Douglas F. Nixon

Abstract

Human endogenous retroviruses (HERVs) are abundant sequences that persist within the human genome as remnants of ancient retroviral infections. These sequences became fixed and accumulate mutations or deletions over time. HERVs have affected human evolution and physiology by providing a unique repertoire of coding and non-coding sequences to the genome. In healthy individuals, HERVs participate in immune responses, formation of syncytiotrophoblasts and cell-fate specification. In this Review, we discuss how endogenized retroviral motifs and regulatory sequences have been co-opted into human physiology and how they are tightly regulated. Infections and mutations can derail this regulation, leading to differential HERV expression, which may contribute to pathologies including neurodegeneration, pathological inflammation and oncogenesis. Emerging evidence demonstrates that HERVs are crucial to human health and represent an understudied facet of many diseases, and we therefore argue that investigating their fundamental properties could improve existing therapies and help develop novel therapeutic strategies.

Sections

Division of Infectious Diseases, Department of Medicine, Weill Cornell Medicine, New York, NY, USA. e-mail: nid4009@med.cornell.edu; dnixon@med.cornell.edu

Introduction

Retroviruses infect a host by fusing with the cell membrane and releasing into the cytosol a capsid complex packaged with the necessary cofactors required for integration of their RNA genome into the host genome[1]. In the cytosol, viral reverse transcriptases (RTases) produce double-stranded DNA (dsDNA) copies of the viral genome, which travel to the nucleus and integrate in the host genome[1] (Fig. 1a). Once integrated, the locus of retroviral genes is termed a 'provirus'. Full-length proviruses contain coding sequences for retroviral proteins, a primer binding site and flanking long terminal repeats (LTRs) that possess transcription factor binding sites (TFBs), which aid in the hijacking of host RNA polymerase machinery[2,3]. Retroviruses that integrate into a germ cell may also be transmitted to host progeny[2,3]. Over generations of humans and ancestors, waves of retroviral infections have introduced proviruses into our germline cells that accumulate point mutations and deletions and undergo recombination, which fixate them to a given locus by restricting their capacity to form infectious particles and prohibit their retrotransposition[2,3] (Fig. 1b). These mutated proviruses are termed endogenous retroviruses (ERVs) and encompass large fractions of mammalian genomes, where they provide a diverse array of regulatory and coding sequences that possess physiological roles[2,3]. Human ERVs (HERVs) constitute an estimated 8% of the human genome[4,5] and remain understudied in comparison to the 1–2% of the genome annotated as canonical coding genes. In this Review, we use 'ERV' when discussing human and non-human studies and 'HERV' when discussing studies conducted on human-specific ERVs.

The phylogenetic classification of HERVs is based primarily upon homology to the infecting ancient retroviral genus. HERVs are divided into class I (Gammaretrovirus and Epsilonretrovirus genera), class II (Betaretrovirus genus) and class III (Spumaretrovirus genus)[6] (Supplementary Table 1). About 0.5% of LTR elements in the human genome remain unclassifiable or are believed to descend from errantiviruses and are known as Ty3 or mdg4 (refs. 6,7). HERV classes I through III are further broken down into eleven supergroups and 39 clades based upon homology between their coding and non-coding sequences[6] (Supplementary Table 1).

Owing to the threat posed by expression of retroviral genes to genome integrity and somatic-cell homoeostasis, the activity of ERVs is tightly regulated by a full complement of epigenetic modifications[8]. By targeting specific retroviral nucleic acid motifs[9] and retroviral characteristics such as intronless nascent mRNAs[10], host cells continuously surveil the activity of these elements with repressive epigenetic modifications. Insights into how our cells control the expression of ERVs have primarily come from mechanistic studies performed in model organisms. Although mouse ERVs are distinct from HERVs in their origin, sequence, genetic composition, distribution and phylogeny and in having the ability to form virus-like particles[2,11], the study of the fundamental molecular properties of ERVs has provided insights into the processes by which HERVs influence human health. The usefulness of model organisms is limited, however, and these insights must be interpreted accordingly, as mouse ERVs possess intact proviruses that can form infectious particles. Our current understanding of ERVs suggests that their transcription is repressed by DNA methylation[12], histone modifications[9] and non-canonical histone variants[13]. Post-transcriptional repression of HERV RNA occurs through RNA-destabilizing modifications[14], RNA degradation by surveillance machinery[10] and germline-specific[15] Piwi-interacting RNA (piRNA)-mediated degradation[16]. This silencing of ERV elements begins during the early stages of embryogenesis[17], and distinct patterns of ERV element activity are then observed throughout embryonic development[18–21] and in fully differentiated tissues[22–24].

In this Review, we provide a comprehensive discussion of HERVs and their observed activity throughout the human lifespan. We first discuss methods for the quantification of HERV RNAs and then delve into their roles and activity in sexual reproduction and development. Next, we discuss the proposed mechanisms by which HERV activity can be modulated by microorganisms, and how their deregulation contributes to pathology of certain diseases. We conclude by discussing the upcoming direction of HERV research and the importance of their study.

Methods of quantifying HERV expression

HERV RNAs undergo the post-transcriptional processing stages of 5′ capping, splicing and 3′ polyadenylation (formation of poly(A) tail)[25], which enable their capture when preparing libraries for next-generation sequencing (NGS). Although commonly used selection of poly(A)-containing RNAs does allow for the enrichment of HERV RNAs, it is unclear whether some library preparation techniques contain biases that unknowingly hinder the recovery of HERV RNA reads owing to the quality of source-tissue preservation, the temporal stability of the RNAs, their extraction techniques or sequencing platform biases[26]. Owing to high homology between and within proviral loci, their quantification is complicated because sequencing reads from HERV RNAs often align with multiple loci[27]. This multimapping phenomenon confounds the assignment of HERV reads when analysing their expression from NGS datasets, resulting in low-definition data. Long-read RNA sequencing techniques can increase the likelihood of capturing unique nucleic acid motifs and provide the most straightforward way to limit multimapping[28,29], but these techniques are currently expensive.

Several methods exist that improve HERV RNA quantification in short-read NGS by stratifying HERV reads into calculable groups or by implementing quantitative strategies that better approximate read assignments[30,31]. For single-cell RNA-sequencing analyses, pipelines such as scTE[22], which minimizes multimapping to accurately depict transposable element classes by sacrificing locus specificity, and soloTE[32], which retains locus specificity through the implementation of an expectation–maximization algorithm, have been developed to accurately calculate the expression of ambiguous transposable elements and HERVs. For bulk RNA-seq analyses of HERV expression, pipelines such as ERVmap[33], which improves the clarity of HERV-specific reads by excluding polymorphic reads and multimapping reads that are specific to conserved regions; Telescope[34] and SQuIRE (software for quantifying interspersed repeat expression)[35], which apply an expectation–maximization algorithm to provide locus-specific definition in otherwise discarded multimapping reads; SalmonTE[36], which applies a variant of stochastic collapsed Bayesian inference followed by constructing equivalence classes for inferred reads that are then improved upon with an expectation–maximization algorithm; and hervQuant[37], which is a computational workflow that relies upon direct base-pair alignments to the reference genome, have been implemented to accurately estimate HERV expression. In addition to variance in the calculative approaches of the different bioinformatics tool, there are also several annotations for HERVs, which vary considerably[6,33,34,38–44]. Thus far, HERV quantification remains an iterative process, which will benefit from future standardizations of nomenclature, quantification strategies and annotations.

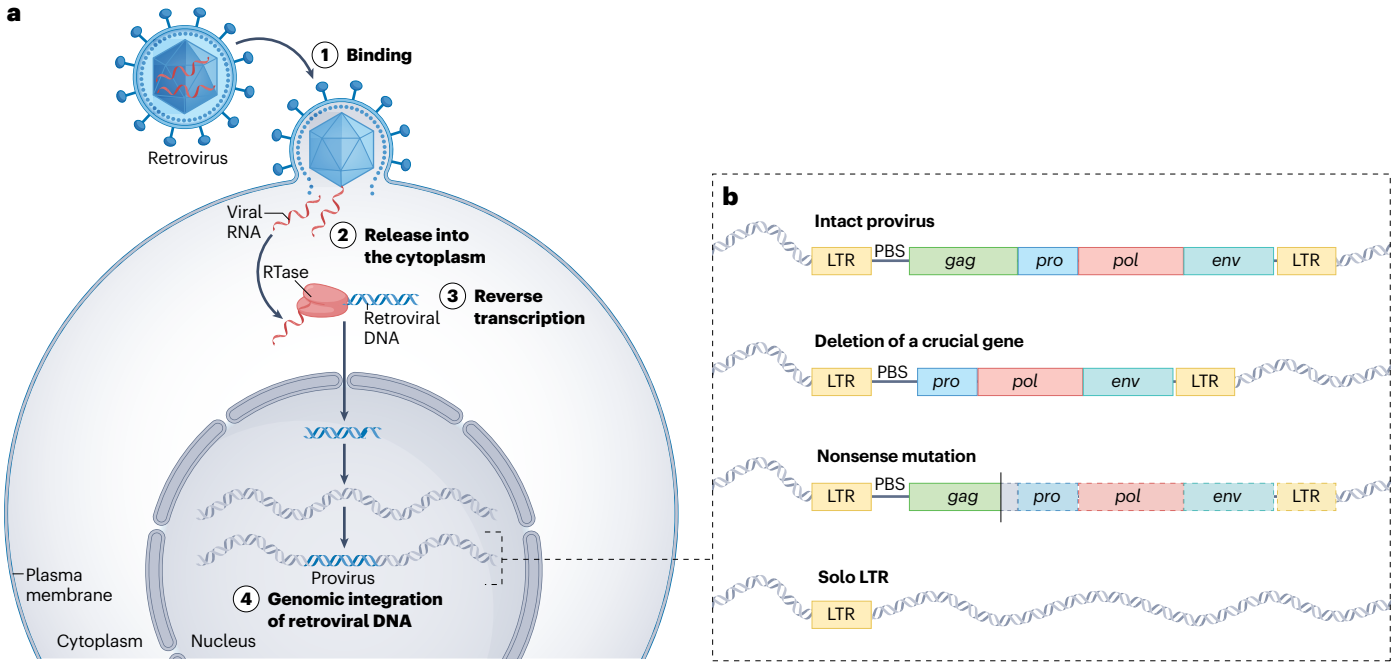

Fig. 1 | The endogenization of a retrovirus. a, Retroviruses bind to and fuse with a target cell (1) to release a capsid complex (2) that contains RNA copies of the retroviral genome, packaged with the factors (not shown) required for integration into the host genome. In the cytosol, viral reverse transcriptases (RTases) produce a DNA molecule of the retroviral RNA genome (3), which is trafficked into the nucleus for genomic integration (4), thereby forming a provirus. **b**, Proviruses can undergo deletions, point mutations, and/or recombination (not shown) that yield endogenized retroviral genes incapable of producing infectious particles. Internal recombination between flanking long terminal repeats (LTRs) of a full-length provirus can yield a solo LTR without coding regions. The nonsense mutation is depicted as a vertical line. *env*, envelope gene; *gag*, group-specific antigen gene; PBS, primer-binding site; *pol*, polymerase gene; *pro*, protease gene.

Expression of HERVs throughout development

A few studies have begun to characterize HERV expression patterns in healthy tissues and cell types[22–24]. HERV RNAs can be detected in primordial germ cells (PGCs)[45] and then traced throughout early embryogenesis and somatogenesis as unique transcriptional profiles. These choreographed changes in HERV activity at the DNA, RNA and protein levels affect embryonic development, and in somatic tissues, this activity may contribute to cell identity (Fig. 2).

HERVs in germ cells and embryogenesis

HERV LTRs function as regulatory elements that coordinate the expression of cell potency factors throughout sexual reproduction and embryonic development[19,46]. The specification of PGCs appears to partly rely on transcriptional networks coordinated through the activation of HERV LTRs, primarily those belonging to the HERVK–LTR5Hs subtypes[45]. Oocytes that arise from PGCs display activation of HERVK14–LTR14 LTRs[19], and upon fertilization, the expression of HERV elements and host genes is largely coordinated by TFBs within HERVK14–LTR14, HERVK–LTR5Hs, HERVH–LTR7, HERV9–LTR12, HERVL–MLT2 and THE1A LTRs[19] (Fig. 2a). In addition to functioning as regulatory elements, some HERVH transcripts function in embryonic stem cells as long non-coding RNAs[18], which recruit the pluripotency transcription factor OCT4 (also known as POU5F1) and its cofactor p300 to genomic HERVH–LTR7 domains, thereby enhancing the transcription of pluripotency factors genes cis-regulated by HERVH LTRs[18]. In mice, the expression of ERV transcripts occurs so early in embryogenesis that they are detectable at the first cell division[47]. As totipotent cells of the morula divide into the pluripotent cells that make up a polarized blastocyst, the transcription machinery is rewired to begin suppressing the embryonically active HERV LTR TFBs of younger human mouse mammary tumour virus-like (HML) and HERVHF supergroups[19,45,48,49]. From a pluripotent blastocyst onwards, an increasingly complex choreography of ERV LTR activity emerges in the endoderm, mesoderm and ectoderm[19,49].

In addition to HERV functions at the DNA and RNA levels, the placental expression of HERV proteins is required for embryonic development[50] (Fig. 2b). The fusogenic HERVW envelope protein Syncytin-1 has been co-opted to function as a surface receptor that facilitates cell–cell fusion during syncytiotrophoblast formation[50]. The syncytiotrophoblast forms a multinucleated barrier between the maternal and embryonic circulatory systems that protects the developing embryo from pathogens and facilitates the transfer of growth factors and nutrients that are crucial for embryonic development[50,51]. Syncytin-A deletion in mice demonstrates that impaired syncytiotrophoblast formation retards embryonic growth and vascular development, resulting in early lethality[51]. Other HERV proteins are highly expressed in embryonic stem cells and the placenta, such as Suppressyn[52] and Rec[20] (Fig. 2b). These proteins possess antiviral activity and could function as restriction factors that protect gestating offspring from viral infections. Collectively, the facilitation of placental barrier formation by Syncytin is crucial for human reproduction[50,51], and the activity of multiple ERV elements occurs in specific spatiotemporal patterns throughout the establishment of progenitor cell potency[18,20,45,46], polarization of multicellular structures[19,20,28,47] and embryonic development[20,49–52]. Whether and to what extent the repeatedly observed spatiotemporally coordinated HERV activity contributes to the development and protection of human embryos is not yet clear.

HERVs in somatic cells

Following conception, HERV activity fluctuates in response to and the expression of factors that regulate early embryogenesis[19,46]. HERV elements are then generally silenced by the deposition of epigenetic modifications[8]; however, many remain active in terminally differentiated cells[22–24]. In somatic tissues, thousands of HERV RNAs are transcribed, the majority of which can be detected at multiple body sites[23]. In contrast to a subset of ubiquitously expressed HERV RNAs (Fig. 2c), other HERV RNAs are expressed in a tissue-specific manner at abundances estimated to range between approximately 0.19 and 1.91% of all poly(A) transcripts[23]. Analysis of global HERV expression profiles suggests that, of the tested tissues, the brain, pituitary gland, testis, liver, heart, blood and muscle tissues possess expression patterns distinct enough to differentiate them from the other tissues by the application of multidimensional scaling[23] (Fig. 2d). Confounding factors such as sex, ethnicity and age also appear to influence the expression of HERVs[23]. Analyses of the most evolutionarily young HERVKs (HML-2) found 37 proviruses to be transcriptionally active in 54 body sites, with tissues composing the central nervous system (CNS) displaying the highest HERV transcript abundances[24]. The expression of many of these elements in somatic cells is hypothesized to be driven by two primary mechanisms. The primary mechanism by which HERVs are transcribed in somatic tissues is 'leaky expression' resulting from the transcription of proximal genes, and the secondary mechanism is by activation of their own 5′ LTR[24]. Prior to their endogenization, ancient retroviruses probably coordinated host transcription in target cells through regulatory elements in their LTRs. Current hypotheses suggest that the activation of HERV RNAs by their LTR region is a remnant of their infectious history, in which the LTR regions would direct cell type-specific transmission and spreading. Following endogenization of a retrovirus, the cell type-specific activity of HERV RNAs probably reflects a carryover of the signals required for retroviral replication. Although the emerging understanding of global HERV transcription during homoeostasis suggests the existence of many undiscovered roles for this activity, the extent by which HERV RNAs may contribute to physiology remains unclear.

Exogenous stimuli that determine HERV expression

The microbiota is a collective of host-associated microorganisms, and like HERVs, it is crucial for human health[53]. HERVs (or ERVs) are now thought to interact with the microbiota[54,55]. Host-associated

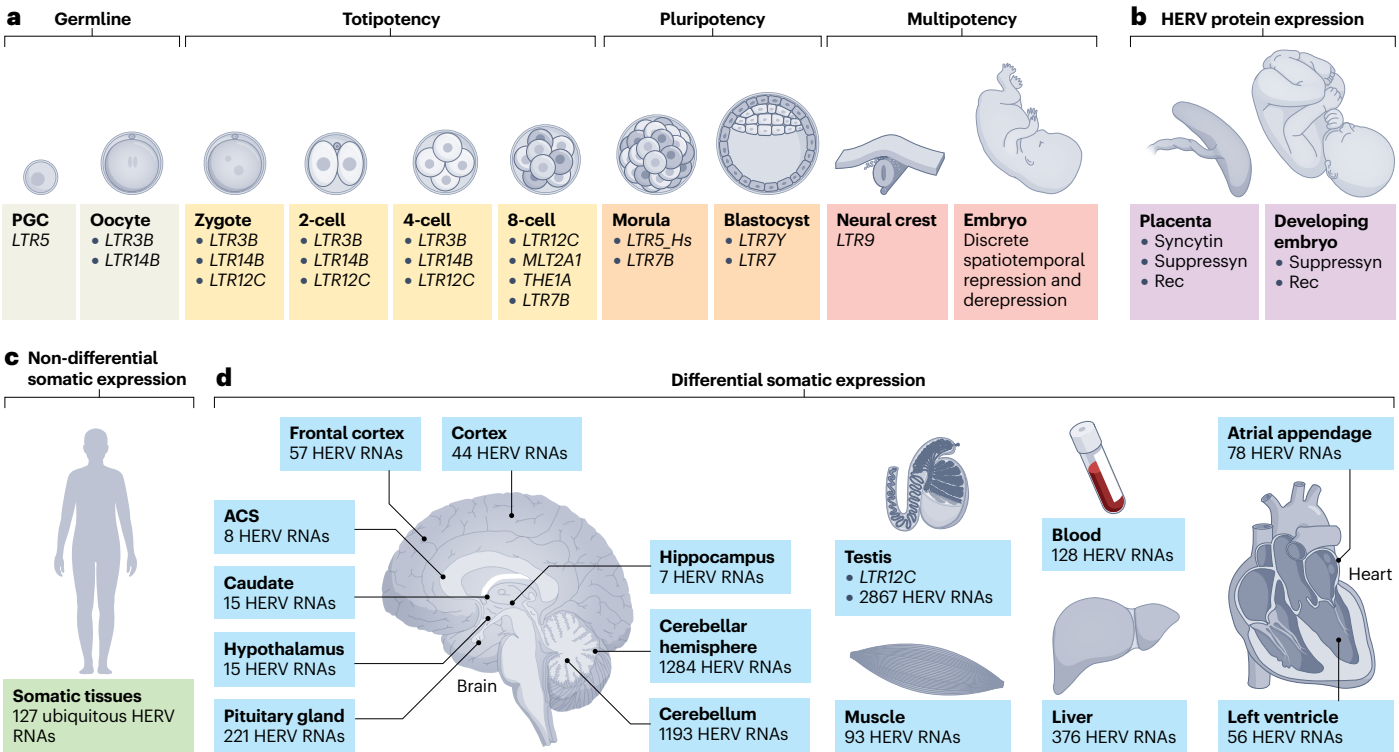

Fig. 2 | Spatiotemporal patterns of human endogenous retrovirus activity throughout development. a, Various patterns of gene transcription driven by the activity of human endogenous retrovirus (HERV) promoter elements can be traced from progenitor germ cells (PGCs)[45] through oocytes[19] to the early stages of embryogenesis, leading to the formation of a polarized blastocyst[19,45,48,49]. As cells of the embryo transition to multipotency, more complex patterns of HERV-driven transcriptional activity can be observed[19,49]. Long terminal repeats (LTRs) that are transcriptionally active in different embryogenesis stages are indicated. **b**, Placentation relies on the expression of the fusogenic HERV protein Syncytin[50,51]. The HERV proteins Suppressyn[52] and Rec[20] are also expressed at high levels in the placenta and can be detected also in the developing embryo. **c**, Following differentiation into somatic tissues, tens of HERV promoter regions remain active to promote the expression of 127 ubiquitously expressed HERV RNAs[23]. **d**, Anatomically distinct regions of the brain, testis, liver, skeletal muscle, blood and heart highly express specific HERV RNAs[23]. LTR12C contributes to the unique HERV expression profile of the testis[23]. ACS, anterior cingulate cortex; MER, medium reiterated frequency repeats; MLT, mammalian LTR transposon; MST, mammalian LTR transposons named for responsiveness to *MST*II restriction enzyme; THE, transposon-like human element.

viruses can trigger pattern recognition receptors (PRRs) and can also directly activate HERV expression by hijacking host transcriptional and translational machinery[56]. Host-associated prokaryotes transduce the expression of HERVs through PRRs[55,57–61].

Transcriptional activation of HERVs by exogenous viruses

Viruses replicate by hijacking the molecular machinery of the host cell[56]. The resulting rewiring of the host transcriptome and translatome is met with resistance by the host in the form of antiviral immunity[62] and stress-induced cell senescence[63]. For example, in response to vaccina virus infection, host cells utilize a specific exapted HERV LTR for the activation of the AIM2 inflammasome[64], which results in rapid derepression of interferon response genes, also allowing the expression of proximal HERVs[65].

During infections, HERV RNA expression is sensitive to the transcriptional reprogramming that occurs in the host cells (Supplementary Table 2). Infection and transformation of B cells by Epstein–Barr virus activates ERVL–MaLR (mammalian apparent LTR retrotransposons), HERVL and ERV1 LTRs throughout the genome[66]. The Epstein–Barr virus transcriptionally transactivates the expression of the HERVK18 envelope protein, which has superantigen activity[67]. Similarly, Kaposi sarcoma-associated herpesvirus (KSHV, also known as HHV-8) directly activates the expression of the HERVK envelope-derived proteins Rec and NP9, which aid virus propagation and potentially contribute to oncogenic processes[68]. Mechanistically, the latent KSHV accessory proteins latency-associated nuclear antigen (LANA) and viral FLICE protein (vFLIP) activate the host MAPK and NF-κB signalling pathways, respectively, to promote the expression of HERVK envelope transcripts[68]. KSHV-induced Rec and NP9 expression promotes the invasiveness and tumorigenicity of infected endothelial cells through pathways dependent on VEGFR1 signalling[68].

The human T-lymphotropic virus type 1 (HTLV-1) cofactor Tax is a transcription transactivator of host genes that promote the expression of the HTLV-1 provirus[69]. Exposure of T-cells to Tax also regulates host gene expression by activating HERVH and HERVW LTRs[69]. Similarly, the human immunodeficiency virus type 1 (HIV-1) cofactor Tat transactivates the transcription of HIV-1 proviruses[70]. In addition to elevating the expression of HIV-1 RNA, Tat promotes transcription at HERVK LTRs to upregulate the expression of HERVK RNAs by binding to NF-κB. HIV-1 may also influence HERVK expression post-transcriptionally through activity of the cofactor Rev, which binds to Rev response elements (RREs) in HIV-1 RNAs and facilitates their shuttling out of the nucleus[71]. The HIV-1 Rev–RRE nuclear export system is structurally similar to HERVK Rec (a functional homologue of Rev) and Rec response elements (RcRE)[71]. During HIV-1 infection, this homology may allow HIV-1 Rev to bind HERVK RcREs and facilitate their export from the nucleus[71]. Collectively, virus-induced changes in cellular transcription drastically alter the activity of HERV elements[72,73].

Crosstalk between ERVs and prokaryotes

Toll-like receptors (TLRs) are PRRs[74]. In mice, activation of TLRs by prokaryotes induces considerable changes in the expression of ERVs[54,59]. Lipoteichoic acid-mediated activation of TLR2 (ref. 55), lipopolysaccharide-mediated activation of TLR4 (refs. 57,60,61) and flagella-mediated activation of TLR5 (ref. 58) orchestrate ERV activity at the RNA level. The differential expression of ERVs in response to microbial cues affects the inflammatory capacity of immune-cell subsets and their local microenvironment[55,58]. In mouse models, this change in immunological capacity is primarily dependent on ERV RTase production of retrovirus-like dsDNA, which activates the pro-inflammatory factor cyclic GMP–AMP synthase (cGAS)[55]. These complex interactions between bacteria, their mammalian host and its endogenous retroelements participate in TLR2-mediated[55] and TLR5-medated[58] inflammatory pathways. Early-stage investigation of ERVs as mediators of host–microorganism interactions has identified these genetic elements as fundamental to immunity[54].

Deregulation of HERVs in disease

Genetics, environment and age can all affect cellular homoeostasis[75,76]. Chronic dysregulation of cellular homoeostasis may derepress HERVs, with consequent pathologies characterized by inadvertent cytotoxicity[77], immunity activation[78] and/or cell senescence[79,80]. Deregulation of HERV activity has been found in neurodegeneration[81], autoimmunity inflammation[78] and oncogenesis[82]. Although our current understanding of HERVs is based primarily on repeated observations that provide strong correlations between their activity and pathogenesis, it is now important to support these hypotheses with mechanistic studies that can conclusively define the effects of HERV activity on pathogenesis. These studies are difficult to conduct owing to the discrepancies between the composition of ERVs in humans and in model organisms, and to the limited availability of reagents that can accurately estimate the activity of HERV elements.

Ageing-associated neurodegeneration

Neurodegenerative disorders are defined by CNS-specific atrophy occurring in a chronic progressive manner[83]. Neurodegeneration presents in clinically distinct pathologies with variations in aetiology and severity. Typical clinical features of advanced-stage neurodegeneration include impairments to cognitive function, deteriorated motor function and wasting of tissues innervated by afflicted brain regions[83]. Neurodegenerative disorders possess no common aetiology[83]. Alzheimer's disease[7,84,85] and amyotrophic lateral sclerosis (ALS)[77,86–89] are two clinically distinct neurodegenerative disorders that, in certain cases, demonstrate deregulation of HERV activity at protein and RNA levels (Fig. 3). This HERV deregulation is hypothesized to contribute to pathogenesis following the initiation of neurodegeneration[81]. The overt expression of pathological HERVK and HERVW envelope proteins may also disrupt neuroplasticity and neural development in neurodegenerative and neuropsychiatric disorders[90–93], suggesting that HERV derepression promotes the initiation of disorders associated with their expression.

Recent studies suggest that many of the processes that underly ageing-associated pathogenicity are reinforced by the re-expression of certain ERVs in senescent cells[79,80]. Evolutionary younger HERVs with intact capsid-coding genes, such as certain HERVK proviruses, may form non-infectious retrovirus-like particles[94] when derepressed in senescent cells, which might activate innate immunity through DNA receptors such as cGAS to drive ageing-associated processes in senescent and non-senescent cells alike[79]. This process led to the hypothesis that ERVs are a cause of ageing-associated pathogeneses, based on findings collected primarily from animal models, post-mortem tissues and disease-relevant human cell lines[79,80].

Post-mortem analysis of CNS tissues collected from individuals with ALS found that the RNA processing protein TAR DNA-binding protein 43 (TDP-43) is typically found in aggregates that lead to loss of its functionality[89]. The RNA processing roles provided by TDP-43 are crucial to regulating the activity of transposable elements, and feedback from HERV overexpression may further impair TDP-43 functionality

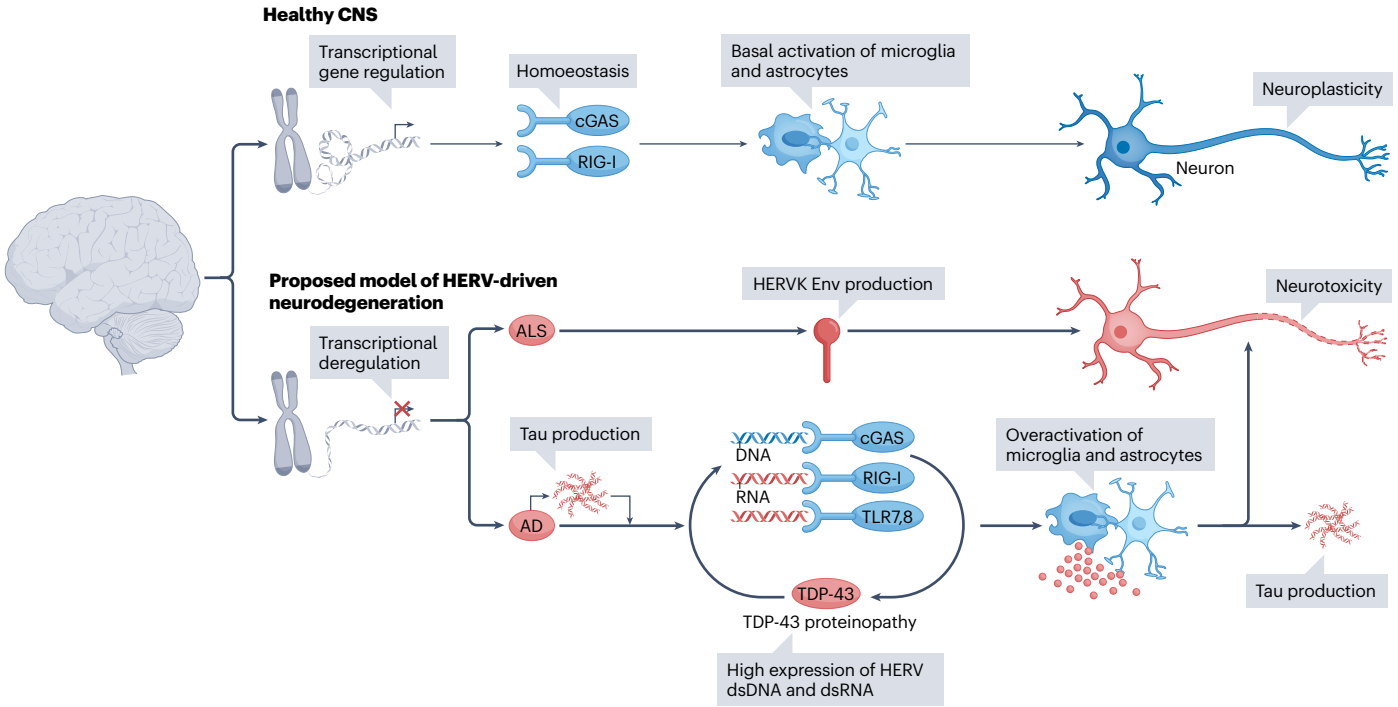

Fig. 3 | Proposed model for human endogenous retrovirus contributions to ageing-associated neurodegeneration. In a healthy central nervous system (CNS), transcription regulation maintains the repression of human endogenous retrovirus (HERV). Homeostasis in the CNS is characterized by basal activation of glia cell subsets, which sustain an environment conducive to neuroplasticity. Neurodegeneration involves disruptions to cellular homoeostasis, which enable the expression of certain HERVs. Transcriptional deregulations of HERVs are proposed to contribute to hallmark pathologies of neurodegenerative disorders, primarily by promoting the accumulation and thus aggregation of tau protein, TAR DNA-binding protein 43 (TDP-43) proteinopathy and neurotoxicity. In amyotrophic lateral sclerosis (ALS), the overexpression of HERVK envelope (Env) proteins may contribute to motor neuron cell death. In Alzheimer disease (AD), the derepression of HERV is associated with pathological tau production. The overexpression of retrovirus-like cytosolic HERV double-stranded DNA (dsDNA) and dsRNA may agonize the intracellular nucleic acid receptors cyclic GMP–AMP synthase (cGAS), Toll-like receptor 7 (TLR7), TLR8 and retinoic acid-inducible gene I (RIG-I). The overabundance of HERV nucleic acids may then reinforce TDP-43 proteinopathy, further contributing to HERV derepression by forming a positive feedback loop conducive to neurodegeneration. The accumulation of retrovirus-like HERV dsDNA and dsRNA in the CNS of individuals with Alzheimer disease is hypothesized to contribute to neurotoxic immune activation in microglia and astrocytes and worsen the severity of disease.

and drive a mutual sustainment between the two[7] (Fig. 3). Dysfunctional aggregation of TDP-43 may cause accumulation of a neurotoxic HERVK envelope protein[77,86,87] and RTase[88,89,95] in the CNS of some individuals with ALS. However, these studies are controversial[96], and other studies have failed to replicate elevation in HERVK activity in ALS[97,98].

Studies in laboratory models of Alzheimer disease have shown that the establishment of neurofibrillary tangles is conducive to the overexpression of cytosolic dsDNA and dsRNA stemming from HERVs and other transposable elements, which may promote neuroinflammation[93,99] and contribute to subsequent tauopathy[85,100]. Post-mortem analyses of CNS tissues from individuals with Alzheimer disease found a significant overabundance of RNAs from multiple HERV families, including those from HERVK, HERVH and HERVL[84,101]. The associations between pathological burden and overabundant transposable element RNAs in post-mortem Alzheimer disease tissues suggest that short interspersed nuclear elements (SINEs) and HERVs possess understudied roles in human tauopathy[100]. The overabundance of HERV transcription in the CNS of post-mortem Alzheimer disease tissues might be driven by impairments in the transcriptional and post-transcriptional regulatory networks that traditionally silence the neuroinflammatory overproduction of ERVs, as demonstrated by piRNA depletion leading to chromatin relaxation in *Drosophila melanogaster* models of tauopathy[84]. Although the expression and regulatory activity of human piRNAs are believed to be predominantly germline-specific[15], these studies demonstrate the potential for subsequent deregulations in ERV activity to mediate tauopathy.

Collectively, the ageing-associated derepression of HERV activity likely contributes to the inflammatory component of cell senescence-relevant pathologies, and deregulated HERV expression may further contribute to the neurotoxicity and tauopathy observed in certain neurodegenerative disorders (Fig. 3).

Pathological inflammation

The aetiology of inflammatory disorders depends on many recurring genetic and environmental factors[102]. In autoimmunity, HERV activation may worsen inflammation symptoms by expressing inflammatory proteins or potentially by forming retrovirus-like dsDNA and dsRNA in endosomes and the cytosol[78]. HERVs are hypothesized to contribute to the pathological inflammation in multiple sclerosis[103,104], psoriasis[55], rheumatoid arthritis[105], systemic lupus erythematosus[106,107], pemphigus vulgaris[108] and Aicardi–Goutières syndrome (AGS)[109,110], amongst others.

In multiple sclerosis, a HERVW envelope protein is believed to drive encephalomyelitis through immune hyper-activation[104]. However, a phase II clinical trial has shown that an antibody that blocks the activity of this envelope protein isoform has no effect on acute inflammation[111]. In AGS[110], colorectal cancers associated with inflammatory bowel disease[112], systemic lupus erythematosus[113] and psoriasis[55], the delivery of RTase inhibitors can ameliorate inflammatory pathologies stemming from cytosolic dsDNA production. Moreover, a clinical trial testing the efficacy of RTase inhibitors in AGS has had positive results and the treatment is a possible therapeutic strategy[110]. However, it is currently unclear how much, if any, of this cytosolic dsDNA production stems from HERVs compared to other, more active endogenous retroelements, such as long interspersed nuclear element (LINEs) and SINEs, which compose the much larger fractions of the human genome, with LINEs being the putative target of RTase inhibition. Our current understanding of HERV-derived cDNA production remains speculative, as it is unclear how much, if at all, they contribute to observations of DNA sensing-driven pathologies. Our understanding of HERV activity in immunity suggests that further investigation into the physiological basis of HERV-mediated inflammation is required for the development of novel therapeutic strategies. Microorganisms remain one of the strongest risk factors for autoimmunity, with microbiota composition and viral infections having lasting consequences to systemic immunity[114]. The immunity axes of host–virus and host–bacteria are likely dependent on the transactivation of HERV transcription, and we suggest that further investigation of HERVs in the context of dysbiosis-mediated autoimmune inflammation is needed[54].

Glossary

Dysbiosis
The state of disbalance in the composition and diversity of microorganisms that constitute the microbiota of a host organism.

Expectation–maximization
An algorithmic computational approach that iteratively defines the maximum likelihood for given estimations based on latent variables, commonly used to improve upon estimations provided by probabilistic models of data.

Inflammasome
Multisubunit complex that assemble in the cytosol in response to inflammatory stimuli. Following their assembly, inflammasomes transduce immunological signals.

Inflammatory disorders
Conditions in which unregulated and typically self-targeting inflammation contributes to pathogenesis. Inflammatory disorders generally demonstrate cyclic cascades of immune responses that further reinforce incipient inflammation.

Neurofibrillary tangles
Pathological insoluble aggregates of hyperphosphorylated tau protein in the central nervous system.

Pattern recognition receptors
Invariant innate immunity receptors that recognize and detect molecular signals that commonly arise from pathogen invasion or cellular damage.

Retrovirus-like particles
Assembled endogenous retroviral particles that are morphologically similar to an infectious retrovirus but are replication incompetent and thus non-infectious.

Superantigen
Molecules that possess nonspecific stimulatory capacity of adaptive immunity cell subsets.

Syncytiotrophoblast
A barrier layer of multinucleated epithelial cells that separates the maternal and embryonic circulatory systems.

Ty3 or mdg4
A phylogeny of LTR-possessing retroelements. The previously used nomenclature for these elements relied on the word 'gypsy'; going forward, these elements should be referred to as 'Ty3' or 'mdg4'.

Oncogenesis

Cancers arise on individual bases from complex aetiologies that draw from genetic, lifestyle and environmental risk factors[115]. Investigation into putative roles for non-coding genomic elements during oncogenesis has repeatedly demonstrated that their deregulation may contribute pathogenesis of certain malignancies[82,116] (Fig. 4a).

TFBs in HERV LTRs form a potent reservoir of cis-regulatory elements that can influence host transcription[48]. Whereas the regulatory capacity of some HERV LTRs has been exapted to benefit germ cells and somatic cells, most of these elements are thought to remain dormant through the deposition of repressive epigenetic modifications. Oncogenic conditions can disrupt host gene regulation, allowing for the derepression of dormant HERV LTRs[117]. The derepression of HERV LTRs can then affect host gene transcription and, in certain cancers, can cis-activate the transcription of full-length, truncated or chimeric isoforms of oncogenes, a process termed 'onco-exaptation'[117] (Fig. 4b). In Hodgkin lymphoma, derepression of ERV1 LTRs and of ERVL–MaLR LTRs drives ectopic expression of the oncogenes interferon regulatory factor 5 (*IRF5*)[118] and macrophage colony-stimulating factor 1 receptor (*CSF1R*)[119], respectively.

Onco-exaptation of specific LTR promoters can yield the production of truncated oncoproteins that may further drive pathogenesis, such as activation of anaplastic lymphoma kinase (*ALK*) by an ERVL promoter in melanoma[120] and receptor tyrosine-protein kinase erb-B4 (*ERBB4*) by an ERVL–MaLR promoter in ALK-negative anaplastic large-cell lymphoma[121]. Chimeras between transcripts of oncogenes and HERVs, such as the oncogenic fatty acid-binding protein 7 (*FABP7*)–LTR2 chimera, which is activated by an ERV1 promoter in diffuse large B cell lymphoma[122], and calbindin (*CALB1*)–HERVH chimeras driven by HERVH promoters in prostate cancer[123] and in lung squamous cell carcinoma[124], are also driven by onco-exaptation.

Although the expression of HERVs is not thought to be capable of driving cellular transformation alone, seminal studies have characterized HERV RNA activity across large pan-cancer cohorts, finding HERV expression to be largely upregulated in accordance with disease severity in solid tumours and haematological malignancies[125–127]. A recent preprint reveals that, by incorporating HERV expression data into the definition of cancer subclasses, individuals with certain haematological malignancies could be provided with a more-accurate prognosis[125].

Cancer-associated HERV RNAs can possess open reading frames that, owing to cancer-associated disruptions in RNA metabolism, can be translated throughout oncogenesis into tumour-specific antigens (TSAs)[37,127–135]. Although canonical splicing occurs between two exons of a gene, splicing often goes awry in malignant conditions[136], resulting in non-canonical trans-splicing between exons and overabundant HERV RNAs, thereby forming junction between exons and transposable element (JET) mRNAs, which are translated to form recurrent TSAs[131,132,135] (Fig. 4c).

Certain HERV envelope proteins are overexpressed in the tumour microenvironment[137]. In Kaposi sarcoma, KSHV-infected endothelial cells overexpress the HERV proteins Rec and NP9[68]. Studies in laboratory models of Kaposi sarcoma have demonstrated that the

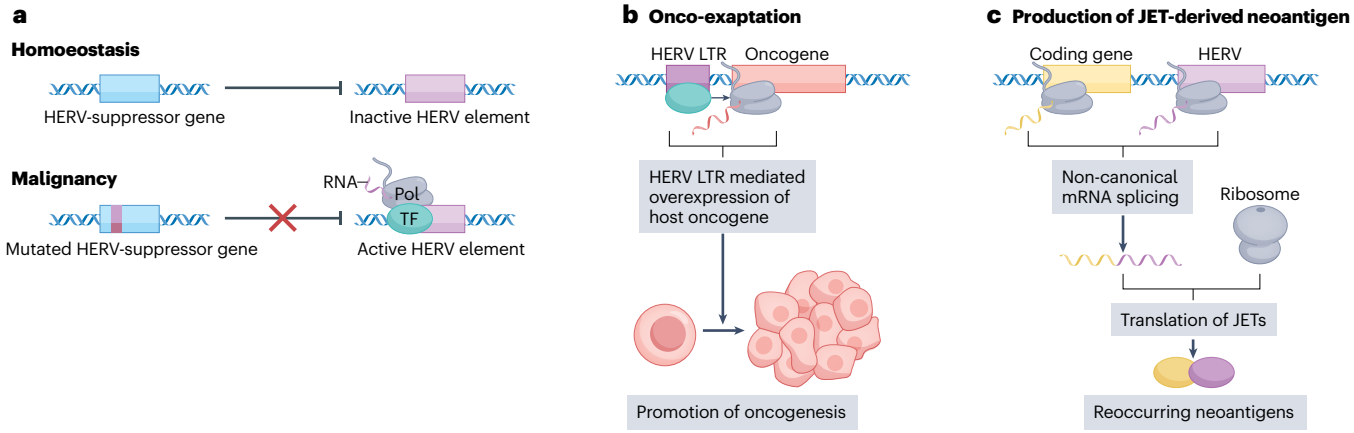

Fig. 4 | Deregulation of human endogenous retroviruses in cancer. **a**, Mutations is genes that supress human endogenous retrovirus (HERV) expression enable interactions between HERVs, RNA polymerase II (Pol) and transcription factors (TF). **b**, The derepression of HERV long terminal repeats (LTRs) entails binding of host transcription factors to the LTRs and cis-activation of downstream oncogenes, a process termed 'onco-exaptation'. **c**, Disruptions in host pre-mRNA splicing during oncogenesis can permit the formation of mRNAs with junctions between exons and transposable elements (JETs). The translation of non-canonical JET mRNAs spliced from host mRNAs and HERV mRNAs results in the production of recurrent neoantigens found in certain malignancies.

overexpression of Rec and NP9 permits invasiveness through exploitation of host VEGFR1 signalling[68], suggesting that Rec and NP9 have roles in oncogenesis. In individuals with glioblastoma, HML-2 envelope proteins are overabundant in the CNS, and the expression of capsid, envelope and RTase RNAs correlates with survival[138]. The activity of HML-2 elements in glioblastoma appears to induce OCT4 activity, which promotes the stemness of neural progenitor-like cells. Disruption of HML-2 activity with CRISPR interference and RTase delivery could disrupt the stem cell niche of glioblastoma models, demonstrating putative roles for deregulated HERVK activity in contributing to glioblastoma pathologies[138]. Whereas other retroelements, such as LINEs, are hypothesized to produce retrovirus-like dsDNA and dsRNA that trigger antiviral innate immunity receptors, which in turn promote conditions that are conducive to tumour development[112,139], it is currently unclear how much, if any, retrovirus-like dsDNA is produced by models that aim to replicate HERV activity and RTase activity on HERV RNAs. It is possible that RTase inhibition disrupts LINE RTase activity, which then prevents the sustainment or initiation of inflammatory cascades that upregulate potentially pathogenic HERV elements. Further studies are therefore needed to discern how much, if any, of the pathological RTase activity observed in cancers and inflammatory disorders is attributable to HERVs in comparison to other retroelements.

The microenvironment of solid tumours hinders antigen accessibility and immunogenicity[140]. HERV-targeting immunotherapies could improve upon existing treatments for certain cancers[141]. Specifically, HERV-derived TSAs have been identified as alternative antigens expressed in bladder urothelial carcinoma[130], clear cell renal cell carcinoma[37], colon adenocarcinoma[130], head and neck squamous cell carcinoma[130], lung adenocarcinoma[128,130,132], breast cancer[142] and lung squamous cell carcinoma[130,132]. In vitro and in vivo models of humoral responses to HERV TSAs in lung adenocarcinoma and breast cancer have shown that HERVK envelope-specific antibodies can exert antitumour effects[128,142], and HERVK envelope protein-specific chimeric antigen receptor (CAR) T cell delivery demonstrated antitumour effects in xenografted models of breast cancer in mice[143]. Thus, the development of HERV-specific immunotherapies may provide novel treatment options for certain cancers[37,116,128–134,141–143].

Conclusion and future perspective

HERVs represent a large and structurally unique component of the human genome, whose activity appears to affect health and disease[2,3]. Choreographed activities of HERVs are observed at the early stages of sexual reproduction[19,20] and then throughout somatogenesis[22,23]. In fully differentiated cells, HERVs LTRs have been co-opted to dictate global transcription networks[48] and their expression may mediate immune system–environment interactions[54]. Animal models that aim to replicate HERV activity are imperfect because ERV sequences are highly stratified across Mammalia owing to post-speciation endogenization events and mutations. Furthermore, the ambiguity of HERV elements complicates validation techniques that rely on antibodies or primers. Computational approaches that best estimate HERV abundances from NGS output currently present the best method for their study; however, these methods are expensive, labour intensive and often limited in their retrievable conclusions. These issues have complicated studies that aim to determine causal roles of HERVs in disease and development.

Nevertheless, repeated observations have suggested possible mechanisms by which HERVs affect the aetiology of poorly understood oncological[116], neurological[81] and immunological[78] pathologies. It is now important to develop research methodologies that could accurately validate any hypothesized causal associations between deregulated HERV expression and human diseases. The validation of HERV involvement in human diseases may lead to new types of therapies. This exciting prospect could very well revolutionize the field of pharmacology; however, it will require much more study of the vast 'dark matter' of the human genome.

Published online: 23 October 2023

References

1. Nisole, S. & Saïb, A. Early steps of retrovirus replicative cycle. *Retrovirology* **1**, 9 (2004).
2. Jern, P. & Coffin, J. M. Effects of retroviruses on host genome function. *Annu. Rev. Genet.* **42**, 709–732 (2008).

3. de Parseval, N. & Heidmann, T. Human endogenous retroviruses: from infectious elements to human genes. *Cytogenet. Genome Res.* **110**, 318–332 (2005).
Together with Jern and Coffin (2008), this paper is a seminal review article that discusses in detail the foundational studies of endogenous retroviruses.
4. Nurk, S. et al. The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
5. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
6. Vargiu, L. et al. Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology* **13**, 7 (2016).
This paper is a comprehensive characterization and annotation of the complex HERV structures scattered throughout the human genome.
7. Chang, Y.-H. & Dubnau, J. Endogenous retroviruses and TDP-43 proteinopathy form a sustaining feedback driving intercellular spread of *Drosophila* neurodegeneration. *Nat. Commun.* **14**, 966 (2023).
8. Dopkins, N. et al. A field guide to endogenous retrovirus regulatory networks. *Mol. Cell* **82**, 3763–3768 (2022).
9. Yang, B. et al. Species-specific KRAB-ZFPs function as repressors of retroviruses by targeting PBS regions. *Proc. Natl Acad. Sci. USA* **119**, e2119415119 (2022).
10. Garland, W. et al. Chromatin modifier HUSH co-operates with RNA decay factor NEXT to restrict transposable element expression. *Mol. Cell* **82**, 1691–1707.e8 (2022).
11. Bannert, N. & Kurth, R. The evolutionary dynamics of human endogenous retroviral families. *Annu. Rev. Genom. Hum. Genet.* **7**, 149–173 (2006).
12. Turelli, P. et al. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res.* **24**, 1260–1270 (2014).
13. Groh, S. et al. Morc3 silences endogenous retroviruses by enabling Daxx-mediated histone H3.3 incorporation. *Nat. Commun.* **12**, 5996 (2021).
14. Chelmicki, T. et al. m6A RNA methylation regulates the fate of endogenous retroviruses. *Nature* **591**, 312–316 (2021).
15. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202 (2006).
16. Ha, H. et al. A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements. *BMC Genom.* **15**, 545 (2014).
17. Whitelaw, E. & Martin, D. I. K. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat. Genet.* **27**, 361–365 (2001).
18. Lu, X. et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat. Struct. Mol. Biol.* **21**, 423–425 (2014).
19. Göke, J. et al. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* **16**, 135–141 (2015).
20. Grow, E. J. et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221–225 (2015).
21. Barakat, T. S. et al. Functional dissection of the enhancer repertoire in human embryonic stem cells. *Cell Stem Cell* **23**, 276–288.e8 (2018).
22. He, J. et al. Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scTE. *Nat. Commun.* **12**, 1456 (2021).
23. She, J. et al. The landscape of hervRNAs transcribed from human endogenous retroviruses across human body sites. *Genome Biol.* **23**, 231 (2022).
24. Burn, A., Roy, F., Freeman, M. & Coffin, J. M. Widespread expression of the ancient HERV-K (HML-2) provirus group in normal human tissues. *PLoS Biol.* **20**, e3001826 (2022).
Together with She et al. (2022), this paper provides an in-depth, locus-specific atlas of HERV RNA expression in various healthy human tissues.
25. Coffin, J. M. et al. (eds) *Retroviruses* (Cold Spring Harbor Laboratory, 1997).
26. Robasky, K., Lewis, N. E. & Church, G. M. The role of replicates for error mitigation in next-generation sequencing. *Nat. Rev. Genet.* **15**, 56–62 (2014).
27. Treangen, T. J. & Salzberg, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* **13**, 36–46 (2011).
28. Berrens, R. V. et al. Locus-specific expression of transposable elements in single cells with CELLO-seq. *Nat. Biotechnol.* **40**, 546–554 (2021).
This paper is a cutting-edge bioinformatics pipeline that provides the highest definition of expression of transposable elements from long-read single-cell RNA sequencing.
29. Troskie, R.-L. et al. Long-read cDNA sequencing identifies functional pseudogenes in the human transcriptome. *Genome Biol.* **22**, 146 (2021).
30. Lanciano, S. & Cristofari, G. Measuring and interpreting transposable element expression. *Nat. Rev. Genet.* **21**, 721–736 (2020).
31. Lerat, E. Recent bioinformatic progress to identify epigenetic changes associated to transposable elements. *Front. Genet.* **13**, 891194 (2022).
32. Rodríguez-Quiroz, R. & Valdebenito-Maturana, B. SoloTE for improved analysis of transposable elements in single-cell RNA-Seq data using locus-specific expression. *Commun. Biol.* **5**, 1063 (2022).
33. Tokuyama, M. et al. ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. *Proc. Natl Acad. Sci. USA* **115**, 12565–12572 (2018).
34. Bendall, M. L. et al. Telescope: characterization of the retrotranscriptome by accurate estimation of transposable element expression. *PLoS Comput. Biol.* **15**, e1006453 (2019).
35. Yang, W. R., Ardeljan, D., Pacyna, C. N., Payer, L. M. & Burns, K. H. SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Res.* **47**, e27 (2019).
36. Jeong, H.-H., Yalamanchili, H. K., Guo, C., Shulman, J. M. & Liu, Z. in *Biocomputing 2018* Vol. 23, 168–179 (World Scientific, 2018).
37. Smith, C. C. et al. Endogenous retroviral signatures predict immunotherapy response in clear cell renal cell carcinoma. *J. Clin. Invest.* **128**, 4804–4820 (2018).
38. Tristem, M. Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. *J. Virol.* **74**, 3715–3730 (2000).
39. Andersson, M. L. et al. Diversity of human endogenous retrovirus class II-like sequences. *J. Gen. Virol.* **80**, 255–260 (1999).
40. Bao, W., Kojima, K. K. & Kohany, O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
41. Paces, J. et al. HERVd: the human endogenous retroviruses database: update. *Nucleic Acids Res.* **32**, D50 (2004).
42. Becker, J. et al. A comprehensive hybridization model allows whole HERV transcriptome profiling using high density microarray. *BMC Genom.* **18**, 286 (2017).
43. Garazha, A. et al. New bioinformatic tool for quick identification of functionally relevant endogenous retroviral inserts in human genome. *Cell Cycle* **14**, 1476–1484 (2015).
44. Hubley, R. et al. The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **44**, D81–D89 (2016).
45. Xiang, X. et al. Human reproduction is regulated by retrotransposons derived from ancient Hominidae-specific viral infections. *Nat. Commun.* **13**, 463 (2022).
46. Kunarso, G. et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **42**, 631–634 (2010).
47. Kigami, D., Minami, N., Takayama, H. & Imai, H. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos1. *Biol. Reprod.* **68**, 651–654 (2003).
48. Fueyo, R., Judd, J., Feschotte, C. & Wysocka, J. Roles of transposable elements in the regulation of mammalian transcription. *Nat. Rev. Mol. Cell Biol.* **23**, 481–497 (2022).
49. Pontis, J. et al. Hominoid-specific transposable elements and KZFPs facilitate human embryonic genome activation and control transcription in naive human ESCs. *Cell Stem Cell* **24**, 724–735.e5 (2019).
50. Mi, S. et al. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785–789 (2000).
51. Dupressoir, A. et al. Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc. Natl Acad. Sci. USA* **106**, 12127–12132 (2009).
52. Frank, J. A. et al. Evolution and antiviral activity of a human protein of retroviral origin. *Science* **378**, 422–428 (2022).
53. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* **19**, 55–71 (2021).
54. Dopkins, N. et al. How human endogenous retroviruses interact with the microbiota in health and disease. *Trends Microbiol.* **30**, 812–815 (2022).
55. Lima-Junior, D. S. et al. Endogenous retroviruses promote homeostatic and inflammatory responses to the microbiota. *Cell* **184**, 3794–3811.e19 (2021).
56. Walsh, D. & Mohr, I. Viral subversion of the host protein synthesis machinery. *Nat. Rev. Microbiol.* **9**, 860–875 (2011).
57. Mommert, M. et al. Dynamic LTR retrotransposon transcriptome landscape in septic shock patients. *Crit. Care* **24**, 96 (2020).
58. Dopkins, N. et al. Endogenous reverse transcriptase inhibition attenuates TLR5-mediated inflammation. *mBio* **14**, e0328022 (2023).
59. Young, G. R., Mavrommatis, B. & Kassiotis, G. Microarray analysis reveals global modulation of endogenous retroelement transcription by microbes. *Retrovirology* **11**, 59 (2014).
60. Mommert, M. et al. LTR-retrotransposon transcriptome modulation in response to endotoxin-induced stress in PBMCs. *BMC Genom.* **19**, 522 (2018).
61. Rookhuizen, D. C. et al. Induction transposable element expression is central to innate sensing. Preprint at *bioRxiv* https://doi.org/10.1101/2021.09.10.457789 (2021).
62. Yan, N. & Chen, Z. J. Intrinsic antiviral immunity. *Nat. Immunol.* **13**, 214–222 (2012).
63. Kohli, J., Veenstra, I. & Demaria, M. The struggle of a good friend getting old: cellular senescence in viral responses and therapy. *EMBO Rep.* **22**, e52243 (2021).
64. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**, 1083–1087 (2016).
65. Platanitis, E. et al. Interferons reshape the 3D conformation and accessibility of macrophage chromatin. *iScience* **25**, 103840 (2022).
66. Leung, A. et al. LTRs activated by Epstein–Barr virus-induced transformation of B cells alter the transcriptome. *Genome Res.* **28**, 1791–1798 (2018).
67. Sutkowski, N., Conrad, B., Thorley-Lawson, D. A. & Huber, B. T. Epstein-Barr virus transactivates the human endogenous retrovirus HERV-K18 that encodes a superantigen. *Immunity* **15**, 579–589 (2001).
68. Dai, L. et al. Transactivation of human endogenous retrovirus K (HERV-K) by KSHV promotes Kaposi's sarcoma development. *Oncogene* **37**, 4534–4545 (2018).
69. Toufaily, C., Landry, S., Leib-Mosch, C., Rassart, E. & Barbeau, B. Activation of LTRs from different human endogenous retrovirus (HERV) families by the HTLV-1 tax protein and T-cell activators. *Viruses* **3**, 2146–2159 (2011).
70. Gonzalez-Hernandez, M. J. et al. Expression of human endogenous retrovirus type K (HML-2) is activated by the Tat protein of HIV-1. *J. Virol.* **86**, 7790–7805 (2012).
71. O'Carroll, I. P. et al. Structural mimicry drives HIV-1 Rev-mediated HERV-K expression. *J. Mol. Biol.* **432**, 166711 (2020).
72. Chen, J., Foroozesh, M. & Qin, Z. Transactivation of human endogenous retroviruses by tumor viruses and their functions in virus-associated malignancies. *Oncogenesis* **8**, 6 (2019).
73. Kyriakou, E. & Magiorkinis, G. Interplay between endogenous and exogenous human retroviruses. *Trends Microbiol.* **31**, 933–946 (2023).

74. Li, D. & Wu, M. Pattern recognition receptors in health and diseases. *Signal Transduct. Target. Ther.* **6**, 291 (2021).
75. Anisimova, A. S. et al. Multifaceted deregulation of gene expression and protein synthesis with age. *Proc. Natl Acad. Sci. USA* **117**, 15581–15590 (2020).
76. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33**, 245–254 (2003).
77. Li, W. et al. Human endogenous retrovirus-K contributes to motor neuron disease. *Sci. Transl. Med.* **7**, 307ra153 (2015).
78. Kassiotis, G. The immunological conundrum of endogenous retroelements. *Annu. Rev. Immunol.* **41**, 99–125 (2023).
This paper is a comprehensive overview of the diverse mechanisms by which endogenous retroviruses are involved in immunity.
79. Liu, X. et al. Resurrection of endogenous retroviruses during aging reinforces senescence. *Cell* **186**, 287–304.e26 (2023).
This study demonstrates that activation of HERV-like particles may contribute to the inflammatory processes that underly ageing-related pathogenesis.
80. Zhang, H. et al. Nuclear lamina erosion-induced resurrection of endogenous retroviruses underlies neuronal aging. *Cell Rep.* **42**, 112593 (2023).
81. Küry, P. et al. Human endogenous retroviruses in neurological diseases. *Trends Mol. Med.* **24**, 379–394 (2018).
82. Jansz, N. & Faulkner, G. J. Endogenous retroviruses in the origins and treatment of cancer. *Genome Biol.* **22**, 147 (2021).
83. Gan, L., Cookson, M. R., Petrucelli, L. & La Spada, A. R. Converging pathways in neurodegeneration, from genetics to mechanisms. *Nat. Neurosci.* **21**, 1300–1309 (2018).
84. Sun, W., Samimi, H., Gamez, M., Zare, H. & Frost, B. Pathogenic tau-induced piRNA depletion promotes neuronal death through transposable element dysregulation in neurodegenerative tauopathies. *Nat. Neurosci.* **21**, 1038–1048 (2018).
85. Ramirez, P. et al. Pathogenic tau accelerates aging-associated activation of transposable elements in the mouse central nervous system. *Prog. Neurobiol.* **208**, 102181 (2022).
86. Steiner, J. P. et al. Human endogenous retrovirus K envelope in spinal fluid of amyotrophic lateral sclerosis is toxic. *Ann. Neurol.* **92**, 545–561 (2022).
87. Garcia-Montojo, M. et al. Antibody response to HML-2 may be protective in amyotrophic lateral sclerosis. *Ann. Neurol.* **92**, 782–792 (2022).
88. Viola, M. V., Frazier, M., White, L., Brody, J. & Spiegelman, S. RNA-instructed DNA polymerase activity in a cytoplasmic particulate fraction in brains from Guamanian patients. *J. Exp. Med.* **142**, 483–494 (1975).
89. Tam, O. H. et al. Postmortem cortex samples identify distinct molecular subtypes of ALS: retrotransposon activation, oxidative stress, and activated glia. *Cell Rep.* **29**, 1164–1177.e5 (2019).
90. Padmanabhan Nair, V. et al. Activation of HERV-K(HML-2) disrupts cortical patterning and neuronal differentiation by increasing NTRK3. *Cell Stem Cell* **28**, 1566–1581.e8 (2021).
91. Johansson, E. M. et al. Human endogenous retroviral protein triggers deficit in glutamate synapse maturation and behaviors associated with psychosis. *Sci. Adv.* **6**, eabc0708 (2020).
92. Jönsson, M. E., Garza, R., Johansson, P. A. & Jakobsson, J. Transposable elements: a common feature of neurodevelopmental and neurodegenerative disorders. *Trends Genet.* **36**, 610–623 (2020).
93. Jönsson, M. E. et al. Activation of endogenous retroviruses during brain development causes an inflammatory response. *EMBO J.* **40**, e106423 (2021).
94. Krebs, A.-S. et al. Molecular architecture and conservation of an immature human endogenous retrovirus. *Nat. Commun.* **14**, 5149 (2023).
95. Douville, R., Liu, J., Rothstein, J. & Nath, A. Identification of active loci of a human endogenous retrovirus in neurons of patients with amyotrophic lateral sclerosis. *Ann. Neurol.* **69**, 141–151 (2011).
96. Garcia-Montojo, M., Li, W. & Nath, A. Technical considerations in detection of HERV-K in amyotrophic lateral sclerosis: selection of controls and the perils of qPCR. *Acta Neuropathol. Commun.* **7**, 101 (2019).
97. Mayer, J. et al. Transcriptional profiling of HERV-K(HML-2) in amyotrophic lateral sclerosis and potential implications for expression of HML-2 proteins. *Mol. Neurodegener.* **13**, 39 (2018).
98. Garson, J. A. et al. Quantitative analysis of human endogenous retrovirus-K transcripts in postmortem premotor cortex fails to confirm elevated expression of HERV-K RNA in amyotrophic lateral sclerosis. *Acta Neuropathol. Commun.* **7**, 45 (2019).
99. Dembny, P. et al. Human endogenous retrovirus HERV-K(HML-2) RNA causes neurodegeneration through Toll-like receptors. *JCI Insight* **5**, e131093 (2020).
100. Guo, C. et al. Tau activates transposable elements in Alzheimer's disease. *Cell Rep.* **23**, 2874–2880 (2018).
101. Dawson, T. et al. Locus specific endogenous retroviral expression associated with Alzheimer's disease. *Front. Aging Neurosci.* **15**, 1186470 (2023).
102. Pisetsky, D. S. Pathogenesis of autoimmune disease. *Nat. Rev. Nephrol.* **19**, 509–524 (2023).
103. Perron, H. et al. Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. *Proc. Natl Acad. Sci. USA* **94**, 7583–7588 (1997).
104. Garson, J. A., Tuke, P. W., Giraud, P., Paranhos-Baccala, G. & Perron, H. Detection of virion-associated MSRV-RNA in serum of patients with multiple sclerosis. *Lancet* **351**, 33 (1998).
105. Nakagawa, K., Brusic, V., McColl, G. & Harrison, L. C. Direct evidence for the expression of multiple endogenous retroviruses in the synovial compartment in rheumatoid arthritis. *Arthritis Rheum.* **40**, 627–638 (1997).
106. Ogasawara, H. et al. Quantitative analyses of messenger RNA of human endogenous retrovirus in patients with systemic lupus erythematosus. *J. Rheumatol.* **28**, 533–538 (2001).
107. Treger, R. S. et al. The lupus susceptibility locus Sgp3 encodes the suppressor of endogenous retrovirus expression SNERV. *Immunity* **50**, 334–347.e9 (2019).
108. Semsari, H. et al. Association of human endogenous retrovirus-W (HERV-W) copies with pemphigus vulgaris. *Curr. Mol. Med.* https://doi.org/10.2174/1566524023666230418114152 (2023).
109. Beck-Engeser, G. B., Eilat, D. & Wabl, M. An autoimmune disease prevented by anti-retroviral drugs. *Retrovirology* **8**, 91 (2011).
110. Rice, G. I. et al. Reverse-transcriptase inhibitors in the Aicardi–Goutières syndrome. *N. Engl. J. Med.* **379**, 2275–2277 (2018).
111. Hartung, H.-P. et al. Efficacy and safety of temelimab in multiple sclerosis: results of a randomized phase 2b and extension study. *Mult. Scler.* **28**, 429–440 (2022).
112. Rajurkar, M. et al. Reverse transcriptase inhibition disrupts repeat element life cycle in colorectal cancer. *Cancer Discov.* **12**, 1462–1481 (2022).
113. Ukadike, K. C. et al. Expression of L1 retrotransposons in granulocytes from patients with active systemic lupus erythematosus. *Mob. DNA* **14**, 5 (2023).
114. Bach, J.-F. The effect of infections on susceptibility to autoimmune and allergic diseases. *N. Engl. J. Med.* **347**, 911–920 (2002).
115. Wu, S., Zhu, W., Thompson, P. & Hannun, Y. A. Evaluating intrinsic and non-intrinsic cancer risk factors. *Nat. Commun.* **9**, 3490 (2018).
116. Reid Cahn, A., Bhardwaj, N. & Vabret, N. Dark genome, bright ideas: recent approaches to harness transposable elements in immunotherapies. *Cancer Cell* **40**, 792–797 (2022).
117. Babaian, A. & Mager, D. L. Endogenous retroviral promoter exaptation in human cancer. *Mob. DNA* **7**, 24 (2016).
118. Babaian, A. et al. Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma. *Oncogene* **35**, 2542–2546 (2016).
119. Lamprecht, B. et al. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat. Med.* **16**, 571–579 (2010).
120. Wiesner, T. et al. Alternative transcription initiation leads to expression of a novel ALK isoform in cancer. *Nature* **526**, 453–457 (2015).
121. Scarfò, I. et al. Identification of a new subclass of ALK-negative ALCL expressing aberrant levels of ERBB4 transcripts. *Blood* **127**, 221–232 (2016).
122. Lock, F. E. et al. Distinct isoform of FABP7 revealed by screening for retroelement-activated genes in diffuse large B-cell lymphoma. *Proc. Natl Acad. Sci. USA* **111**, E3534–E3543 (2014).
123. Liu, A. Y. & Abraham, B. A. Subtractive cloning of a hybrid human endogenous retrovirus and calbindin gene in the prostate cell line PC3. *Cancer Res.* **51**, 4107–4110 (1991).
124. Attig, J. et al. Human endogenous retrovirus onco-exaptation counters cancer cell senescence through calbindin. *J. Clin. Invest.* **133**, e164397 (2023).
125. Singh, B. et al. Locus specific human endogenous retroviruses reveal new lymphoma subtypes. Preprint at *bioRxiv* https://doi.org/10.1101/2023.06.08.544208 (2023).
126. Steiner, M. C. et al. Locus-specific characterization of human endogenous retrovirus expression in prostate, breast, and colon cancers. *Cancer Res.* **81**, 3449–3460 (2021).
127. Alcazer, V. et al. HERVs characterize normal and leukemia stem cells and represent a source of shared epitopes for cancer immunotherapy. *Am. J. Hematol.* **97**, 1200–1214 (2022).
128. Ng, K. W. et al. Antibodies against endogenous retroviruses promote lung cancer immunotherapy. *Nature* **616**, 563–573 (2023).
This study demonstrates the possibly tumourigenic effects of HERV protein expression and its immunotherapeutic potential.
129. Saini, S. K. et al. Human endogenous retroviruses form a reservoir of T cell targets in hematological cancers. *Nat. Commun.* **11**, 5660 (2020).
130. Bonaventura, P. et al. Identification of shared tumor epitopes from endogenous retroviruses inducing high-avidity cytotoxic T cells for cancer immunotherapy. *Sci. Adv.* **8**, eabj3671 (2022).
131. Burbage, M. et al. Epigenetically controlled tumor antigens derived from splice junctions between exons and transposable elements. *Sci. Immunol.* **8**, eabm6360 (2023).
132. Merlotti, A. et al. Noncanonical splicing junctions between exons and transposable elements represent a source of immunogenic recurrent neo-antigens in patients with lung cancer. *Sci. Immunol.* **8**, eabm6359 (2023).
This study provides a molecular basis for the selection of endogenous retroviruses that can form recurrent neoantigens in various malignancies.
133. Cuevas, M. V. R. et al. BamQuery: a proteogenomic tool to explore the immunopeptidome and prioritize actionable tumor antigens. *Genome Biol.* **24**, 188 (2023).
134. Ehx, G. et al. Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes. *Immunity* **54**, 737–752.e10 (2021).
135. Shah, N. M. et al. Pan-cancer analysis identifies tumor-specific antigens derived from transposable elements. *Nat. Genet.* **55**, 631–639 (2023).
136. Wang, E. & Aifantis, I. RNA splicing and cancer. *Trends Cancer* **6**, 631–644 (2020).
137. Ko, E.-J. et al. Expression profiles of human endogenous retrovirus (HERV)-K and HERV-R Env proteins in various cancers. *BMB Rep.* **54**, 368–373 (2021).
138. Shah, A. H. et al. Human endogenous retrovirus K contributes to a stem cell niche in glioblastoma. *J. Clin. Invest.* **133**, e167929 (2023).

139. Chiappinelli, K. B. et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* **162**, 974–986 (2015).
140. Marofi, F. et al. CAR T cells in solid tumors: challenges and opportunities. *Stem Cell Res. Ther.* **12**, 81 (2021).
141. Smith, C. C. et al. Alternative tumour-specific antigens. *Nat. Rev. Cancer* **19**, 465–478 (2019).
142. Wang-Johanning, F. et al. Immunotherapeutic potential of anti-human endogenous retrovirus-K envelope protein antibodies in targeting breast tumors. *J. Natl Cancer Inst.* **104**, 189–210 (2012).
143. Zhou, F. et al. Chimeric antigen receptor T cells targeting HERV-K inhibit breast cancer and its metastasis through downregulation of Ras. *Oncoimmunology* **4**, e1047582 (2015).

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41580-023-00674-z.

Peer review information *Nature Reviews Molecular Cell Biology* thanks Stéphane Depil, Paola Bonaventura and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.