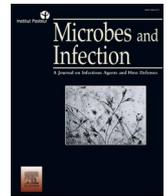


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Microbes and Infection

journal homepage: www.elsevier.com/locate/micinf

HERV-W ENV transcription in B cells predicting symptomatic COVID-19 and risk for long COVID can express a full-length protein despite stop codon in mRNA from chromosome X via a ribosome readthrough

Joanna Brunel ^a, Julien Paganini ^b, Melissa Galloux ^b, Benjamin Charvet ^a, Hervé Perron ^{a,*}

^a GeNeuro Innovation, 60A Avenue Rockefeller, 69008, Lyon, France

^b Xegen, 15 Rue Dominique Piazza, 13420, Gemenos, France

ARTICLE INFO

Keywords:

COVID-19
Long COVID
Post-acute sequelae of COVID-19
PASC
Post-COVID
Endogenous retroviruses
HERV
HERV-W
HERV-W ENV
RNAseq
Atypical memory B cell
Chromosome X
Stop codon
Inactivating mutation
Ribosome readthrough

ABSTRACT

The human genome comprises 8 % of endogenous retroviruses (HERVs). Though HERVs contribute to physiological functions, copies retained pathogenic potential. The HERV-W ENV protein was shown expressed in patients with worse COVID-19 symptoms and post-COVID syndrome. A significant detection of the mRNA encoding HERV-W ENV from patients with COVID-19 in B cells from RNAseq reads obtained from peripheral blood mononuclear cells. This data stratified with increased COVID-19 symptoms or with post-acute sequelae of COVID-19 (long COVID) after 3 months. The HERV-W ENV-U3R RNA was confirmed to display the best alignment with chromosome X ERVWE2 locus. However, a stop codon precluding its translation was re-addressed after recent understandings of ribosome readthrough mechanisms. Experimental results evidenced that this HERV gene can effectively express a full-length protein in the presence of molecules allowing translation via a readthrough mechanism at the ribosome level. Results not only confirm HERV-W ENV RNA origin in these patients but show for the first time how a defective HERV copy can be translated into a complete protein when specific factors make it possible at the ribosome level. The present proof of concept now requires further studies to identify the factors involved in this newly understood mechanism, following SARS-CoV-2 exposure.

1. Introduction

The COVID-19 pandemic has created what could be described as a secondary pandemic, with the onset and persistence of many symptoms after the viral infection is resolved, affecting about 10–20 % of patients after COVID-19 [1–3]. This is now a major concern for millions of patients around the world, and many of them evolve toward chronicity [4]. This post-infectious syndrome has been referred to as post-acute sequelae of COVID-19 (PASC), post-COVID syndrome/condition or, more commonly now, long COVID (LC) [5]. The appearance of so many cases within a short delay after COVID-19, the frequency, and the important number of symptoms revealed the involvement of a viral infection in the appearance of a post-infectious syndrome much different from the fatigue of a convalescence phase [6,7].

Beyond, it questions the etiopathogenesis of chronic pathologies whose initial or underlying cause remains unknown and whose epidemiology tells us that they may be associated with infections contracted

before their onset [8–11]. However, unlike COVID-19, it seems that the delay between the activating role of an infectious agent and the appearance of major or detectable clinical symptoms can be much longer than for the onset of post-COVID symptoms. This probably explains the difficulty of linking certain infections with post-infectious pathologies observed later when the agent that induced their development in susceptible individuals (linked to genetics, comorbidity or exposure to certain environmental factors), is no longer or rarely present [12–15].

In this context, studies that we have conducted for many years, along with those of other groups, have shown that infectious agents can dysregulate the epigenetic control and, also, activate the abnormal expression of endogenous retroviral elements -HERVs- [16–51]. HERVs constitute 8 % of the human genome [52] and, altogether with transposable or retro-transposable (TE) elements they represent about a half of the human DNA, now referred to as the "dark genome" [53–55]. This must be compared to the small proportion (1–3%) of sequences

* Corresponding author.

E-mail address: hperron@geneuro.com (H. Perron).

<https://doi.org/10.1016/j.micinf.2024.105431>

Received 21 June 2024; Received in revised form 2 October 2024; Accepted 15 October 2024

Available online 16 October 2024

1286-4579/© 2024 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved, including those for text and data mining, AI training, and similar technologies.

encoding somatic and physiological proteins of the human body [56–58].

The high number of highly homologous copies of the different families of TEs and HERVs inserted into human DNA, as well as their copy-number variations in human individuals [23,59–63], create a complexity that often exceeds the possibilities of an analysis of precisely localized copies in one or more chromosomal sites. This nonetheless remains possible at the level of families involved in their activation by an infectious agent or in pathologies evolving over the long term. Furthermore, when a HERV protein is abnormally produced and causes pathogenic effects, its related HERV family can also be detected and characterized [18,20,21,64].

This is why, after it had been shown that the retroviral envelope protein of the HERV-W family (W-ENV, from the previously named LM7 or MSRV retrovirus after its discovery in multiple sclerosis) had pathogenic effects on immune and neurobiological functions [21,24,31,39,65,66], but also that HERV-W ENV genes could be abnormally activated by viruses or parasites inducing its long-lasting expression [25,35,38,40], the question of its deleterious expression was raised in the context of COVID-19. In the absence of prior neutralizing immunity, COVID-19 was characterized by an hyperactivation of innate immunity causing a "cytokine storm", in parallel with lymphocyte deficiency and impairment of adaptive immunity functions that are essential for infection control [67]. The possible role of the W-ENV protein, also called pHERV-W ENV for its role in the pathogenesis of several diseases [22,27,43,68], was thus envisaged. W-ENV is a potent agonist of the TLR4 receptor [39,69] and, therefore, of the pro-inflammatory activation of innate immunity. In addition, and although this is conditional on the prior activation of TLR4 signaling pathways, this protein showed superantigen-like activity on T cells, in which it had never been observed to be expressed [65].

In a first study carried out during the first wave of the COVID-19 pandemic, the transcriptional activation and HERV-W ENV protein expression was demonstrated from the onset of infection in peripheral blood lymphoid cells and, for the first time, in T cells [70]. With a retrospective follow-up of the patients studied, this study showed that a high percentage of lymphocytes expressing W-ENV predicted a severe evolution of COVID-19 with cytokine storm and a need for oxygen or even intubation in intensive care. A second study [17] subsequently showed that: (i) *in vitro* exposure of healthy control peripheral blood mononuclear cells (PBMC) cells to SARS-CoV-2 could induce transcriptional activation and then protein expression of W-ENV in about 20 % of PBMC from normal blood donors, (ii) in early cohorts collected during the PCR diagnosis of COVID-19 the soluble W-ENV protein was found in the same proportion (about 20 %) of sera from patients, but their further clinical evolution was not known (iii) that the proportion of patients with soluble W-ENV in the serum increased to 100 % in patients hospitalized for severe forms in an intensive care unit (ICU) during successive epidemic waves with different variants and that (iv) autopsy tissues of COVID-19 patients who died of severe forms showed high W-ENV expression in cells close to, but different from, those that expressed SARS-CoV-2 antigens; namely: alveolar macrophages and pulmonary subepithelial lymphoid infiltrates; endothelial cells of certain blood vessels in all analyzed organs and the only positive cells in cardiac and peri-cardiac tissues; cells associated with intra-vascular blood clots; lymphoid infiltrates of the epithelial tissues and villi of the digestive tract and, massively, in lymphoid structures associated with mucous membranes; finally, microglial cells of the cerebral parenchyma with a complete absence of SARS-CoV-2 antigen in the central nervous tissue whereas, in sections at the level of the cribriform plate separating the olfactory bulb from the nasal mucosa, SARS-CoV-2 proteins were strongly detected in neighboring mucosal cells as well as W-ENV in lymphoid infiltrates.

Subsequently, the discovery of a post-COVID (PASC) or long COVID (LC) syndrome raised the question of a possibly persistent W-ENV expression following its initial activation. Its known effects involving the

persistent activation of cerebral microglia and various immune dysfunctions, was therefore calling for further research in LC. Initial results showed that about one-third of patients with LC had a detectable W-ENV protein in their serum and, therefore, that the expression of this protein persisted at least in certain tissues with soluble W-ENV secretion found in peripheral blood [71–73], even if no longer in circulating lymphocyte cells as observed during the acute COVID-19 infection. It was therefore necessary to detect and confirm the abnormal activation of this expression and, if possible, from the early SARS-CoV-2 infection phase. This should anticipate W-ENV expression that has already been shown to be pejorative for the course and complications of COVID-19 itself [70] and to be involved in a significant proportion of patients who developed LC associating a persistent expression of this pathogenic W-ENV protein [72,73].

To firstly address this pathogenic expression, we were interested in studies carried out on the PRESCO cohort of COVID-19 patients whose blood samples were collected early during SARS-CoV-2 infection and who were clinically followed up to the post-infectious phase with or without an LC diagnosis at this stage [74]. While immunological and multi-omics analyses showed different profiles without effective specifications for targeted therapeutic approaches [75,76], the clinical follow-up of a large subgroup of patients made it possible to identify three clusters with three divergent evolution of symptoms over a follow-up of several months [74]: a group (cluster 1) whose symptoms regressed, with spontaneous improvement, a group (cluster 2) whose symptoms persisted but remained stable in number and intensity and a group (cluster 3) whose symptoms worsened with the appearance of new symptoms not observed during the previous visits. This favorable, stable or pejorative evolution has grouped together symptoms related to complications in the infectious phase, but also related to their number and severity in patients with long COVID.

In the present study, we first analyzed the sequences extracted from samples with available RNAseq data of patients who had a clinical follow-up up to the last post-infection visit (visit 5), to confirm the analyzable coverage of the sequences of interest and to compare their expression levels in all the blood cell phenotypes studied. This was compared between patients classified in the three clinical clusters as well as in the patients diagnosed as LC (PASC) or non-LC (no-PASC) at the end of this follow-up. The pilot study on a few cases led us to analyze the large amount of data available from RNA sequencing during visit 1 and 2 (early infection) of all patients followed clinically and classified in cluster 3 of pejorative course in visit 5, versus cluster 1 of favorable spontaneous course. We also analyzed the complications (reported symptoms) corresponding to the COVID-19 infection as considered within clinical clusters, separately from the post-infectious status with or without LC.

Then having confirmed the mRNA sequence associated with this pathogenic expression of HERV-W ENV, we re-addressed the question of the chromosomal copy at the origin of this RNA transcription. As previously suggested, the best homology over the entire sequence was supposed to be located on chromosome Xq22.3, but the corresponding open reading frame (orf) was confirmed to have a premature stop codon that could only produce an N-terminally truncated protein [77], which was not compatible with the molecular weight and biochemical characteristics of the HERV-W ENV protein analyzed from MS brain lesions [20] and the labelling with antibodies targeting distant epitopes of its amino-acid sequence [34]. This locus, now named ERVWE2, was also confirmed to retain this stop codon in all healthy and MS individuals in a dedicated study on human DNA [78]. This led to debate on hypotheses consisting in a transcriptional interference between copies from multiple chromosomes, in the presence of non-ubiquitous copy(ies) in genetically susceptible individuals, or in the induction of somatic DNA rearrangements allowing the complete HERV-W protein(s) translation [79,80].

In this study, we have shown for the first time that the RNA corresponding to this ERVWE2 sequence can be translated into a full-length HERV-W ENV protein as described in pathogenic conditions, provided

that specific helper molecules interact with the ribosomal machinery to allow a readthrough translation replacing the stop codon by a tryptophan. How one HERV-W copy may encode this pathogenic ENV protein was experimentally shown. This new understanding of an unusual human gene transcription is analyzed from the structure of the RBM41 gene intron and exon across which ERVWE2 copy is inserted.

2. Materials and methods

To conduct a study on the PRESCO cohort, we retrieved RNA sequencing data from the blood mononuclear cell subpopulations (PBMCs) of patients that were sorted and sequenced with the Verily's Immune Profiler platform [76]. Our analyses focused on all the sequences of the HERV-W genome, including those of the gene encoding the W-ENV protein (*ENV* according to the nomenclature of human genes, or *env*, according to the virological nomenclature), with the expected structure of its mRNA which includes the U3R part of the flanking repeat region (LTR) where the polyadenylation signal is located (Fig. 1).

2.1. RNAseq data and analyses

2.1.1 Presco project data: The study used the RNA sequencing data of the Presco Project (clinicaltrials.gov NCT04388813) on samples analyzed with Verily's Immune Profiler platform, which included RNA sequencing of samples [74–76]. Briefly, this platform begins with the isolation of 25 immune cell subsets (5 myeloid cell subsets, 7 B cell subsets, 10 T cell subsets, 2 Natural Killer cell subsets, and a bulk peripheral blood mononuclear cells sample) from a starting material of approximately 10 million cryo-preserved PBMCs per individual. RNA-seq is performed for each of the 25 subsets. RNA-seq was performed using a SMART-Seq2-based procedure optimized for use on hundreds of cells.

We decided to analyze two subsets of data according to the associated meta information. The composition of datasets is provided below.

2.1.1. RNAseq data from the PRESCO cohort

Data are made available upon requests sent at immuneprofiler@verily.com.

2.1.2. Reference sequence dataset

A reference data of HERV-W sequences of interest was built at fasta format. Sequences were retrieved from NCBI. A fasta file was created for each and a global fasta file including all sequence was also created. List of sequence, details and accession are provided in Supplementary Files.

2.1.3. Reads extraction

Reads extraction from the PRESCO cohort [76] was done in two steps. Firstly, a non-stringent analysis was done to retrieve all potentials reads of interest. This analysis was performed using bowtie2 software [81]. This step was done using a local mode alignment. The fasta file containing all HERV-W sequences was used as reference. An example of command line used is provided in Supplementary files. The second step retrieved reads associated with each individual reference, using a more stringent analysis with end-to-end alignment. The same bowtie2 software was used but with an end-to-end alignment mode and, as reference, a fasta file with the sequence of interest only. Input reads used for this analysis correspond to the reads obtained during the first step.

2.1.4. Read quantification

For each file of reads obtained after extraction analysis, we used the fastqc software [82,83] and the multiQC software [84] to look at the reads features and to obtain the number of reads per file. We automatically retrieved this information using a simple algorithm developed internally to summarize all results in a table as input data for statistical

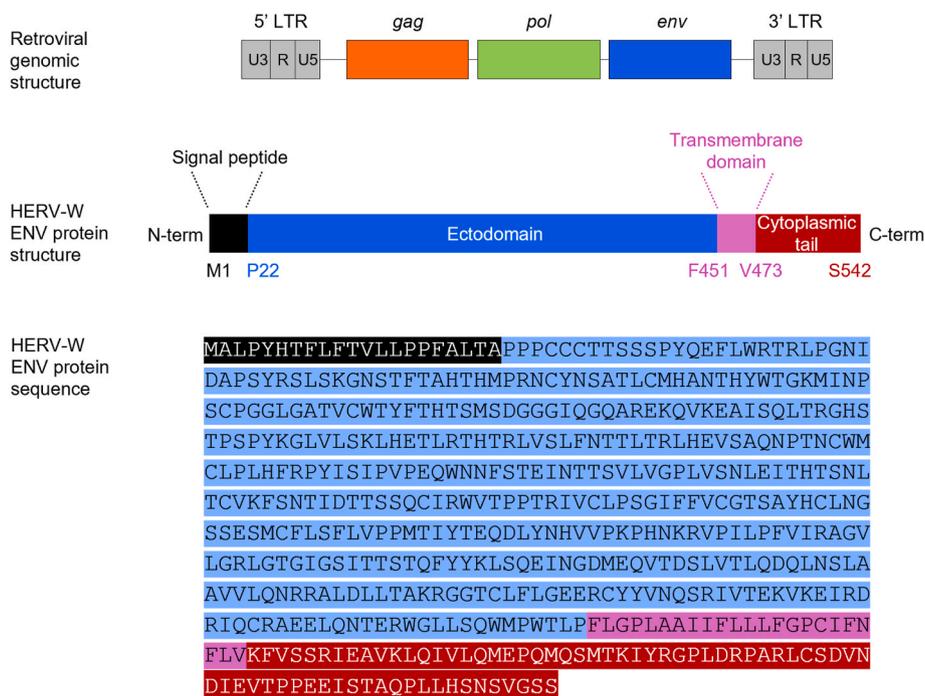


Fig. 1. HERV-W Nucleotide structure and envelope protein.

Top: The nucleotide structure of a retroviral genomic RNA obtained from HERV-W (previously named MSRV) original sequences isolated from multiple sclerosis [68] comprises long terminal repeats (LTR; RU5 region in 5' and U3R region with polyadenylation signal for mRNAs in 3') flanking the sequences of the *gag* gene-encoded proteins (Matrix, capsid and nucleocapsid), the *pol* gene-encoded enzymes (protease, reverse-transcriptase and integrase) followed by the *ENV* gene-encoded envelope protein. Middle: The different domains of the envelope protein are represented: signal peptide (propeptide in black), ectodomain in blue, the transmembrane domain encompassing the membrane-anchoring sequences in pink and the cytoplasmic tail in dark red. Bottom: The amino acid sequence of HERV-W ENV protein (GenBank: AAK18189.1).

analyses.

2.2. Statistical analyses

2.2.1. PASC and no-PASC comparison for all samples

For each gene, the frequency of reads was expressed as the number of reads for a specific HERV-W gene per one million of total RNAseq reads of the sorted cells [76] from each sample. To investigate the difference in read frequency, a statistical analysis was conducted using a two-tailed independent samples *t*-test to compare mean read frequency between PASC and no-PASC patients for each gene and each cell phenotype. Analysis was done using R version 4.2.2 (A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>) and plots were generated using ggplot2 version 3.4.4 [85] and ggpvr version 0.6.0 packages [86].

2.2.2. Difference in ratio of read frequency for ENV gene in aMBc cells between clusters

The ratios of ENV reads frequency between aMBc cells and the mean frequency of all other PBMC phenotypes [75] were computed for each sample from patients previously identified to belong to a clinical cluster with favorable (cluster 1) or pejorative (cluster 3) symptoms evolution after COVID-19, as well as with or without PASC diagnosis [74]. Patients classified as “stable” (Cluster2) were not taken into account, since representing an intermediate group with potential heterogeneity in its longer-term evolution, whereas clusters 1 and 3 were already identified with significant variations in their clinical parameters at the time of the last sampling. Then, all possible ratios were used as a threshold to define two groups: a high ratio group and a low ratio group. For each threshold, a Fisher’s exact test was conducted to determine the odds ratio (OR) of belonging to either cluster 3 or cluster 1, based on whether the participant fell into the high-ratio or low-ratio group. Fisher’s exact test was chosen due to the small sample size and the categorical nature of the data per sorted PBMC phenotype, ensuring accurate analysis of group differences.

The threshold giving the highest OR with p-value lower than 0.05 was selected as a good candidate to discriminate cluster 3 and cluster 1 patients in all analyses. The positive likelihood ratio (LR+) was also computed using this threshold. The positive likelihood ratio represents the likelihood of having an HERV-W ENV reads ratio in aMBc versus all other phenotypes above the threshold, therefore corresponding the high-ratio group for patients in cluster 3 compared to the likelihood of being in the high-ratio group for patients in cluster 1. The analysis was done using a custom R script.

2.3. Cells for transfection experiments

Human embryonic kidney (HEK) 293T cells (ATCC, CRL-3216) were grown in DMEM (Gibco, 41965-039) supplemented with 10 % of heat inactivated (30 min at 56 °C) fetal bovine serum (FBS) (ATCC, 30–2020) and 2 mM L-Glutamine (ThermoFisher Scientific, 25030024) in 5 % CO₂ incubator at 37 °C and were tested negative for mycoplasma spp.

2.3.1. Transfection

293T cells were cultured in 6-well plates (2.5.10⁵ cells/well) or in 8-well Nunc® Lab-Tek Chamber slide system (Merck, C7182) (5.10³ cells/well) precoated with Matrigel® (Corning, 354277). 24 h after seeding, cells were transfected with 3 µg (6-well plates) or 0.3 µg (Lab-Tek chamber) of plasmid using Lipofectamine 2000 transfection reagent (Invitrogen, 11668-019). Three plasmids were transfected: pCMV-HERV-W ENV encompassing the complete ORF of pHERV-W ENV (HERV-W ENV clone, 542 amino acids GenBank no. AF331500.1) followed by U3R sequence, pCMV-HERV-W ENV-39STOP containing a single nucleotidic substitution introducing a STOP codon instead of the Tryptophan at amino acid position 39 (GeNeuro) or pCMV-Empty vector which shares

the same backbone as the two previous plasmids but does not encode any protein. Plasmids and inserts are described in [Supplementary Fig. 1](#).

2.3.2. Treatment with nonsense mutations readthrough activity molecules

2 h after transfection, 2,6-Diaminopurine (Merck, 247847-1G) previously resuspended at 50 mM in DMSO, G418 disulfate salt previously resuspended at 144 mM in sterile water (Merck, A1720) or Clitocine (CliniSciences, HY-118341, in DMSO) were added to the culture medium.

DMSO SHAM treatment condition is determined by the volume used for the highest dose of 2,6-Diaminopurine or Clitocine used in the experiment presented.

2.3.3. Immunofluorescence microscopy (IF)

48 h post transfection, cells were washed once with 200 µL of Phosphate-Buffered Saline 1X (prepared from PBS 10X pH 7.4, Gibco, 70011-044) and fixed in 4 % paraformaldehyde solution (Alfa Aesar, J61899) for 5 min. Cells were further washed three times with 200 µL of PBS 1X and permeabilized with 200 µL of PBS 1X supplemented with 0.2 % Tween 20 (Merck, P7949) for 15 min. After 30 min of incubation in 200 µL of blocking solution made with PBS 1X, 0.2 % Tween 20 and 2.5 % horse serum (ATCC, 30–2040), cells were incubated with 50 µg/mL of anti-HERV-W ENV (GeNeuro, GN_mAb_Env02, murine antibody) diluted in blocking solution during 1 h. After 3 washes with 200 µL of PBS 1X, cells were incubated for 1 h with 1 µg/mL Alexa Fluor 488 goat anti-mouse IgG antibody (Invitrogen, A11029) and 0.1 µg/mL DAPI (Invitrogen, D3571) diluted in blocking solution. After 3 washes with 200 µL of PBS 1X, cells were mounted using Fluoromount-G mounting medium (Southern Biotech, 0100-01). Pictures were acquired on NIKON Eclipse TS2R microscope and analyzed on ImageJ software.

2.3.4. Protein extraction

48 h post transfection, cells were harvested and lysed in RIPA 1X buffer (prepared from RIPA 10X, Millipore, 20–188) supplemented with 1 % Fos-Choline-16 (Anatrace, F316S 1 GM) and protease inhibitor cocktail (Roche, 04693132001). Lysates were incubated for 2 h at 25 °C with gentle agitation (120 rpm). After 10 min of centrifugation at 10,000 g and 4 °C, supernatants were collected. Total protein amount was evaluated using Pierce™ 660 nm Protein Assay Reagent (Thermo Scientific, 22660).

2.3.5. Automated capillary western blot detection of HERV-W ENV

HERV-W ENV antigen detection was analyzed on the Wes device using Simple Western technology an automated capillary-based size sorting and immunolabeling system (ProteinSimple™). All procedures were performed with manufacturer’s reagents according to their manual. Briefly, 1500 µg/mL of protein lysate was mixed with fluorescent master mix and heated at 95 °C for 5 min. The samples, blocking reagent, wash buffer, 20 µg/mL of primary antibody anti-HERV-W ENV (GeNeuro, GN_mAb_Env02, murine antibody), secondary anti-mouse HRP-coupled antibody (ProteinSimple, 042–205) and chemiluminescent substrate were dispensed into microplate. Protein samples were loaded into individual capillaries on a 25 capillary cartridge (12–230 kDa or 66–440 kDa separation matrix) provided by the manufacturer. Protein separation and immunodetection was performed automatically on individual capillaries using default settings. The global signal (AUC) containing the glycosylated monomer HERV-W ENV (electrophoregram peak about 120 kDa) and the hexamer HERV-W ENV (electrophoregram peak about 440 kDa) were both measured using Compass software for Wes device (ProteinSimple/Biotechne).

2.3.6. Quantitative RT-PCR (RT-qPCR)

48 h post transfection, cells were harvested and total RNA was extracted using NucleoSpin RNA Mini Kit (Macherey Nagel, 740955) according to the manufacturer’s protocol.

200 ng of DNase-treated RNA were reverse-transcribed into cDNA

using iScript cDNA Synthesis Kit (Bio-Rad, 1708891) according to the manufacturer's protocol. An amount of 5 ng of initial RNA in RT reaction has been used to quantitatively evaluate the transcriptional level of HERV-W *ENV* gene by RT-qPCR. The assays were performed in a StepOnePlus instrument (Applied Biosystems) using Platinum SYBR Green (Invitrogen, 11744-500). The housekeeping Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used to normalize the results.

The RT-qPCR was performed using following primers: HERV-W *ENV* forward primer "fwd" [5'-GTATGTCTGATGGGGTGGAG-3'] and reverse primer "rev" [5'-CTAGTCCTTTGTAGGGCTAGAG-3']; GAPDH "fwd" [5'-CACCCACTCCTCCACCTTTGAC-3'] and "rev" [5'-GTCCAC-CACCCTGTTGCTGTAG-3'].

Each experiment was completed with a melting curve analysis to confirm the specificity of amplification and the lack of any non-specific product and primer dimer. The conditions of amplification generated by

the software of StepOnePlus platform are the following: holding stage for 10 min at 95 °C, 40-cycle stage with 15 s at 95 °C followed by 1 min at 60 °C and melt curve stage with 15 s at 95 °C, 1 min at 60 °C and increase by increments of 0.3 °C until 95 °C is reached for 15 s. Quantification was performed using the threshold cycle (Ct) comparative method: the relative expression was calculated as follow: $2^{-[\Delta Ct(\text{sample}) - \Delta Ct(\text{calibrator})]}$, where $\Delta Ct(\text{sample}) = [Ct(\text{HERV-W } ENV) - Ct(\text{GAPDH})]$ and the $\Delta Ct(\text{calibrator})$ is the mean of ΔCt of "Empty vector" transfected cells for each treatment condition.

3. Results

a. Expression of HERV-W *ENV-U3R* mRNA in PACS patients

Following previous studies on the PRESCO cohort [74], a first

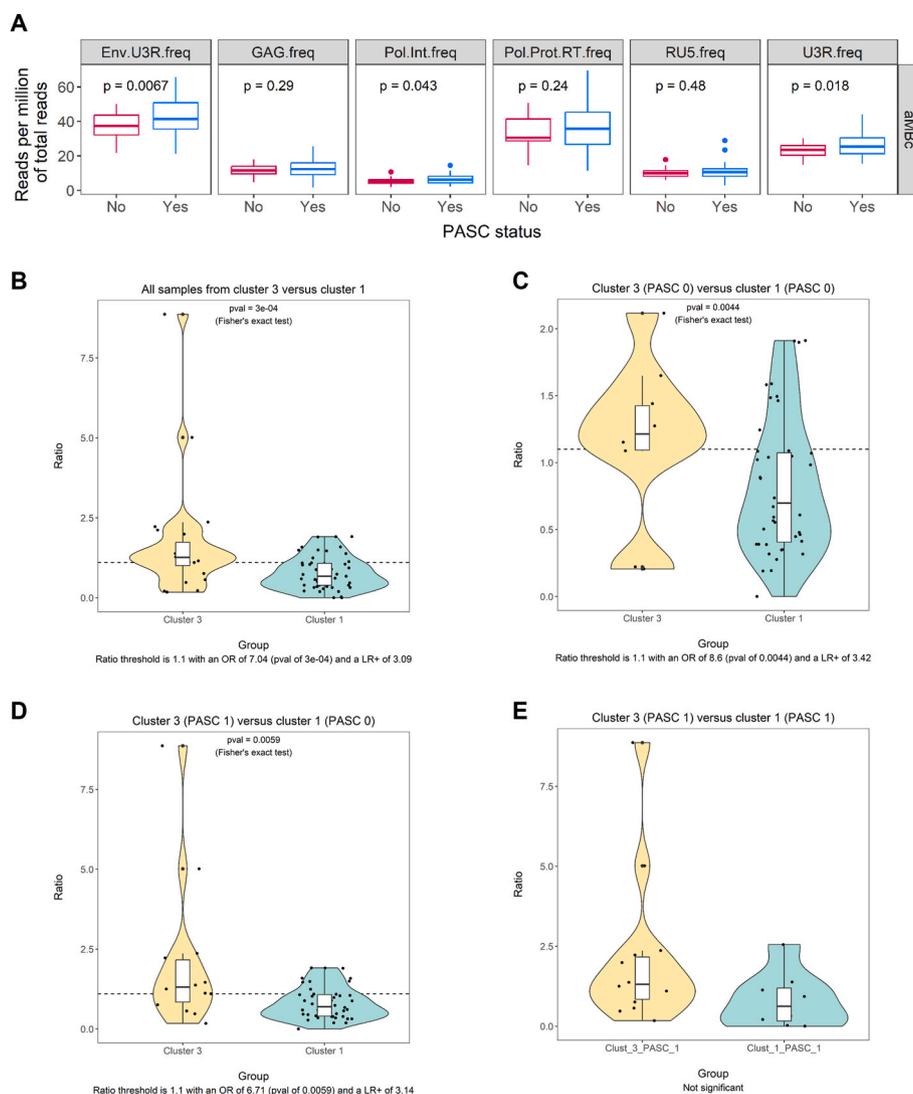


Fig. 2. Atypical memory B-cells RNAseq analysis from early COVID-19 samples. HERV-W genes reads and comparison versus all other PBMC subtypes for of *ENV-U3R* reads in patients from post-infection groups 1 and 3, later diagnosed with PASC and no-PASC.

A: Number of reads mapping to different genes or regions of HERV-W sequence in atypical memory B-cell PBMC phenotype (aMBC) from patients with or without Post-acute COVID sequelae. The RNAseq reads of aMBC from earliest visit 1 at COVID-19 diagnosis to the last visit 5 post-infection at PASC diagnosis (CDC criteria) were pooled per patient and compared between those diagnosed with PASC (Long COVID) versus no PASC. Reads were mapped to HERV-W sequences (Cf. Fig. 1). Env-U3R (expected mRNA encoding the envelope protein), gag gene (GAG), pol gene region encoding the integrase (pol-int.) or the protease and reverse-transcriptase region (pol-Prot.RT), 5' LTR (RU5) and 3' LTR (U3R) sequences. B-E: Ratio of reads mapping Env-U3R between aMBC cells versus all other phenotypes, in samples from patients with early SARS-CoV-2 infection: 2B. comparison of samples from cluster 3 versus all those from cluster 1. 2C comparison of samples from patients without PASC, from cluster 3 versus patients from cluster 1. 2D comparison of samples from patients with PASC in cluster 3 versus those with no PASC in cluster 1 taken as controls without PASC. 2E comparison of all samples from patients later diagnosed with PASC according to CDC criteria, from cluster 3 versus cluster 1.

analysis was performed on the samples from which RNA seq data were available over the successive visits for clinical follow-up, from visit 1 at early COVID-19 diagnosis to Visit 5 at 3 months post-infection.

Reads from all PBMC phenotypes aligning to HERV-W genomic sequences (HERV-W reference sequences in Supplementary information) were extracted and categorized per gene according to the protocol illustrated in Supplementary Fig. 2.

The number of reads were first analyzed per million of total RNA seq reads pooled from all PBMC phenotypes. As illustrated in Fig. 2A and detailed in Supplementary Fig. 3, the number of reads mapping to different genes or regions of HERV-W sequence varied with the PBMC phenotype and the post-infectious diagnosis of Post-acute COVID sequelae (PASC/long COVID). It clearly appeared that, when comparing samples from patients with PASC versus those with no PASC, the most elevated number of reads mapped to the ENV-U3R region in atypical memory B cells (aMbc). Reads showed a good coverage of the HERV-W ENV gene followed by the U3R sequence with highly homologous reads (end-to-end alignment), which corresponds to the expected mRNA since the poly-adenylation signal resides in the 3' U3 sequences. No such level and significant difference was seen for the other HERV-W genes (gag, pol) or 5'LTR/RU5 sequences, though occasional and isolated differences between PASC and no-PASC samples were found in certain PBMC phenotypes for HERV-W pol, gag and RU5 sequences. These were associated with significantly elevated ENV-U3R reads in unswitched naïve B cells (BnUS) and central memory CD4 T cells (T4Cm), but with lower significance or relatively lower reads frequency (Cf. Supplementary Fig. 3). Thus, HERV-W sequences with dominant ENV-encoding RNAseq reads appeared to be differentially expressed in these three PBMC phenotypes versus others when comparing PASC and no-PASC patients from visit 1 and 2 during the early SARS-CoV-2 infection until visit 5, after COVID-19 infectious phase.

Globally, the RNAseq reads from pooled PBMC phenotypes showed a good coverage of ENV-U3R sequences, though relatively more abundant reads aligned to ENV 3' end. This was even increased for the U3R sequence, which may result from the repeated sequences in this region (from retroviral Long-terminal repeat-LTR) but may also indicate variable thermodynamics in reads amplification between regions of the ENV-U3R sequence.

This prompted us to further seek potentially significant differences from a higher number of samples with RNAseq reads analyzed at visit 1 and 2 ($n = 97$), while only few samples were made available with data from visit 5 ($n = 6$). Among them, the PRESCO cohort comprised 80 patients with available RNAseq data from blood samples drawn during early COVID-19 (Visits 1 and 2) along with clinical follow up allowing both PASC diagnosis and retrospective analysis of the number and severity of COVID-19 symptoms. Patients' clinical evolution and symptoms until last visit had made it possible to identify three clusters of patients irrespective of PASC diagnosis [74]: one of patients who spontaneously improved (cluster 1), one of patients with persisting but stable symptoms (cluster 2) and one of patients with an increased number and worsening symptoms (cluster 3). As expected, most patients diagnosed with PASC were diagnosed in cluster 3, representing most cases with numerous symptoms.

The analysis of the relative frequency of reads in RNA from FACS-sorted PBMC subtypes from all patients at visit 1 and 2 showed that the most frequent HERV-W reads were still mapped to ENV-U3R sequences. They were then significantly detected in PASC or patients from cluster 3, within aMbc phenotype, which could be differentiated from all other sorted phenotypes. BnUs B and T4cm showed no difference, unlike in previous analyses in which they had rather low frequency of reads but were significantly different in PASC versus no-PASC when considering reads from post-infection visit 5. The reads aligning with ENV sequence only or ENV-U3R sequences also revealed comparable in aMbc. We therefore focused on the ratio of ENV-U3R reads in aMbc versus the pooled reads from all other PBMC phenotypes. This represented a ratio between an early ENV-expressing PBMC sub-otype with potential HERV-

W transcription, versus all others. This ratio also provided intra-sample normalization for comparison between samples from all patients. In the following analyses, samples from cluster 2 were not included, since representing an intermediate cohort that may require longer term follow-up to differentiate clinically defined evolution toward improved or worsened symptoms and their number.

In Fig. 2B, the comparison of ENV-U3R reads ratio between all samples from cluster 3 versus those from cluster 1 showed a quite complete differentiation between the clusters: OR of 7.04 and a LR+ of 3.09, with a threshold at 1.1. The statistical difference of ENV-U3R reads between the two clusters was highly significant ($p = 3e-04$; Fischer exact test).

Fig. 2C compared patients from clusters 3 and 1, all of which were not later diagnosed with PASC and showed a very good discrimination between clusters. OR of 8.6 ($p = 0.0044$; Fischer exact test) and a LR+ of 3.42 with a threshold at 1.1, were found. Given that these samples were obtained during initial COVID-19, this difference should be due to the diverging evolution of symptoms and disease severity of the SARS-CoV-2 infection period between cluster 1 and cluster 3. This is consistent with HERV-W envelope RNA and protein expression in peripheral blood lymphocytes already described to be predicting, and associated with, pejorative disease evolution in early and late infection phases of COVID-19 across pandemic waves with various SARS-CoV-2 variants [17,70].

Fig. 2D compared patients with PASC in cluster 3 versus those with no PASC in cluster 1, taken as controls, at least without diagnosed PASC at Visit 5 and without persisting major symptom from acute COVID-19. A very significant difference was confirmed to link PASC to the early detection of an elevated ratio of reads mapping to HERV-W ENV-U3R: OR of 6.71 ($p = 0.0059$; Fischer exact test) and a LR+ of 3.14, with a ratio threshold at 1.1. This positive detection corresponds to the RNA transcription over the threshold in aMbc from the early days of COVID-19 in patients later diagnosed with PASC, with very elevated values in few cases.

Fig. 2E similarly compared patients later diagnosed with PASC according to CDC criteria from cluster 3 versus PASC patients from cluster 1 and revealed no significant difference between them. This suggested a stronger association of ENV-U3R RNA expression with PASC than with the global clinical evolution considered to define the present clusters. A better discrimination provided by RNAseq data could thus be relevant when few PASC patients are found in the clinically defined cluster 1, though most of them were in cluster 3.

b. Identification of W ENV encoding region

Further to these results showing HERV-W ENV transcriptional activity with reads mapping specifically to HERV-W sequences cloned from retrovirus-like particles of MS cell cultures and body fluids (originally named MSRV-ENV) [65,68], the best possible alignment on human chromosomes still corresponded to the previously described ERVWE2 copy [78] that had been shown to encode a truncated protein when using appropriate vectors and conditions for an *in vitro* transcription [77].

As shown in Fig. 3A, the reference HERV-W ENV sequence cloned from MS (GenBank ID: AF331500.1) is best aligned to a region in chromosome Xq22.3 (chrX: 106295765–106297364), with 1600/1629 aligned nucleotides and 98,5 % identity. This region corresponds to *Homo sapiens* RNA binding motif protein 41 (RBM41), Genbank sequence ID NG_016385.2. HERV-W ENV sequence is shown to correspond to the reverse strand of the RNA binding motif protein (RBM41) gene, starting from the end of intron 6 and ending within the beginning of exon 7. The RBM41 locus and the neighboring genes are presented in Fig. 3B, whereas Fig. 3C presents the positions of the 6 introns and 7 exons of RBM41 gene, all of which are used to generate alternatively spliced RNAs. The aligned nucleotide sequences also show that the entire ENV-U3R mRNA sequence has 1936/2012 (96.22 %) identities from Chromosome X positions 107052123–107054134. Thus, in Fig. 4,

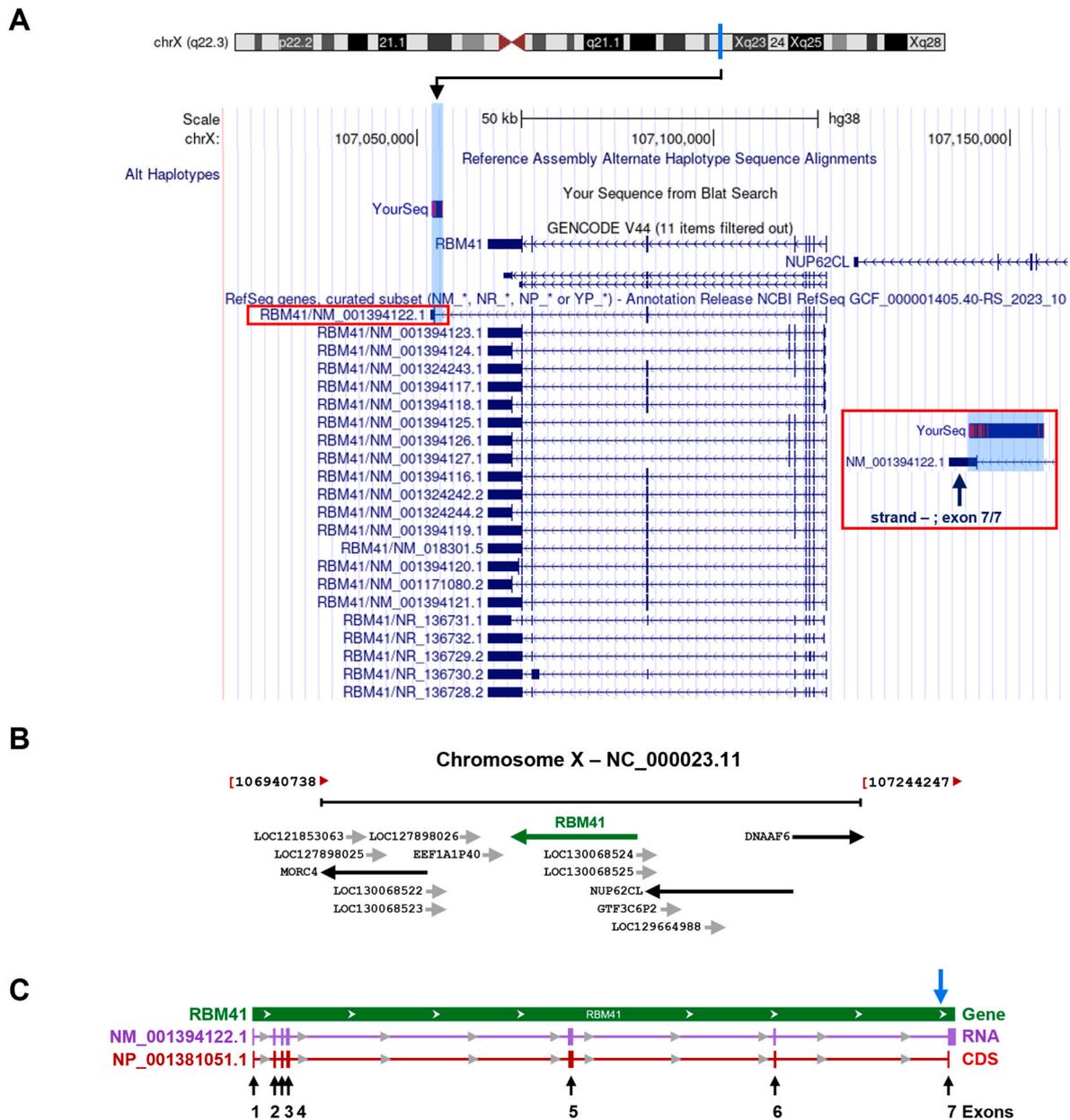


Fig. 3. BLAT search of HERV-W ENV sequence using UCSC Genome Browser on Human (GRCh38/hg38).

A. HERV-W ENV sequence is located (1600 nt aligned /1629 nt total ; 98,5 % identity) on the reverse strand of genomic chromosome X (chrX:107052926-107052997) on the region of RBM41 gene. The RNA reference RBM41/NM_001394122.1 boxed in red corresponds to the transcript variant 12 comprising exon 7 that contains the 3' end of ENV-U3R sequence that is associated with the mRNA structure found to be expressed in COVID-19 aMbc. The protein isoform 12 expressed by this spliced RNA is predicted to be part of U12-type spliceosome complex. (provided by Alliance of Genome Resources, Apr 2022). The red square on the right shows, with blue background, the genomic region encompassing the HERV-W ENV sequence in an antisense orf starting in intron 6 and ending in exon 7. **B.** Localization of RBM41 gene on chromosome X, adapted from NCBI: <https://www.ncbi.nlm.nih.gov/gene/55285>. **C.** Genomic region (blue arrow indicates HERV-W ENV sequence alignment location), transcript and product of RBM41. RNA: mRNA-RNA binding motif protein 41, transcript variant 12. Protein: RNA-binding protein 41 isoform 12. Location: complement (107,052,102...107,118,822). CDS length: 1,095 nt. Protein length: 364 aa.

the complete HERV-W ENV-U3R sequence is found in this locus and therefore can generate a transcript corresponding to the mRNA structure on which RNA seq reads of aMbc from COVID-19 patients who developed PASC or had pejorative evolution (cluster 3) were significantly mapped. Moreover, as highlighted in yellow, TATA boxes are present upstream of the HERV-W ENV orf within the RBM41 intron 6, but with an antisense orientation since presented in the 5' region of the ATG trinucleotide from which the translation of the HERV-W envelope can start. TATA and CAAT regulatory elements, with CAP site and Poly A signal in HERV-W (previously named MSRV) U3R sequences, along with strong promoter activity of MS-derived U3R clones, have also been

described [68]. The additional presence of TATA boxes in RBM41 intron 6, upstream from this HERV-W ENV locus is indicating that a transcription is made possible from this HERV gene copy, independently from RBM41 normal RNA expression. This can be initiated within the RBM41 intron 6 where the ATG and the major part of the 5' region of HERV-W ENV sequence is found, followed by the 3' region within the RBM41 exon 7 where the downstream U3R region comprising promoter and regulatory sequences is located.

The aligned proteins respectively encoded by the reference sequence on chromosome X, also named ERVWE2, and the HERV-W sequence, previously named MSRV, cloned and sequenced from an MS sample [41]

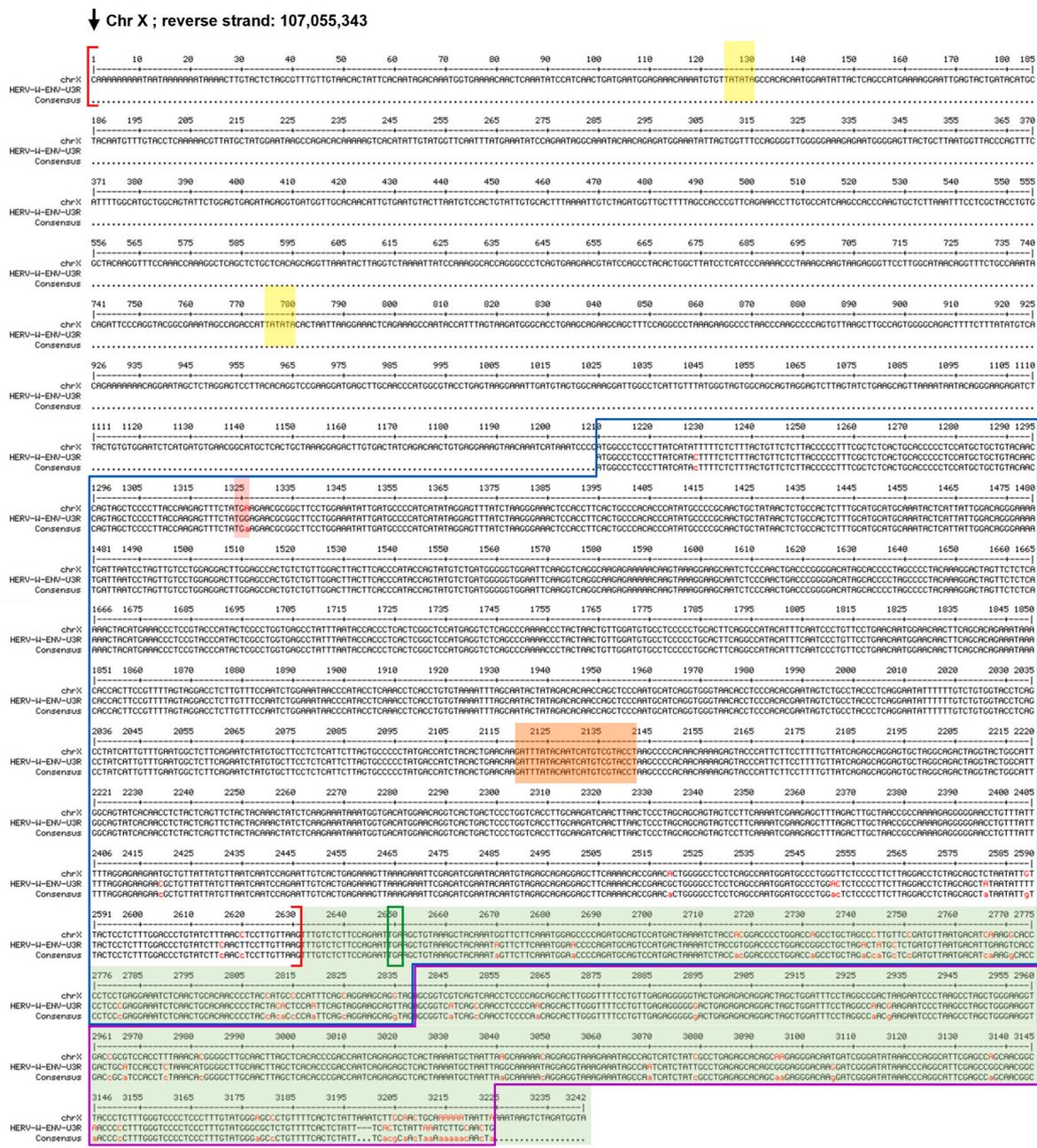


Fig. 4. Aligned nucleotide sequences of Chr X portion 107052102–107055343 (reverse strand) with HERV-W ENV-U3R RNA sequence encoding the pathogenic HERV-W-ENV protein.

-TATA box motifs within chromosome X intron are highlighted in yellow; regulatory sequences in U3R, i.e., TATA boxes, CAAT element, Cap site and Poly A signal have already been described⁶⁸. -HERV-W ENV encoding nucleotides are boxed in light blue; nucleotides encoding tryptophane in position 39 of the HERV-W envelope protein and corresponding to the stop codon in chromosome X, are highlighted in light red; nucleotides encoding the epitope specifically recognized by Env02 monoclonal antibody used for immunofluorescence microscopy examination and immunocapillary western blot are highlighted in light orange; U3R sequences are boxed in purple. -RBM41 intron 6 partial sequence are presented between red brackets in the aligned clone; RBM41 exon 7 sequence is highlighted in light green in the aligned clone ; the stop codon in the chromosome X corresponding to the terminal stop codon of RBM41orf for the isoform 12 is boxed in green.

are shown in [Supplementary Fig. 4](#). A stop codon is interrupting the orf from the ERVWE2 locus and the corresponding RNA might be translated into a truncated protein from an ATG trinucleotide encoding a methionine downstream the position of tryptophane 39 as shown *in vitro* [77]. This in-frame stop codon within the ERVWE2 locus precluded the expression of a full-length protein with the specific biochemical and biophysical properties characterized in native HERV-W ENV antigen extracted from, e.g., MS brain lesions [20]. However, we recently understood that such stop codons are used for the regulation of protein expression at the ribosome level in prokaryotes [87,88] and that some mechanisms can also promote ribosome readthrough in vertebrates [89–92]. Moreover, molecules were proposed for therapeutic strategies promoting such ribosome readthrough in human genetic diseases with stop codons impairing the production of physiological proteins [93–95]. Therefore, such ribosome readthrough might be possible for HERV genes in human cells. We then conceived experiments to find conditions that could provide the proof of concept for a full length HERV-W ENV protein expressed from this ERVWE2 sequence in human cells, despite an in-frame stop codon.

c. Ribosomal readthrough experiments.

To experimentally test this hypothesis, we generated a plasmid with the corresponding ERVWE2 stop codon in the HERV-W ENV orf. In parallel, we have used the HERV-W ENV plasmid encoding the complete envelope protein cloned from an MS sample and used to produce a recombinant protein (GenBank ID: AF331500.1) [41]. The constructs are schematically presented in [Supplementary Fig. 1](#). The sequences were all confirmed by plasmid sequencing. Of note, though the cleavage of the signal peptide is observed with the full-length HERV-W envelope from plasmid expression, the Furin cleavage site is inactivated by mutations found in the original clones from MS [65,68] and confirmed within the HERV-W ENV proteins extracted and analyzed from MS brain lesions [20]. The produced full-length envelope protein therefore no longer contains the propeptide, but still retains both surface (SU) and transmembrane (TM) units.

HEK cells were transfected with either plasmid construct and cultured in medium with different molecules shown to have a possible effect on ribosome readthrough of in-frame stop codons, further lysed and analyzed with Simple Western immunocapillary platform (Proteinsimple, CA, USA), as detailed in Materials and Methods. In [Fig. 5](#), results from transfections in the presence of three molecules already shown to display such an effect in humans, 2,6-Diaminopurine (DAP) [95], G418 [96] and clitocine [97] are presented in parallel with results of the transfection in normal medium (no treatment). DMSO at the same concentration as used to solubilize 2,6-DAP or clitocine, was also compared as a SHAM control. As can be seen with the electrophoregram from the Simple Western immunocapillary analysis, a peak of protein immunolabelled with a specific anti-HERV-W ENV mouse monoclonal antibody (GN_mAb_Env02, Geneuro SA, Switzerland) migrating at the expected molecular weight of the mature (without propeptide) glycosylated HERV-W ENV monomer, was seen in all conditions with the W ENV plasmid. The transfection with W ENV-39 STOP plasmid containing the stop codon showed no protein expression at all in normal medium, even under the form of a truncated protein with lower molecular weight as previously made possible *in vitro* [77]. This form would have been detected since the position of the recognized epitope is located within this truncated product (Cf. [Fig. 4](#) and [Supplementary Fig. 4](#)). No expression was either observed with W ENV-39 STOP plasmid in the presence of G418 or clitocine. However, in the presence of DAP, the expected peak of the mature and glycosylated complete HERV-W ENV monomer was observed with 100 μ M DAP added in the culture medium ([5D](#)). A dose-response experiment with DAP at increasing concentrations was then performed and, as seen in [5F](#), an effect ranging from 50 μ M to 150 μ M with an optimal concentration of 100 μ M DAP for the expression of HERV-W ENV protein was confirmed for the plasmid with

the stop codon, along with good reproducibility between experiment duplicates. To confirm results of the analysis of the denatured protein from transfected cells as previously performed, immunocytology microscopy analysis was performed with the same monoclonal antibody (GN_mAb_Env02) using 50 μ M concentration of DAP. In [5G](#), immunofluorescent staining was observed in all conditions with the W ENV plasmid, whereas the same positivity and pattern of expression with strong labelling at the membrane level was only observed in the presence of DAP with the W ENV-39STOP plasmid.

Because the HERV-W ENV antigen extracted from MS brain lesions had revealed to be present *in vivo* with an oligomer conformation (hexamer) characterizing a secreted soluble form of this pathogenic expression, which could be reproduced with the recombinant protein expressed from the present W ENV clone [20], we looked for this high-molecular weight form on appropriate capillaries. Of note, this hexameric form resisted strong denaturing conditions used for the simple western capillary analysis and other western-blotting formats. In [Fig. 6A](#), both HERV-W ENV monomer around 120 KDa and hexamer around 440 KDa could be detected in cells transfected with W ENV-39STOP plasmid and 75 μ M DAP, similarly to what was obtained with the W ENV plasmid in normal medium. A dose-response analysis showed that the monomer was similarly detected with W ENV at 20–75 μ M of DAP, but slightly increased compared to SHAM condition with DMSO. However, the decrease at 100 μ M could indicate that some cell toxicity was encountered at this concentration. In these conditions, the monomer was increasingly expressed from 20 to 75 μ M DAP with W ENV-39STOP plasmid. Though non-significant difference was seen at 100 μ M concentration, the mean + SD value of 2–8 independent capillaries from 2 to 4 independent experiments was reduced relatively to previous concentrations ([6B](#)). The same dose-response analysis at the level of the hexamer molecular weight also revealed an increasing production from 20 to 75 μ M DAP with W ENV-39STOP plasmid, whereas a potential cytotoxicity was confirmed with a clear decrease at 100 μ M ([6C](#)). Finally, because differences were confirmed to result from factors acting at the translational level, we performed a study of the transcriptional levels to compare the RNA expressed from the different conditions of transfection with both W ENV and W ENV-39STOP plasmids. No significant difference between the transcriptional levels of each plasmid was observed in all conditions, thereby confirming that differences observed at the protein level are not influenced by the transcription of either sequence in transfected cells.

4. Discussion

In a first part, this study completes the ones already performed on clinical, immunological and other multi-omics parameters from the PRESCO cohort as previously described [74–76], by focusing on RNAseq data mapping the HERV-W sequences. The rationale for this study in patients relied upon the facts that HERV-W was previously shown to encode an immune- and neuro-pathogenic envelope protein [24,98] whose expression in lymphoid cells and release as soluble antigen in peripheral blood (i) could be induced by SARS-CoV-2 [17] and (ii) correlated with COVID-19 severity and/or a subgroup of patients with long COVID (PASC) [72,73].

The results from this RNAseq analysis show that HERV-W RNA reads mapping to the envelope-encoding mRNA reference sequence, originally cloned from cultured cells or CSF of patients with multiple sclerosis (MS), were most significantly elevated in atypical memory B cells (aMBC) of early SARS-CoV-2 infection in patients who were later diagnosed with Long COVID (Cf. [Supplementary Fig. 5](#)), but also in those who experienced worsened COVID-19 and late symptoms.

Interestingly, B lymphocytes have long-been shown to be permissive for this pathogenic HERV-W expression in MS and other diseases [24,43,99–102]. In MS, the role and effects of HERV-W ENV in the central nervous system lesional process has now been deciphered and revealed to rely upon a lifelong-lasting expression in microglia oriented toward

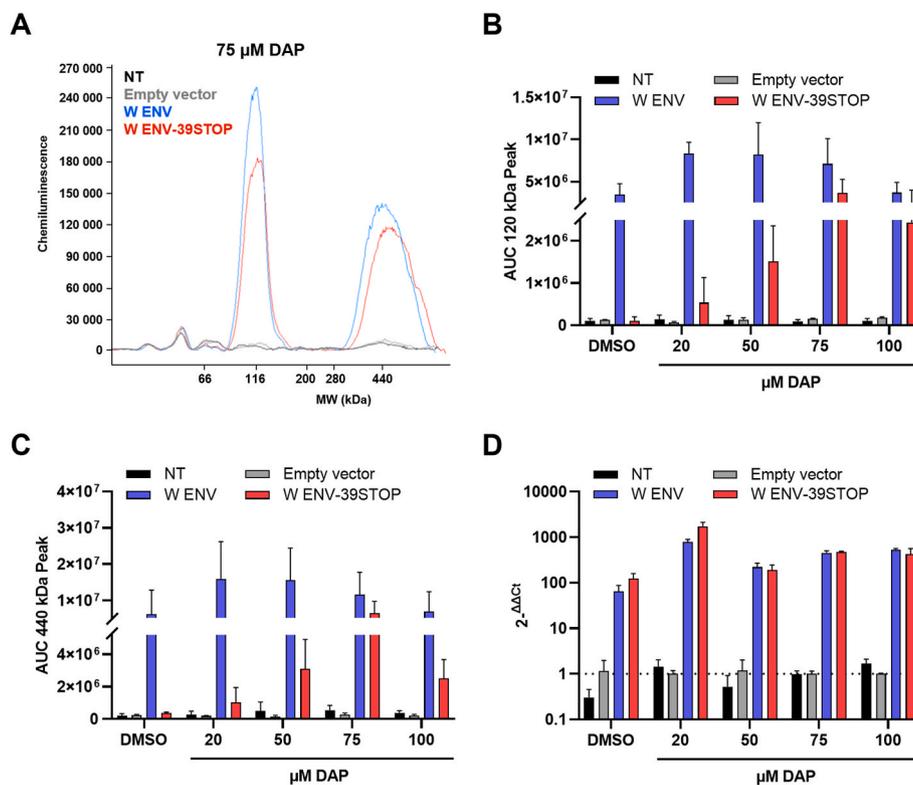


Fig. 6. HERV-W ENV protein induced by stop codon readthrough can still form hexamer.

Non-transfected 293T cells, 293T cells transfected with the empty vector or with the plasmid coding for HERV-W ENV or for HERV-W ENV-39STOP construct were treated with DMSO or DAP range from 20 to 100 μ M. (A-C) 48 hours post transfection, total proteins were extracted, soluble fractions were incubated with a second buffer to improve hexamer extraction and HERV-W ENV expression analyzed by Simple Western on 66-440 kDa size separation matrices. Results are presented as migration electrophoregram (A, example of 75 μ M DAP) where the 120 kDa peak corresponds to full length HERV-W ENV glycosylated monomer expression and the 440 kDa peak to the hexameric form of the protein or as quantitative interpretation based on calculation of the area under curve (AUC) of 120 kDa peak (B, mean+SD of 2 to 8 independent capillaries from 2 to 4 independent experiments) or 440 kDa peak (C, mean+SD of 2 to 8 independent capillaries from 2 to 4 independent experiments). (D) 48 hours post-transfection, cells were harvested and total RNA extracted. RT-qPCR was performed to determine the transcriptional level of HERV-W ENV gene for each transfection and DAP treatment condition. Results are presented as the mean+SD of $2^{-\Delta\Delta C_t}$ from three independent experiments.

immune-aggressive phenotype causing demyelination and neurodegeneration, but also an inhibition of remyelination by oligodendrocyte precursor cells in MS [22,32,103]. This has recently also been recapitulated within a mouse model transgenic for the HERV-W ENV-U3R human sequences [16]. Since HERV-W activation has also been shown to be induced via epigenetic dysregulation and transactivation by exogenous viruses such as EBV and HHV6 [25,102], observations made with SARS-CoV-2 now indicate a more global post-infectious phenomenon leading to long-lasting and chronic diseases mediated by such an HERV protein. Moreover, the role of B-cells is also known to be critical in MS [104,105]. It is therefore of the greatest interest to learn that B lymphocytes infected with Epstein-Barr virus (EBV) undergo differentiation into aMBC and that aMBC are pathogenically associated with numerous immune disorders [106]. Thus, HERV-W pathogenic expression induced in such B cells in “genetically susceptible” individuals, as seen with PBMC exposure to SARS-CoV-2 [17], is likely to be an early feature of related pathogenic post-infectious dynamics. EBV and SARS-CoV-2 infections appear to display common features linked to HERV-W activation and conditions of expression, though global post-infectious kinetics and even some molecular genetics mechanisms are expected to differ.

The potential diagnostic and therapeutic windows of opportunity are therefore that of the period when COVID-19 diagnosis is made in normal conditions of medical practice, which was also shown in an independent cohort and using a different approach consisting in detecting the HERV-W ENV protein in the serum of patients at the time of SARS-CoV-2 qRT-PCR diagnosis [17]. However, this previous study made on a large Spanish cohort and, independently, on additional French cases, did not

identify the early mRNA expression in a characterized PBMC subtype since dealing with the antigen released in serum and since cytofluorometric analyses, which also focused on protein expression, were made quite later in the course of COVID-19. A percentage of B-cells without further phenotyping still appeared to be positive in less severe patients but T-cells were seen to dominantly express at significant levels along with markers of exhaustion at this stage and/or in most severe patients.

Thus, the present study confirms that the window of opportunity is that of early COVID-19 and is made possible from the early diagnosis by these different methods. It also shows that the very nature and fate of atypical memory B-cells is not trivial and is meaningful when dominantly found to express HERV-W ENV-U3R mRNA. This early detection of specific RNA in aMBC is likely to provide a predictive biomarker of pejorative COVID-19, particularly in patients at risk or with undiagnosed comorbidities whatever the SARS-CoV-2 strains, vaccine or previous infection status. This is also consistent with peculiar findings on aMBC in COVID-19 showing the prominent expansion of aMBC in severe COVID-19 [107], suggested to represent one of the critical drivers in disease severity [108].

However, given the present public health burden caused by the post-COVID syndromes affecting millions of individuals worldwide and their continuous incidence despite the rising global immunity and the emergence of less pathogenic coronavirus strains, this predictive detection from early infection paves the way to therapeutic prevention with precision medicine targeting this pathogenic HERV-W expression. Such biomarker-based therapeutic prevention could be made possible using the Verily’s Immune Profiler platform [76] or using adapted protocols to

perform a similar quantification of the HERV-W ENV-encoding mRNA in aMBC from SARS-CoV-2 early infected patients. Alternatively and as previously shown, the detection of the soluble HERV-W ENV hexamer in serum with the Simple Western technology could also be used as a biomarker for the same purpose [17]. Consequently, the prevention of COVID-19 complications and, most of all, of post-COVID condition known as PASC or Long COVID remains a challenge that is likely to be addressed from early infection with a biomarker-based targeted strategy. Even if the persistent expression of W-ENV should not be relevant for all cases, it represents a major subgroup of patients with complex and highly heterogeneous post-COVID symptoms for which a biomarker-based precision medicine is made possible.

In a second part and, further to this early mRNA expression specifically observed in aMBC from COVID-19 patients, the present study re-addressed the question of the involved chromosomal HERV-W gene copy in the light of new knowledge on the regulation of mRNA at the ribosome level. SARS-CoV-2 was shown to induce HERV-W ENV protein expression in peripheral blood mononuclear cells (PBMC) from healthy blood donors maintained in normal cell-culture medium, with RT-qPCR detecting a peak of RNA transcription at 2h post-exposure and with detectable protein expression by immunocytology in macrophage and lymphocyte cells at 48h post-exposure. Apart from the fact that some (epi)genetic factors may explain why only about 20 % of healthy donor's PBMC did express HERV-W ENV when exposed to SARS-CoV-2 [17], the present results confirming a complete mRNA sequence transcription in clinically worsening COVID-19 patients implicitly raised the question of the origin of this protein expression.

The results of the present study show for the first time that an HERV protein encoded by a gene with a stop codon can be expressed in human cells in presence of a specific “helper” molecule. Moreover, it confirmed that the reference HERV-W ENV sequence on which RNAseq reads from COVID-19 patients mapped, best aligned to the previously mentioned ERVWE2 locus on chromosome X [77–79]. This remains the only known chromosomal HERV-W copy of HERV-W envelope gene from the present databases with this nearly perfect nucleotide sequence identity. Therefore, the previously considered ERVWE2 impossibility to encode an mRNA that can be translated into the full-length protein with characteristics identified from MS brain lesions [20,34], is now understood to be untrue. This, since the nucleotide sequence with its in-frame stop codon has been shown to yield the full-length HERV-W ENV protein forming hexamers as shown for the MS-related protein *in vivo* and *in vitro*. So, though an early triggering event with delayed clinical emergence in MS is likely to be provided by an EBV infection that might have caused somatic rearrangements or mutations over a long period of time [80,102,109] followed by “boosting” agents or by alternative triggering infections [25,38], the quite immediate induction of HERV-W ENV expression caused in lymphoid cells from healthy individuals by SARS-CoV-2 suggests a direct effect on an HERV-W copy present in human chromosomes without delayed effects of potentially induced DNA and/or RNA rearrangements. Thus, the present results showing that a regulation of HERV mRNA expression with in-frame stop codon(s) can be achieved at the ribosome translation level in human cells as reported in other species and human conditions, elucidates the origin of the corresponding HERV-W ENV protein production, at least the one involved in COVID-19. Interestingly, all sequence motifs making it possible to induce the autonomous transcription of this HERV-W gene copy from chromosome X, independently from RBM41 gene transcription have now been identified. Thus, HERV-W ENV transcription can start from upstream TATA boxes in RBM41 intron 6 ending into its exon 7. The presence of HERV-W U3R regulatory sequences in ERVWE2 locus is allowing interactions with viral encoded or virus-induced molecules as previously reviewed [26], but may also be involved in a potent (epi) genetic repression of its transcription in about 80 % of “non-responding” individuals (versus about 20 % of “responders”) [17]. This last aspect calls for important studies of corresponding sequence polymorphism and epigenetics, among many others linked to the chromosome X.

In conclusion, RNAseq data from COVID-19 patients along with the renewed analysis of ERVWE2 locus in chromosome X and the provided experimental proof of concept that its in-frame stop codon is not precluding the production of HERV-W ENV protein, elucidates a new possible mechanism at the ribosome translation level. This can occur when a “helper” molecule is present that makes a readthrough process possible for the corresponding mRNA and therefore suggests that SARS-CoV-2 directly or indirectly provides this key factor. Moreover, as the activation of the ERVWE2 copy transcription is seen possible whenever epigenetically feasible, it also suggests that SARS-CoV-2 would first contribute to trigger regulatory elements leading to its autonomous transcription, while allowing its downstream translation. Studies on such mechanisms induced by SARS-CoV-2, and, also, other infections with triggering viruses such as Epstein-Barr virus (EBV) or other Herpesviridae, are now needed to further elucidate viral factors involved in such post-infectious HERV activation.

CRediT authorship contribution statement

Joanna Brunel: Writing – original draft, Methodology, Investigation, Formal analysis. **Julien Paganini:** Writing – review & editing, Software, Methodology, Investigation, Formal analysis, Data curation. **Melissa Galloux:** Software, Methodology, Formal analysis, Data curation. **Benjamin Charvet:** Writing – review & editing, Validation, Methodology, Formal analysis. **Hervé Perron:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Conceptualization.

Acknowledgements

This study was supported by Geneuro-Innovation SAS, France and by the European HERVCOV project (research program “Horizon Europe”, Grant agreement No 101057302).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.micinf.2024.105431>.

References

- [1] Harris E. Millions of US children experience range of long COVID effects. *JAMA* 2024.
- [2] McCorkell L, Peluso MJ. Long COVID research risks losing momentum - we need a moonshot. *Nature* 2023;622(7983):457–60.
- [3] Graham F. Daily briefing: at least 65 million people have long COVID. *Nature* 2023.
- [4] Nehme M, Braillard O, Chappuis F, CoviCare Study T, Guessous I. The chronification of post-COVID condition associated with neurocognitive symptoms, functional impairment and increased healthcare utilization. *Sci Rep* 2022;12(1):14505.
- [5] WHO. Long COVID in the WHO European region. 2022.
- [6] Silva Andrade B, Siqueira S, de Assis Soares WR, de Souza Rangel F, Santos NO, Dos Santos Freitas A, et al. Long-COVID and post-COVID health complications: an up-to-date review on clinical conditions and their possible molecular mechanisms. *Viruses* 2021;13(4).
- [7] Liew F, Efstathiou C, Fontanella S, Richardson M, Saunders R, Swieboda D, et al. Large-scale phenotyping of patients with long COVID post-hospitalization reveals mechanistic subtypes of disease. *Nat Immunol* 2024;25(4):607–21.
- [8] Huang J, Tengvall K, Bomfim Lima I, Hedstrom AK, Butt J, Brenner N, et al. Genetics of immune response to Epstein-Barr virus: prospects for multiple sclerosis pathogenesis. *Brain* 2024.
- [9] Rippee-Brooks MD, Wu W, Dong J, Pappolla M, Fang X, Bao X. Viral infections, are they a trigger and risk factor of alzheimer's disease? *Pathogens* 2024;13(3).
- [10] Zhao J, Xia F, Jiao X, Lyu X. Long COVID and its association with neurodegenerative diseases: pathogenesis, neuroimaging, and treatment. *Front Neurol* 2024;15:1367974.
- [11] Lemos JRN, Hirani K, von Herrath M. Immunological and virological triggers of type 1 diabetes: insights and implications. *Front Immunol* 2023;14:1326711.
- [12] Krogvold L, Genoni A, Puggioni A, Campani D, Richardson SJ, Flaxman CS, et al. Live enteroviruses, but not other viruses, detected in human pancreas at the onset of type 1 diabetes in the DiViD study. *Diabetologia* 2022;65(12):2108–20.

- [13] Vehik K, Lynch KF, Wong MC, Tian X, Ross MC, Gibbs RA, et al. Prospective virome analyses in young children at increased genetic risk for type 1 diabetes. *Nat Med* 2019;25(12):1865–72.
- [14] Agostini S, Mancuso R, Baglio F, Cabinio M, Hernis A, Guerini FR, et al. Lack of evidence for a role of HHV-6 in the pathogenesis of Alzheimer's disease. *J Alzheimers Dis* 2016;49(1):229–35.
- [15] Peppercorn K, Edgar CD, Kleffmann T, Tate WP. A pilot study on the immune cell proteome of long COVID patients shows changes to physiological pathways similar to those in myalgic encephalomyelitis/chronic fatigue syndrome. *Sci Rep* 2023;13(1):22068.
- [16] Gruchot J, Lewen I, Dietrich M, Reiche L, Sindi M, Hecker C, et al. Transgenic expression of the HERV-W envelope protein leads to polarized glial cell populations and a neurodegenerative environment. *Proc Natl Acad Sci U S A* 2023;120(38):e2308187120.
- [17] Charvet B, Brunel J, Pierquin J, Iampietro M, Decimo D, Queruel N, et al. SARS-CoV-2 awakens ancient retroviral genes and the expression of proinflammatory HERV-W envelope protein in COVID-19 patients. *iScience* 2023;26(5):106604.
- [18] Steiner JP, Bachani M, Malik N, DeMarino C, Li W, Sampson K, et al. Human endogenous retrovirus K envelope in spinal fluid of amyotrophic lateral sclerosis is toxic. *Ann Neurol* 2022;92(4):545–61.
- [19] Tamouza R, Meyer U, Foiselle M, Richard JR, Wu CL, Boukouaci W, et al. Identification of inflammatory subgroups of schizophrenia and bipolar disorder patients with HERV-W ENV antigenemia by unsupervised cluster analysis. *Transl Psychiatry* 2021;11(1):377.
- [20] Charvet B, Pierquin J, Brunel J, Gorter R, Quetard C, Horvat B, et al. Human endogenous retrovirus type W envelope from multiple sclerosis demyelinating lesions shows unique solubility and antigenic characteristics. *Virology* 2021.
- [21] Johansson EM, Bouchet D, Tamouza R, Ellul P, Morr AS, Avignone E, et al. Human endogenous retroviral protein triggers deficit in glutamate synapse maturation and behaviors associated with psychosis. *Sci Adv* 2020;6(29):eabc0708.
- [22] Kremer D, Gruchot J, Weyers V, Oldemeier L, Gottle P, Healy L, et al. pHERV-W envelope protein fuels microglial cell-dependent damage of myelinated axons in multiple sclerosis. *Proc Natl Acad Sci U S A* 2019;116(30):15216–25.
- [23] Thomas J, Perron H, Feschotte C. Variation in proviral content among human genomes mediated by LTR recombination. *Mobile DNA* 2018;9:36.
- [24] Kury P, Nath A, Creange A, Dolei A, Marche P, Gold J, et al. Human endogenous retroviruses in neurological diseases. *Trends Mol Med* 2018;24(4):379–94.
- [25] Charvet B, Reynaud JM, Gourru-Lesimple G, Perron H, Marche PN, Horvat B. Induction of proinflammatory multiple sclerosis-associated retrovirus envelope protein by human herpesvirus-6A and CD46 receptor engagement. *Front Immunol* 2018;9:2803.
- [26] Medina J, Charvet B, Leblanc P, Germi R, Horvat B, Marche PN, et al. [Endogenous retroviral sequences in the human genome can play a physiological or pathological role]. *Med Sci* 2017;33(4):397–403.
- [27] Levit S, Medina J, Joanou J, Demolder A, Queruel N, Reant K, et al. An ancestral retroviral protein identified as a therapeutic target in type-1 diabetes. *JCI Insight* 2017;2(17).
- [28] Faucard R, Madeira A, Gehin N, Authier FJ, Panaite PA, Lesage C, et al. Human endogenous retrovirus and neuroinflammation in chronic inflammatory demyelinating polyradiculoneuropathy. *EBioMedicine* 2016;6:190–8.
- [29] Duperray A, Barbe D, Raguenez G, Weksler BB, Romero IA, Couraud PO, et al. Inflammatory response of endothelial cells to a human endogenous retrovirus associated with multiple sclerosis is mediated by TLR4. *Int Immunol* 2015;27(11):545–53.
- [30] Dolei A, Garson JA, Arru G, Clerici M, Germi R, Marche PN, et al. Multiple sclerosis-associated retrovirus and related human endogenous retrovirus-W in patients with multiple sclerosis. *J Neuroimmunol* 2014;266(1–2):87–8.
- [31] Perron H, Dougier-Reynaud HL, Lomparski C, Popa I, Firouzi R, Bertrand JB, et al. Human endogenous retrovirus protein activates innate immunity and promotes experimental allergic encephalomyelitis in mice. *PLoS One* 2013;8(12):e80128.
- [32] Kremer D, Schichel T, Forster M, Tzekova N, Bernard C, van der Valk P, et al. Human endogenous retrovirus type W envelope protein inhibits oligodendroglial precursor cell differentiation. *Ann Neurol* 2013;74(5):721–32.
- [33] Garcia-Montojo M, Dominguez-Mozo M, Arias-Leal A, Garcia-Martinez A, De las Heras V, Casanova I, et al. The DNA copy number of human endogenous retrovirus-W (MSRV-type) is increased in multiple sclerosis patients and is influenced by gender and disease severity. *PLoS One* 2013;8(1):e53623.
- [34] Perron H, Germi R, Bernard C, Garcia-Montojo M, Deluen C, Fariñelli L, et al. Human endogenous retrovirus type W envelope expression in blood and brain cells provides new insights into multiple sclerosis disease. *Mult Scler* 2012;18(12):1721–36.
- [35] Perron H, Bernard C, Bertrand JB, Lang AB, Popa I, Sanhadji K, et al. Endogenous retroviral genes, herpesviruses and gender in multiple sclerosis. *J Neurol Sci* 2009;286(1–2):65–72.
- [36] Dolei A, Perron H. The multiple sclerosis-associated retrovirus and its HERV-W endogenous family: a biological interface between virology, genetics, and immunology in human physiology and disease. *J Neurovirol* 2009;15(1):4–13.
- [37] Yao Y, Schroder J, Nellaker C, Bottmer C, Bachmann S, Yolken RH, et al. Elevated levels of human endogenous retrovirus-W transcripts in blood cells from patients with first episode schizophrenia. *Gene Brain Behav* 2007.
- [38] Ruprecht K, Obojes K, Wengel V, Gronen F, Kim KS, Perron H, et al. Regulation of human endogenous retrovirus W protein expression by herpes simplex virus type 1: implications for multiple sclerosis. *J Neurovirol* 2006;12(1):65–71.
- [39] Rolland A, Jouvin-Marche E, Viret C, Faure M, Perron H, Marche PN. The envelope protein of a human endogenous retrovirus-W family activates innate immunity through CD14/TLR4 and promotes Th1-like responses. *J Immunol* 2006;176(12):7636–44.
- [40] Nellaker C, Yao Y, Jones-Brando L, Mallet F, Yolken RH, Karlsson H. Transactivation of elements in the human endogenous retrovirus W family by viral infection. *Retrovirology* 2006;3:44.
- [41] Lafon M, Jouvin-Marche E, Marche PN, Perron H. Human viral superantigens: to be or not to be transactivated? *Trends Immunol* 2002;23(5):238–9. ; author reply 239.
- [42] Yolken RH, Karlsson H, Yee F, Johnston-Wilson NL, Torrey EF. Endogenous retroviruses and schizophrenia. *Brain Res Brain Res Rev* 2000;31(2–3):193–9.
- [43] Perron H, Garson JA, Bedin F, Beseme F, Paranhos-Baccala G, Komurian-Pradel F, et al. Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. The Collaborative Research Group on Multiple Sclerosis. *Proc Natl Acad Sci U S A* 1997;94(14):7583–8.
- [44] Menard A, Amouri R, Michel M, Marcel F, Brouillet A, Belliveau J, et al. Gliotoxicity, reverse transcriptase activity and retroviral RNA in monocyte/macrophage culture supernatants from patients with multiple sclerosis. *FEBS Lett* 1997;413(3):477–85.
- [45] Westarp ME, Ferrante P, Perron H, Bartmann P, Kornhuber HH. Sporadic ALS/MND: a global neurodegeneration with retroviral involvement? *J Neurol Sci* 1995;129(Suppl):145–7.
- [46] Perron H, Lalande B, Gratacap B, Laurent A, Genoulat O, Geny C, et al. Isolation of retrovirus from patients with multiple sclerosis. *Lancet* 1991;337(8745):862–3.
- [47] Perron H, Geny C, Laurent A, Mouriquand C, Pellat J, Perret J, et al. Leptomeningeal cell line from multiple sclerosis with reverse transcriptase activity and viral particles. *Res Virol* 1989;140(6):551–61.
- [48] Li X, Wu X, Li W, Yan Q, Zhou P, Xia Y, et al. HERV-W ENV induces innate immune activation and neuronal apoptosis via linc01930/cGAS Axis in recent-onset schizophrenia. *Int J Mol Sci* 2023;24(3).
- [49] Yan Q, Wu X, Zhou P, Zhou Y, Li X, Liu Z, et al. HERV-W envelope triggers abnormal dopaminergic neuron process through DRD2/PP2A/AKT1/GSK3 for schizophrenia risk. *Viruses* 2022;14(1).
- [50] Huang W, Li S, Hu Y, Yu H, Luo F, Zhang Q, et al. Implication of the env gene of the human endogenous retrovirus W family in the expression of BDNF and DRD3 and development of recent-onset schizophrenia. *Schizophr Bull* 2011;37(5):988–1000.
- [51] Huang WJ, Liu ZC, Wei W, Wang GH, Wu JG, Zhu F. Human endogenous retroviral pol RNA and protein detected and identified in the blood of individuals with schizophrenia. *Schizophr Res* 2006;83(2–3):193–9.
- [52] Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet* 2012;13(4):283–96.
- [53] Boeke JD, Burns KH, Chiappinelli KB, Classon M, Coffin JM, DeCarvalho DD, et al. Proceedings of the inaugural dark genome symposium: november 2022. *Mobile DNA* 2023;14(1):18.
- [54] Liu S, Huang J, Zhou J, Chen S, Zheng W, Liu C, et al. NAP-seq reveals multiple classes of structured noncoding RNAs with regulatory functions. *Nat Commun* 2024;15(1):2425.
- [55] Eisenstein M. Startups probe hidden viruses in the 'dark genome' to treat disease. *Nat Biotechnol* 2024;42(4):539–41.
- [56] Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* 2017;18(2):71–86.
- [57] Medina J, Perron H. [DNA sequences from mobile genetic elements, a hidden half of the human genome]. *Med Sci* 2017;33(2):151–8.
- [58] Payer LM, Burns KH. Transposable elements in human genetic disease. *Nat Rev Genet* 2019.
- [59] Wildschutte JH, Williams ZH, Montesin M, Subramanian RP, Kidd JM, Coffin JM. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc Natl Acad Sci U S A* 2016;113(16):E2326–34.
- [60] Hennis AG, Henaff E, Jiang E, Eisenberg AR, Carson JR, Villasante CM, et al. Genomic DNA transposition induced by human PGBD5. *Elife* 2015;4.
- [61] Marchi E, Kanapin A, Magiorkinis G, Belshaw R. Unfixed endogenous retroviral insertions in the human population. *J Virol* 2014;88(17):9529–37.
- [62] Chuong EB, Feschotte C. Evolution. Transposons up the dosage. *Science* 2013;342(6160):812–3.
- [63] Shin W, Lee J, Son SY, Ahn K, Kim HS, Han K. Human-specific HERV-K insertion causes genomic variations in the human genome. *PLoS One* 2013;8(4):e60605.
- [64] Halcrow PW, Quansah DNK, Kumar N, Steiner JP, Nath A, Geiger JD. HERV-K (HML-2) envelope protein induces mitochondrial depolarization and neurotoxicity via endolysosome iron dyshomeostasis. *J Neurosci* 2024;44(14).
- [65] Perron H, Jouvin-Marche E, Michel M, Ounanian-Paraz A, Camelo S, Dumon A, et al. Multiple sclerosis retrovirus particles and recombinant envelope trigger an abnormal immune response in vitro, by inducing polyclonal Vbeta16 T-lymphocyte activation. *Virology* 2001;287(2):321–32.
- [66] Firouzi R, Rolland A, Michel M, Jouvin-Marche E, Hauw JJ, Malcus-Vocanson C, et al. Multiple sclerosis-associated retrovirus particles cause T lymphocyte-dependent death with brain hemorrhage in humanized SCID mice model. *J Neurovirol* 2003;9(1):79–93.
- [67] Zhou T, Su TT, Mudiando T, Wang J. Immune asynchrony in COVID-19 pathogenesis and potential immunotherapies. *J Exp Med* 2020;217(10).
- [68] Komurian-Pradel F, Paranhos-Baccala G, Bedin F, Ounanian-Paraz A, Sodoyer M, Ott C, et al. Molecular cloning and characterization of MSRV-related sequences associated with retrovirus-like particles. *Virology* 1999;260(1):1–9.
- [69] Madeira A, Burgelin I, Perron H, Curtin F, Lang AB, Faucard R. MSRV envelope protein is a potent, endogenous and pathogenic agonist of human toll-like receptor 4: relevance of GNBAC1 in multiple sclerosis treatment. *J Neuroimmunol* 2016;291:29–38.

- [70] Balestrieri E, Minutolo A, Petrone V, Fanelli M, Iannetta M, Malagnino V, et al. Evidence of the pathogenic HERV-W envelope expression in T lymphocytes in association with the respiratory outcome of COVID-19 patients. *EBioMedicine* 2021;66:103341.
- [71] Tamouza R, Meyer U, Lucas A, Richard JR, Nkam I, Pinot A, et al. Patients with psychosis spectrum disorders hospitalized during the COVID-19 pandemic unravel overlooked SARS-CoV-2 past infection clustering with HERV-W ENV expression and chronic inflammation. *Transl Psychiatry* 2023;13(1):272.
- [72] Gimenez-Orenga K, Pierquin J, Brunel J, Charvet B, Martin-Martinez E, Perron H, et al. HERV-W ENV antigenemia and correlation of increased anti-SARS-CoV-2 immunoglobulin levels with post-COVID-19 symptoms. *Front Immunol* 2022;13:1020064.
- [73] Charvet B, Lucas A, Pierquin J, Brunel J, Fried S, Bernis C, et al. Blood biomarkers-defined subgroups show heterogeneity in post-acute COVID-19 syndrome: a rationale for precision medicine. Cold Spring Harbor Laboratory; 2023.
- [74] Chen C, Parthasarathy S, Leung JM, Wu MJ, Drake KA, Ridaura VK, et al. Distinct temporal trajectories and risk factors for Post-acute sequelae of SARS-CoV-2 infection. *Front Med* 2023;10:1227883.
- [75] Leung JM, Wu MJ, Kheradpour P, Chen C, Drake KA, Tong G, et al. Early immune factors associated with the development of post-acute sequelae of SARS-CoV-2 infection in hospitalized and non-hospitalized individuals. *Front Immunol* 2024; 15:1348041.
- [76] Drake KA, Talantov D, Tong GJ, Lin JT, Verheijden S, Katz S, et al. Multi-omic profiling reveals early immunological indicators for identifying COVID-19 Progressors. *Clin Immunol* 2023;256:109808.
- [77] Roebke C, Wahl S, Laufer G, Stadelmann C, Sauter M, Mueller-Lantsch N, et al. An N-terminally truncated envelope protein encoded by a human endogenous retrovirus W locus on chromosome Xq22.3. *Retrovirology* 2010;7(1):69.
- [78] do Olival GS, Faria TS, Nali LH, de Oliveira AC, Casseb J, Vidal JE, et al. Genomic analysis of ERVWE2 locus in patients with multiple sclerosis: absence of genetic association but potential role of human endogenous retrovirus type W elements in molecular mimicry with myelin antigen. *Front Microbiol* 2013;4:172.
- [79] Ruprecht K, Mayer J. On the origin of a pathogenic HERV-W envelope protein present in multiple sclerosis lesions. *Proc Natl Acad Sci U S A* 2019;116(40): 19791–2.
- [80] Kremer D, Perron H, Kury P. Reply to Ruprecht and Mayer: unearthing genomic fossils in the pathogenesis of multiple sclerosis. *Proc Natl Acad Sci U S A* 2019; 116(40):19793–4.
- [81] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9(4):357–9.
- [82] Vashishtha K, Gaud C, Andrews S, Krueger C. Librarian: a quality control tool to analyse sequencing library compositions. *F1000Res* 2022;11:1122.
- [83] Trivedi UH, Cezard T, Bridgett S, Montazam A, Nichols J, Blaxter M, et al. Quality control of next-generation sequencing data without a reference. *Front Genet* 2014;5:111.
- [84] Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016;32(19): 3047–8.
- [85] Wickham H. *ggplot2: elegant graphics for data analysis*. In: Use R! Cham: Springer International Publishing; Imprint: Springer; 2016. p. 1. online resource (XVI), 260 pages 232 illustrations, 140 illustrations in color.
- [86] Kassambara A. *ggpubr: 'ggplot2' Based Publication Ready Plots*. R package version 2023;0.6.0.
- [87] Fan Y, Evans CR, Barber KW, Banerjee K, Weiss KJ, Margolin W, Igoshin OA, Rinehart J, Ling J. Heterogeneity of stop codon readthrough in single bacterial cells and implications for population fitness. *Mol Cell* 2017;67:826–836.e5.
- [88] Zhang H, Lyu Z, Fan Y, Evans CR, Barber KW, Banerjee K, Igoshin OA, Rinehart J, Ling J. Metabolic stress promotes stop-codon readthrough and phenotypic heterogeneity, vol. 117. *Proceedings of the National Academy of Sciences*; 2020. p. 22167–72.
- [89] Liu Z, Wang J, Shi Y, Yee BA, Terrey M, Zhang Q, et al. Seryl-tRNA synthetase promotes translational readthrough by mRNA binding and involvement of the selenocysteine incorporation machinery. *Nucleic Acids Res* 2023;51(19): 10768–81.
- [90] Biziaev N, Sokolova E, Yanvarev DV, Toropygin IY, Shuvalov A, Egorova T, et al. Recognition of 3' nucleotide context and stop codon readthrough are determined during mRNA translation elongation. *J Biol Chem* 2022;298(7):102133.
- [91] Karki P, Carney TD, Maracci C, Yatsenko AS, Shcherbata HR, Rodnina MV. Tissue-specific regulation of translational readthrough tunes functions of the traffic jam transcription factor. *Nucleic Acids Res* 2022;50(11):6001–19.
- [92] Mangkalaphiban K, Fu L, Du M, Thrasher K, Keeling KM, Bedwell DM, et al. Extended stop codon context predicts nonsense codon readthrough efficiency in human cells. *Nat Commun* 2024;15(1):2486.
- [93] Lombardi S, Testa MF, Pinotti M, Branchini A. Translation termination codons in protein synthesis and disease. *Adv Protein Chem Struct Biol* 2022;132:1–48.
- [94] Chen J, Thrasher K, Fu L, Wang W, Aghamohammadzadeh S, Wen H, et al. The synthetic aminoglycoside ELX-02 induces readthrough of G550X-CFTR producing superfunctional protein that can be further enhanced by CFTR modulators. *Am J Physiol Lung Cell Mol Physiol* 2023;324(6):L756–70.
- [95] Leroy C, Spelier S, Essonghe NC, Poix V, Kong R, Gizzi P, et al. Use of 2,6-diaminopurine as a potent suppressor of UGA premature stop codons in cystic fibrosis. *Mol Ther* 2023;31(4):970–85.
- [96] Azimov R, Abuladze N, Sassani P, Newman D, Kao L, Liu W, et al. G418-mediated ribosomal read-through of a nonsense mutation causing autosomal recessive proximal renal tubular acidosis. *Am J Physiol Ren Physiol* 2008;295(3):F633–41.
- [97] Friesen WJ, Trotta CR, Tomizawa Y, Zhuo J, Johnson B, Sierra J, et al. The nucleoside analog cliticine is a potent and efficacious readthrough agent. *RNA* 2017;23(4):567–77.
- [98] Perron H, Lang A. The human endogenous retrovirus link between genes and environment in multiple sclerosis and in multifactorial diseases associating neuroinflammation. *Clin Rev Allergy Immunol* 2010;39(1):51–61.
- [99] Haahr S, Sommerlund M, Moller-Larsen A, Mogensen S, Andersen HM. Is multiple sclerosis caused by a dual infection with retrovirus and Epstein-Barr virus? *Neuroepidemiology* 1992;11(4–6):299–303.
- [100] Sommerlund M, Pallesen G, Moller-Larsen A, Hansen HJ, Haahr S. Retrovirus-like particles in an Epstein-Barr virus-producing cell line derived from a patient with chronic progressive myelopathy. *Acta Neurol Scand* 1993;87(2):71–6.
- [101] Perron H, Firouzi R, Tuke P, Garson JA, Michel M, Beseme F, et al. Cell cultures and associated retroviruses in multiple sclerosis. Collaborative Research Group on MS. *Acta Neurol Scand Suppl* 1997;169:22–31.
- [102] Mameli G, Poddighe L, Mei A, Uleri E, Sotgiu S, Serra C, et al. Expression and activation by Epstein Barr virus of human endogenous retroviruses-W in blood cells and astrocytes: inference for multiple sclerosis. *PLoS One* 2012;7(9):e44991.
- [103] van Horssen J, van der Pol S, Nijland P, Amor S, Perron H. Human endogenous retrovirus W in brain lesions: rationale for targeted therapy in multiple sclerosis. *Mult Scler Relat Disord* 2016;8:11–8.
- [104] Gaetani L, Salvadori N, Brachelente G, Sperandei S, Di Sabatino E, Fiacca A, et al. Intrathecal B cell activation and memory impairment in multiple sclerosis. *Mult Scler Relat Disord* 2024;85:105548.
- [105] Marti Z, Ruder J, Thomas OG, Bronge M, De La Parra Soto L, Gronlund H, et al. Enhanced and cross-reactive in vitro memory B cell response against Epstein-Barr virus nuclear antigen 1 in multiple sclerosis. *Front Immunol* 2024;15:1334720.
- [106] SoRelle ED, Dai J, Reinoso-Vizcaino NM, Barry AP, Chan C, Luftig MA. Time-resolved transcriptomes reveal diverse B cell fate trajectories in the early response to Epstein-Barr virus infection. *Cell Rep* 2022;40(9):111286.
- [107] Oliviero B, Varchetta S, Mele D, Mantovani S, Cerino A, Perotti CG, et al. Expansion of atypical memory B cells is a prominent feature of COVID-19. *Cell Mol Immunol* 2020;17(10):1101–3.
- [108] Lee T, Kim Y, Kim HJ, Ha NY, Lee S, Chin B, et al. Acute surge of atypical memory and plasma B-cell subsets driven by an extrafollicular response in severe COVID-19. *Front Cell Infect Microbiol* 2022;12:909218.
- [109] Ruprecht K, Perron H. Exposure to infant siblings during early life and risk of multiple sclerosis. *JAMA* 2005;293(17):90. 2089; author reply 2089.