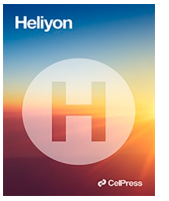
Heliyon 9 (2023) e21283

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Human endogenous retroviruses of the HERV-K (HML-2) family are expressed in the brain of healthy individuals and modify the composition of the brain-infiltrating immune cells

Tapio Nevalainen [a,b,c,1,*], Arttu Autio [a,b,1], Mikko Hurme [a,b,c]

[a] Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
[b] Gerontology Research Center (GEREC), Tampere, Finland
[c] Tampere University Hospital, Finland

ARTICLE INFO

Keywords:
HERV-K
Brain
T cells
CIBERSORTx
GTEx

ABSTRACT

Human endogenous retroviruses (HERVs) are remnants of ancient retroviral infections in the human genome. RNA expression of individual HERVs has frequently been observed in various pathologic conditions, but some activity can also be seen in healthy individuals, e.g. in the blood. To quantitate the basal expression levels of HERVs in the brain, we now used high-throughput sequencing-based metagenomic analysis to characterize the expression profiles of the HERV-K (HML-2) family proviruses in different brain regions of healthy brain tissue. To this end, RNA-seq data from the Genotype-Tissue Expression (GTEx) project was used. The GTEx project is a public resource to study tissue-specific gene expression and regulation, consisting of a large selection of sequenced samples from different tissues. The GTEx data used in this study consisted of 378 samples taken from 13 brain regions from 55 individuals. The data demonstrated that out of 99 intact proviruses in the family 58 were expressed, but the expression profiles were highly divergent and there were no significant differences in the expression profiles between the various anatomic regions of the brain. It is known that the brain contains a variety of infiltrating immune cells, which are probably of great importance both in the normal defense mechanisms as well as in the various pathogenic processes. Digital cytometry (CIBERSORTx) was used to quantify the proportions of the infiltrating immune cells in the same brain samples. Six most abundant (>5 % of the total population) cell types were observed to be CD4 memory resting T cells, M0 macrophages, plasma cells, CD8 T cells, CD4 memory activated T cells, and monocytes. Analysis of the correlations between the individual HERVs and infiltrating cell types indicated that a cluster of 6 HERVs had a notable correlation signature between T cell type infiltrating cell proportions and HERV RNA expression intensity. The correlations between inflammatory type infiltrating cells were negative or weak. Taken together, these data indicate that the expression of HERVs is associated with a T cell type immunity.

* Corresponding author.Faculty of Medicine and Health Technology, Tampere University, Arvo Ylpön katu 34, 33520 Tampere, Finland.
E-mail addresses: tapio.nevalainen@tuni.fi (T. Nevalainen), arttu.autio@tuni.fi (A. Autio), mikko.hurme@tuni.fi (M. Hurme).
[1] Co-first authorship.

https://doi.org/10.1016/j.heliyon.2023.e21283
Received 18 January 2023; Received in revised form 10 May 2023; Accepted 18 October 2023
Available online 20 October 2023

1. Introduction

Human endogenous retroviruses (HERVs) are retroviral sequences that have accumulated in the genome during evolution as a result of ancient retroviral infections. It has been estimated that as much as 8 % of human DNA consist of HERVs [1]. Structurally, intact HERV element consists of three open reading frames (ORF) that code for Gag, Pro, Pol and Env proteins, and are flanked by long terminal repeat (LTR) sequences [2]. LTRs can act as transcriptional promoters, initiating the transcription of HERV ORFs, and HERV transcription is tightly regulated by epigenetic modifications, such as H3K9 histone trimethylation and DNA methylation [3]. Recent research has indicated, that HERV transcription can be activated by environmental factors [4] and by exogenous viruses EBV [5,6], CMV [7], and HIV [8]. Despite transcriptional silencing, HERV RNA expression has been detected in all human tissues [9]. HERV RNA expression originates from a subset of proviruses, as due to truncations and mutational decay, the majority of HERV proviruses no longer encode functional genes [10].

Though the biological impact of some HERV proviruses is known, the potential effects of most proviruses are still unclear. It has been shown that HERVs contribute to physiological processes, such as immunomodulation in the placenta and cell fusion events that give rise to syncytiotrophoblasts [11]. It is also well known that HERV LTRs can regulate the expression of neighboring genes [12]. In addition to these, due to their resemblance to exogenous nucleic acids, HERV RNA has the potential to activate innate immune system via Toll like receptors [13].

Increased expression, or even protein production, of HERVs have been detected both in blood or brain tissue in neuroinflammatory diseases such as Alzheimer's disease, multiple sclerosis, and amyotrophic lateral sclerosis [14]. HERV families HERV-K (HML-2) and HERV-W have especially been associated with such neurodegenerative diseases [14]. However, regardless of the intensive research activity in this field, their role in the pathogenesis is still enigmatic, as they could be anything from "innocent bystanders" to active inducers of inflammation.

There are several difficulties in the analysis of the possible functions and significancy of the HERVs in different clinical conditions. Firstly, the classical polymerase chain reaction -based techniques allow only the quantitation of the expression of a small number of HERVs. Secondly, there are several mechanisms induced by HERVs and thirdly, as the HERV-induced effects are probably mediated by the encoded proteins, their abundance should also be measured. To overcome the limitations of the traditional PCR, whole-genome RNA-sequencing allows the analysis of expression of large number of individual HERV proviruses. This is based on the available DNA sequence data of individual HERVs in a given family. Quantitation of matching sequenced RNA in the sample gives the information about location and expression levels [15]. We have recently used this approach to analyze the role of HERV-K and HERV-W HERV families in the aging of the human immune system [16,17].

As HERVs might have a pathogenic role in the neuroinflammatory disease we decided to characterize the RNA expression profiles of the HERV-K (HML-2) family proviruses in different regions of the brain from healthy brain tissue samples. This would give information about the baseline expression. To analyze their possible effects on the immune system functions, the correlations between their levels of expression and the composition of the brain-infiltrating immune cells were calculated.

2. Material and methods

2.1. Origin of the raw data

This work utilizes polyA-enriched RNA-sequencing data obtained from Genotype-Tissue Expression (GTEx) Project (dbGaP accession number phs000424. v8. p2) [18]. Selected data consists of non-diseased brain samples from different regions of the brain.

2.2. Donors and samples

A total of 378 RNA sequenced brain samples were utilized in this work. The samples are region-specific, originating from 13 brain regions. The number of samples for each region are as follows: amygdala (n = 24), anterior cingulate cortex (n = 24), caudate (n = 35), cerebellar hemisphere (n = 37), cerebellum (n = 39), cortex (n = 31), frontal cortex (n = 31), hippocampus (n = 28), hypothalamus (n = 25), nucleus accumbens (n = 34), putamen (n = 30), spinal cord (n = 20) and substantia nigra (n = 20). The samples originate from 55 sample donors, with at most one sample per brain region per individual. The average number of samples from a single individual was 6.9 (Std deviation 3.8, min 1, max 13). Sample donors were between 21 and 70 years of age, with the mean age being 56.9 years (Std deviation 11.1). Of the sample donors, 37 were male and 18 were female (male mean age 56.1 years, female mean age 58.14 years). All sample donors were surgical patients or post-mortem donors [18]. More detailed information on eligibility criteria and sequencing of the biological samples is available elsewhere [18,19].

2.3. Origin of HERV annotations

The HERV-K (HML-2) annotations originate from a curated list provided by Subramanian et al. [10]. These annotations were chosen for their focus on full-length, well preserved HERV proviruses.

2.4. Bioinformatics pipeline

All the computational procedures were run in Puhti supercomputer of CSC (Espoo, Finland). Raw RNA-sequencing data were

transferred to Puhti from Sequence Read Archive using SRA Toolkit (v2.10.8). Illumina Universal Adapters and low-quality bases (Phred score <20) were trimmed using TrimGalore (v0.6.4; https://github.com/FelixKrueger/TrimGalore; May 10, 2021). PRINSEQ (lite v0.20.4) [20] was utilized to perform further quality preprocessing such that read length was ≥50 nucleotides, mean quality score ≥25, proportion of ambiguous bases was ≤1 %, and DUST score ≤7. In addition, all duplicate reads were removed. Quality scores were obtained with FastQC (v0.11.8; https://www.bioinformatics.babraham.ac.uk/projects/fastqc; May 10, 2021). Read alignment against human reference genome (GCF_000001405.26_GRCh38_genomic.fna from NCBI) was performed with STAR (v2.7.1a) [21]. HTseq [22] was used to obtain the read counts and normalization was performed with DESeq2 default normalization [23]. Read count value ≥ 16 was considered as reliable HERV-K expression.

2.5. Deconvolution analysis

The proportions of different immune cell types in each sample was estimated using deconvolution analysis with the tool CIBERSORTx [24]. From a sample containing a mixed cell population, CIBERSORTx can estimate the relative abundances of predefined cell types using gene expression data and known connections between genes and cell types. CIBERSORTx utilizes machine learning in the form of support vector regression when determining cell proportions. To evaluate deconvolution performance, CIBERSORTx provides an empirical p-value. This is achieved by comparing the resulting cell type fractions with fractions that would have been obtained by random chance [25]. CIBERSORT has been found to perform well, when its accuracy has been evaluated with immunohistochemistry [26] and simulated datasets [27].

For our deconvolution analysis, we ran CIBERSORTx with TPM normalized gene expression values from the RNA sequenced brain samples as the mixture matrix. The cell types of interest and their marker genes are defined in a signature matrix file. We used as signature matrix the default LM22 dataset, which can estimate the proportions of 22 functionally defined human hematopoietic subsets, based on 547 signature genes [28]. A more detailed description of the LM22 dataset can be found in the original CIBERSORT publication [29]. We used CIBERSORTx batch correction, which is done prior to deconvolution analysis. For the CIBERSORTx significance analysis, we set the number of permutations to 1000.

3. Results

Expression intensities of 99 intact HERV-K HML-2 proviruses at 13 different brain regions (amygdala, anterior cingulate cortex, caudate, cerebellar hemisphere, cerebellum, cortex, frontal cortex, hippocampus, hypothalamus, nucleus accumbens, putamen, spinal cord, and substantia nigra) were quantified from GTEx RNA-seq data. Out of these, 58 proviruses were considered expressed (median read count of at least 16) in at least one brain region. Expression intensity was measured as normalized RNA-seq read counts. Fig. 1 shows the distributions of total HERV-K (HML-2) expression intensities, at each brain region. Total HERV-K (HML-2) expression intensity (obtained by adding together read counts of all HERV-K (HML-2) proviruses) was observed to be somewhat higher in the cerebellum when compared to other brain areas. RNA expression intensities of individual HERV-K (HML-2) proviruses for each brain region are shown in Supplementary Table 1.

To investigate how individual HERV-K (HML-2) proviruses contribute to the total HERV-K expression shown in Fig. 1, HERV-K (HML-2) RNA expression intensities of individual proviruses were analyzed at each brain region. Heatmap in Fig. 2 shows the

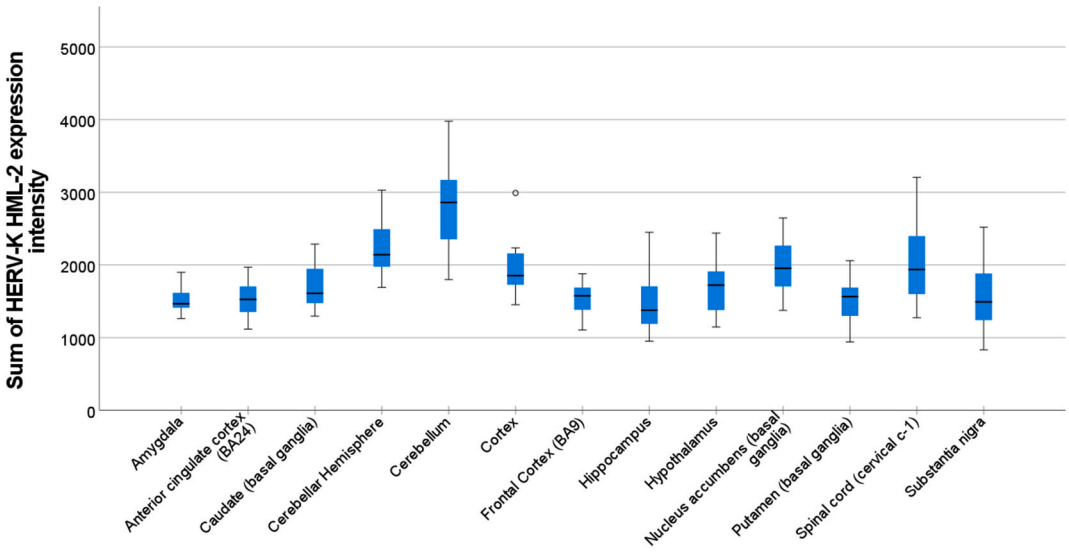


Fig. 1. Distributions of total HERV-K (HML-2) (the sum of all expressed proviruses) RNA expression intensities at each brain region. Sample size varied for each brain region (detailed in materials and methods).

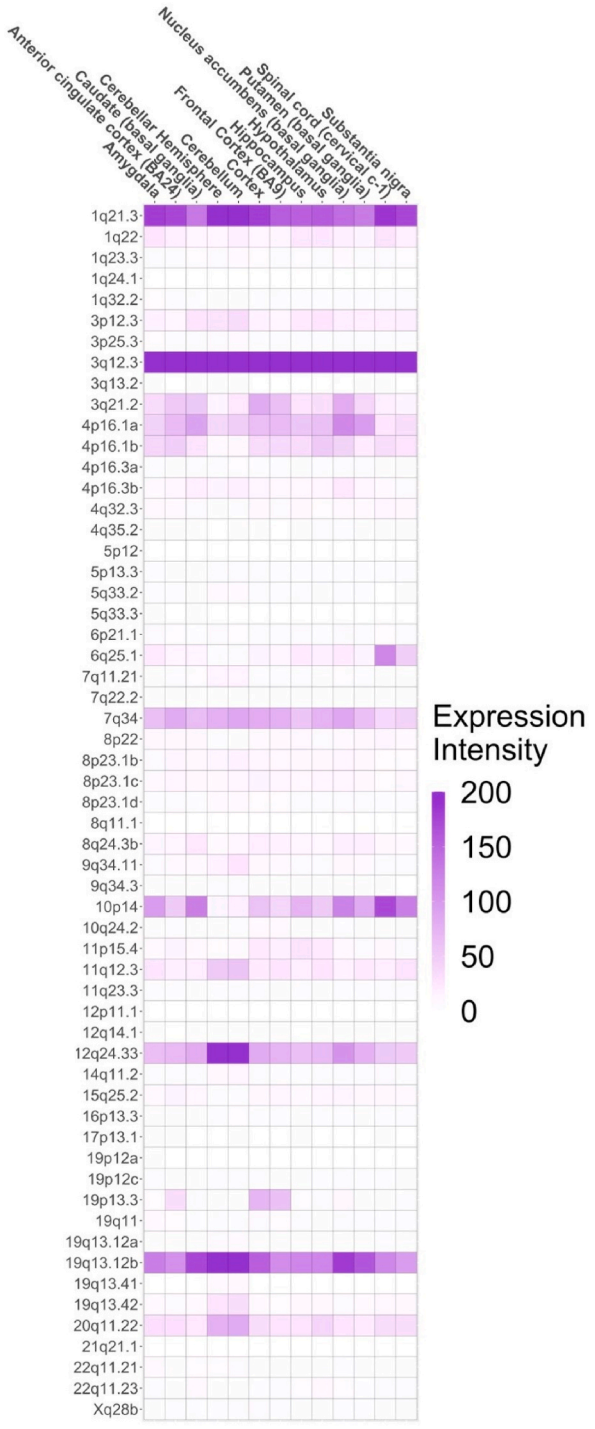


Fig. 2. Mean RNA expression intensities of 58 expressed HERV-K (HML-2) proviruses at different brain regions. Expression intensity is capped at 200 to give better resolution for lower expression values. Sample size varied for each brain region (detailed in materials and methods).

mean RNA expression intensity of the expressed 58 HERV-K (HML-2) proviruses. It can be observed that the total HERV-K (HML-2) expression is dominated by three highly expressing proviruses (1q21.3, 3q12.3, and 19q13.12 b). In case of most of the proviruses, the expression levels were considerably low. See Supplementary Fig. 1 for log-scale version of the heatmap with no read count capping, which has better resolution for low-end read count values.

There was variation within the expression of individual HERV-K (HML-2) proviruses between different brain regions. Cerebellum and cerebellar hemisphere especially showed distinct expression profiles for specific HERV-K (HML-2) proviruses. For example, in the

case of 12q24.33, the expression intensity was higher in cerebellar hemisphere and cerebellum compared to other brain areas, while conversely the expression intensity of 10p14 was lower.

To study the possible biological effects of the expressed proviruses, their association with the composition of the brain infiltrating immune cells was analyzed. To this end, we utilized previously determined cell type deconvolution proportions obtained with digital cytometry [30]. The used tool CIBERSORTx allows the quantitation of proportions of 22 immune cell types with the default signature matrix, based on transcriptomic data. The data in Fig. 3 shows the overall proportions of the immune cells in the brain (without brain region separation). It can be observed that CD4 memory resting T cells, M0 macrophages, plasma cells, CD8 T cells, CD4 memory activated T cells, and monocytes were the most abundant cell types (>5 % of the total population). The proportions of these 6 cell types were relatively similar at each brain region as shown in our previous work [30]. All digital cytometry data are represented in Supplementary Table 2.

Fig. 4 shows the correlations between RNA expression intensity of HERV-K (HML-2) proviruses and infiltrating immune cell proportions in the brain. Only those cell types and proviruses that have at least one significant correlation, are included in the figure (Correlation coefficients and p-values for all cell types and proviruses are included in Supplementary Table 3. The clustering in the heatmap shows that a subset of 6 HERV-K (HML-2) proviruses (5q33.2, 7q11.21, 9q34.11, 12q24.33, 19q13.12 b, and 19q13.41) seem to have a similar correlation profile in terms of CD4 memory resting T cells (high negative correlation), CD8 T cells (high positive correlation), and follicular helper T cells (high positive correlation). Table 1 shows the correlation coefficients and Bonferroni adjusted p-values for these associations.

4. Discussion

HERVs are known to contribute to immunomodulation, but the mechanisms involved are unclear. The association between HERVs and neurodegenerative diseases, which have an inflammatory component, supports the idea that HERVs might modulate the functioning of immune system cells in the brain. Understanding how HERV proviruses interact with immune cells in the brain in normal conditions could increase understanding of what their contributions might be in disease states. Of particular interest is the recently integrated HERV-K (HML-2) family, as its proviruses are better preserved and thus more likely to have biological significance. In addition, members of HERV-K (HML-2) family vary in intactness, and it is not clear which individual proviruses are relevant in terms of their biological impact. Therefore, study of individual proviruses of the HERV-K (HML-2) family, and their potential

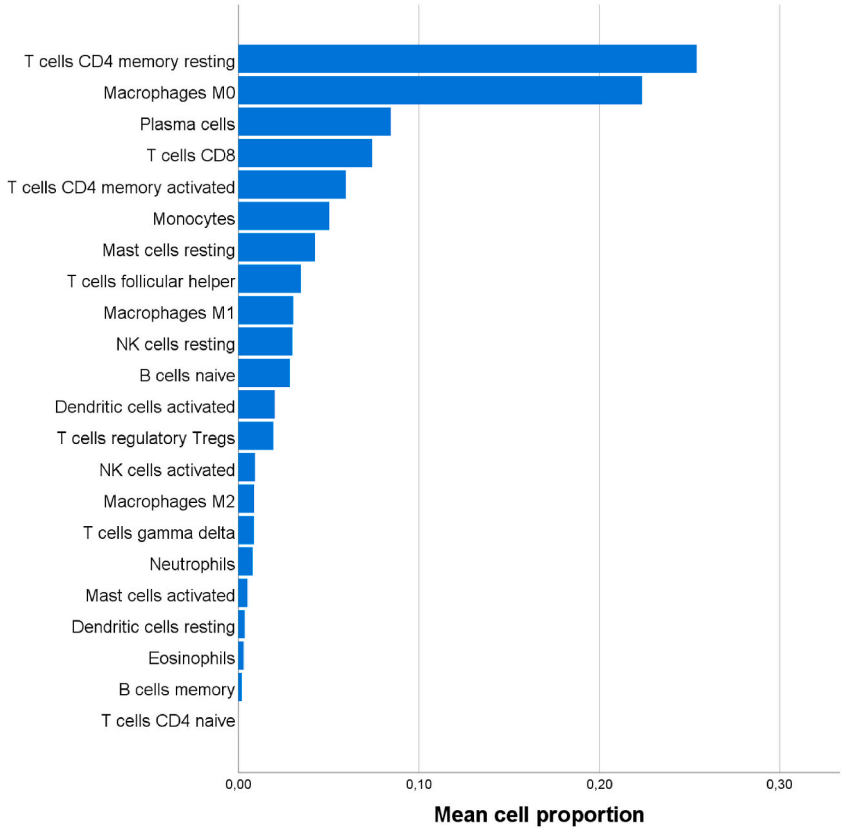


Fig. 3. Mean cell proportions of all 22 quantified infiltrating immune cell types in the brain (without brain region separation) from previously published data. Number of samples is 378.

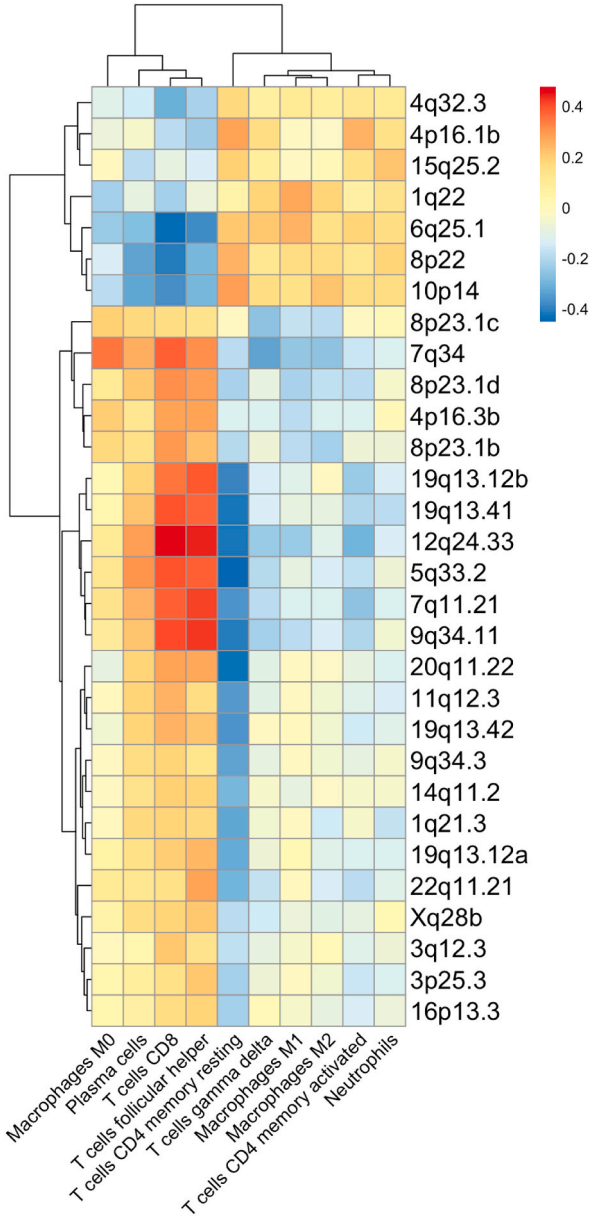


Fig. 4. Spearman's rank correlation between immune cell proportions and HERV-K (HML-2) provirus RNA expression intensity. Only the cell types and proviruses that have at least one significant correlation, are included. Number of samples used to calculate each correlation coefficient is 378.

Table 1
Spearman's correlation coefficients and Bonferroni adjusted p-values for 6 HERV-K (HML-2) proviruses that were identified as having a similar correlation profile with CD8 T cells, CD4 memory resting T cells, and follicular helper T cells.

	T cells CD8		T cells CD4 memory resting		T cells follicular helper	
	r	p	r	p	r	p
5q33.2	0.402	5.58E-13	−0.451	2.87E-17	0.379	3.14E-11
7q11.21	0.379	2.98E-11	−0.365	3.09E-10	0.414	5.55E-14
9q34.11	0.410	1.21E-13	−0.407	1.90E-13	0.424	8.01E-15
12q24.33	0.477	8.19E-20	−0.423	8.94E-15	0.448	5.54E-17
19q13.12b	0.354	1.81E-09	−0.392	2.92E-12	0.393	2.77E-12
19q13.41	0.396	1.58E-12	−0.421	1.58E-14	0.372	9.50E-11

immunomodulatory effect on immune cells present in the brain, could increase our understanding of many neurodegenerative diseases.

In this study, we quantified the expression HERV-K (HML-2) proviruses in multiple regions of the brain to investigate the associations of these proviruses with the numbers of infiltrating immune cells. We have previously quantified infiltrating immune cells proportions in the brain with digital cytometry using the tool CIBERSORTx, in the context of aging [30]. Now we have found that based on RNA-sequencing data of proviral expression, the expression levels of six HERV-K (HML-2) proviruses (at 5q33.2, 7q11.21, 9q34.11, 12q24.33, 19q13.12 b, and 19q13.41) significantly correlated with T cell subsets. More specifically, these six proviruses correlated positively with the proportions of CD8 T cells and follicular helper T cells, and negatively with the proportions of CD4 memory resting T cells (Table 1). It could be that in the case of these proviruses, their proviral products (RNA or protein) may have immunomodulatory effects affecting the numbers of these immune cells. It has been shown that HERV-K (HML-2) RNA can act as a ligand for Toll Like Receptors thus contributing to immune system homeostasis locally [13]. As a result, altered brain immune system homeostasis could activate specific peripheral immune cells to relocate to brain. In terms of HERV-K protein products, it has been shown that they can act as antigens that are recognized by T cells [31].

Even though HERV-K (HML-2) proviruses have high sequence similarity, mutational decay and intactness of the open reading frames affect the potential of proviral RNA and protein products to act as ligands or antigens. The level of expression of individual proviruses is also affected by epigenetic silencing [32]. This could explain why this cell type effect is seen in these six proviruses, yet not in others. The provirus 19q13.12 b has been reported to be expressed in almost all tissue types [33]. There has been reported to be a seemingly intact *env* gene at the locus 19q13.41 [34], which could explain the cell type association of the corresponding provirus through immunogenicity. HERV-K (HML-2) 12q24.33 RNA has been previously shown to be upregulated in the hepatoblastoma cells and associated with activation of leukocytes and leukocyte mediated immunity [35]. The protein products of this provirus are possibly providing immunogenic epitopes for the immune system, and this is in line with our results.

The T cell subsets that were found to be associated with HERV-K (HML-2) expression, CD4 memory cells and CD8 cells, are typical in the induction of immunity in viral infections. Brain infections can be caused by several neurotropic viruses, such as poliovirus, coxsackievirus and enterovirus but also by re-activation of the viruses in the latent form, e.g. herpesviruses. The induction of immunity in these infections is well characterized [36]. Thus, our data suggest that certain HERV-K (HML-2) proviruses in the brain can affect the central nervous system similarly to exogenous viral infection. As the sample donors were without any sign of active infection, it is unlikely that the induction of HERVs was caused by an exogenous virus. Therefore, the quantitated CD8 cells might represent CD8 memory cells that have been induced by past infections. CD4 memory resting T cells are involved in the maintenance of long-term immunological memory, but it is still rather ill-defined and may differ from the other CD4 memory cells in its tissue distribution and development [37,38]. Follicular helper T cells have been shown to be associated with autoimmunity and inflammation [39].

The results shown here may also be related to the mechanisms observed in the CNS disease ALS, as HERV-K (HML-2) has been associated with it. Moreover, there is substantial evidence showing that HERV-K (HML-2) is involved in the pathogenesis of this disease [14]. Therefore, it could be speculated that the effect of HERVs on CD8 and CD4 memory resting cells in healthy tissue as described here would represent a baseline of blood-brain barrier leakage and additional or stronger signals would be required to cause the overt disease.

The data shown in this report indicates that there are differences between brain anatomic regions in both total HERV-K (HML-2) expression (Fig. 1) and the expression of specific HERV-K (HML-2) proviruses (Fig. 2). Our results indicate an elevated HERV-K (HML-2) provirus expression in the cerebellum and cerebellar hemisphere, corroborating the findings of a study by Burn et al. [33]. The cerebellum and cerebellar hemispheres have also been reported to have gene expression patterns that are distinct from all other brain regions [40]. Differences in transcriptomes partly reflect the differences in epigenetic regulation, which would also affect the expression of HERV proviruses. This could be due to opening of chromatin, making HERV promoters accessible, or by the increased read-through transcription driven by the genes upstream from the HERVs.

As the studied samples are derived from healthy tissue, the intensity of the HERV-K (HML-2) proviral expression may be related to the local physiologic processes in the brain, e.g. epigenetic inhibition, cell growth and differentiation. In this way the data are characterizing the normal baseline expression of the HERV-K (HML-2) proviruses in different brain regions. Assuming that expression of specific HERV-K (HML-2) proviruses can have immunomodulatory effects, variation from this baseline could be associated with disease conditions.

The results shown in this report give a good example of why high-throughput sequencing –based bioinformatic tools offer several advantages to study the role of HERVs, when compared to the older PCR-based approaches [15], as only a minority of the studied proviruses demonstrated the immune effects described. Next-generation sequencing has made it possible to study the expression of a large collection of proviruses at once, and powerful new alignment software, such as STAR [21], can differentiate more precisely the origin of RNA fragment between duplicate proviruses located at different parts of the genome.

There are some limitations to our study that should be taken into consideration when interpreting the results. Firstly, only the proportions of infiltrating leukocyte types were quantified, not resident brain cells. Thus, the proportions mentioned here only reflect proportions in the quantified subset of cells. Using a more comprehensive set of cell types in the deconvolution could provide more accurate results in terms of cell type proportion estimates. Reference profiles combining tissue-resident and tissue-infiltrating cellular subsets are not currently available [41]. Secondly, HERV RNA may or may not result in HERV protein products. Though HERV expression is thought to have biological significance through RNA as well, it is especially the protein products, such as Env, that are associated with certain diseases. Our RNA-seq based approach therefore cannot differentiate whether the observed associations relate to HERV RNA or protein products. Thus, quantifying the HERV-K (HML-2) protein levels in the similar setting would provide more information on the nature of the association. Thirdly, it should also be noted, that though the GTEx derived data is from non-diseased

tissue sites, the samples were gathered post-mortem. This could affect the estimated cell type proportions and our analysis of their association with HERV-K (HML-2) expression.

In conclusion, our results indicate that there is an association between the expression of proviruses of the HERV-K (HML-2) family and proportions of T cell subsets. Moreover, this effect was seen only in the case of specific proviruses of HERV-K (HML-2) family. This implies that individual HERV-K (HML-2) proviruses affect the proportions of infiltrating immune cells in the brain of healthy individuals. It could be speculated that the observed HERV-K (HML-2) expression profile indicates a baseline state, variation from which could contribute to pathologies in the brain. As the expression of the individual HERV-K (HML-2) proviruses has been noted in several neurodegenerative diseases, we think that this transcriptomic approach, and the potential functional effects described here, could be informative in searching for the pathogenetic mechanisms in neurodegenerative diseases.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The raw RNA-seq data of the GTEx project analyzed in this work can be accessed for research purposes through the database of Genotypes and Phenotypes (dbGaP) system. The dbGaP accession number for the project is phs000424. v8. p2. Access to GTEx protected data, which includes the raw sequencing data, requires an approved dbGaP application.

Funding

This work was financially supported by research funding provided by the Tampere Tuberculosis foundation (MH); the Competitive State Research Financing of the Expert Responsibility Area of Tampere University Hospital (MH); the Finnish Cultural Foundation, Pirkanmaa Regional Fund (TN); the Finnish Cultural Foundation (AA); and the Emil Aaltonen Foundation (AA).

Authors' contributions

TN and AA contributed to data analysis and co-wrote the paper. MH designed the experiment, co-wrote the paper and supervised the research. All authors read and approved the final manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health (commonfund.nih.gov/GTEx). Additional funds were provided by the NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. Donors were enrolled at Biospecimen Source Sites funded by NCI\Leidos Biomedical Research, Inc. Subcontracts to the National Disease Research Interchange (10XS170), Roswell Park Cancer Institute (10XS171), and Science Care, Inc. (X10S172). The Laboratory, Data Analysis, and Coordinating Center (LDACC) was funded through a contract (HHSN268201000029C) to the The Broad Institute, Inc. Biorepository operations were funded through a Leidos Biomedical Research, Inc. Subcontract to Van Andel Research Institute (10ST1035). Additional data repository and project management were provided by Leidos Biomedical Research, Inc. (HHSN261200800001E). The Brain Bank was supported supplements to University of Miami grant DA006227. Statistical Methods development grants were made to the University of Geneva (MH090941 & MH101814), the University of Chicago (MH090951, MH090937, MH101825, & MH101820), the University of North Carolina - Chapel Hill (MH090936), North Carolina State University (MH101819), Harvard University (MH090948), Stanford University (MH101782), Washington University (MH101810), and to the University of Pennsylvania (MH101822). The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000424. v8. p2.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e21283.

List of abbreviations

HERV	Human Endogenous Retrovirus
HERV-K	Youngest and most active group of HERVs
HML-2	Subgroup of HERV-K
GTEx	The Genotype-Tissue-Expression project
CD8	T cell surface marker for cytotoxic subset of T cells
ORF	Open reading frame
LTR	Long terminal repeat
EBV	Epstein-Barr virus
CMV	Cytomegalovirus
ALS	Amyotrophic lateral sclerosis
RNA-Seq	RNA sequencing
CSC	Finnish IT center for science
SRA	Sequence Read Archive
STAR	Spliced Transcripts Alignment to a Reference, sequence aligner
GRCh	Genome Reference Consortium Human genome
NCBI	The National Center for Biotechnology Information
TPM	Transcripts per million
CD4	T cell surface marker for helper subset of T cells
PCR	The polymerase chain reaction

References

[1] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, et al., Initial sequencing and analysis of the human genome, Nature 409 (6822) (2001 Feb) 860–921.
[2] J. Mao, Q. Zhang, Y.S. Cong, Human endogenous retroviruses in development and disease, Comput. Struct. Biotechnol. J. 19 (2021) 5978–5986.
[3] F.K. Geis, S.P. Goff, Silencing and transcriptional regulation of endogenous retroviruses: an overview, Viruses 12 (8) (2020 Aug).
[4] M. Zhang, J.Q. Liang, S. Zheng, Expressional activation and functional roles of human endogenous retroviruses in cancers, Rev. Med. Virol. 29 (2) (2019 Mar), e2025.
[5] J. Gruchot, D. Kremer, P. Küry, Neural cell responses upon exposure to human endogenous retroviruses, Front. Genet. 10 (2019) 655.
[6] V. Gröger, A. Emmer, M.S. Staege, H. Cynis, Endogenous retroviruses in nervous system disorders, Pharmaceuticals 14 (1) (2021 Jan).
[7] A. Assinger, K.C. Yaiw, I. Göttesdorfer, C. Leib-Mösch, C. Söderberg-Nauclér, Human cytomegalovirus (HCMV) induces human endogenous retrovirus (HERV) transcription, Retrovirology 10 (2013 Nov) 132.
[8] A.C. van der Kuyl, HIV infection and HERV expression: a review, Retrovirology 9 (2012 Jan) 6.
[9] J.M. Yi, H.M. Kim, H.S. Kim, Expression of the human endogenous retrovirus HERV-W family in various human tissues and cancer cells, J. Gen. Virol. 85 (Pt 5) (2004 May) 1203–1210.
[10] R.P. Subramanian, J.H. Wildschutte, C. Russo, J.M. Coffin, Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses, Retrovirology 8 (2011 Nov) 90.
[11] U. Kämmerer, A. Germeyer, S. Stengel, M. Kapp, J. Denner, Human endogenous retrovirus K (HERV-K) is expressed in villous and extravillous cytotrophoblast cells of the human placenta, J. Reprod. Immunol. 91 (1–2) (2011 Sep) 1–8.
[12] R. Rebollo, M.T. Romanish, D.L. Mager, Transposable elements: an abundant and natural source of regulatory sequences for host genes, Annu. Rev. Genet. 46 (2012) 21–42.
[13] P. Dembny, A.G. Newman, M. Singh, M. Hinz, M. Szczepek, C. Krüger, et al., Human endogenous retrovirus HERV-K(HML-2) RNA causes neurodegeneration through Toll-like receptors, JCI insight 5 (7) (2020 Apr).
[14] C. Römer, Viruses and endogenous retroviruses as roots for neuroinflammation and neurodegenerative diseases, Front. Neurosci. 15 (2021), 648629.
[15] M.P. Pisano, N. Grandi, E. Tramontano, High-throughput sequencing is a crucial tool to investigate the contribution of human endogenous retroviruses (HERVs) to human biology and development, Viruses 12 (6) (2020 Jun).
[16] T. Nevalainen, A. Autio, B.H. Mishra, S. Marttila, M. Jylhä, M. Hurme, Aging-associated patterns in the expression of human endogenous retroviruses, PLoS One 13 (12) (2018), e0207407.
[17] A. Autio, T. Nevalainen, B.H. Mishra, M. Jylhä, H. Flinck, M. Hurme, Effect of aging on the transcriptomic changes associated with the expression of the HERV-K (HML-2) provirus at 1q22, Immun. Ageing 17 (2020) 11.
[18] J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, et al., The genotype-tissue expression (GTEx) project, Nat. Genet. 45 (6) (2013 Jun 29) 580–585.
[19] L.J. Carithers, K. Ardlie, M. Barcus, P.A. Branton, A. Britton, S.A. Buia, et al., A novel approach to high-quality postmortem tissue procurement: the GTEx project, Available from, Biopreserv. Biobanking 13 (5) (2015 Oct 1) 311–317, https://doi.org/10.1089/bio.2015.0032. :.
[20] R. Schmieder, R. Edwards, Quality control and preprocessing of metagenomic datasets, Bioinformatics 27 (6) (2011 Mar) 863–864.
[21] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, et al., STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (1) (2013 Jan) 15–21.
[22] S. Anders, P.T. Pyl, W. Huber, HTSeq–a Python framework to work with high-throughput sequencing data, Bioinformatics 31 (2) (2015 Jan) 166–169.
[23] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (12) (2014) 550.
[24] A.M. Newman, C.B. Steen, C.L. Liu, A.J. Gentles, A.A. Chaudhuri, F. Scherer, et al., Determining cell type abundance and expression from bulk tissues with digital cytometry, Nat. Biotechnol. 37 (7) (2019 Jul) 773–782.
[25] C.B. Steen, C.L. Liu, A.A. Alizadeh, A.M. Newman, Profiling cell type abundance and expression in bulk tissues with CIBERSORTx, Methods Mol. Biol. 2117 (2020) 135–157.
[26] E. Patrick, M. Taga, A. Ergun, B. Ng, W. Casazza, M. Cimpean, et al., Deconvolving the contributions of cell-type heterogeneity on cortical gene expression, PLoS Comput. Biol. 16 (8) (2020 Aug), e1008120.
[27] G.J. Sutton, D. Poppe, R.K. Simmons, K. Walsh, U. Nawaz, R. Lister, et al., Comprehensive evaluation of deconvolution methods for human brain gene expression, Nat. Commun. 13 (1) (2022 Mar) 1358.
[28] B. Chen, M.S. Khodadoust, C.L. Liu, A.M. Newman, A.A. Alizadeh, Profiling tumor infiltrating immune cells with CIBERSORT, Methods Mol. Biol. 1711 (2018) 243–259.

[29] A.M. Newman, C.L. Liu, M.R. Green, A.J. Gentles, W. Feng, Y. Xu, et al., Robust enumeration of cell subsets from tissue expression profiles, Nat. Methods 12 (5) (2015 May) 453–457.
[30] T. Nevalainen, A. Autio, M. Hurme, Composition of the infiltrating immune cells in the brain of healthy individuals: effect of aging, Immun. Ageing 19 (1) (2022 Oct) 45.
[31] B. Xue, L.A. Sechi, D.J. Kelvin, Human endogenous retrovirus K (HML-2) in Health and disease, Front. Microbiol. 11 (2020) 1690.
[32] T.P. Hurst, G. Magiorkinis, Epigenetic control of human endogenous retrovirus expression: focus on regulation of long-terminal repeats (LTRs), Viruses 9 (6) (2017 May).
[33] A. Burn, F. Roy, M. Freeman, J.M. Coffin, Widespread expression of the ancient HERV-K (HML-2) provirus group in normal human tissues, PLoS Biol. 20 (10) (2022 Oct), e3001826.
[34] P. Villesen, L. Aagaard, C. Wiuf, F.S. Pedersen, Identification of endogenous retroviral reading frames in the human genome, Retrovirology 1 (2004 Oct) 32.
[35] D.F. Grabski, A. Ratan, L.R. Gray, S. Bekiranov, D. Rekosh, M.L. Hammarskjold, et al., Upregulation of human endogenous retrovirus-K (HML-2) mRNAs in hepatoblastoma: identification of potential new immunotherapeutic targets and biomarkers, J. Pediatr. Surg. 56 (2) (2021 Feb) 286–292.
[36] S. Prasad, J.R. Lokensgard, Brain-resident T cells following viral infection, Viral Immunol. 32 (1) (2019) 48–54.
[37] S. Hojyo, D. Tumes, A. Murata, K. Tokoyoda, Multiple developmental pathways lead to the generation of CD4 T-cell memory, Int. Immunol. 32 (9) (2020 Sep) 589–595.
[38] K.K. McKinstry, T.M. Strutt, S.L. Swain, The potential of CD4 T-cell memory, Immunology 130 (1) (2010 May) 1–9.
[39] L. Qin, T.C. Waseem, A. Sahoo, S. Bieerkehazhi, H. Zhou, E.V. Galkina, et al., Insights into the molecular mechanisms of T follicular helper-mediated immunity and pathology, Front. Immunol. 9 (2018) 1884.
[40] Y. Hang, M. Aburidi, B. Husain, A.R. Hickman, W.L. Poehlman, F.A. Feltus, Exploration into biomarker potential of region-specific brain gene co-expression networks, Sci. Rep. 10 (1) (2020 Oct), 17089.
[41] A.R. Marderstein, M. Uppal, A. Verma, B. Bhinder, Z. Tayyebi, J. Mezey, et al., Demographic and genetic factors influence the abundance of infiltrating immune cells in human tissues, Nat. Commun. 11 (1) (2020 May) 2213.