HERV activation segregates ME/CFS from fibromyalgia while defining a novel nosologic entity

Karen Giménez-Orenga[1], Eva Martín-Martínez[2], Lubov Nathanson[3], Elisa Oltra[4]*

[1]Escuela de Doctorado, Catholic University of Valencia, Valencia, Spain.

[2]National Health Service, Manises Hospital, Valencia, Spain.

[3]Institute for Neuro-Immune Medicine, Dr. Kiran C. Patel College of Osteopathic Medicine, Nova Southeastern University, Fort Lauderdale, FL 33328, USA.

[4]Department of Pathology, School of Health Sciences, Catholic University of Valencia, Valencia, Spain.

*Corresponding author. Email: elisa.oltra@ucv.es

Running title: HERV segregates ME/CFS and FM

Abstract

Research of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) and Fibromyalgia (FM), two acquired chronic illnesses affecting mainly females, has failed to ascertain their frequent co-appearance and etiology. Despite prior detection of human endogenous retrovirus (HERV) activation in these diseases, the potential biomarker value of HERV expression profiles for their diagnosis, and the relationship of HERV expression profiles with patient immune systems and symptoms had remained unexplored. By using HERV-V3 high-density microarrays (including over 350k HERV elements and more than 1500 immune-related genes) to interrogate the transcriptomes of peripheral blood mononuclear cells from female patients diagnosed with ME/CFS, FM or both, and matched healthy controls (n=43), this study fills this gap of knowledge. Hierarchical clustering of HERV expression profiles strikingly allowed perfect participant assignment into four distinct groups: ME/CFS, FM, co-diagnosed, or healthy, pointing at a potent biomarker value of HERV expression profiles to differentiate between these hard-to-diagnose chronic syndromes. Differentially expressed HERV-immune-gene modules revealed unique profiles for each of the four study groups and highlighting decreased γδ T cells, and increased plasma and resting CD4 memory T cells, correlating with patient symptom severity in ME/CFS. Moreover, activation of HERV sequences coincided with enrichment of binding sequences targeted by transcription factors which recruit SETDB1 and TRIM28, two known epigenetic silencers of HERV, in ME/CFS, offering a mechanistic explanation for the findings. Unexpectedly, HERV expression profiles appeared minimally affected in co-diagnosed patients denoting a new nosological entity with low epigenetic impact, a seemingly relevant aspect for the diagnosis and treatment of this prevalent group of patients.

Key words: Fibromyalgia, Human endogenous retrovirus (HERV), Immune system, Myalgic encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS)

INTRODUCTION

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS), classified by the WHO with the ICD-11 8E49 code as a postviral fatigue syndrome, and fibromyalgia (FM) (ICD-11 MG30.0 for chronic primary pain) (Harrison et al., 2021), are chronic, disabling, acquired diseases, characterized by complex symptomatology that affects multiple organs (Bateman et al., 2021; Wolfe et al., 2010). Diagnosis of ME/CFS and FM continues to be based on the clinical assessment of unspecific symptoms, such as debilitating fatigue, generalized pain, cognitive impairment or intestinal, sleep, and immune disturbances (Carruthers et al., 2011, 2003; Wolfe et al., 2016, 2010, 1990). Their frequent co-diagnosis drove the hypothesis of a "single syndrome" and promoted the search for common or differentiating factors (Abbi and Natelson, 2013; Natelson, 2019; Wessely et al., 1999). However, despite molecular support for ME/CFS and FM constituting different entities (Groven et al., 2021; Light et al., 2012; Nepotchatykh et al., 2023), their frequent joint appearance (over 50% in females) (Castro-Marrero et al., 2017) remains enigmatic.

Extensive research into ME/CFS and FM has not yet been able to ascertain their origin and pathophysiology. However, several environmental factors have been suggested as triggering agents (Chu et al., 2019; Furness et al., 2018; Tschopp et al., 2023). Viral infections are, particularly, gaining momentum after the emergence of a type of persistent post-viral syndrome with symptoms that closely resemble those in ME/CFS (Komaroff and Lipkin, 2023) and FM (Clauw and Calabrese, 2023), which has been defined as post-COVID-19 condition (ICD-11 RA02) (Soriano et al., 2022). The prevalence of post-COVID-19 cases meeting ME/CFS

criteria has been estimated at 58% (Jason and Dorri, 2022) raising concerns of a health, social, and economic burden of unprecedented dimensions.

Viral infections trigger dramatic changes in host cells gene expression, affecting their metabolism and epigenetic landscape (Liu et al., 2020). Although functionally not well understood, these epigenetic changes involve de-repression of human endogenous retroviruses (HERV) (Macchietto et al., 2020) sequences that were incorporated from exogenous viral infections during evolution and that currently represent about 8% of our genome (Giménez-Orenga and Oltra, 2021a). In fact, SARS-CoV-2 infection alters PBMC ERV expression in human, monkey and mice expressing human ACE2 receptors, associating with immune response activation and histone modification genes in severe COVID-19 cases (Guo, X., et al, 2024), and the presence of HERV proteins in postmortem tissues of lungs, heart, gastrointestinal tract, brain olfactory bulb, and nasal mucosa from COVID-19 patients (Charvet et al., 2023).

Although still poorly understood, advances in sequencing technology have aided in the discovery of the wide range of physiological processes in which HERV participate, including coordination of the immune response (Kassiotis, 2023), interaction with microbiota (Dopkins et al., 2022; Lima-Junior et al., 2021), or shaping neurological functions (Ferrari et al., 2021), among other. Derangement of HERV expression associates with disease (e.g. multiple sclerosis (Gruchot et al., 2023), systemic lupus erythematosus (Khadjinova et al., 2021), or post-COVID-19 condition (Giménez-Orenga et al., 2022). Particularly, the expression of HERV-encoded proteins has been shown to stimulate an immune response present in autoimmunity (Gruchot et al., 2023; Khadjinova et al., 2021). In addition, HERV influence on pathophysiology extends to their long terminal repeats (LTR) regulatory regions, shaping host gene expression (Ito et al., 2017).

Previous studies detected overexpression of some HERV families in immune cells of ME/CFS (Rodrigues et al., 2019) and FM (Ovejero et al., 2020) at the transcript level. However, the use of directional RT-qPCR (reverse transcription followed by quantitative polymerase chain reaction)-based approaches using degenerated primers limited the detection of affected genomic loci. In this study, we elevated these analyses to obtain genome-wide HERV profiles (HERV expression profiles) of patients having received ME/CFS, FM diagnosis, or both, hypothesizing that HERV expression profiles may reveal distinct underlying pathomechanisms that justify their classification as separate diseases, and/or common fingerprints explaining their joint high prevalence. To test our hypothesis, we scrutinized HERV and gene transcriptomes by using high-density microarray technology in a selected cohort of female patients carefully phenotyped by a single expert clinician into three defined groups of patients: patients fulfilling Canadian and International ME/CFS diagnosis criteria (Carruthers et al., 2011, 2003), patients fulfilling FM diagnosis ACR (American College of Rheumatology) criteria (Wolfe et al., 2016, 2010, 1990), or patients complying with both, ME/CFS and FM diagnosis criteria, being called co-diagnosed from now on. The results of comparing HERV profiles within PBMCs across these three patient groups as well as to HERV profiles of matched healthy subjects allowed, not only for the separation of patients from healthy individuals, but also for perfect discrimination of patients into three distinct disease groups, suggesting distinct subjacent pathomechanisms. Unexpectedly, HERV expression profiles appeared minimally affected in co-diagnosed patients denoting a new nosological entity with low epigenetic impact. In addition, HERV profiling also exposed some commonalities between the FM and ME/CFS subgroups and marked quantitative differences within the ME/CFS group that correlated with immune disturbances and patient symptomatology, aspects that may well set new criteria for the differential diagnosis of FM and ME/CFS as well as for patient subtyping with expected impact in precision medicine programs.

RESULTS

Demographics and Clinical Characteristics of Participants

The study compared a total of 43 female subjects: 8 ME/CFS cases, 10 FM, 16 co-diagnosed cases, and 9 matched healthy controls. The average age for participants was 54±3 years (range 50-58) for ME/CFS subjects, 50±5 years (range 42-58) for FM, 47±15 years (range 22-70) for co-diagnosed individuals, and 51±6 (range 43-61) for controls (**Figure S1A, Table S1**). Questionnaires used for patient symptom phenotyping included Fibromyalgia Impact Questionnaire (FIQ) (Burckhardt et al., 1991), Multi Fatigue Inventory (MFI) (Smets et al., 1995) for general fatigue, and Short-Form-36 Health Survey (SF-36) (MCHORNEY et al., 1993) for quality-of-life assessment (**Figure S1A**). Average total and subdomain scores are shown in **Table 1** for each patient group, while itemized questionnaire scores are provided in **Table S1**. According to total FIQ scores, after considering that a score <39 indicates mild affection, ≥39 through ≤59 is assigned to moderately affected patients, and >59 to severe affection (Rivera and González, 2004); around 88% of cases in the studied cohort correspond with severe cases while 12% are moderate. No statistically significant difference between patient groups was observed for any questionnaire domain except for MFI Physical Fatigue (p=0.011) and Reduced Motivation (p=0.038), and only between the ME/CFS and co-diagnosed groups.

HERV signatures discriminate ME/CFS, FM, and co-diagnosed groups from healthy controls

Genome-wide HERV expression profiles for each of the four study groups (three disease groups: ME/CFS, FM, and co-diagnosed, plus one control group corresponding to healthy participants), using custom high-density Affymetrix HERV-V3 microarrays (Becker et al., 2017), showed that a set of 489 HERV (502 probesets) is differentially expressed (DE) between at least two of the groups (FDR<0.1 and |Log2FC|>1) (**Fig. 1**; **Table S3**), confirming

dysregulation of particular sets of HERV elements in the immune systems of ME/CFS, FM, and co-diagnosed cases as compared to healthy controls. Volcano plots show the number of HERV elements found significantly over or under-expressed for each of the comparisons, evidencing an enhanced dysregulation of HERV in ME/CFS as compared to FM, co-diagnosed, or healthy control groups (**Fig. 1A**). Consistently, hierarchical clustering of samples based on DE HERV fingerprints clearly differentiate these pathologies from healthy subjects, strikingly segregating ME/CFS from all other compared groups (**Fig. 1B**); while hierarchical clustering of DE HERV sequences revealed four main clusters consisting of loci specifically overexpressed (Cluster 1, 251 HERV, 293 probesets) or underexpressed (Cluster 2, 119 HERV, 163 probesets) in ME/CFS, loci specifically overexpressed in FM (Cluster 3, 68 HERV, 74 probesets), and loci underexpressed in all three pathologies (Cluster 4, 54 HERV, 64 probesets). DE of the HERV loci for each study group was corroborated by analysis of HERV expression levels encompassed in each cluster, as illustrated with violin plots (**Fig. 1B**), with some of the changes being validated by the alternative RT-qPCR approach, as shown in **Fig. S1B.**

Closer examination of Clusters 1 and 2 indicated that the ME/CFS group could be subdivided into two subgroups showing differential levels of HERV expression profiles (**Fig. 1C).** While both ME/CFS subgroups (subgroup 1 and subgroup 2) show clear upregulation (Cluster 1, $p<0.0001$) or downregulation (Cluster 2, $p<0.0001$) of several HERV loci as compared to all the other study groups, they show quantitative differences, with the ME/CFS subgroup 2 exhibiting a more pronounced dysregulation than subgroup 1, particularly affecting the downregulated Cluster 2 (**Fig. 1C)**. Interestingly, subgroup 2 ME/CFS patients present more accused levels of fatigue ($p<0.05$) and reduced motivation as compared to the co-diagnosed group and the ME/CFS subgroup 1 (**Table S4**) indicating that quantitative HERV fingerprint differences might help to assess the degree of ME/CFS severity.

In line with our findings, unsupervised principal component analysis (PCA) of DE HERV loci supports perfect discrimination of samples by study group and differentiates the two identified ME/CFS subgroups (**Fig. 1D**). While principal component 1 (PC1) perfectly segregates the two ME/CFS subgroups from FM, co-diagnosed, and control groups; principal component 2 (PC2) allows partition of FM, co-diagnosed, and control samples, but not so much of ME/CFS from co-diagnosed cases. Thus, PC1 explains better the variance observed across groups (29.1%) than PC2 (9.3%), suggesting a prominent role of DE HERV loci in ME/CFS pathology (**Fig. 1E**). In summary, these results show that ME/CFS, FM, and their co-diagnosis present unique HERV expression profiles capable of discriminating patient subtypes from healthy subjects, providing potential biomarkers for their differential diagnosis and stratification. ME/CFS shows the greatest HERV dysregulation which, importantly, associates with patient symptomatology, suggesting a prominent role of DE HERV in the pathomechanism of this disease, while the co-diagnosed group presents the closest HERV expression profiles to those of healthy controls indicating much less involvement of epigenetic changes in these patients, thus, defining a different nosological entity from ME/CFS or FM.

Dysregulated HERV families are disease-specific

Since preferential HERV-family derangements may indicate particularly altered functions in disease (Ito et al., 2017), we next examined the family composition of the identified DE HERV clusters (**Fig. 2**). Overall, ME/CFS (Clusters 1 and 2) outstands with the largest heterogeneity with up to 66 HERV dysregulated families, contrasting with only 22 families in FM (Cluster 3) and 16 in Cluster 4, the latter representing DE HERV common to all patient groups (**Table S5, Fig. S2**).

Among DE HERV families 8 appear to be common across all clusters, and 6 of them represent the most abundantly deregulated (>2.5%) (MLT1, MST, THE1, HERV16, ERVL, and

HERVL33) (**Fig. 2A, Table S5**), suggesting an unspecific association with all disease groups. Within these families, however, MLT1 appears as the only significantly downregulated in all patient groups ($p<0.01$) (**Fig. S2**), accounting for more than a third of the DE HERV in each cluster (**Fig. 2A, Table S5**), a finding that may be, at least partially, due to the fact of MLT1 being the most abundant HERV family in the human genome (**Fig. 2A**).

Interestingly, unique sets of DE HERV families were found for each disease (**Table S5**). Up to 21 families, comprising HERV-H, HUERSP3, and RRHERVI among other, appeared specifically upregulated in Cluster 1 (**Fig. 2B, Table S5, Fig. S2**), whilst 12 families, including HERV-P, HERV-Rb, and HERV-XA34, were specifically silenced in Cluster 2, both strikingly DE in ME/CFS patients from subgroup 2 (**Fig. 2C, Table S5, Fig. S2**). Regarding FM, three families consisting of HERV-T, LTR84, and PRIMA41, were found specifically upregulated (Cluster 3) (**Fig. 2D, Table S5, Fig. S2**). Five families were commonly downregulated across ME/CFS, FM, and co-diagnosed groups, including HERV-W, HERVK14C, and MER67 (**Fig. 2E, Table S5, Fig. S2**). Contrary to the literature (Ovejero et al., 2020), reporting upregulation of HERV-W family in FM, our results showed downregulation of HERV-W family in all patient groups, more accentuated in FM and co-diagnosed groups than ME/CFS with respect to healthy levels (**Fig. 2E**). In all, these results reveal family-specific derangement of HERV expression in immune cells of FM and ME/CFS but not in co-diagnosed cases.

In light of the potential involvement of HERV-encoded proviral proteins in disease development (Giménez-Orenga and Oltra, 2021a) and to deepen the analysis of disease-specific HERV fingerprints, we then determined the proportion of these HERV fingerprints corresponding to solitary and proviral elements, as well as the different subdomains of the HERV structure according to probe detection in the microarrays i.e., LTR containing regulatory sequences *vs.* gag, pol, and env protein-coding genes. Although functional information of these

sequences is quite limited, the nature of the detected HERV motifs may inform of potential regulatory mechanisms leading to or perpetuating disease status (e.g., toxic env proteins or non-coding regulatory sequences). In agreement with hg38 HERV genomic data, it was found that the great majority of the dysregulated HERV loci corresponded to solitary LTR elements, which are known to be the most abundant in our genome, with few exceptions corresponding to proviral HERV (**Fig. 2F, Fig. S3**).

Lastly, given the potential of HERV LTRs to influence the expression of neighboring genes through the regulatory elements they hold, we examined the genomic context of the DE HERV loci by cluster (**Fig. 2G, Fig. S4**). Although not allocating significant association to promoter regions, repressed loci in ME/CFS (Cluster 2) and overexpressed loci in FM (Cluster 3) were significantly associated with gene 3' endings, and silenced loci in the three pathologies (Cluster 4) significantly associated with intronic regions. These results suggest that the molecular mechanisms by which disease-specific deregulated HERV loci could be participating in the development or maintenance of disease likely involve changes in transcript turnover or availability of alternative splicing events. In addition, DE analysis of intergenic solo LTRs show that they are underexpressed in ME/CFS (Fig. S5), pointing at the participation of epigenetic regulatory mechanisms in the disease.

HERV fingerprints associate with abnormal immune gene expression

Since disease-associated HERV derangement could indicate, even mediate, differences in global gene expression, we proceeded to analyze potential associations between disease-associated HERV fingerprints and a set of genes involved in eight pathways (immunity, inflammation, cancer, central nervous system disorders, differentiation, telomere maintenance, chromatin structure, and gag-like genes) detected by probes present in this same HERV-V3 high-density microarrays (Becker et al., 2017). Differential transcript analysis revealed 368

genes (1,037 probesets) DE between at least two study groups (FDR<0.1 and |log2FC|>1) (**Table S6)**. In line with DE HERV expression, ME/CFS immune cells showed a more accused dysregulation in gene expression profiles compared to FM´s and co-diagnosed groups (**Fig. S6A**). Hierarchical clustering of samples and PCA based on DE genes, however, did not cluster all samples by study group as HERV signatures did, with the only exception of ME/CFS subgroup 2, which exhibited a markedly different gene expression profile further supporting its distinction from the remaining study groups (**Fig. S6B**).

To uncover potentially affected physiologic functions linked to DE HERV, we studied the relationships between DE HERV and gene transcriptomes. First, we performed hierarchical clustering of DE sequences and identified sets of HERV-genes with correlated expression patterns (modules) (**Fig. 3**) and then conducted functional analysis of DE genes by gene ontology (GO) for each of the modules identified. Hierarchical clustering analysis yielded eight modules with different HERV and gene compositions exhibiting DE patterns across study groups. In particular, 37 HERV (19 from Module 1 and 18 from Module 6) with decreased expression levels across the three disease groups, correlated with several genes involved in the differentiation of immune cells and regulation of cell-cell adhesion (**Fig. 3A)**. Remarkably, ME/CFS exhibited upregulation of 258 HERV loci with expression levels that correlated with 34 genes involved in alpha-beta T cell activation and T-helper 17 cell commitment-related genes (Module 3). In addition, 78 HERV downregulated in ME/CFS correlated with pathogen detection-related genes (**Fig. 3B,** GO:0016045, GO:0016032, GO:0098543), suggesting a special involvement of the immune system in the disease. Lastly, several HERV and genes exhibited upregulated expression in FM whilst appearing downregulated in ME/CFS. These FM-specific DE expressed genes are involved in telomere structure organization and maintenance (Modules 2 and 5), cellular response to glucocorticoids (Module 7), and regulation

of NF-kB signaling (Module 8) (**Fig. 3C**), indicating divergent immune states in ME/CFS and FM with a differential impact in inflammatory processes.

Co-expression of DE HERV and immune response genes predominantly occur by mechanisms other than co-transcription

To further investigate the relationship between HERV dysregulation and immune abnormalities, we mapped the genomic localization of HERV and genes within each module, looking for genomic overlaps that could explain their correlation by co-expression. Surprisingly, the results showed no major overlaps of DE HERV and their correlated DE genes (**Fig. 4A**), with distances that separate them at least 100 kbp (**Fig. 4B, Table S7**), thereby suggesting overall independent expression of DE HERV and DE genes. As an exception, it was noticed that some elements from the MLT1 family co-localized with DE genes within a 30 kbp window (**Table S7**). For example, the element MLT1_5q32, from Module 1, appears located 18114 bp upstream of the *CD74* gene, involved in MHC class II antigen processing, or the element MLT1_4q24, belonging to Module 6, lies within *NFκB1* first intron. Both HERV loci and their co-localized genes exhibited downregulation in ME/CFS, FM, and co-diagnosed groups ($p<0.05$). Likewise, an exception was noticed for the MLT1_8q12.1 and MLT1_Xp22.3 HERV elements in Module 2, located within *LYN*'s tenth intron and *TLR8*-long isoform's second exon, respectively. Their genomic co-localization may explain that their specific lower expression levels in the ME/CFS subgroup 2 occur as result of co-transcription (**Fig. 4C**, $p<0.05$).

Since common *trans*-regulatory mechanisms may explain correlation by co-expression in the absence of co-transcription, we analyzed DE HERV in each module for potential enrichment in transcription factor binding sites (TFBS) by comparing the annotated TFBS within dysregulated HERV genomic regions against the full set of TFBS in the genome (based on publicly available

human ChIP-seq data provided by ReMap2022) (**Fig. 4D**). Then, we performed GO analysis for HERV-gene sets enriched TFBS to infer their function (**Fig. S7**). Interestingly enough, the results showed enrichment of binding sites for transcription factors involved in the immune response in modules 1 through 7, further supporting the role of immune disturbances in ME/CFS, FM, and co-diagnosed cases, while no significant enrichment of TFBS was found in module 8. In addition, HERV downregulated across the three patient groups (Modules 1 and 6) were enriched in TFBS involved in angiogenesis, while those upregulated in FM but downregulated in ME/CFS (Modules 2, 5, and 7), were enriched in TFBS for miRNA transcription and glucocorticoid signaling. It is worth mentioning that HERV upregulated in ME/CFS (Module 3) were markedly enriched with binding sites for chromatin remodeling factors, from which SETDB1 and TRIM28 stand out as key epigenetic repressors of HERV expression and CTCF for its importance in the establishment of topologically associating domains in chromosomes. These results therefore support the possibility that DE HERV expression occurs independently, but in coordination with their correlated DE genes by yet to define mechanisms, likely driven by and/or driving epigenetic changes in immune cells of ME/CFS patients.

DE HERV strongly associates with plasma and T-cell levels

Given the detected relationships between HERV dysregulation and immunological disturbances, we investigated differences in immune cell populations across the study groups by performing CIBERSORTx analysis on normalized microarray gene expression data. CIBERSORTx is an analytical tool that provides an estimation of cell type abundance using gene expression data (Newman et al., 2019). The results point at a decrease in γδ T cells as well as an increase in resting CD4 T memory and plasma cells ($p<0.05$) in ME/CFS patients from subgroup 2 (**Fig. 5A**). To further understand the relationship between gene signatures and

the potential differences in immune cell populations with DE HERV, we analyzed correlations across HERV and genes included in Modules 1 through 8 and immune cell proportions, finding that dysregulated HERV loci moderately to strongly correlate with plasma cell and resting CD4 memory T cells increase, and γδ T cell population decrease ($|R|>0.4$, $p<0.01$) (**Fig. 5B**). Curiously, HERV and genes upregulated in ME/CFS subgroup 2 (module 3), negatively correlate with γδ T cell abundance (e.g., LTR65_20p11.21, $R=0.57$, $p=6.3e-05$) and positively correlate with resting CD4 memory T cells, and plasma cell abundances (e.g., LTR65_20p11.21, $R=0.57$, $p=6.3e-05$; MER49_11q13.1, $R=-0.67$, $p=1e-06$) (**Fig. 5C**), while the tendency of correlation appears reversed for the other modules. These results suggest a potential relationship between HERV expression and immune cell fractions in a subgroup of ME/CFS patients.

HERV and immune alterations associate with ME/CFS diagnosis

Lastly, in an effort to detect potential relationships between deregulated HERV and genes, and patient diagnosis, we analyzed the association between DE of HERV and genes belonging to each module and the diagnosis received. Interestingly, activation of HERV and genes involved in immune response (Module 3) seem to strongly associate with diagnosis of ME/CFS (subgroup 2) ($|R|>0.89$, $p<0.001$) (**Fig. 6A**). In fact, top correlated genes and HERV with ME/CFS diagnosis (**Fig. 6B**), showed the strongest association with disease symptoms (**Fig. 6C**), particularly with some domains of the MFI and SF-36 questionnaires, including physical and mental fatigue ($p<0.05$). These results suggest the participation of these HERV and genes in ME/CFS disease while uncovering their potential as biomarkers.

DISCUSSION

Researchers have strived for decades to elucidate the etiology and pathophysiology of ME/CFS and FM, with disease complexity and patient heterogeneity hindering this task. With no biomarkers available and unclear pathophysiology, ME/CFS and FM are often misdiagnosed, further complicating the identification of effective treatments. In this study, we tried to overcome patient heterogeneity by studying a very finely-phenotyped cohort diagnosed by a single expert clinician, representing one of the few studies (Nepotchatykh et al., 2023) including samples of patients diagnosed with either ME/CFS, FM, or both (co-diagnosed group) by two different clinical criteria (Canadian (Carruthers et al., 2003) and International Consensus (Carruthers et al., 2011) criteria, and the 1990 (Wolfe et al., 1990) and 2011 (Wolfe et al., 2016, 2010) American College of Rheumatology (ACR) criteria, respectively). Due to previously reported sex-associated differences in these diseases, only female patients were included, limiting the external validity of the results in the male population while minimizing potential sex-associated bias.

Triggering agents for ME/CFS and FM, such as infections, stress, or trauma (Chu et al., 2019; Furness et al., 2018), seem capable of disrupting the epigenetic mechanisms constraining HERV expression (Rangel et al., 2022). Dysregulation of HERV has been proposed to contain more distinctive features than the host transcriptome itself in response to viral infections (Marston et al., 2021), and it is becoming detected in additional complex diseases (Giménez-Orenga and Oltra, 2021a), offering a new perspective to study their physiopathology. In fact, HERV profiling could constitute an indirect approach (surrogate marker) to evidence epigenetic derangements and 3D alterations of the chromatin structure with obvious functional consequences. Previous research revealed aberrant activation of HERV in ME/CFS (Rodrigues et al., 2019) and FM (Ovejero et al., 2020). However, these two studies consisting of broad amplification of HERV by degenerate primer sets, did not allow for specific HERV loci identification. In this study, we fill this gap of knowledge by genome-wide analyzing annotated

HERV sequences in the human genome and their relationship with immune gene expression and patient symptoms. The results show for the first time a comparison of peripheral blood mononuclear cells (PBMC) HERV expression profiles of ME/CFS, FM, and co-diagnosed patient groups as compared to HERV expression profiles of healthy subjects. Despite the limited number of samples (8 ME/CFS cases, 10 FM, 16 comorbid cases, and 9 matched healthy controls), our results show a specific HERV fingerprint for each disease that allows perfect discrimination into three distinct patient groups and their separation from healthy participants, opening the possibility of HERV expression profiles-based diagnostic of these symptom-related, diseases. Methodology that could be extended to the study of other chronic diseases such as post-COVID-19 condition. Particularly interesting will be to determine whether post-COVID-19 cases present different HERV fingerprints that justify commonalities to ME/CFS and/or FM, or if by contrast show an evolution closer to the more prevalent co-diagnosed state. Worth noting are the results obtained in the co-diagnosed group which, unexpectedly, presented a completely different HERV profile to those identified for ME/CFS and for FM while strikingly closer to that of healthy controls. Findings that support a distinct subjacent pathomechanism for the co-diagnosed group of patients and, therefore, the identification of a novel nosologic entity.

In addition, our study reveals enhanced HERV dysregulation in the ME/CFS group when compared to FM, co-diagnosed cases, or healthy controls and significant quantitative differences within a more severely affected ME/CFS subgroup (subgroup 2) supporting inter-individual HERV expression variability towards patient subtyping, with potential relevance for patient tailored treatments, and potential for HERV expression levels to define degree of severity in ME/CFS. The study also defined a set of DE HERV commonly downregulated in all three patient groups (ME/CFS, FM, and their co-diagnosis, Cluster 4), a deeper study of which may help explain clinical commonalities between these conditions. In fact, the lack of a

specific association of Cluster 4 DE HERV with a patient-study group opens the possibility of their derangement in additional complex and chronic diseases.

Analysis of DE HERV loci at family level within each cluster interestingly showed the involvement of different HERV families in each of the study groups, with ME/CFS outstanding by the largest heterogeneity, including deregulation in families such as HERV-H or HERV-P. Deregulation of HERV-H and HERV-P families has been previously linked to disease (Bergallo et al., 2019; Morris et al., 2019; Yi et al., 2006) with potential involvement in immune function performance (Goodchild et al., 1993; Kulski et al., 1999; Mager et al., 1999; Mangeney et al., 2001; Matsuzawa et al., 2021). Another HERV family consistently linked to disease is the HERV-W (Dolei et al., 2014; Giménez-Orenga et al., 2022; Kremer et al., 2013), whose upregulation has been reported in FM (Ovejero et al., 2020). Contrarily, our results evidenced downregulation of HERV-W family in all patient groups, particularly the HERV-W-LTRSU3_19q13.42_st element. Dissimilar results may be due to the use of different techniques to evaluate family expression (degenerate primers in RT-qPCR vs. probes in microarray), or accuracy in patient diagnosis. Strikingly, our study showed that most downregulated HERV in ME/CFS, FM, and co-diagnosed groups belonged to a sole HERV family, the MLT1, usually found enriched near DE genes (Bogdan et al., 2020) contributing to the coding of conserved lncRNAs (Ramsay et al., 2017), whether these lncRNA play a role in these diseases is merely speculative at this point.

In accordance with genomic data (Consortium et al., 2001), most deregulated HERV loci consisted of solitary LTRs, known for their ability to influence gene expression through their regulatory sequences (Ito et al., 2017). For example, a HERV-H solitary LTR located at the *HHLA2* locus serves as its main polyadenylation signal (Goodchild et al., 1993; Mager et al., 1999), allowing the encoding of the HHLA2 protein, which regulates CD4 and CD8 T cell

functions along with antigen-presenting cell proliferation (Zhao et al., 2013). Furthermore, it is known that non-allelic homologous recombination events leading to solitary LTRs create inter-individual variation, not only by removing coding potential but also by altering the *cis*-regulatory or transcriptional activity of the affected particular genomic regions (Rebollo et al., 2012; Thomas et al., 2018) even causing microdeletions (Hermetz et al., 2012; Sanchez-Valle et al., 2010; Shuvarikov et al., 2013), which offer possibilities for exploring familial risk factors of these diseases.

Analysis of differential gene expression in the exact same samples provided additional evidence of HERV-associated immune changes in the ME/CFS group as compared to FM, co-diagnosed, or healthy control groups, further supporting the existence of two ME/CFS subgroups differing, once more, at the quantitative level only.

Although genomic localization of correlated HERV and genes suggested major mechanisms other than co-transcription, it should be noted that gene probes included in the microarray do not fully picture the complete mRNA transcriptome, and therefore the possibility that other DE genes locate nearer to deregulated HERV cannot be ruled out. Furthermore, long-range interactions between gene promoters and LTR regulatory elements have been described for some HERV families, including MLT1, MER21, MER41, or LTR54 (Ito et al., 2017), reported as DE in this study. In this regard, it has been shown that the HERV-H family, upregulated in ME/CFS subgroup 2, can influence gene expression over long genomic distances, for example by creating topologically associated domains (Zhang et al., 2019), perhaps explaining at least some of our findings of DE immune genes being co-expressed with DE HERV. Under the hypothesis that independently transcribed DE HERV could be regulated by the same *trans*-acting mechanisms as DE genes, we interestingly found enrichment for TFBS regulating the immune response within DE HERV, suggesting *trans*-mediated coordination of deranged

HERV and gene expression. Furthermore, HERV upregulated in ME/CFS were markedly enriched in chromatin remodeling factor binding sites, including SETDB1 and TRIM28, both being key epigenetic repressors of HERV expression (Giménez-Orenga and Oltra, 2021b), pointing to the impairment of TRIM28/SETDB1 axis as the underlying mechanism of HERV upregulation in ME/CFS. In support of this possibility, influenza A virus, a suspected trigger of ME/CFS (Magnus et al., 2015), has been shown to activate HERV expression by impairing the repressor TRIM28/SETDB1 axis (Schmidt et al., 2019).

As the literature endorses changes in immune cell subsets in ME/CFS (Brenu et al., 2014, 2011; Curriu et al., 2013), we aimed to indirectly measure potentially skewed immune cell type ratios in our study groups with the use of CIBERSORTx software (Newman et al., 2019). The detected increase in plasma cells and resting CD4 T memory cells in ME/CFS subgroup 2, suggests the presence or the history of previous infections stimulating the immune system in these subjects, coinciding with greater dysregulation in immune-related genes and HERV upregulation and increased severity of symptoms. On another hand, this same subgroup of patients exhibited lower levels of γδ T cells, an event reported to our knowledge in other autoimmune diseases like systemic lupus erythematosus (SLE) and psoriasis (Fay et al., 2016; Wang et al., 2012). It has been described that γδ T cells mediate the immune response against viral infections like EBV, HHV-6, or CMV (Zhao et al., 2018), which together with natural killer cell dysfunction reported in the mentioned autoimmune diseases (Henriques et al., 2013; Kim et al., 2019) but also in ME/CFS (Eaton-Fitch et al., 2019), suggests an impaired cytotoxic function rendering patients with lower counts of this cell type, becoming more susceptible to viral reactivation. Interestingly, anomalies in immune cell abundance highly correlated with observed HERV expression patterns, indicating that HERV deregulation may either reflect expression derangement in those cell subpopulations or somehow influence immune cell ratios. The analysis of PBMC, rather than isolated subpopulations, is limited at distinguishing between

these non-exclusive possibilities. In addition, we found strong associations between the HERV elements MER31_3p26.1 and HERV16_3p21.31 expression correlating with physical and mental fatigue according to health status assessment instruments MFI (Smets et al., 1995) and SF-36 (MCHORNEY et al., 1993), where higher scores indicate a worse or better health status, respectively. In this regard, both HERV elements, belonging to module 3, i.e., upregulated in ME/CFS subgroup 2, positively and negatively correlated with MFI and SF-36 subdomains, respectively, indicating that increased expression of MER31_3p26.1 and HERV16_3p21.31 elements may affect patient health status. Interestingly enough, previous correlations between HERV expression and physical fatigue are reported in the post-viral syndrome Post-COVID-19 condition (Giménez-Orenga et al., 2022), which shares symptomatology with ME/CFS (Komaroff and Lipkin, 2023). Further study of MER31_3p26.1 and HERV16_3p21.31 elements and their potential pathogenic function seem granted.

Overall, this study not only shows the potential of HERV expression profiles as biomarkers for ME/CFS, FM, and for the definition of ME/CFS and FM co-diagnosis, but also exposes their potential involvement in immune anomalies and connection with patient symptoms in these diseases, with etiological potential. Our results reveal family-specific HERV deregulation in ME/CFS and FM immune cells and their association with changes in pathogen-detection genes and immune cell subsets. Recent findings by Marston et al. (Marston et al., 2021) demonstrated specific disruption of HERV families in response to specific viral infection agents. Perhaps a database enabling the comparison of disease HERV expression profiles with those induced by exogenous infections could help reveal the triggering agents behind diseases of suspected viral etiology, such as ME/CFS and FM.

In summary, this study pioneers the comparison of female ME/CFS, FM and co-diagnosed HERV expression profiles showing the superior diagnostic potential of HERV fingerprints over

immune transcriptomes, with clinical implications, importantly providing an effective tool for the objective differential diagnosis of these hard-to-diagnose co-occurring diseases. The potential mechanisms point at underlying epigenetic disturbances strikingly derepressing HERV silencing, particularly in ME/CFS. Whether these disturbances primarily affect certain immune subpopulations, estimated decreased for γδ T cells, or increased for plasma and resting CD4 memory T cells, correlating with patient symptom severity in ME/CFS, or reflect the presence of DEHERV profiles in immune cell subpopulations biased in these patient groups is an aspect that awaits further work.

MATERIALS AND METHODS

Study design. This cross-sectional observational study was approved by the Public Health Research Ethics Committee DGSP-CSISP of Valencia, núm. 20190301/12, Valencia, Spain. The study included a total of 43 female patients invited to participate from local patient associations who were clinically diagnosed with ME/CFS (n=8), FM (n=10), or both (n=16) (National Biobank Registry Ref. 0006024) and 9 healthy control individuals' population-matched for age and BMI (National Biobank Registry Ref. 0006034). Patients were diagnosed by an FM and ME/CFS specialized clinician (Hospital de Manises, Valencia, Spain) using the 1990 (Wolfe et al., 1990) and 2011 (Wolfe et al., 2016, 2010) American College of Rheumatology (ACR) criteria for FM and the Canadian (Carruthers et al., 2003) and International Consensus (Carruthers et al., 2011) criteria for ME/CFS diagnosis. Patients with health problems other than FM and ME/CFS were excluded from the study. Individuals with any similar o related pathology, including a medical history of chronic pain and/or fatigue, or serious health complications, were excluded from control group, as well as medicated healthy controls. Written informed consent was obtained from all study participants and patient health status was also evaluated with the use of standardized questionnaires, including the FIQ case

report form (Burckhardt et al., 1991; Rivera and González, 2004), the MFI questionnaire (Smets et al., 1995), and the quality-of-life SF-36 instrument (MCHORNEY et al., 1993). Participating patients agreed to withdraw medication at least 12 h prior to blood draw. Microarray analysis of RNA extracted from peripheral blood mononuclear cells was performed along with complementary bioinformatic analysis to evaluate differences in HERV profiles among patients with potential application as biomarkers or indicators of the underlying pathomechanisms.

Isolation of peripheral blood mononuclear cells and total RNA extraction. Up to 10 mL of whole blood were collected via venipuncture after a 12 h overnight fasting in K2EDTA tubes (Becton Dickinson, Franklin Lakes, NJ, USA) and processed within 2 hours by dilution at 1:1 (v/v) ratio in phosphate-buffered saline solution (PBS) with layering on top of 1 volume of Ficoll-Paque Premium (GE Healthcare, Chicago, IL, USA) and separation by density centrifugation at 500× g for 30 minutes (20°C, brakes off). The PBMC layer was isolated and washed with PBS and resuspended in red blood cell lysis buffer (155 mM NH4Cl, 10 mM NaHCO3, 0.1 mM EDTA, and pH 7.4), kept on ice for 5 min, and centrifuged (20 C at 500 g for 10 min), to remove contaminating erythrocytes. The washed pellets were adjusted to a final concentration of 10^7 cells/mL in freezing medium (90 % FBS, 10 % DMSO), aliquoted, and deeply frozen in liquid nitrogen until use. Total RNA was extracted using RNeasy Mini Kit (Qiagen, MD, USA) according to the manufacturer's instructions. RNA quality was assessed using Agilent TapeStation 4200 (Agilent). All RNA samples had an RNA Integrity number (RIN) above 7.

Transcriptome analysis by microarray. To minimize batch effects the samples were anonymized (blinded to the operator) and scrambled across groups. HERV transcriptome was scrutinized using custom high-density HERV-V3 microarrays (Becker et al., 2017), capable to discriminate 174,852 HERV elements, 179,142 MaLR elements, and putative active 1,072

LINE-1 elements at the locus level, in addition to detecting a set of 1,559 genes involved in eight potentially relevant cellular pathways (immunity, inflammation, cancer, central nervous system affections, differentiation, telomere maintenance, chromatin structure, and gag-like genes). These custom HERV-V3 arrays also allow to discern the different subdomains within proviral sequences: 3' or 5' LTRs and *gag/pol/env* regions of proviral HERV and solitary LTRs originated after non-allelic recombination events between the two LTRs causing complete proviral elimination(Thomas et al., 2018). To analyze HERV transcriptomes of PBMC-derived RNA samples, cDNA was synthesized and amplified from 45 ng of RNA using the Ovation Pico WTA System V2 kit (Nugen) according to manufacturer's instructions. The resulting amplified ssDNA was purified using the QIAquick purification kit (Qiagen, MD, USA). Total DNA concentration was measured with NanoDrop 2000 spectrophotometer (Thermo Scientific) and quality was assessed on the Bioanalyzer 2100. Five micrograms of purified ssDNA were enzymatically fragmented into 50–100 bp fragments and biotin-labeled with the Encore Biotin Module kit from Nugen according to manufacturer's instructions. The resulting target was mixed with standard hybridization controls and B2 oligonucleotides following supplier recommendations. The hybridization cocktail was heat-denatured at 95 ºC for 2 min, incubated at 50 ºC for 5 min, and centrifuged at 16,000 g for 5 min to pellet the residual salts. HERV-V3 microarrays were pre-hybridized with 200 µL of hybridization buffer and placed under stirring (60 rpm) in an oven at 50 C for 10 min. Hybridization buffer was then replaced by denatured hybridization cocktail. Hybridization was performed at 50 °C for 18 h in the oven under constant stirring (60 rpm). Washing and staining were carried out according to manufacturer´s protocol, using a fluidic station (GeneChip fluidic station 450, Affymetrix). Arrays were finally scanned using a fluorometric scanner (GeneChip scanner 3000 7G, Affymetrix). Biotin purification, hybridization, and reading steps were performed by Sampled (Piscataway, NJ 08854, USA). Overall, 1,397,352 probes were detected, the vast majority of them

corresponding to HERV (1,290,800 probesets), followed by genes (103,724 probesets), and LINE-1 (2,828 probesets).

Identification of DE HERV and genes. All bioinformatic analysis were performed with RStudio software version 4.2.1. Microarray CEL files were processed and analyzed using R oligo package (Carvalho and Irizarry, 2010). Data were normalized, adjusted for background noise, and summarized using the RMA (Robust Multi-Array) algorithm. DE analysis was performed using limma R package (Ritchie et al., 2015), considering DE those probes with a "Benjamin-Hochberg" (BH) adjusted *p* value (FDR)<0.1 and an absolute $\log_2$ fold-change>1. The analysis results were presented in volcano plots using EnhancedVolcano R package (Blighe et al., n.d.).

Clustering analysis. DE HERV or genes (FDR<0.1 and |Log2FC|>1) were used to perform clustering analysis. Principal component analysis (PCA) was performed to analyze the behavior of samples under study based on DE HERV expression profiles by using factoextra R package (Kassambara and Mundt, n.d.). Samples were plotted according to the first two PCA principal components and the contribution of variables to each of the principal components was represented in barplots. Scaled expression data of DE HERV or gene probesets were used to cluster samples and probes. Data was represented in heatmap plots by using pheatmap R package (Kolde et al., n.d.). Rows and columns were clustered using Euclidean and Pearson correlation distances, respectively. HERV clusters were visually identified based on heatmap results and extracted from dendrograms using the pheatmap implemented function. Mean probe expression by group was calculated for all probes encompassed in each cluster and their distribution and expression level were visualized by box and violin plots. Average-linkage agglomerative hierarchical clustering with Pearson correlation distance was performed on normalized expression data to identify modules of HERV and genes with correlated expression

patterns across samples. Modules of HERV-genes were produced by applying a dynamic tree cut of dendrograms using a minimum cluster size of 110 HERV or genes, as described by Marston et al., 2021 (Marston et al., 2021).

Enrichment analysis. Microarray probes were annotated based on the HERV-V3 annotation file kindly provided by bioMérieux, allowing retrieval of HERV subdomain (gag, pol, env, or LTR), family, and genomic location of each probe. Family group information for solitary LTR probes was added based on Dfam and EnHERV databases. Original hg37 HERV-V3 coordinates were converted into hg38 genomic coordinates by using UCSC LiftOver tool (Kuhn et al., 2013) and the corresponding chromosome band information was added. The genomic context of the HERV loci was assessed with Goldmine R package (Bhasin and Ting, 2016). Enrichment analysis for HERV family, HERV subdomain, and genomic context was performed for the HERV encompassed in each cluster by comparing their frequency in the cluster to their corresponding on hg38 genomic data according to HERV-V3 microarray annotation (Becker et al., 2017). The frequency of families, HERV subdomains, and genomic context (intron, exon, intergenic, promoter, or 3' end) for the DE HERV encompassed in each cluster and in the overall hg38 genomic data were represented as stacked bar plots with ggplot2 R package (Wickham, 2016). To assess for transcription factor binding site (TFBS) enrichment, the ReMapEnrich Shiny 1.4 web interface was used providing DE HERV loci genomic coordinates as input. Enriched TFBS were represented as a heatmap using pheatmap R package (Kolde et al., n.d.), with colors defining modules in which the enrichment was found, and no enrichments were shown as blank.

Overrepresentation functional analysis. Genes encompassed in modules identified by hierarchical clustering as well as TF binding to identified enriched TFBS in HERV of each module were functionally analyzed by Gene Ontology (GO) using the R package clusterProfiler

(Wu et al., 2021) with an adjusted p-value cutoff of 0.05. For each module, the top 5 GO terms as ranked by increasing adjusted p-value were visualized in a heatmap using pheatmap R package (Kolde et al., n.d.). Statistical significance of each GO term as adjusted p-values was represented by a color gradient, with darker colors representing more significant terms in the module. Terms sharing significant overrepresentation in other modules (adjusted $p<0.05$) were colored on heatmaps accordingly.

Evaluation of overlaps between HERV and gene sequences. Genomic coordinates of HERV and genes encompassed in each module were transformed to genomic range objects (Granges) using the GenomicRanges R package (Lawrence et al., 2013). They were subsequently plotted over a linear chromosomal representation with karyoploteR package (Gel and Serra, 2017), which displays the density of sequences throughout chromosomes. The genomic distance between HERV and genes in the same module was determined with a specific function included in GenomicRanges package and displayed as density plots.

Measurement of immune cell relative amounts. Digital cytometry analysis was performed with CIBERSORTx (Newman et al., 2019) software on normalized non-logarithmic gene expression data for all samples. The default LM22 leukocyte gene signature matrix(Newman et al., 2015) was used as a reference. LM22 was generated using Affymetrix HGU133A microarray data distinguishing 22 human hematopoietic cell phenotypes, including different types of T cells, B cells (naïve and memory), plasma cells, NK cells, monocytes, dendritic cells and some myeloid subsets (Newman et al., 2015). Output results were visually inspected for potential differences in immune cell counts and significant differences were assessed by Welch's t-test with adjusted p-value<0.05 from specific immune cell types. Results were plotted as violin plots.

Trait-measure and symptom association. For each module extracted by hierarchical clustering the eigengenes were calculated with the use of the function moduleEigengenes implemented in WGCNA package (Langfelder and Horvath, 2008). A matrix of normalized expression data and a list assigning each sequence to each module was provided as input. Calculated module eigengenes were then correlated to external microarray data, including immune cell proportion values provided by CIBERSORTx, questionnaire scores for FIQ, MFI, and SF-36, or the clinical diagnosis. This last variable was binarized with the use of binarizeCategorialColumns. Results were plotted by using pheatmap package (Kolde et al., n.d.) with boxes showing Pearson correlation values and associated p-values between gene expression levels of each module and the disease trait. A correlation value of 1 (orange or green) and -1 (blue) were used to quantify the strongest positive and negative correlations, respectively, while 0 (white) was assigned for no correlation.

Real-time quantitative Polymerase Chain Reaction (RT-qPCR). DE of specific HERV sequences was validated using the same RNA samples analyzed by microarray (n=8 controls, n=8 ME/CFS, n=7 FM, and n=14 co-diagnosed). Reverse transcription was performed using High-Capacity cDNA reverse Transcription kit (Applied Biosystems, Waltham, MA, USA, cat. 4308228), with 1µg of total RNA according to manufacturer's guidelines. cDNAs were used for Real-time PCR using PowerUP Sybr Green Master Mix (Applied Biosystems, cat. 100029283) and a Lightcycler LC480 instrument (Roche, Penzberg, Germany). Standard amplification conditions were applied, including a single hotstart polymerase preactivation cycle at 94 ◦ C for 15 min, up to 45 amplification cycles, each one consisting of three steps: denaturation at 95 ◦ C for 15 s, annealing at 60 ◦ C for 30 s and extension at 70 ◦ C for 30 s. Sequences of specific primers used are detailed in **Table S8**. GAPDH levels were used for relative quantification of the RNAs amplified, and $2^{-\Delta Ct}$ analysis to calculate expression was applied.

Statistical analysis. All statistical analyses were done in R v4.2.1. Data distributions were tested for normality. Normally distributed data were tested using two-tailed unpaired Student's t-tests; non-normal data were analyzed with non-parametric statistical test, as detailed. For enrichment analyses, we used a Fisher exact test to calculate p-value, considering enriched in the provided list if an adjusted p-value (FDR) was less than 0.05. We chose to use Fisher's exact test in the analysis of contingency tables to compare ME/CFS and healthy group, as it is more appropriate for small sample sizes in comparison to the chi-square test or G-test of independence.

DECLARATIONS

Ethics approval and consent to participate: This study has been performed in accordance with the Declaration of Helsinki. It was approved by the Public Health Research Ethics Committee DGSP-CSISP of Valencia, núm. 20190301/12, Valencia, Spain. All participants signed an informed consent to participate in this study.

Availability of data and materials: All data associated with the study are available in the paper or supplementary materials, except for HERV-V3 microarray annotation provided by bioMérieux under material transfer agreement. Microarray data will be made available from the NCBI GEO database at publication. Bioinformatic analysis codes will be made available upon request.

Competing interests: 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.

Funding: This study was funded by an ME Research UK (SCIO charity number SC036942) grant, and by Generalitat Valenciana CIAICO/2021/103 grant to EO. KG-O is supported by the

Generalitat Valenciana ACIF2021/179 grant. Funders were not involved in any of the research stages.

Authors' contributions: KG-O: investigation and data gathering, data analysis and interpretation, figure drawing, writing original draft, and manuscript review. EM-M: patient diagnosis and manuscript review. LN: data interpretation and manuscript review. EO: conceptualization, funding acquisition, supervision, formal analysis, investigation, data curation, writing original draft and manuscript review. All authors approved the submitted version.

Acknowledgments: we particularly thank all the patients who participated in the study and Dr. Vicente Serra (Umivale,Valencia, Spain) for his help in the recruitment of healthy volunteers. We also acknowledge bioMérieux (Ain, France) for allowing the use of their custom-made HERV-V3 high-density microarrays and Sampled LTD (Piscataway, NJ, USA) for assisting us in the analysis by microarray.

REFERENCES

Abbi B, Natelson BH. 2013. Is chronic fatigue syndrome the same illness as fibromyalgia: evaluating the 'single syndrome' hypothesis. *QJM: Int J Med* **106**:3–9. doi:10.1093/qjmed/hcs156Bateman L, Bested AC, Bonilla HF, Chheda BV, Chu L, Curtin JM, Dempsey TT, Dimmock ME, Dowell TG, Felsenstein D, Kaufman DL, Klimas NG, Komaroff AL, Lapp CW, Levine SM, Montoya JG, Natelson BH, Peterson DL, Podell RN, Rey IR, Ruhoy IS, Vera-Nunez MA, Yellman BP. 2021. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Essentials of Diagnosis and Management. *Mayo Clin Proc* **96**:2861–2878. doi:10.1016/j.mayocp.2021.07.004

Becker J, Pérot P, Cheynet V, Oriol G, Mugnier N, Mommert M, Tabone O, Textoris J, Veyrieras J-B, Mallet F. 2017. A comprehensive hybridization model allows whole HERV transcriptome profiling using high density microarray. *Bmc Genomics* **18**:286. doi:10.1186/s12864-017-3669-7

Bergallo M, Mareschi K, Montanari P, Calvi C, Berger M, Bini I, Daprà V, Galliano I, Fagioli F. 2019. Human endogenous retrovirus, HERV-P and HERV-R in pediatric leukemia patients. *J Hematop* **12**:51–56. doi:10.1007/s12308-019-00352-0

Bhasin JM, Ting AH. 2016. Goldmine integrates information placing genomic ranges into meaningful biological contexts. *Nucleic Acids Res* **44**:5550–5556. doi:10.1093/nar/gkw477

Blighe K, Rana S, Lewis M. n.d. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. https://github.com/kevinblighe/EnhancedVolcano

Bogdan L, Barreiro L, Bourque G. 2020. Transposable elements have contributed human regulatory regions that are activated upon bacterial infection. *Philosophical Transactions Royal Soc B* **375**:20190332. doi:10.1098/rstb.2019.0332

Brenu EW, Driel ML van, Staines DR, Ashton KJ, Ramos SB, Keane J, Klimas NG, Marshall-Gradisnik SM. 2011. Immunological abnormalities as potential biomarkers in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. *J Transl Med* **9**:81. doi:10.1186/1479-5876-9-81

Brenu EW, Huth TK, Hardcastle SL, Fuller K, Kaur M, Johnston S, Ramos SB, Staines DR, Marshall-Gradisnik SM. 2014. Role of adaptive and innate immune cells in chronic fatigue syndrome/myalgic encephalomyelitis. *Int Immunol* **26**:233–242. doi:10.1093/intimm/dxt068

Burckhardt CS, Clark SR, Bennett RM. 1991. The fibromyalgia impact questionnaire: development and validation. *J Rheumatology* **18**:728–33.

Carruthers BM, Jain AK, Meirleir KLD, Peterson DL, Klimas NG, Lerner AM, Bested AC, Flor-Henry P, Joshi P, Powles ACP, Sherkey JA, Sande MI van de. 2003. Myalgic

Encephalomyelitis/Chronic Fatigue Syndrome. *J Chronic Fatigue Syndr* **11**:7–115. doi:10.1300/j092v11n01_02

Carruthers BM, Sande MI van de, Meirleir KLD, Klimas NG, Broderick G, Mitchell T, Staines D, Powles ACP, Speight N, Vallings R, Bateman L, Baumgarten-Austrheim B, Bell DS, Carlo-Stella N, Chia J, Darragh A, Jo D, Lewis D, Light AR, Marshall-Gradisbik S, Mena I, Mikovits JA, Miwa K, Murovska M, Pall ML, Stevens S. 2011. Myalgic encephalomyelitis: International Consensus Criteria. *J Intern Med* **270**:327–338. doi:10.1111/j.1365-2796.2011.02428.x

Carvalho BS, Irizarry RA. 2010. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**:2363–2367. doi:10.1093/bioinformatics/btq431

Castro-Marrero J, Faro M, Aliste L, Sáez-Francàs N, Calvo N, Martínez-Martínez A, Sevilla TF de, Alegre J. 2017. Comorbidity in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis: A Nationwide Population-Based Cohort Study. *Psychosomatics* **58**:533–543. doi:10.1016/j.psym.2017.04.010

Charvet B, Brunel J, Pierquin J, Iampietro M, Decimo D, Queruel N, Lucas A, Encabo-Berzosa MDM, Arenaz I, Marmolejo TP, Gonzalez AI, Maldonado AC, Mathieu C, Küry P, Flores-Rivera J, Torres-Ruiz F, Avila-Rios S, Salgado Montes de Oca G, Schoorlemmer J, Perron H, Horvat B. 2023. SARS-CoV-2 awakens ancient retroviral genes and the expression of proinflammatory HERV-W envelope protein in COVID-19 patients. *iScience* **26**(5):106604. doi: 10.1016/j.isci.2023.106604.

Chu L, Valencia IJ, Garvert DW, Montoya JG. 2019. Onset Patterns and Course of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Front Pediatr* **7**:12. doi:10.3389/fped.2019.00012

Clauw DJ, Calabrese L. 2023. Rheumatology and Long COVID: lessons from the study of fibromyalgia. *Ann Rheum Dis* ard-2023-224250. doi:10.1136/ard-2023-224250

Consortium IHGS, Research: WI for BR Center for Genome, Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond Christina, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Centre: TS, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Center WUGS, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Institute: UDJG, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng J-F, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Center: BC of MHGS, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Center: RGS, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y,

Taylor T, UMR-8030: G and C, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Biotechnology: D of GA Institute of Molecular, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Center: GS, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Center: BGIG, Yang H, Yu J, Wang J, Huang G, Gu J, Biology: MSC The Institute for Systems, Hood L, Rowen L, Madan A, Qin S, Center: SGT, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Technology: U of OAC for G, Roe BA, Chen F, Pan H, Genetics: MPI for M, Ramser J, Lehrach H, Reinhardt R, Center: CSHL Lita Annenberg Hazen Genome, McCombie WR, Bastide M de la, Dedhia N, Biotechnology: GRC for, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen H-C, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JGR, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AFA, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang S-P, Yeh R-F, Health: S management: NHGRI US National Institutes of, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Center: SHG, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Center: U of WG, Olson MV, Kaul R, Raymond Christopher, Medicine: D of MB Keio University School of, Shimizu N, Kawasaki K, Minoshima S, Dallas: U of TSMC at, Evans GA, Athanasiou M, Schultz R, Energy: O of S US Department of, Patrinos A, Trust: TW, Morgan MJ. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–921. doi:10.1038/35057062

Curriu M, Carrillo J, Massanella M, Rigau J, Alegre J, Puig J, Garcia-Quintana AM, Castro-Marrero J, Negredo E, Clotet B, Cabrera C, Blanco J. 2013. Screening NK-, B- and T-cell phenotype and function in patients suffering from Chronic Fatigue Syndrome. *J Transl Med* **11**:68. doi:10.1186/1479-5876-11-68

Dolei A, Garson JA, Arru G, Clerici M, Germi R, Marche PN, Perron H. 2014. Multiple sclerosis-associated retrovirus and related human endogenous retrovirus-W in patients with multiple sclerosis. *J Neuroimmunol* **266**:87–88. doi:10.1016/j.jneuroim.2013.11.009

Dopkins N, O'Mara MM, Singh B, Marston JL, Bendall ML, Nixon DF. 2022. How human endogenous retroviruses interact with the microbiota in health and disease. *Trends Microbiol*. doi:10.1016/j.tim.2022.05.011

Eaton-Fitch N, Preez S du, Cabanas H, Staines D, Marshall-Gradisnik S. 2019. A systematic review of natural killer cells profile and cytotoxic function in myalgic encephalomyelitis/chronic fatigue syndrome. *Syst Rev* **8**:279. doi:10.1186/s13643-019-1202-6

Fay NS, Larson EC, Jameson JM. 2016. Chronic Inflammation and γδ T Cells. *Front Immunol* **7**:210. doi:10.3389/fimmu.2016.00210

Ferrari R, Grandi N, Tramontano E, Dieci G. 2021. Retrotransposons as Drivers of Mammalian Brain Evolution. *Life* **11**:376. doi:10.3390/life11050376

Furness PJ, Vogt K, Ashe S, Taylor S, Haywood-Small S, Lawson K. 2018. What causes fibromyalgia? An online survey of patient perspectives. *Heal Psychol Open* **5**:2055102918802683. doi:10.1177/2055102918802683

Gel B, Serra E. 2017. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**:3088–3090. doi:10.1093/bioinformatics/btx346

Giménez-Orenga K, Oltra E. 2021a. Human Endogenous Retrovirus as Therapeutic Targets in Neurologic Disease. *Pharm* **14**:495. doi:10.3390/ph14060495

Giménez-Orenga K, Oltra E. 2021b. Human Endogenous Retrovirus as Therapeutic Targets in Neurologic Disease. *Pharm* **14**:495. doi:10.3390/ph14060495

Giménez-Orenga K, Pierquin J, Brunel J, Charvet B, Martín-Martínez E, Perron H, Oltra E. 2022. HERV-W ENV antigenemia and correlation of increased anti-SARS-CoV-2 immunoglobulin levels with post-COVID-19 symptoms. *Front Immunol* **13**:1020064. doi:10.3389/fimmu.2022.1020064

Goodchild NL, Wilkinson DA, Mager DL. 1993. Recent Evolutionary Expansion of a Subfamily of RTVL-H Human Endogenous Retrovirus-like Elements. *Virology* **196**:778–788. doi:10.1006/viro.1993.1535

Groven N, Reitan SK, Fors EA, Guzey IC. 2021. Kynurenine metabolites and ratios differ between Chronic Fatigue Syndrome, Fibromyalgia, and healthy controls. *Psychoneuroendocrinology* **131**:105287. doi:10.1016/j.psyneuen.2021.105287

Gruchot J, Lewen I, Dietrich M, Reiche L, Sindi M, Hecker C, Herrero F, Charvet B, Weber-Stadlbauer U, Hartung H-P, Albrecht P, Perron H, Meyer U, Küry P. 2023. Transgenic expression of the HERV-W envelope protein leads to polarized glial cell populations and a neurodegenerative environment. *Proc Natl Acad Sci* **120**:e2308187120. doi:10.1073/pnas.2308187120

Guo X, Zhao Y, You F. 2024. Identification and characterization of endogenous retroviruses upon SARS-CoV-2 infection. *Front Immunol*.**15**:1294020. doi: 10.3389/fimmu.2024.1294020.

Harrison JE, Weber S, Jakob R, Chute CG. 2021. ICD-11: an international classification of diseases for the twenty-first century. *BMC Méd Inform Decis Mak* **21**:206. doi:10.1186/s12911-021-01534-6

Henriques A, Teixeira L, Inês L, Carvalheiro T, Gonçalves A, Martinho A, Pais ML, Silva JAP da, Paiva A. 2013. NK cells dysfunction in systemic lupus erythematosus: relation to disease activity. *Clin Rheumatol* **32**:805–813. doi:10.1007/s10067-013-2176-8

Hermetz KE, Surti U, Cody JD, Rudd MK. 2012. A recurrent translocation is mediated by homologous recombination between HERV-H elements. *Mol Cytogenet* **5**:6. doi:10.1186/1755-8166-5-6

Ito J, Sugimoto R, Nakaoka H, Yamada S, Kimura T, Hayano T, Inoue I. 2017. Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. *Plos Genet* **13**:e1006883. doi:10.1371/journal.pgen.1006883

Jason LA, Dorri JA. 2022. ME/CFS and Post-Exertional Malaise among Patients with Long COVID. *Neurol Int* **15**:1–11. doi:10.3390/neurolint15010001

Kassambara A, Mundt F. n.d. Factoextra: extract and Visualize the Results of Multivariate Data Analyses. https://rpkgs.datanovia.com/factoextra/index.html

Kassiotis G. 2023. The Immunological Conundrum of Endogenous Retroelements. *Annu Rev Immunol* **41**:99–125. doi:10.1146/annurev-immunol-101721-033341

Khadjinova AI, Wang X, Laine A, Ukadike K, Eckert M, Stevens A, Bengtsson AA, Lood C, Mustelin T. 2021. Autoantibodies against the envelope proteins of endogenous retroviruses K102 and K108 in patients with systemic lupus erythematosus correlate with active disease. *Clin Exp Rheumatol* **40**:1306–1312. doi:10.55563/clinexprheumatol/2kg1d8

Kim JH, Park K, Lee SB, Kang S, Park JS, Ahn CW, Nam JS. 2019. Relationship between natural killer cell activity and glucose control in patients with type 2 diabetes and prediabetes. *J Diabetes Investig* **10**:1223–1228. doi:10.1111/jdi.13002

Kolde R, Lizee A, Metsalu T. n.d. Pheatmap: Pretty heatmaps. https://github.com/raivokolde/pheatmap

Komaroff AL, Lipkin WI. 2023. ME/CFS and Long COVID share similar symptoms and biological abnormalities: road map to the literature. *Front Med* **10**:1187163. doi:10.3389/fmed.2023.1187163

Kremer D, Schichel T, Förster M, Tzekova N, Bernard C, Valk P, Horssen J, Hartung H, Perron H, Küry P. 2013. Human endogenous retrovirus type W envelope protein inhibits oligodendroglial precursor cell differentiation. *Ann Neurol* **74**:721–732. doi:10.1002/ana.23970

Kuhn RM, Haussler D, Kent WJ. 2013. The UCSC genome browser and associated tools. *Brief Bioinform* **14**:144–161. doi:10.1093/bib/bbs038

Kulski JK, Gaudieri S, Inoko H, Dawkins RL. 1999. Comparison Between Two Human Endogenous Retrovirus (HERV)-Rich Regions Within the Major Histocompatibility Complex. *J Mol Evol* **48**:675–683. doi:10.1007/pl00006511

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *Bmc Bioinformatics* **9**:559–559. doi:10.1186/1471-2105-9-559

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol* **9**:e1003118. doi:10.1371/journal.pcbi.1003118

Light AR, Bateman L, Jo D, Hughen RW, VanHaitsma TA, White AT, Light KC. 2012. Gene expression alterations at baseline and following moderate exercise in patients with Chronic

Fatigue Syndrome and Fibromyalgia Syndrome. *J Intern Med* **271**:64–81. doi:10.1111/j.1365-2796.2011.02405.x

Lima-Junior DS, Krishnamurthy SR, Bouladoux N, Collins N, Han S-J, Chen EY, Constantinides MG, Link VM, Lim AI, Enamorado M, Cataisson C, Gil L, Rao I, Farley TK, Koroleva G, Attig J, Yuspa SH, Fischbach MA, Kassiotis G, Belkaid Y. 2021. Endogenous retroviruses promote homeostatic and inflammatory responses to the microbiota. *Cell* **184**:3794-3811.e19. doi:10.1016/j.cell.2021.05.020

Liu X, Hong T, Parameswaran S, Ernst K, Marazzi I, Weirauch MT, Bass JIF. 2020. Human Virus Transcriptional Regulators. *Cell* **182**:24–37. doi:10.1016/j.cell.2020.06.023

Macchietto MG, Langlois RA, Shen SS. 2020. Virus-induced transposable element expression up-regulation in human and mouse host cells. *Life Sci Alliance* **3**:e201900536. doi:10.26508/lsa.201900536

Mager DL, Hunter DG, Schertzer M, Freeman JD. 1999. Endogenous Retroviruses Provide the Primary Polyadenylation Signal for Two New Human Genes (HHLA2 and HHLA3). *Genomics* **59**:255–263. doi:10.1006/geno.1999.5877

Magnus P, Gunnes N, Tveito K, Bakken IJ, Ghaderi S, Stoltenberg C, Hornig M, Lipkin WI, Trogstad L, Håberg SE. 2015. Chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) is associated with pandemic influenza infection, but not with an adjuvanted pandemic influenza vaccine. *Vaccine* **33**:6173–6177. doi:10.1016/j.vaccine.2015.10.018

Mangeney M, Parseval N de, Thomas G, Heidmann T. 2001. The full-length envelope of an HERV-H human endogenous retrovirus has immunosuppressive properties. *J Gen Virol* **82**:2515–2518. doi:10.1099/0022-1317-82-10-2515

Marston JL, Greenig M, Singh M, Bendall ML, Duarte RRR, Feschotte C, Iñiguez LP, Nixon DF. 2021. SARS-CoV-2 infection mediates differential expression of human endogenous retroviruses and long interspersed nuclear elements. *Jci Insight* **6**:e147170. doi:10.1172/jci.insight.147170

Matsuzawa A, Lee J, Nakagawa S, Itoh J, Ueda MT, Mitsuhashi S, Kochi Y, Kaneko-Ishino T, Ishino F. 2021. HERV-Derived Ervpb1 Is Conserved in Simiiformes, Exhibiting Expression in Hematopoietic Cell Lineages Including Macrophages. *Int J Mol Sci* **22**:4504. doi:10.3390/ijms22094504

MCHORNEY CA, WARE J, RACZEK A. 1993. The MOS 36-Item Short-Form Health Survey (SF-36). *Med Care* **31**:247–263. doi:10.1097/00005650-199303000-00006

Morris G, Maes M, Murdjeva M, Puri BK. 2019. Do Human Endogenous Retroviruses Contribute to Multiple Sclerosis, and if So, How? *Mol Neurobiol* **56**:2590–2605. doi:10.1007/s12035-018-1255-x

Natelson BH. 2019. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and Fibromyalgia: Definitions, Similarities, and Differences. *Clin Ther* **41**:612–618. doi:10.1016/j.clinthera.2018.12.016

Nepotchatykh E, Caraus I, Elremaly W, Leveau C, Elbakry M, Godbout C, Rostami-Afshari B, Petre D, Khatami N, Franco A, Moreau A. 2023. Circulating microRNA expression signatures accurately discriminate myalgic encephalomyelitis from fibromyalgia and comorbid conditions. *Sci Rep-uk* **13**:1896. doi:10.1038/s41598-023-28955-9

Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. 2015. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**:453–457. doi:10.1038/nmeth.3337

Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D, Diehn M, Alizadeh AA. 2019. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* **37**:773–782. doi:10.1038/s41587-019-0114-2

Ovejero T, Sadones O, Sánchez-Fito T, Almenar-Pérez E, Espejo JA, Martín-Martínez E, Nathanson L, Oltra E. 2020. Activation of Transposable Elements in Immune Cells of Fibromyalgia Patients. *Int J Mol Sci* **21**:1366. doi:10.3390/ijms21041366

Ramsay L, Marchetto MC, Caron M, Chen S-H, Busche S, Kwan T, Pastinen T, Gage FH, Bourque G. 2017. Conserved expression of transposon-derived non-coding transcripts in primate stem cells. *BMC Genom* **18**:214. doi:10.1186/s12864-017-3568-y

Rangel SC, Silva MD da, Silva AL da, Santos J de MB dos, Neves LM, Pedrosa A, Rodrigues FM, Trettel C dos S, Furtado GE, Barros MP de, Bachi ALL, Romano CM, Nali LHDS. 2022. Human endogenous retroviruses and the inflammatory response: A vicious circle associated with health and illness. *Front Immunol* **13**:1057791. doi:10.3389/fimmu.2022.1057791

Rebollo R, Miceli-Royer K, Zhang Y, Farivar S, Gagnier L, Mager DL. 2012. Epigenetic interplay between mouse endogenous retroviruses and host genes. *Genome Biol* **13**:R89. doi:10.1186/gb-2012-13-10-r89

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**:e47–e47. doi:10.1093/nar/gkv007

Rivera J, González T. 2004. The Fibromyalgia Impact Questionnaire: a validated Spanish version to assess the health status in women with fibromyalgia. *Clin Exp Rheumatol* **22**:554–60.

Rodrigues LS, Nali LH da S, Leal COD, Sabino EC, Lacerda EM, Kingdon CC, Nacul L, Romano CM. 2019. HERV-K and HERV-W transcriptional activity in myalgic encephalomyelitis/chronic fatigue syndrome. *Autoimmun Highlights* **10**:12. doi:10.1186/s13317-019-0122-8

Sanchez-Valle A, Wang X, Potocki L, Xia Z, Kang SL, Carlin ME, Michel D, Williams P, Cabrera-Meza G, Brundage EK, Eifert AL, Stankiewicz P, Cheung SW, Lalani SR. 2010. HERV-mediated genomic rearrangement of EYA1 in an individual with branchio-oto-renal syndrome. *Am J Méd Genet Part A* **152A**:2854–2860. doi:10.1002/ajmg.a.33686

Schmidt N, Domingues P, Golebiowski F, Patzina C, Tatham MH, Hay RT, Hale BG. 2019. An influenza virus-triggered SUMO switch orchestrates co-opted endogenous retroviruses to stimulate host antiviral immunity. *Proc Natl Acad Sci* **116**:17399–17408. doi:10.1073/pnas.1907031116

Shuvarikov A, Campbell IM, Dittwald P, Neill NJ, Bialer MG, Moore C, Wheeler PG, Wallace SE, Hannibal MC, Murray MF, Giovanni MA, Terespolsky D, Sodhi S, Cassina M, Viskochil D, Moghaddam B, Herman K, Brown CW, Beck CR, Gambin A, Cheung SW, Patel A, Lamb AN, Shaffer LG, Ellison JW, Ravnan JB, Stankiewicz P, Rosenfeld JA. 2013. Recurrent HERV-H-Mediated 3q13.2–q13.31 Deletions Cause a Syndrome of Hypotonia and Motor, Language, and Cognitive Delays. *Hum Mutat* **34**:1415–1423. doi:10.1002/humu.22384

Smets EMA, Garssen B, Bonke B, Haes JCJMD. 1995. The multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *J Psychosom Res* **39**:315–325. doi:10.1016/0022-3999(94)00125-o

Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV, Condition WCCDWG on P-C-19. 2022. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis* **22**:e102–e107. doi:10.1016/s1473-3099(21)00703-9

Thomas J, Perron H, Feschotte C. 2018. Variation in proviral content among human genomes mediated by LTR recombination. *Mobile Dna-uk* **9**:36. doi:10.1186/s13100-018-0142-3

Tschopp R, König RS, Rejmer P, Paris DH. 2023. Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS): A preliminary survey among patients in Switzerland. *Heliyon* **9**:e15595. doi:10.1016/j.heliyon.2023.e15595

Wang L, Kang N, Zhou J, Guo Y, Zhang X, Cui L, Ba D, He W. 2012. Downregulation of CD94/NKG2A Inhibitory Receptor on Decreased γδ T Cells in Patients with Systemic Lupus Erythematosus. *Scand J Immunol* **76**:62–69. doi:10.1111/j.1365-3083.2012.02705.x

Wessely S, Nimnuan C, Sharpe M. 1999. Functional somatic syndromes: one or many? *Lancet* **354**:936–939. doi:10.1016/s0140-6736(98)08320-2

Wickham H. 2016. ggplot2, Elegant Graphics for Data Analysis. *R*. doi:10.1007/978-3-319-24277-4

Wolfe F, Clauw DJ, Fitzcharles M, Goldenberg DL, Katz RS, Mease P, Russell AS, Russell IJ, Winfield JB, Yunus MB. 2010. The American College of Rheumatology Preliminary Diagnostic Criteria for Fibromyalgia and Measurement of Symptom Severity. *Arthrit Care Res* **62**:600–610. doi:10.1002/acr.20140

Wolfe F, Clauw DJ, Fitzcharles M-A, Goldenberg DL, Häuser W, Katz RL, Mease PJ, Russell AS, Russell IJ, Walitt B. 2016. 2016 Revisions to the 2010/2011 fibromyalgia diagnostic criteria. *Semin Arthritis Rheum* **46**:319–329. doi:10.1016/j.semarthrit.2016.08.012

Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, Tugwell P, Campbell SM, Abeles M, Clark P, Fam AG, Farber SJ, Fiechtner JJ, Franklin CM, Gatter

RA, Hamaty D, Lessard J, Lichtbroun AS, Masi AT, Mccain GA, Reynolds WJ, Romano TJ, Russell IJ, Sheon RP. 1990. The american college of rheumatology 1990 criteria for the classification of fibromyalgia. *Arthritis Rheumatism* **33**:160–172. doi:10.1002/art.1780330203

Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**:100141. doi:10.1016/j.xinn.2021.100141

Yi J-M, Kim H-M, Kim H-S. 2006. Human endogenous retrovirus HERV-H family in human tissues and cancer cells: expression, identification, and phylogeny. *Cancer Lett* **231**:228–239. doi:10.1016/j.canlet.2005.02.001

Zhang Y, Li T, Preissl S, Amaral ML, Grinstein JD, Farah EN, Destici E, Qiu Y, Hu R, Lee AY, Chee S, Ma K, Ye Z, Zhu Q, Huang H, Fang R, Yu L, Belmonte JCI, Wu J, Evans SM, Chi NC, Ren B. 2019. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat Genet* **51**:1380–1388. doi:10.1038/s41588-019-0479-7

Zhao R, Chinai JM, Buhl S, Scandiuzzi L, Ray A, Jeon H, Ohaegbulam KC, Ghosh K, Zhao A, Scharff MD, Zang X. 2013. HHLA2 is a member of the B7 family and inhibits human CD4 and CD8 T-cell function. *Proc Natl Acad Sci* **110**:9879–9884. doi:10.1073/pnas.1303524110

Zhao Y, Lin L, Xiao Z, Li M, Wu X, Li W, Li X, Zhao Q, Wu Y, Zhang H, Yin J, Zhang L, Cho CH, Shen J. 2018. Protective Role of γδ T Cells in Different Pathogen Infections and Its Potential Clinical Application. *J Immunol Res* **2018**:5081634. doi:10.1155/2018/5081634

FIGURES AND TABLES

Table 1. Patient health status assessment with FIQ, MFI, and SF-36 (Burckhardt et al., 1991; MCHORNEY et al., 1993; Smets et al., 1995) questionnaires.

	Control (n=9)	**ME/CFS (n=8)**	**FM (n=10)**	**Co-diagnosed (n=16)**
	Mean ± SD [Range]	**Mean ±SD [Range]**	**Mean ±SD [Range]**	**Mean ±SD [Range]**
Age	51 ± 2 [43 – 61]	54 ± 3 [50 – 58]	50 ± 5 [42 – 58]	47 ± 15.48 [22 – 58]
BMI	25.03 ± 2.07 [22.10 – 28.04]	26.29 ± 3.33 [21.34 – 31.22]	25.31 ± 1.73 [23.68 – 28.40]	24.80 ± 4.72 [18.71 – 30.43]
FIQ				
Total FIQ	28.16 ± 13.08 [0 – 2.86]	74.4 ± 4.8 [56.3-85.6]	75.9 ± 3.4 [51.6-92.8]	74.6 ± 2.5 [47.8-96.3]
Function	2.38 ± 0.57 [1.98 – 3.63]	5.2 ± 2.2 [1.7-7.9]	6.8 ± 1.5 [4.3-8.9]	5.2 ± 1.8 [1.7-9.2]
Overall	10.01 ± 0 [10.01 – 10.01]	8.9 ± 3.0 [1.4-10.0]	8.2 ± 3.4 [0-10.0]	7.9 ± 2.7 [1.4-10]
Symptoms	0.32 ± 0.95 [0 - 2.86]	6.6 ± 4.8 [0-10.0]	6.0 ± 3.4 [0-10.0]	6.2 ± 2.5 [2.9-10]
MFI				
General Fatigue	11.56 ± 4.19 [5 – 17]	17.6 ± 3.1 [11-20]	16.3 ± 4.6 [10-20]	13.6 ± 3.2 [7-20]
Physical Fatigue	10.11 ± 4.01 [6 – 17]	18.0 ± 1.9 [16-20]	17.0 ± 3.7 [12-20]	14.1 ± 2.9 [12-20]
Reduced Activity	7.33 ± 2.74 [4 – 12]	15.9 ± 3.4 [12-20]	15.9 ± 4.4 [9-20]	12.9 ± 2.8 [11-20]
Reduced Motivation	7.22 ± 3.31 [4 -15]	15.4 ± 3.7 [9-19]	14.8 ± 4.1 [9-20]	11.6 ± 2.4 [8-16]
Mental Fatigue	7.11 ± 2.93 [4 – 13]	15.5 ± 3.7 [11-20]	15.1 ± 3.8	13.3 ± 3.1 [9-19]
SF-36				
Physical Functioning	86.67 ± 13.23 [65 – 100]	44.5 ± 14.5 [15-60]	31.5 ± 13.9	35.3 ± 14.8 [0-65]
Role Physical	83.33 ± 17.95 [50 – 100]	3.2 ± 8.8 [0-25]	10.6 ± 17.7 [0-43.8]	14.5 ± 21.9 [0-75]
Bodily Pain	58.61 ± 12.63 [45 – 80]	19.3 ± 11.7 [0-32.5]	19.3 ± 17.8 [0-45]	17.3 ± 14.2 [0-45]
General Health	65.56 ± 15.09 [35 – 85]	18.9 ± 11.0 [3.8-35.0]	24.8 ± 16.1 [0-50]	23.4 ± 14.8 [0-45]
Vitality	56.94 ± 9.08 [43.75 – 75.00]	13.6 ± 10.1 [0-30]	11.1 ± 16.8 [0-43.8]	14.5 ± 11.6 [0-40]
Social Functioning	84.72 ± 13.66 [62.50 – 100]	33.6 ± 21.9 [4.0-77.5]	25.0 ± 23.2 [0-62.1]	29.5 ± 25.6 [0-75]
Role Emotional	79.63 ± 20.88 [50 – 100]	12.8 ± 35.3 [0-100]	43.3 ± 37.2 [0-100]	12.8 ± 38.9 [0-100]

Mental Health	71.67 ± 18.54 [40 – 95]	34.9 ± 21.2 [3.4-72.0]	44.7 ± 19.9 [12-70]	45.2 ± 17.1 [15-80]

Abbreviations: FIQ, Fibromyalgia Impact Questionnaire; MFI, Multi Fatigue Inventory; SF-36, Short-Form-36 Health Survey; SD, standard deviation; SE, standard error. Range refers to the possible values in the studied group.

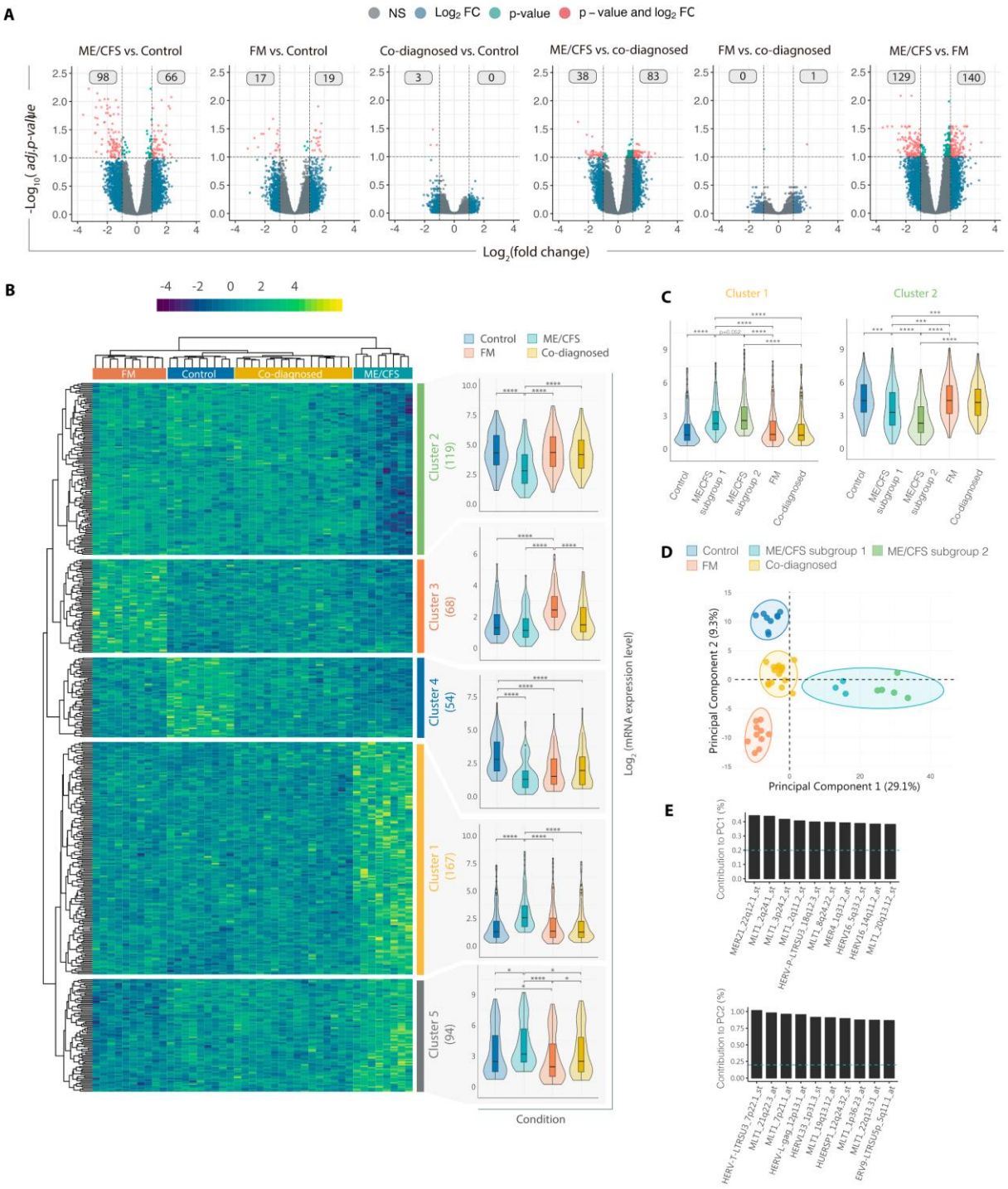

Fig. 1. DE of HERV elements discriminates ME/CFS, FM, and co-diagnosis. (A) Volcano plots showing Log_2(fold changes) and the adjusted p-values for all HERV assessed with HERV-V3 microarray for each set of groups, as indicated. Red dots indicate DE HERV (FDR<0.1 and |log2FC|>1). Grey boxes show numbers of overexpressed or underexpressed HERV elements. **(B)** HERV expression heatmap and cluster analysis of ME/CFS (n=8, green), FM (n=10, orange), co-diagnosed (n=16, yellow), and healthy control (n=9, blue) samples. The

heatmap includes all HERV probes displaying significant DE between at least two of the compared groups (FDR<0.1 and |log2FC|>1). Clusters 1-4 correspond to groups of HERV probes displaying significant over/under-expression in at least one group. Box and violin plots summarize the distribution and expression levels of the DE probes in each cluster per study group. The scaled mean expression value (z score) for each HERV probe is plotted. Box plots show the median z-score value and the first and third quartiles. **(C)** Box and violin plots for Clusters 1 and 2 differentiating ME/CFS subgroups 1 and 2. **(D)** Principal component analysis of DE HERV. **(E)** Bar plot of top ten DE HERV with greater influence/contribution to principal component 1 (PC1) or principal component 2 (PC2). Statistical tests: unpaired two-sample Wilcoxon Test with Benjamin-Hochberg p-value correction (***p<0.001, ****p < 0.0001).

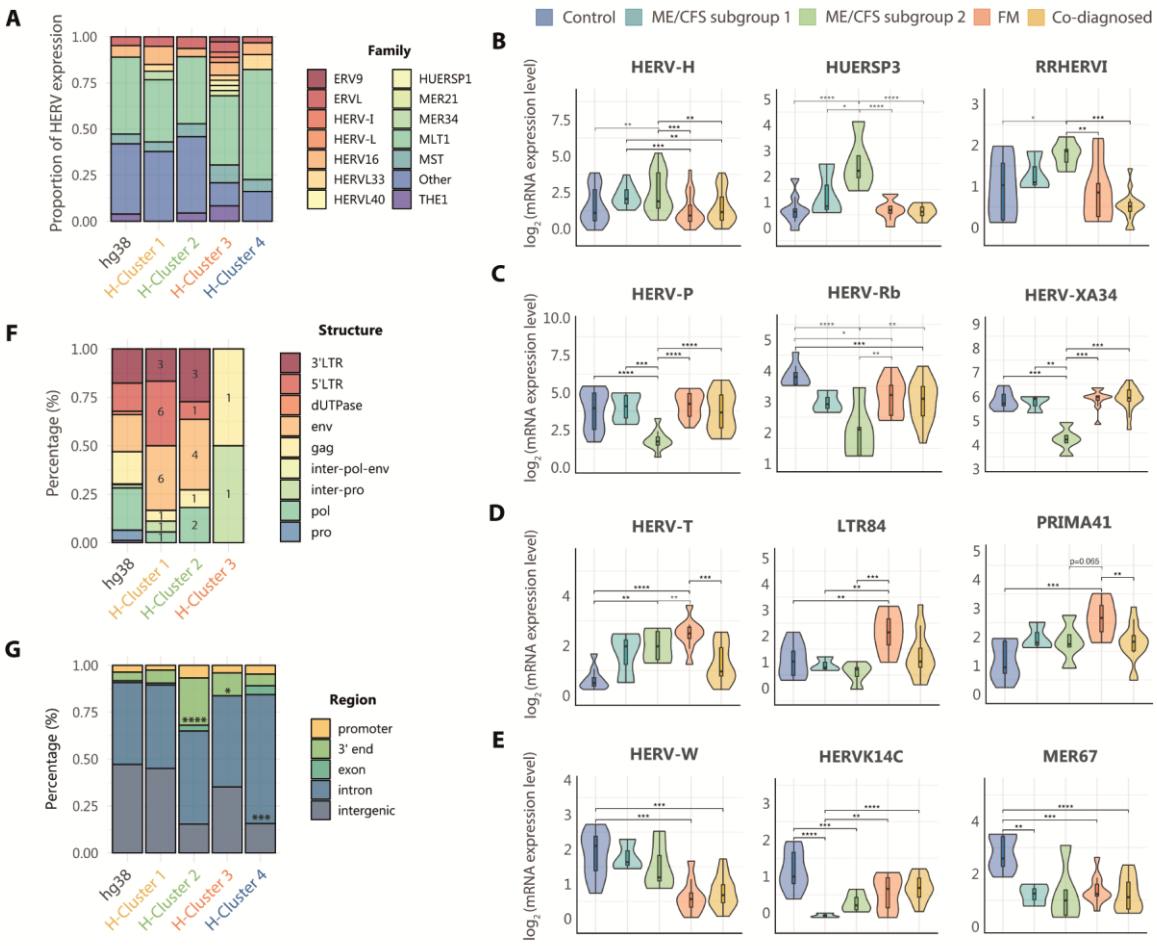

Figure 2. Diverse solitary LTR families are deregulated in ME/CFS, FM, and co-diagnosed conditions. **(A)** Relative contribution of HERV families to each cluster, calculated as the proportion of HERV loci assigned to each family relative to total HERV loci in each cluster. Families with a representation of at least 2.5% are shown. Main HERV families **(B)** downregulated in ME/CFS, FM and co-diagnosed groups **(**Cluster 4), **(C)** upregulated

in ME/CFS (Cluster 1), **(D)** downregulated in ME/CFS (Cluster 2)**,** and **(E)** upregulated in FM (Cluster 3). Box and violin plots summarize distributions and expression levels of the different HERV probes belonging to the same family in the different study groups. The scaled mean expression value (z score) for each HERV probe is plotted. Box plots show the median z-score value and the first and third quartiles. **(F)** Proportion of HERV subdomains expressed by cluster. **(G)** Genomic context of the DE HERV loci by cluster. Statistical tests: Fisher's exact test, T-test or Wilcoxon test with Benjamin-Hochberg p-value correction (*$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$).

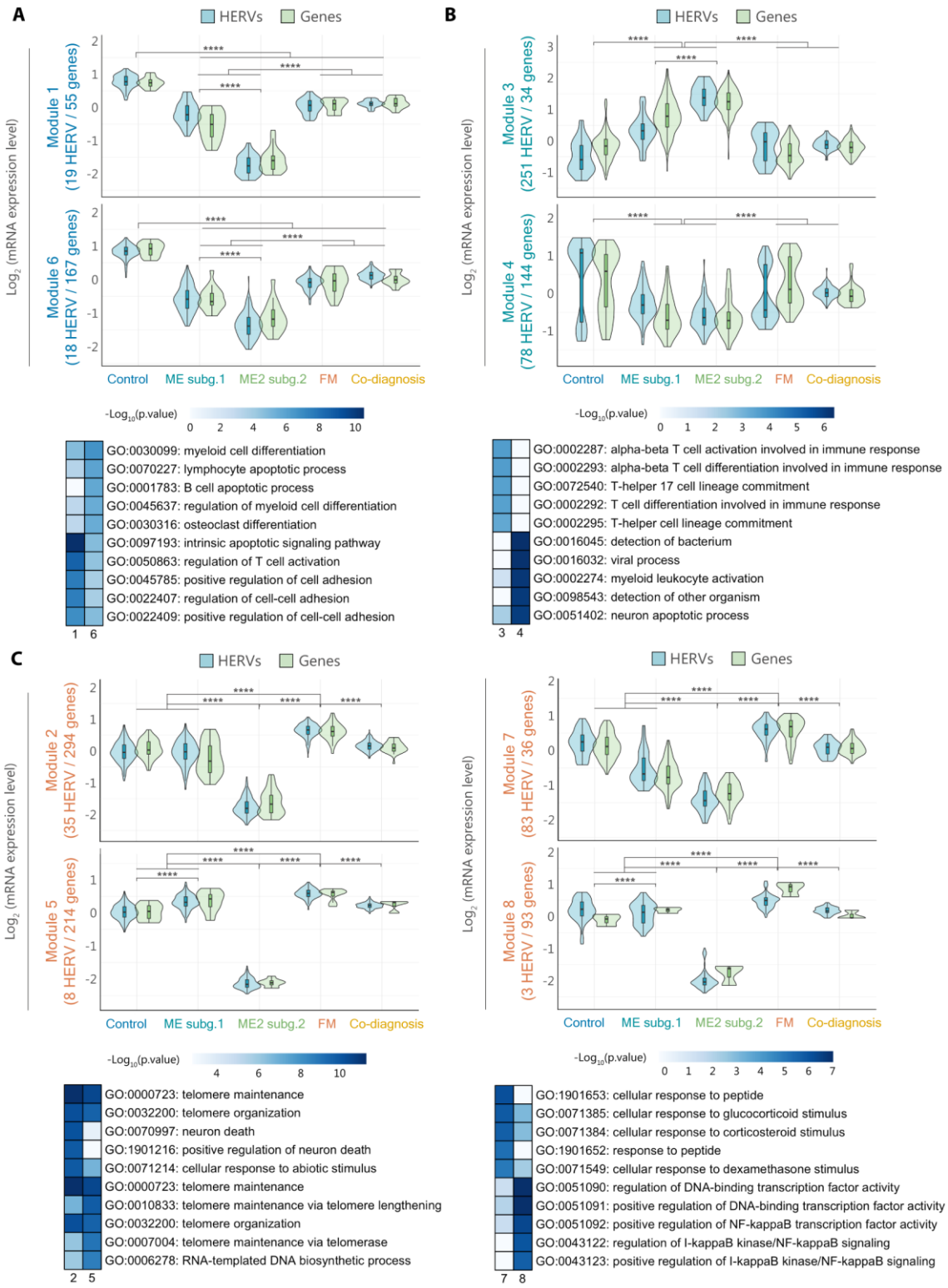

Figure 3. HERV expression correlates with immune response genes. Hierarchical clustering of DE HERV and genes (FDR<0.1 and |log2FC|>1) providing eight modules with highly correlated expression levels. Box and violin plots summarize the distribution and expression level of the different HERV (blue) and gene (green) probes in modules **(A)** 1 and 6, **(B)** 3 and 4, and **(C)** 2, 5, 7, and 8, per study group. The scaled mean expression value (z

score) for each HERV and gene probe sets are plotted. Box plots show the median z score values with first and third quartiles. Statistical tests: Wilcoxon test (*p < 0.05; **p<0.01, ***p<0.001, ****p < 0.0001).

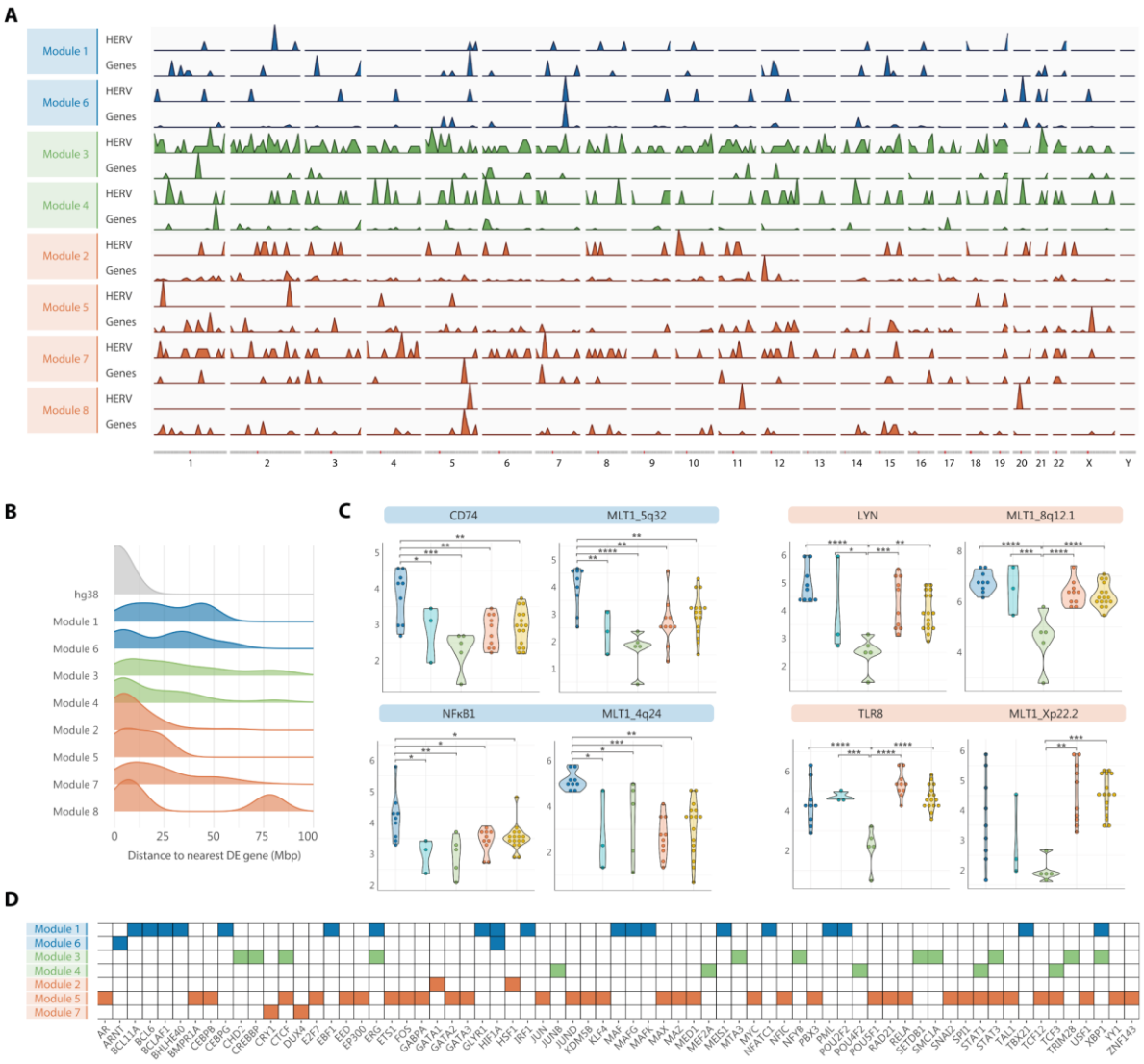

Figure 4. DE HERV sequences and DE immune-response genes are majorly independently transcribed (A) Genomic distribution of HERV, and genes encompassed in modules one through eight. **(B)** Genomic distance between DE HERV and their nearest DE genes per Module, plotted as density curves illustrating the distribution of distances from 0 to 100 million base pair (Mbp). **(C)** Violin dot plots showing the distribution and expression levels of some examples of co-localized HERV loci and genes in ME/CFS subgroup 1 (n=3), ME/CFS subgroup 2 (n=5), FM (n=10), co-diagnosed (n=16) and healthy control (n=9) samples. Blue and orange highlighted headings indicates the pair HERV/gene belonged to modules 1 or 6, or modules 2, 5, 7, or 8, respectively. **(D)** Heatmap of enriched transcription factor binding sites (TFBS) in HERV loci of modules one through seven. Module eight did not show significant enrichment of TFBS.

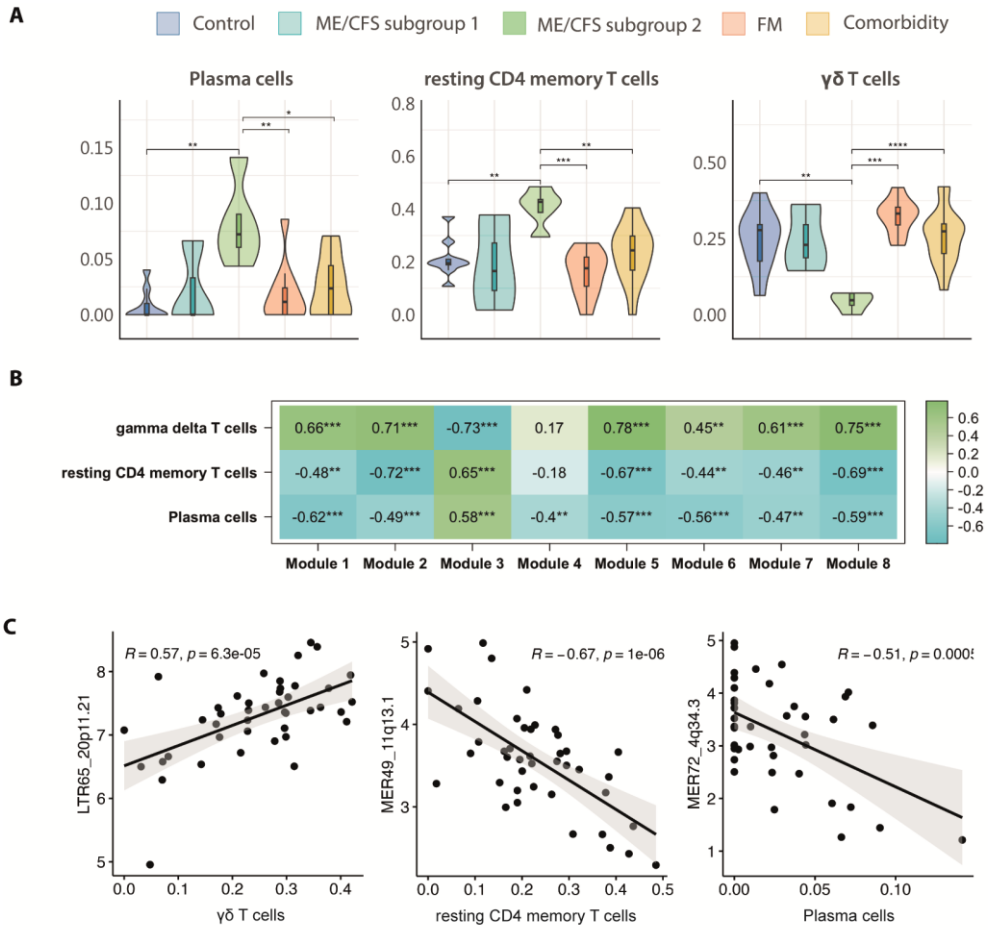

Figure 5. HERV fingerprints correlate with immune cell profiles. (A) Measurements of immune cell proportions by CIBERSORTx per study group (n=8 ME/CFS, n=10 FM, n=16 co-diagnosed, n=9 controls). **(B)** Association between modules and immune cell proportions was evaluated by correlating eigengenes from each module with cell proportion values obtained from CIBERSORTx analysis. Boxes show Pearson correlation values and associated p-values (*, p<0.05; **, p<0.01; ***, p<0.001) between gene expression levels of each module and quantity of cells from each specific type. A value of 1 (green) and -1 (blue) quantify strongest positive and negative correlations, respectively, while 0 (white) shows no correlation. **(C)** Scatter plots between top HERV loci expression levels and immune cell proportion as measured by CIBERSORTx. Pearson correlation coefficients and associated p-values are shown in plots. Correlation 95% confidence intervals are: 0.33 to 0.74 for LTR65_20p11.21 vs. γδ T cells, -0.81 to -0.46 for MER49_11q13.1 vs. resting CD4 memory T cells, and -0.70 to -0.25 for MER72_4q34.3 vs. plasma cells.

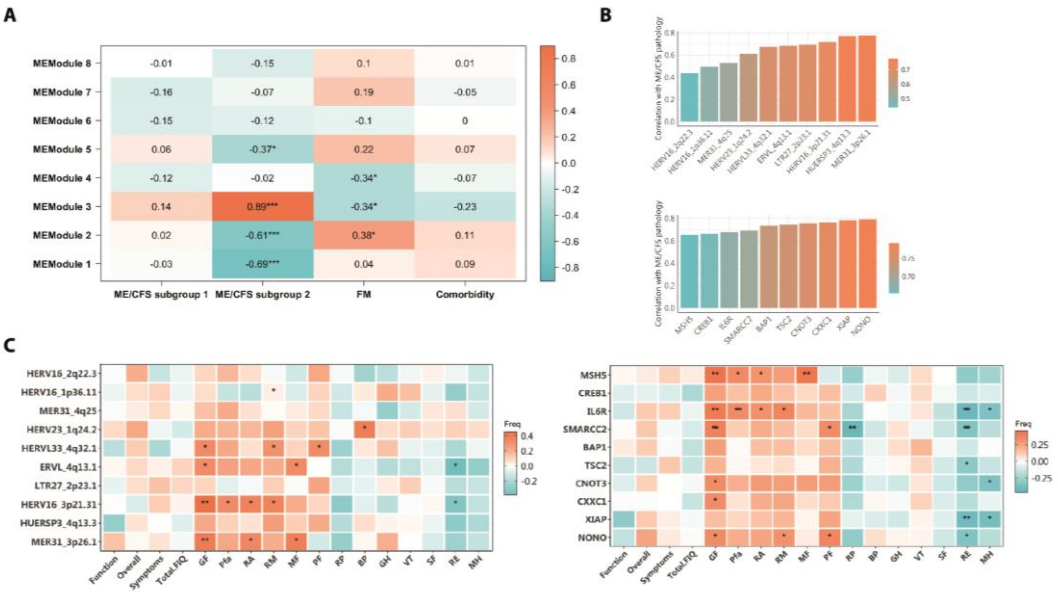

Figure 6. HERV and gene deregulation associates with disease symptomatology (A) Association between modules and disease determined by correlations of eigengenes from each module with binarized disease traits. Boxes show Pearson correlation values and associated p-values (*, p<0.05; **, p<0.01; ***, p<0.001) by module and each disease trait. A value of 1 (orange) and -1 (blue) quantify strongest positive and negative correlations, respectively, while 0 (white) shows no correlation. **(B)** Barplots showing top ten HERV loci (upper) and top ten genes (lower) with highest correlations with ME/CFS disease. Bar height and color display the strength of each correlation, being high orange bars more correlated than lower blue bars. **(C)** Association between top ten HERV loci (left) and top ten genes (right) with disease traits as determined by their correlation with ME/CFS patient symptoms as assessed by questionnaire scores. Boxes show Pearson correlation values and associated p-values (*, p<0.05; **, p<0.01) by DE HERV or DE gene as indicated for FIQ subdomains (Function, Overall and Symptoms) or total FIQ; for MFI subdomains (GF: general fatigue, Pfa: physical fatigue, RA: reduced activity, RM: reduced motivation, and MF: mental fatigue), and for SF-36 subdomains (PF: physical functioning, RP: role physical, BP: bodily pain, GH: general health, VT: vitality, SF: social functioning, RE: role emotional, and MH: mental health).