HERV activation segregates ME/CFS from fibromyalgia and defines a novel nosological entity for patients fulfilling both clinical criteria

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

[1]Escuela de Doctorado, Universidad Católica de Valencia San Vicente Mártir, 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, Universidad Católica de Valencia San Vicente Mártir, Valencia, Spain.

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

Abstract

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) and fibromyalgia (FM) are chronic diseases with poorly understood pathophysiology and diagnosis based on clinical assessment of unspecific symptoms. The recent post-COVID-19 condition, which shares similarities with ME/CFS and FM, has raised concerns about viral-induced transcriptome changes in post-viral syndromes. Viral infections, and other types of stress, are known to unleash human endogenous retroviruses (HERV) repression that if maintained could lead to symptom chronicity. This study evaluated this possibility for ME/CFS and FM on a selected cohort of female patients complying with diagnosis criteria for ME/CFS, FM, or both, and matched healthy controls (n=43). The results show specific HERV fingerprints for each disease, confirming biological differences between ME/CFS and FM. Unexpectedly, HERV profiles segregated patients that met both ME/CFS and FM clinical criteria from patients complying only with ME or FM criteria, while clearly differentiating patients from healthy subjects, supporting that the highly prevalent comorbidity condition must constitute a different nosological entity. Moreover, HERV profiles exposed significant quantitative differences within the ME/CFS group that correlated with differences in immune gene expression and patient symptomatology, supporting ME/CFS patient subtyping and confirming immunological disturbances in this disease. Pending issues include validation of HERV profiles as disease biomarkers of post-viral syndromes and understanding the role of HERV during infection and beyond.

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) *(1)*, are chronic, disabling, acquired diseases, characterized by complex symptomatology that affects multiple organs *(2, 3)*. 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 *(3–7)*. Their frequent co-diagnosis drove the hypothesis of a "single syndrome" and promoted the search for common or differentiating factors *(8–10)*. However, despite molecular support for ME/CFS and FM constituting different entities *(11–13)*, their frequent joint appearance (over 50% comorbidity in females) *(14)* 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 for these orphan diseases *(15–17)*. Particularly, viral infections are gaining momentum after the emergence of a type of persistent post-viral syndrome with symptoms that closely resemble those in ME/CFS *(18)* and FM *(19),* defined as post-COVID-19 condition (ICD-11 RA02) *(20)*. The prevalence of post-COVID-19 cases meeting ME/CFS criteria has been estimated at 58% *(21)* raising concerns of a health, social, and economic burden of unprecedented dimensions with post-viral chronicity.

Viral infections trigger dramatic changes in host cell gene expression, affecting their metabolism and epigenetic landscape *(22)*. Although functionally not well understood, these epigenetic changes involve de-repression of human endogenous retroviruses (HERV) *(23)*, sequences of viral origin that were incorporated during human evolution by exogenous infections and currently represent about 8% of our genome *(24)*. Initially poorly understood, advances in sequencing technology aided in the discovery of the wide range of physiological processes in which they participate, including coordination of the immune response *(25)*, interaction with microbiota *(26, 27)*, or shaping neurological functions *(28)*, among other, revealing them as master regulators of physiological function whose derangement underlies a number of diseases, including multiple sclerosis *(29)*, systemic lupus erythematosus *(30)*, or post-COVID-19 condition *(31)*. Particularly, the expression of HERV-encoded proteins has been shown to stimulate an immune response in the context of autoimmune diseases *(29, 30)*, although HERV influence on pathophysiology extends to their regulatory regions encompassed in their long terminal repeats (LTR), through which host gene expression is coordinated *(32)*.

Previous studies found overexpression of some HERV families in immune cells of ME/CFS *(33)* and FM *(34)* at the transcript level. However, the study and comprehension of HERV in ME/CFS and FM are still at infancy. In this study, we hypothesized that genome-wide HERV profiling of ME/CFS and FM could reveal distinct underlying pathomechanisms that justify their separate classification, and/or common fingerprints explaining their high comorbid prevalence. For this purpose, we scrutinized HERV and gene transcriptomes by 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 *(4, 5)*, patients fulfilling FM diagnosis ACR (American College of Rheumatology) criteria *(3, 6, 7)*, or patients complying with both, ME/CFS and FM diagnosis criteria, which will be called comorbid group, or "comorbidity" from now on. The results of comparing HERV profiles of the immune system across these three patient groups as well as to HERV profiles of healthy matched subjects allowed, not only for the separation of patients from

healthy individuals, but also for perfect discrimination of patients into three distinct disease groups, supporting distinct subjacent pathomechanisms not only for ME/CFS and FM, but, unexpectedly, also for the comorbid group which seems to constitute a novel nosological entity rather than a combination of both diseases. In addition, HERV profiling also exposed 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 criteria for the differential diagnosis of fibromyalgia and ME/CFS as well as for patient subtyping of the latter with expected impact in precision medicine programs.

RESULTS

Demographics and Clinical Characteristics of Participants

The study compared a total of 43 female subjects between the ages of 42 and 58: 8 ME/CFS cases, 10 FM, 16 comorbid cases, and 9 matched healthy controls. The average age for participants was 53±4 years (range 50-58) for ME/CFS subjects, 54±7 years (range 42-66) for FM, 53±11 years (range 22-70) for comorbidity subjects, and 50±6 (range 43-61) for controls (**Figure S1A, Table S1**). Questionnaires used for patient symptom phenotyping included Fibromyalgia Impact Questionnaire (FIQ) *(35)*, Multi Fatigue Inventory (MFI) *(36)* for general fatigue, and Short-Form-36 Health Survey (SF-36) *(37)* 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 the 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 *(38)*; 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 comorbidity groups.

Table 1. Patient health status assessment with FIQ, MFI, and SF-36 *(35–37)* questionnaires.

	ME/CFS (n=8)			FM (n=10)			Comorbidity (n=16)			ME/CFS vs. FM *p*	ME/CFS vs. comorbidity *p*	FM vs. comorbidity *p*
Questionnaire	**Mean**	**SD ± SE**	**Range**	**Mean**	**SD ± SE**	**Range**	**Mean**	**SD ± SE**	**Range**			
FIQ												
Total FIQ	74.4	4.8 ± 3.3	(56.3-85.6)	75.9	3.4 ± 3.9	(51.6-92.8)	74.6	2.5 ± 3.4	(47.8-96.3)	n.s.	n.s.	n.s.
Function	5.2	2.2 ± 0.8	(1.7-7.9)	6.8	1.5 ± 0.5	(4.3-8.9)	5.2	1.8 ± 0.5	(1.7-9.2)	n.s.	n.s.	n.s.
Overall	8.9	3.0 ± 1.0	(1.4-10.0)	8.2	3.4 ± 1.1	(0-10.0)	7.9	2.7 ± 0.7	(1.4-10)	n.s.	n.s.	n.s.
Symptoms	6.6	4.8 ± 1.8	(0-10.0)	6.0	3.4 ± 1.1	(0-10.0)	6.2	2.5 ± 0.6	(2.9-10)	n.s.	n.s.	n.s.
MFI												
General Fatigue	17.6	3.1 ± 1.1	(11-20)	16.3	4.6 ± 1.5	(10-20)	13.6	3.2 ± 0.8	(7-20)	n.s.	0.054	n.s.
Physical Fatigue	18.0	1.9 ± 0.7	(16-20)	17.0	3.7 ± 1.2	(12-20)	14.1	2.9 ± 0.7	(12-20)	n.s.	*	0.082
Reduced Activity	15.9	3.4 ± 1.2	(12-20)	15.9	4.4 ± 1.4	(9-20)	12.9	2.8 ± 0.7	(11-20)	n.s.	0.063	n.s.
Reduced Motivation	15.4	3.7 ± 1.3	(9-19)	14.8	4.1 ± 1.3	(9-20)	11.6	2.4 ± 0.6	(8-16)	n.s.	*	0.067
Mental Fatigue	15.5	3.7 ± 1.3	(11-20)	15.1	3.8 ± 1.2	(9-20)	13.3	3.1 ± 0.8	(9-19)	n.s.	n.s.	n.s.
SF-36												
Physical Functioning	44.5	14.5 ± 5.1	(15-60)	31.5	13.9 ± 4.4	(5-55)	35.3	14.8 ± 3.7	(0-65)	n.s.	n.s.	n.s.
Role Physical	3.2	8.8 ± 3.1	(0-25)	10.6	17.7± 5.6	(0-43.8)	14.5	21.9 ± 5.5	(0-75)	n.s.	n.s.	n.s.
Bodily Pain	19.3	11.7 ± 4.1	(0-32.5)	19.3	17.8 ± 5.6	(0-45)	17.3	14.2 ± 3.6	(0-45)	n.s.	n.s.	n.s.
General Health	18.9	11.0 ± 3.9	(3.8-35.0)	24.8	16.1 ± 5.1	(0-50)	23.4	14.8 ± 3.7	(0-45)	n.s.	n.s.	n.s.
Vitality	13.6	10.1 ± 3.6	(0-30)	11.1	16.8 ± 5.3	(0-43.8)	14.5	11.6 ± 2.9	(0-40)	n.s.	n.s.	n.s.
Social Functioning	33.6	21.9 ± 7.8	(4.0-77.5)	25.0	23.2 ± 7.3	(0-62.1)	29.5	25.6 ± 6.4	(0-75)	n.s.	n.s.	n.s.
Role Emotional	12.8	35.3 ± 12.5	(0-100)	43.3	37.2 ± 11.8	(0-100)	12.8	38.9 ± 9.7	(0-100)	n.s.	n.s.	n.s.
Mental Health	34.9	21.2 ± 7.5	(3.4-72.0)	44.7	19.9 ± 6.3	(12-70)	45.2	17.1 ± 4.3	(15-80)	n.s.	n.s.	n.s.

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. Data were analyzed using parametric t-test or non-parametric Wilcoxon pairwise comparisons if normally distributed or not, respectively. P values were adjusted for multiple comparisons with Benjamini-Hochberg (BH) correction. Statistically significant comparisons are indicated by asterisks as p<0.05.

HERV signatures discriminate ME/CFS, FM, and comorbidity groups from healthy controls

The epigenetic mechanisms constraining HERV can be disrupted by a number of environmental factors, including some (e.g., infections, stress, or toxin exposure) reported as triggering agents for ME/CFS and FM *(15, 16)*. As mentioned, differential expression of some HERV families was reported in ME/CFS *(33)* and FM *(34)* when compared to expression in healthy subjects. However, these studies assessed HERV levels with the use of degenerate primers which does not inform of the affected HERV elements. In this study, we sought to fill this gap by investigating at loci-level resolution the HERV transcriptome in peripheral blood mononuclear cells (PBMCs) of ME/CFS and FM individuals, as well as in individuals presenting comorbidity, as a strategy to identify commonalities and differences between these frequently overlapping diseases.

Genome-wide HERV expression profiles for each of the four study groups (three disease groups: ME/CFS, FM, and their comorbidity, plus one control group corresponding to healthy participants), using custom high-density Affymetrix HERV-V3 microarrays *(39)*, 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 S2**), confirming dysregulation of particular sets of HERV elements in the immune systems of ME/CFS, FM, and comorbidity 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, comorbidity, or healthy control groups (**Fig. 1A**). Consistently, hierarchical clustering of samples based on the DE HERV fingerprints clearly differentiate these pathologies from healthy subjects, strikingly segregating ME/CFS from all other compared groups (**Fig. 1B**). 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). Differential expression of the HERV loci encompassed in each cluster was corroborated by analysis of the expression levels of HERV encompassed in each cluster, as illustrated with violin plots (**Fig. 1B**). Moreover, these changes were validated by real-time quantitative polymerase chain reaction (RT-qPCR) as shown in **Fig. S1B.**

Examination of Clusters 1 and 2 indicate that the ME/CFS group could be subdivided into two subgroups showing differential 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 groups, the ME/CFS subgroup 2 exhibits a more pronounced dysregulation of the HERV ME/CFS fingerprint when compared to subgroup 1 at the quantitative level, particularly affecting the downregulated HERV (Cluster 2, $p<0.0001$) (**Fig. 1C**). Potential differences in patient health status explaining this variability were analyzed, discovering that ME/CFS subgroup 2 presents more accused levels of fatigue ($p<0.05$) and reduced motivation as compared to comorbidity group than ME/CFS subgroup 1 (**Table S3**).

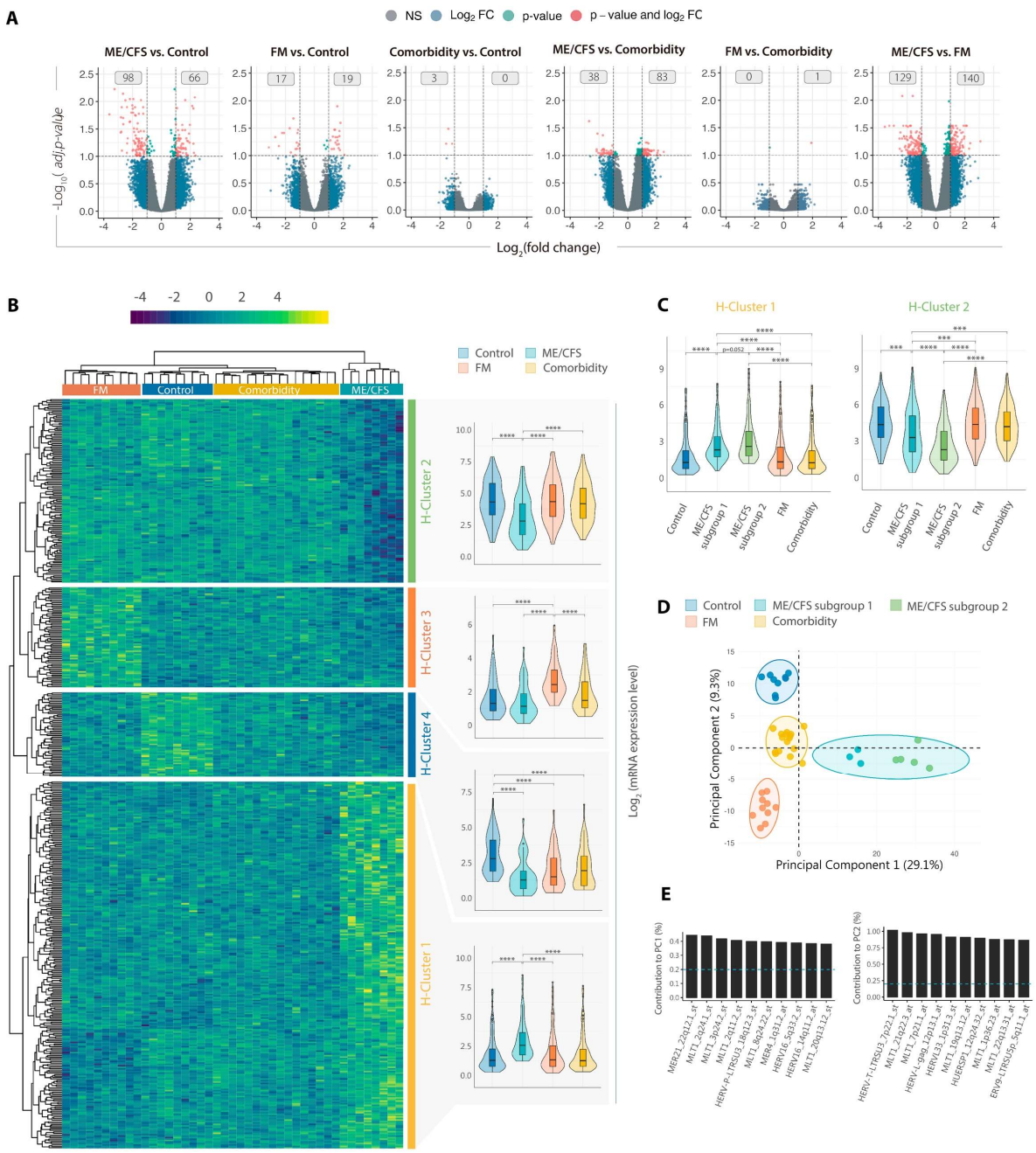

Fig. 1. Differential expression of HERV elements discriminates ME/CFS, FM, and comorbidity. (A) Volcano plots showing Log_2(fold changes) and the adjusted p-values for all HERV assessed with HERV-V3 microarray for each set of groups, as indicated. Red dots indicate differentially expressed (DE) HERV (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), comorbidity (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).

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, comorbidity, and control groups; principal component 2 (PC2)

allows partition of FM, comorbidity, and control samples, but not so much of ME/CFS from comorbidity. 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; while the variation explained by each principal component relies on several HERV loci with comparable contributions of ≤1%, indicating complex dysregulation of different loci may be behind these pathologies (**Fig. 1E**). In summary, these results show that ME/CFS, FM, and their comorbidity 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 comorbidity presents the closest HERV profiles to those of healthy controls indicating much less involvement of HERV in these patients. Thus, based on HERV profiles, comorbidity condition might represent a different nosological entity from ME/CFS or FM.

Dysregulated HERV families are disease-specific

The involvement of HERV in host physiology is based on the specific regulation of cellular functions by different HERV families *(32)*, entailing that preferential HERV-family derangements may indicate particular altered functions in the diseases under study. With this premise, we examined the family composition of the identified HERV clusters (**Fig. 2**). Overall, ME/CFS (Clusters 1 and 2) outstands with the largest heterogeneity with up to 66 HERV families dysregulated, 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 S4, Fig. S2**).

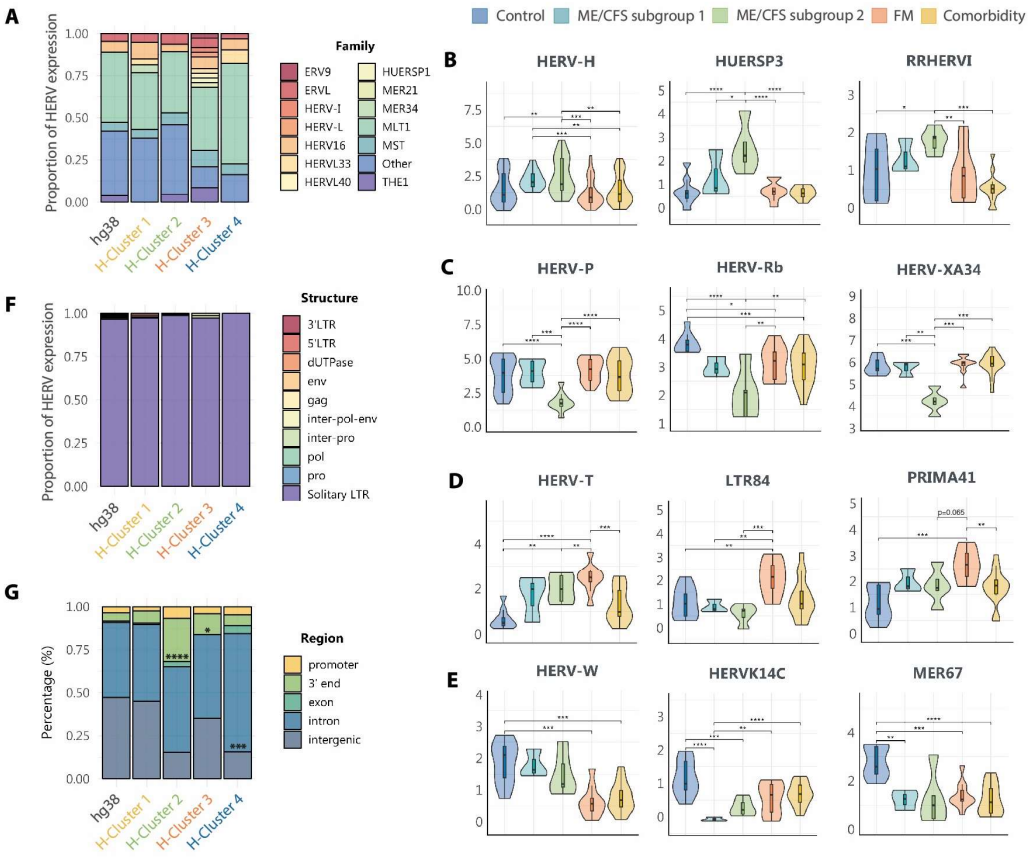

Figure 2. Diverse solitary LTR families are deregulated in ME/CFS, FM, and comorbid conditions. **(A)** Relative contribution of HERV families to each cluster, calculated as the proportion of HERV loci assigned to each family relative to total HERV loci in each cluster. Families with a representation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

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 S4**), 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 S4**), 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 S4**). Up to 21 families, comprising HERV-H, HUERSP3, and RRHERVI among other, appeared specifically upregulated in Cluster 1 (**Fig. 2B, Table S4, 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 S4, Fig. S2**). Regarding FM, three families consisting of HERV-T, LTR84, and PRIMA41, were found specifically upregulated (Cluster 3) (**Fig. 2D, Table S4, Fig. S2**). Five families were commonly downregulated across ME/CFS, FM, and comorbidity, including HERV-W, HERVK14C, and MER67 (**Fig. 2E, Table S4, Fig. S2**). Contrary to literature *(34)*, 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 comorbidity 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 comorbidity.

In light of the potential involvement of HERV-encoded proviral proteins in disease development *(24)* 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., long terminal repeats (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**).

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**). 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.

HERV fingerprints associate with abnormal immune response gene expression

Since disease-associated HERV derangement could mediate differences in 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) by this same HERV-V3 high-density microarray technology. Differential transcript analysis revealed 368 genes (1,037 probesets) DE between at least two study groups (FDR<0.1 and |log2FC|>1) (**Table S5)**. In line with DE HERV expression, ME/CFS immune cells showed a more accused dysregulation in gene expression profiles compared to FM´s and comorbidity groups (**Fig. S3A**). 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. S3B**).

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 RNA 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**), reflecting a common disturbance in cell-cell communication and immune cell maturation in all three disease groups under study. 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 S6**), 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 S6**). 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 comorbidity 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 a result of co-transcription (**Fig. 4C**, $p<0.05$).

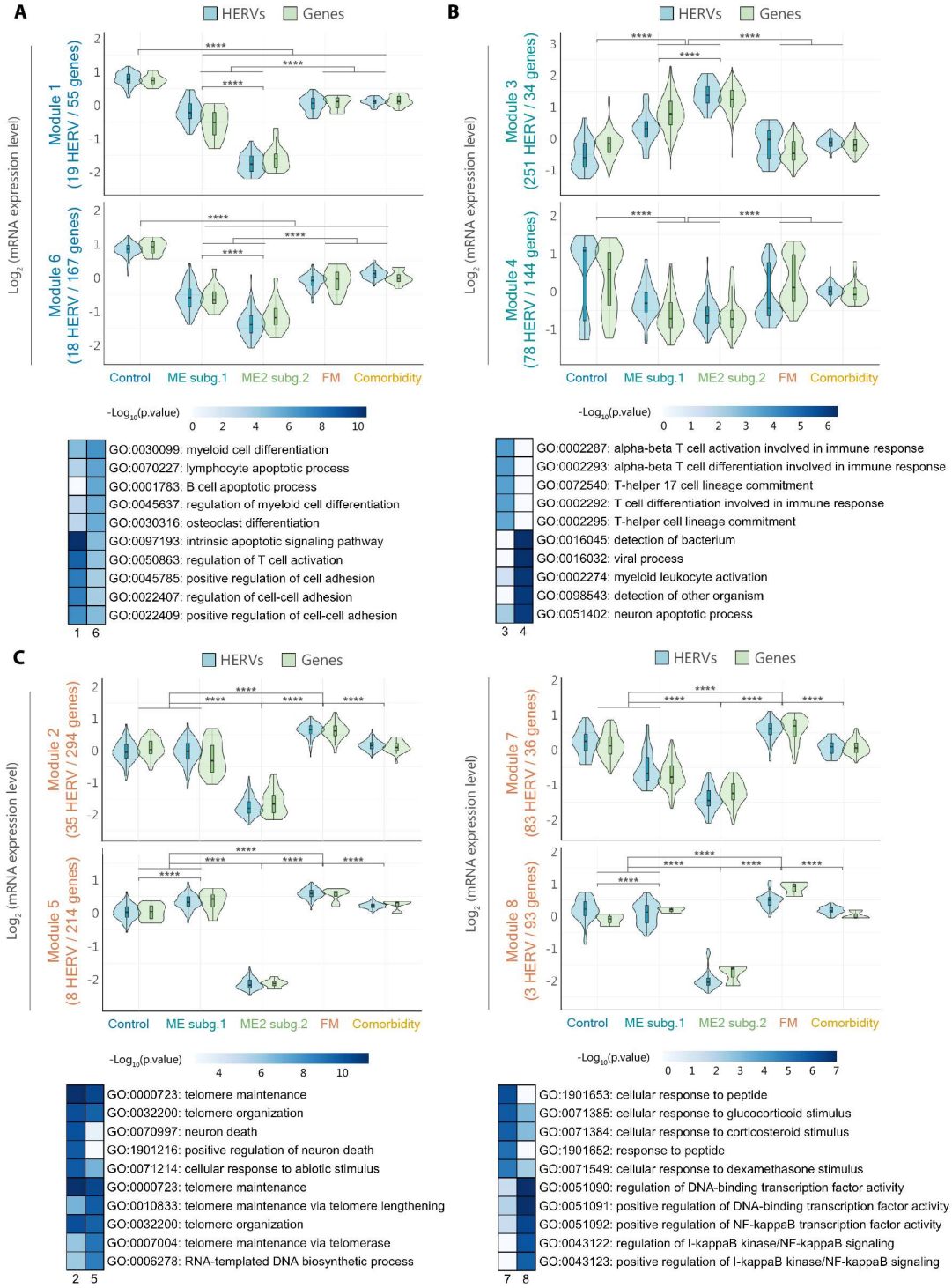

Figure 3. HERV expression correlates with immune response genes. Hierarchical clustering of differentially expressed HERV and genes (FDR<0.1 and |log2FC|>1) providing eight modules with highly correlated expression levels. Box and violin plots summarize the distribution and expression level

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).

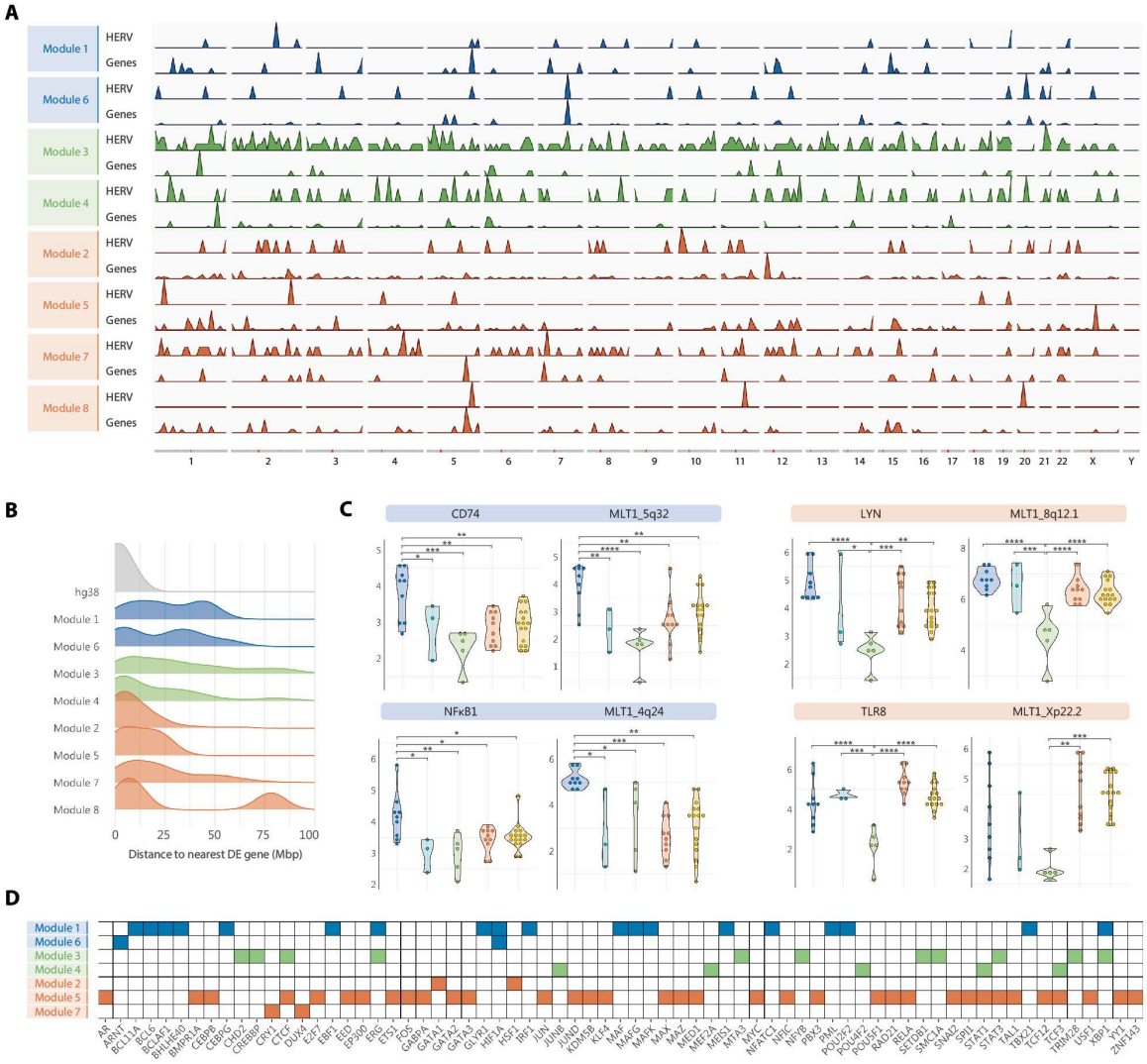

Figure 4. DE HERV sequences and DE immune-response genes are majorly independently transcribed (A) Genomic distribution of HERV, and genes encompassed in modules one through eight. **(B)** Genomic distance between DE HERV and their nearest DE genes per Module, plotted as density curves illustrating the 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), comorbidity (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.

To further search for potential mechanisms leading to HERV-gene DE correlations in the absence of cotranscription or other *cis*-regulatory mechanisms, 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 data provided by ReMap2022) (**Fig. 4D**), as common *trans*-regulatory mechanisms may explain their correlation by co-expression. Then, we performed GO analysis for HERV-gene sets enriched TFBS to infer their function (**Fig. S4**). 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 comorbidity, 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.

Differential HERV expression 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 microarray gene expression normalized data. The results interestingly evidence a decrease in γδ T cells as well as an increase in resting CD4 T memory cells 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 detected 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.

Subhead 7: 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 enough, 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**). Particularly, ME/CFS-specific upregulated elements MER31_3p26.1, HUERSP3_4q13.3, and HERV16_3p21.31 showed the strongest association with disease. Similarly, ME/CFS-specific upregulated genes NONO, XIAP, and CXXC1 were top genes with higher correlation with ME/CFS diagnosis (**Fig. 6B**). In addition, top DE HERV and genes showed significant associations with ME/CFS symptoms, with higher levels of MER31_3p26.1 and HERV16_3p21.31 associating with increased 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 for a subgroup of ME/CFS.

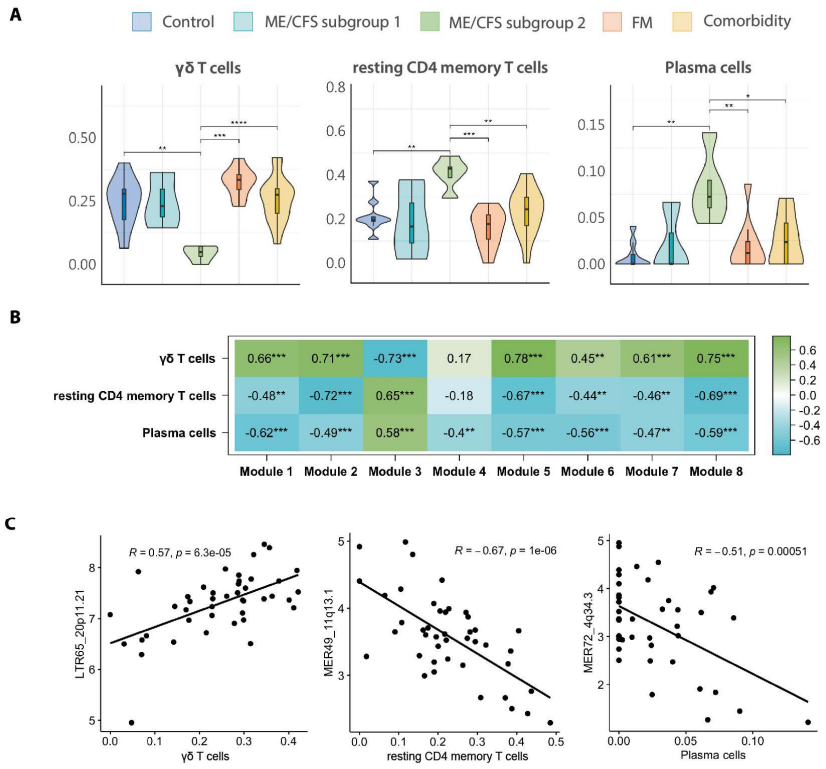

Figure 5. HERV fingerprints correlate with immune cell profiles. (A) Measurements of immune cell proportions by CIBERSORTx per study group (n=8 ME/CFS, n=10 FM, n=16 comorbid, n=9 controls). **(B)** Association between modules and immune cell proportions was evaluated by correlating eigengenes from each 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.

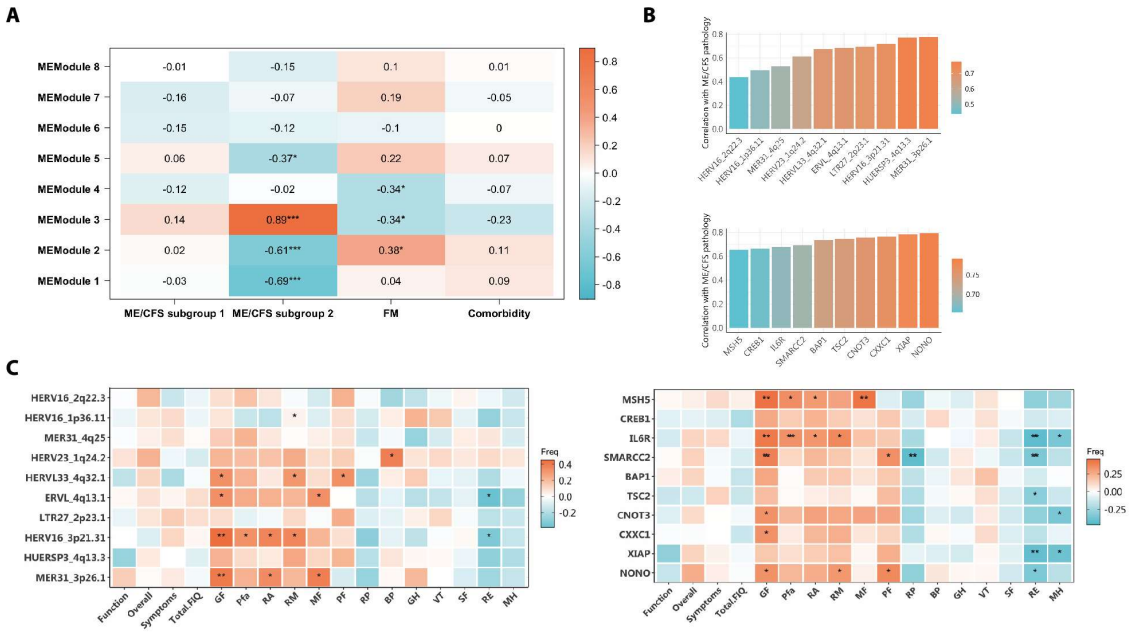

Figure 6. HERV and gene deregulation associates with disease symptomatology (A) Association between modules and disease determined by correlations of eigengenes from each module with binarized disease traits. Boxes show Pearson correlation values and associated p-values (*, p<0.05; **, p<0.01; ***, p<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 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).

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 *(13)* including samples of patients diagnosed with either ME/CFS, FM, or both (comorbid group) by two different clinical criteria (Canadian *(4)* and International Consensus *(5)* criteria, and the 1990 *(7)* and 2011 *(3, 6)* American College of Rheumatology (ACR) criteria, respectively).

Triggering agents for ME/CFS and FM, such as infections, stress, or trauma *(15, 16)*, seem capable of disrupting the epigenetic mechanisms constraining HERV expression *(40)*. Dysregulation of HERV has been proposed to contain more distinctive features than the host transcriptome itself in response to viral infections *(41)*, and it is becoming detected in additional complex diseases *(24)*, offering a new perspective to study their physiopathology. In fact, HERV profiling could constitute an indirect approach 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 *(33)* and FM *(34)*. 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 PBMC HERV transcriptomes (HERVomes) of ME/CFS, FM, and comorbidity patient groups as compared to HERVomes 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 HERVome-based diagnostic of these symptom-related, frequently comorbid diseases. Methodology that could be extended to the study of 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 comorbidity state. Worth noting are the results obtained in the comorbid 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 comorbid 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, comorbidity, 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. The study also defined a set of DE HERV commonly downregulated in all three patient groups (ME/CFS, FM, and their comorbidity, Cluster 4), a deeper study of which could 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 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 *(42–44)* with potential involvement in immune function performance *(45–49)*. Another HERV family consistently linked to disease is the HERV-W *(31, 50, 51)*, whose upregulation has been reported in FM *(34)*. 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 comorbidity groups belonged to a sole HERV family, the MLT1, usually found enriched near differentially expressed genes *(52)* contributing to the coding of conserved lncRNAs *(53)*, which suggests that altered gene expression by downregulation of HERV-derived lncRNAs may be a common denominator to ME/CFS, FM, and comorbid pathophysiologies.

In this regard and in accordance with genomic data *(54)*, most deregulated HERV loci consisted of solitary LTRs, known for their ability to influence gene expression through their regulatory sequences *(32)*. For example, a HERV-H solitary LTR located at the *HHLA2* locus serves as its main polyadenylation signal *(47, 48)*, allowing the encoding of the HHLA2 protein, which regulates CD4 and CD8 T cell functions along with antigen-presenting cell proliferation *(55)*. 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 *(56, 57)* even causing microdeletions *(58–60)*, which perhaps could explain the observed quantitative differences in HERV and gene expression profiles between the two ME/CFS subgroups identified in this study. A possibility worth to be evaluated by continuation studies.

Analysis of differential gene expression in the exact same samples provided advantageous evidence of HERV-associated immune changes in the ME/CFS group as compared to FM, comorbidity, or healthy controls, further supporting the existence of two ME/CFS subgroups. Our analysis of the correlation between deregulated HERV and gene profiles, combined with overrepresentation analysis of HERVome-associated genes, suggests the involvement of aberrantly expressed HERV in immune processes of ME/CFS, FM, and comorbid groups, a finding already reported for other diseases *(41)* and supported by HERV-mediated regulation of immune response *(61, 62)*. Strikingly, all upregulated HERV in ME/CFS (n=251) correlated with genes (n=34) involved in alpha-beta T cell activation and T-helper 17 cell commitment, in good agreement with the literature reporting alterations in these cell populations of ME/CFS *(63–65)*, and which seem especially relevant in autoimmune disorders *(66)*. Furthermore, some downregulated HERV correlated with pathogen detection-related genes, consistent with the suspected infectious onset of ME/CFS and FM *(16, 17)*. In addition, deregulation of other cellular processes formerly connected to these diseases e.g., NF-kB signaling *(67, 68)*, glucocorticoid sensitivity *(69, 70)*, and telomere attrition in immune cells *(71, 72)*, was also found in the ME/CFS and FM cases studied here, coordinated with HERV expression patterns, altogether supporting HERV's role for these orphan diseases.

Since the correlation between HERV and gene expression could be reflecting co-transcription or HERV-mediated *cis* gene regulation, we analyzed the genomic localization of correlated HERV and genes, finding no major overlaps between them, implying HERV transcription must be independent of correlated genes. However, it should be noted that gene probes included in the microarray do not fully picture the complete mRNA transcriptome, leaving opened the possibility that other differentially expressed genes locate nearer to deregulated HERV, scenario in which their modulation by HERV LTR could be possible. Furthermore, long-range interactions between gene promoters and LTR regulatory elements have been described for some HERV families, including MLT1, MER21, MER41, or LTR54 *(32)*, reported as DE in this study. In this regard, it has also been shown that the HERV-H family, upregulated in ME/CFS subgroup 2, has the ability to influence gene expression over long genomic distances, for example by creating topologically associated domains *(73)*, an aspect that cannot be discarded to explain 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 in ME/CFS, FM, and also in the comorbidity group. 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 *(74)*, pointing to the impairment of TRIM28/SETDB1 axis as the underlying cause of HERV upregulation in ME/CFS. In support of this, influenza A virus, a suspected trigger of ME/CFS *(75)*, has been shown to activate HERV expression by impairing the repressor TRIM28/SETDB1 axis *(76)*.

As the literature endorses changes in immune cell subsets in ME/CFS *(77–79)*, we aimed to indirectly measure immune cell type ratios in ME/CFS, FM, and their comorbidity with the use of CIBERSORTx software *(80)*. The detected increase in plasma cells and resting CD4 T memory cells in ME/CFS subgroup 2, suggests the presence, present and/or previous infections stimulating the immune system, coinciding with greater dysregulation in immune-related genes and HERV upregulation in this ME/CFS subgroup 2. 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 *(81, 82)*. It has been described that γδ T cells mediate the immune response against viral infections like EBV, HHV-6, or CMV *(83)*, which together with natural killer cell dysfunction reported in the mentioned autoimmune diseases *(84, 85)* but also in ME/CFS *(86)*, suggests that there is an impaired cytotoxic function rendering patients with lower counts of this cell type 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 particular cell subpopulations or influence immune cell ratios by stimulating an immune response. Complementary experiments are needed to address this question. In addition, we found a strong association between HERV activation and the condition of presenting ME/CFS disease for our subgroup 2, implying once more that the activation of these elements, whose expression associates with an impaired immune response, is tightly linked with the diagnosis of a subgroup of more severely affected ME/CFS cases. Some of the HERV elements with stronger associations with the diagnosis of ME/CFS disease were MER31_3p26.1 and HERV16_3p21.31, whose expression correlated with physical and mental fatigue according to health status assessment instruments MFI *(36)* and SF-36 *(37)*, 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 increased expression of MER31_3p26.1 and HERV16_3p21.31 elements relates to more affected patient health status. Interestingly enough, previous correlation between HERV expression and physical fatigue has been reported in the post-viral syndrome Post-COVID-19 condition *(31)*, which shares symptomatology with ME/CFS *(18)*. Further study of MER31_3p26.1 and HERV16_3p21.31 elements is needed to unveil their implication in disease symptoms.

Overall, this study not only shows the potential of HERV sequences as biomarkers for ME/CFS, FM, and for the definition of comorbidity but also exposes their potential involvement in the immune anomalies reported in these diseases, their etiological potential and their connection with patient symptoms. 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. demonstrated specific disruption of HERV families in response to particular viral infection agents. Perhaps a database enabling the comparison of disease HERV fingerprints with those induced by exogenous infections could help reveal the triggering agents behind a particular disease by studying their associated HERV fingerprints. Such tool could in principle allow for the assignment of an infection route map for ME/CFS, FM and comorbidity with the data obtained from this unique study. Nonetheless, much remains to be discovered about HERV nature and their role in human health and disease. Improved analytical tools to reveal the multifaceted ways in which HERV influence abnormal immune responses by their encoded proviral proteins, ncRNAs or dsRNA molecules are largely needed.

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 comorbidity (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 *(7)* and 2011 *(3, 6)* American College of Rheumatology (ACR) criteria for FM and the Canadian *(4)* and International Consensus *(5)* 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 *(35, 38)*, the MFI questionnaire *(36)*, and the quality-of-life SF-36 instrument *(37)*. 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. HERV transcriptome was scrutinized using custom high-density HERV-V3 microarrays, 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*(56)*. 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 differentially expressed 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 *(87)*. Data were normalized, adjusted for background noise, and summarized using the RMA (Robust Multi-Array) algorithm. Differential expression (DE) analysis was performed using limma R package *(88)*, considering differentially expressed those probes with a "Benjamin-Hochberg" (BH) adjusted p value<0.1 and an absolute $\log_2$ fold-change>1. The analysis results were presented in volcano plots using EnhancedVolcano R package *(89)*.

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 *(90)*. 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 differentially expressed HERV or gene probesets were used to cluster samples and probes. Data was represented in heatmap plots by using pheatmap R package *(91)*. 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 *(41)*.

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 *(92)* and the corresponding chromosome band information was added. The genomic context of the HERV loci was assessed with Goldmine R package *(93)*. 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. 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 *(94)*. 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 *(91)*, 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 *(95)* 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 *(91)*. 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 *(96)*. They were subsequently plotted over a linear chromosomal representation with karyoploteR package *(97)*, 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 *(80)* software on normalized non-logarithmic gene expression data for all samples. The default LM22 leukocyte gene signature matrix *(98)* was used as a reference. 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 *(99)*. 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 *(91)* 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). Differential expression 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 comorbidity). 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 $2^{-\Delta\Delta Ct}$ analysis to calculate fold-change 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.

REFERENCES

1. J. E. Harrison, S. Weber, R. Jakob, C. G. Chute, ICD-11: an international classification of diseases for the twenty-first century. *BMC Méd. Inform. Decis. Mak.* **21**, 206 (2021).

2. L. Bateman, A. C. Bested, H. F. Bonilla, B. V. Chheda, L. Chu, J. M. Curtin, T. T. Dempsey, M. E. Dimmock, T. G. Dowell, D. Felsenstein, D. L. Kaufman, N. G. Klimas, A. L. Komaroff, C. W. Lapp, S. M. Levine, J. G. Montoya, B. H. Natelson, D. L. Peterson, R. N. Podell, I. R. Rey, I. S. Ruhoy, M. A. Vera-Nunez, B. P. Yellman, Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Essentials of Diagnosis and Management. *Mayo Clin. Proc.* **96**, 2861–2878 (2021).

3. F. Wolfe, D. J. Clauw, M. Fitzcharles, D. L. Goldenberg, R. S. Katz, P. Mease, A. S. Russell, I. J. Russell, J. B. Winfield, M. B. Yunus, The American College of Rheumatology Preliminary Diagnostic Criteria for Fibromyalgia and Measurement of Symptom Severity. *Arthrit Care Res* **62**, 600–610 (2010).

4. B. M. Carruthers, A. K. Jain, K. L. D. Meirleir, D. L. Peterson, N. G. Klimas, A. M. Lerner, A. C. Bested, P. Flor-Henry, P. Joshi, A. C. P. Powles, J. A. Sherkey, M. I. van de Sande, Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *J. Chronic Fatigue Syndr.* **11**, 7–115 (2011).

5. B. M. Carruthers, M. I. van de Sande, K. L. D. Meirleir, N. G. Klimas, G. Broderick, T. Mitchell, D. Staines, A. C. P. Powles, N. Speight, R. Vallings, L. Bateman, B. Baumgarten-Austrheim, D. S. Bell, N. Carlo-Stella, J. Chia, A. Darragh, D. Jo, D. Lewis, A. R. Light, S. Marshall-Gradisbik, I. Mena, J. A. Mikovits, K. Miwa, M. Murovska, M. L. Pall, S. Stevens, Myalgic encephalomyelitis: International Consensus Criteria. *J. Intern. Med.* **270**, 327–338 (2011).

6. F. Wolfe, D. J. Clauw, M.-A. Fitzcharles, D. L. Goldenberg, W. Häuser, R. L. Katz, P. J. Mease, A. S. Russell, I. J. Russell, B. Walitt, 2016 Revisions to the 2010/2011 fibromyalgia diagnostic criteria. *Semin. Arthritis Rheum.* **46**, 319–329 (2016).

7. F. Wolfe, H. A. Smythe, M. B. Yunus, R. M. Bennett, C. Bombardier, D. L. Goldenberg, P. Tugwell, S. M. Campbell, M. Abeles, P. Clark, A. G. Fam, S. J. Farber, J. J. Fiechtner, C. M. Franklin, R. A. Gatter, D. Hamaty, J. Lessard, A. S. Lichtbroun, A. T. Masi, G. A. Mccain, W. J. Reynolds, T. J. Romano, I. J. Russell, R. P. Sheon, The american college of rheumatology 1990 criteria for the classification of fibromyalgia. *Arthritis Rheumatism* **33**, 160–172 (1990).

8. S. Wessely, C. Nimnuan, M. Sharpe, Functional somatic syndromes: one or many? *Lancet* **354**, 936–939 (1999).

9. B. H. Natelson, Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and Fibromyalgia: Definitions, Similarities, and Differences. *Clin. Ther.* **41**, 612–618 (2019).

10. B. Abbi, B. H. Natelson, Is chronic fatigue syndrome the same illness as fibromyalgia: evaluating the 'single syndrome' hypothesis. *QJM: Int. J. Med.* **106**, 3–9 (2013).

11. A. R. Light, L. Bateman, D. Jo, R. W. Hughen, T. A. VanHaitsma, A. T. White, K. C. Light, Gene expression alterations at baseline and following moderate exercise in patients with Chronic Fatigue Syndrome and Fibromyalgia Syndrome. *J. Intern. Med.* **271**, 64–81 (2012).

12. N. Groven, S. K. Reitan, E. A. Fors, I. C. Guzey, Kynurenine metabolites and ratios differ between Chronic Fatigue Syndrome, Fibromyalgia, and healthy controls. *Psychoneuroendocrinology* **131**, 105287 (2021).

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

14. J. Castro-Marrero, M. Faro, L. Aliste, N. Sáez-Francàs, N. Calvo, A. Martínez-Martínez, T. F. de Sevilla, J. Alegre, Comorbidity in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis: A Nationwide Population-Based Cohort Study. *Psychosomatics* **58**, 533–543 (2017).

15. P. J. Furness, K. Vogt, S. Ashe, S. Taylor, S. Haywood-Small, K. Lawson, What causes fibromyalgia? An online survey of patient perspectives. *Heal. Psychol. Open* **5**, 2055102918802683 (2018).

16. L. Chu, I. J. Valencia, D. W. Garvert, J. G. Montoya, Onset Patterns and Course of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Front. Pediatr.* **7**, 12 (2019).

17. R. Tschopp, R. S. König, P. Rejmer, D. H. Paris, Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS): A preliminary survey among patients in Switzerland. *Heliyon* **9**, e15595 (2023).

18. A. L. Komaroff, W. I. Lipkin, ME/CFS and Long COVID share similar symptoms and biological abnormalities: road map to the literature. *Front. Med.* **10**, 1187163 (2023).

19. D. J. Clauw, L. Calabrese, Rheumatology and Long COVID: lessons from the study of fibromyalgia. *Ann. Rheum. Dis.* , ard-2023-224250 (2023).

20. J. B. Soriano, S. Murthy, J. C. Marshall, P. Relan, J. V. Diaz, W. C. C. D. W. G. on P.-C.-19 Condition, A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect. Dis.* **22**, e102–e107 (2022).

21. L. A. Jason, J. A. Dorri, ME/CFS and Post-Exertional Malaise among Patients with Long COVID. *Neurol. Int.* **15**, 1–11 (2022).

22. X. Liu, T. Hong, S. Parameswaran, K. Ernst, I. Marazzi, M. T. Weirauch, J. I. F. Bass, Human Virus Transcriptional Regulators. *Cell* **182**, 24–37 (2020).

23. M. G. Macchietto, R. A. Langlois, S. S. Shen, Virus-induced transposable element expression up-regulation in human and mouse host cells. *Life Sci. Alliance* **3**, e201900536 (2020).

24. K. Giménez-Orenga, E. Oltra, Human Endogenous Retrovirus as Therapeutic Targets in Neurologic Disease. *Pharm* **14**, 495 (2021).

25. G. Kassiotis, The Immunological Conundrum of Endogenous Retroelements. *Annu. Rev. Immunol.* **41**, 99–125 (2023).

26. D. S. Lima-Junior, S. R. Krishnamurthy, N. Bouladoux, N. Collins, S.-J. Han, E. Y. Chen, M. G. Constantinides, V. M. Link, A. I. Lim, M. Enamorado, C. Cataisson, L. Gil, I. Rao, T. K. Farley, G. Koroleva, J. Attig, S. H. Yuspa, M. A. Fischbach, G. Kassiotis, Y. Belkaid, Endogenous retroviruses promote homeostatic and inflammatory responses to the microbiota. *Cell* **184**, 3794-3811.e19 (2021).

27. N. Dopkins, M. M. O'Mara, B. Singh, J. L. Marston, M. L. Bendall, D. F. Nixon, How human endogenous retroviruses interact with the microbiota in health and disease. *Trends Microbiol* (2022), doi:10.1016/j.tim.2022.05.011.

28. R. Ferrari, N. Grandi, E. Tramontano, G. Dieci, Retrotransposons as Drivers of Mammalian Brain Evolution. *Life* **11**, 376 (2021).

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

30. A. I. Khadjinova, X. Wang, A. Laine, K. Ukadike, M. Eckert, A. Stevens, A. A. Bengtsson, C. Lood, T. Mustelin, 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 (2021).

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

32. J. Ito, R. Sugimoto, H. Nakaoka, S. Yamada, T. Kimura, T. Hayano, I. Inoue, Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. *Plos Genet* **13**, e1006883 (2017).

33. L. S. Rodrigues, L. H. da S. Nali, C. O. D. Leal, E. C. Sabino, E. M. Lacerda, C. C. Kingdon, L. Nacul, C. M. Romano, HERV-K and HERV-W transcriptional activity in myalgic encephalomyelitis/chronic fatigue syndrome. *Autoimmun Highlights* **10**, 12 (2019).

34. T. Ovejero, O. Sadones, T. Sánchez-Fito, E. Almenar-Pérez, J. A. Espejo, E. Martín-Martínez, L. Nathanson, E. Oltra, Activation of Transposable Elements in Immune Cells of Fibromyalgia Patients. *Int J Mol Sci* **21**, 1366 (2020).

35. C. S. Burckhardt, S. R. Clark, R. M. Bennett, The fibromyalgia impact questionnaire: development and validation. *J Rheumatology* **18**, 728–33 (1991).

36. E. M. A. Smets, B. Garssen, B. Bonke, J. C. J. M. D. Haes, The multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *J Psychosom Res* **39**, 315–325 (1995).

37. C. A. MCHORNEY, J. WARE, A. RACZEK, The MOS 36-Item Short-Form Health Survey (SF-36). *Med Care* **31**, 247–263 (1993).

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

39. J. Becker, P. Pérot, V. Cheynet, G. Oriol, N. Mugnier, M. Mommert, O. Tabone, J. Textoris, J.-B. Veyrieras, F. Mallet, A comprehensive hybridization model allows whole HERV transcriptome profiling using high density microarray. *Bmc Genomics* **18**, 286 (2017).

40. S. C. Rangel, M. D. da Silva, A. L. da Silva, J. de M. B. dos Santos, L. M. Neves, A. Pedrosa, F. M. Rodrigues, C. dos S. Trettel, G. E. Furtado, M. P. de Barros, A. L. L. Bachi, C. M. Romano, L. H. D. S. Nali, Human endogenous retroviruses and the inflammatory response: A vicious circle associated with health and illness. *Front. Immunol.* **13**, 1057791 (2022).

41. J. L. Marston, M. Greenig, M. Singh, M. L. Bendall, R. R. R. Duarte, C. Feschotte, L. P. Iñiguez, D. F. Nixon, SARS-CoV-2 infection mediates differential expression of human endogenous retroviruses and long interspersed nuclear elements. *Jci Insight* **6**, e147170 (2021).

42. J.-M. Yi, H.-M. Kim, H.-S. Kim, Human endogenous retrovirus HERV-H family in human tissues and cancer cells: expression, identification, and phylogeny. *Cancer Lett.* **231**, 228–239 (2006).

43. G. Morris, M. Maes, M. Murdjeva, B. K. Puri, Do Human Endogenous Retroviruses Contribute to Multiple Sclerosis, and if So, How? *Mol. Neurobiol.* **56**, 2590–2605 (2019).

44. M. Bergallo, K. Mareschi, P. Montanari, C. Calvi, M. Berger, I. Bini, V. Daprà, I. Galliano, F. Fagioli, Human endogenous retrovirus, HERV-P and HERV-R in pediatric leukemia patients. *J. Hematop.* **12**, 51–56 (2019).

45. M. Mangeney, N. de Parseval, G. Thomas, T. Heidmann, The full-length envelope of an HERV-H human endogenous retrovirus has immunosuppressive properties. *J. Gen. Virol.* **82**, 2515–2518 (2001).

46. J. K. Kulski, S. Gaudieri, H. Inoko, R. L. Dawkins, Comparison Between Two Human Endogenous Retrovirus (HERV)-Rich Regions Within the Major Histocompatibility Complex. *J. Mol. Evol.* **48**, 675–683 (1999).

47. N. L. Goodchild, D. A. Wilkinson, D. L. Mager, Recent Evolutionary Expansion of a Subfamily of RTVL-H Human Endogenous Retrovirus-like Elements. *Virology* **196**, 778–788 (1993).

48. D. L. Mager, D. G. Hunter, M. Schertzer, J. D. Freeman, Endogenous Retroviruses Provide the Primary Polyadenylation Signal for Two New Human Genes (HHLA2 and HHLA3). *Genomics* **59**, 255–263 (1999).

49. A. Matsuzawa, J. Lee, S. Nakagawa, J. Itoh, M. T. Ueda, S. Mitsuhashi, Y. Kochi, T. Kaneko-Ishino, F. Ishino, HERV-Derived Ervpb1 Is Conserved in Simiiformes, Exhibiting Expression in Hematopoietic Cell Lineages Including Macrophages. *Int. J. Mol. Sci.* **22**, 4504 (2021).

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

51. A. Dolei, J. A. Garson, G. Arru, M. Clerici, R. Germi, P. N. Marche, H. Perron, Multiple sclerosis-associated retrovirus and related human endogenous retrovirus-W in patients with multiple sclerosis. *J Neuroimmunol* **266**, 87–88 (2014).

52. L. Bogdan, L. Barreiro, G. Bourque, Transposable elements have contributed human regulatory regions that are activated upon bacterial infection. *Philosophical Transactions Royal Soc B* **375**, 20190332 (2020).

53. L. Ramsay, M. C. Marchetto, M. Caron, S.-H. Chen, S. Busche, T. Kwan, T. Pastinen, F. H. Gage, G. Bourque, Conserved expression of transposon-derived non-coding transcripts in primate stem cells. *BMC Genom.* **18**, 214 (2017).

54. I. H. G. S. Consortium, W. I. for B. R. Research: Center for Genome, E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, T. S. Centre:, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, W. U. G. S. Center, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, U. D. J. G. Institute:, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.-F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, B. C. of M. H. G. S. Center:, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, R. G. S. Center:, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, G. and C. UMR-8030:, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. of G. A. Biotechnology: Institute of Molecular, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, G. S. Center:, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, B. G. I. G. Center:, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, M. S. C. Biology: The Institute for Systems, L. Hood, L. Rowen, A. Madan, S. Qin, S. G. T. Center:, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, U. of O. A. C. for G. Technology:, B. A. Roe, F. Chen, H. Pan, M. P. I. for M. Genetics:, J. Ramser, H. Lehrach, R. Reinhardt, C. S. H. L. Center: Lita Annenberg Hazen Genome, W. R. McCombie, M. de la Bastide, N. Dedhia, G. R. C. for Biotechnology:, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork,

D. G. Brown, C. B. Burge, L. Cerutti, H.-C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. R. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. A. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S.-P. Yang, R.-F. Yeh, S. management: N. H. G. R. I. Health: US National Institutes of, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, S. H. G. Center:, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, U. of W. G. Center:, M. V. Olson, R. Kaul, C. Raymond, D. of M. B. Medicine: Keio University School of, N. Shimizu, K. Kawasaki, S. Minoshima, U. of T. S. M. C. at Dallas:, G. A. Evans, M. Athanasiou, R. Schultz, O. of S. Energy: US Department of, A. Patrinos, T. W. Trust:, M. J. Morgan, Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

55. R. Zhao, J. M. Chinai, S. Buhl, L. Scandiuzzi, A. Ray, H. Jeon, K. C. Ohaegbulam, K. Ghosh, A. Zhao, M. D. Scharff, X. Zang, HHLA2 is a member of the B7 family and inhibits human CD4 and CD8 T-cell function. *Proc. Natl. Acad. Sci.* **110**, 9879–9884 (2013).

56. J. Thomas, H. Perron, C. Feschotte, Variation in proviral content among human genomes mediated by LTR recombination. *Mobile Dna-uk* **9**, 36 (2018).

57. R. Rebollo, K. Miceli-Royer, Y. Zhang, S. Farivar, L. Gagnier, D. L. Mager, Epigenetic interplay between mouse endogenous retroviruses and host genes. *Genome Biol.* **13**, R89 (2012).

58. K. E. Hermetz, U. Surti, J. D. Cody, M. K. Rudd, A recurrent translocation is mediated by homologous recombination between HERV-H elements. *Mol. Cytogenet.* **5**, 6 (2012).

59. A. Sanchez-Valle, X. Wang, L. Potocki, Z. Xia, S. L. Kang, M. E. Carlin, D. Michel, P. Williams, G. Cabrera-Meza, E. K. Brundage, A. L. Eifert, P. Stankiewicz, S. W. Cheung, S. R. Lalani, HERV-mediated genomic rearrangement of EYA1 in an individual with branchio-oto-renal syndrome. *Am. J. Méd. Genet. Part A* **152A**, 2854–2860 (2010).

60. A. Shuvarikov, I. M. Campbell, P. Dittwald, N. J. Neill, M. G. Bialer, C. Moore, P. G. Wheeler, S. E. Wallace, M. C. Hannibal, M. F. Murray, M. A. Giovanni, D. Terespolsky, S. Sodhi, M. Cassina, D. Viskochil, B. Moghaddam, K. Herman, C. W. Brown, C. R. Beck, A. Gambin, S. W. Cheung, A. Patel, A. N. Lamb, L. G. Shaffer, J. W. Ellison, J. B. Ravnan, P. Stankiewicz, J. A. Rosenfeld, 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 (2013).

61. M. Greenig, HERVs, immunity, and autoimmunity: understanding the connection. *Peerj* **7**, e6711 (2019).

62. E. Russ, S. Iordanskiy, Endogenous Retroviruses as Modulators of Innate Immunity. *Pathogens* **12**, 162 (2023).

63. K. Metzger, M. Frémont, C. Roelant, K. D. Meirleir, Lower frequency of IL-17F sequence variant (His161Arg) in chronic fatigue syndrome patients. *Biochem Bioph Res Co* **376**, 231–233 (2008).

64. T. Yang, Y. Yang, D. Wang, C. Li, Y. Qu, J. Guo, T. Shi, W. Bo, Z. Sun, T. Asakawa, The clinical value of cytokines in chronic fatigue syndrome. *J. Transl. Med.* **17**, 213 (2019).

65. A. L. Smylie, G. Broderick, H. Fernandes, S. Razdan, Z. Barnes, F. Collado, C. Sol, M. A. Fletcher, N. Klimas, A comparison of sex-specific immune signatures in Gulf War illness and chronic fatigue syndrome. *BMC Immunol.* **14**, 29 (2013).

66. K. Yasuda, Y. Takeuchi, K. Hirota, The pathogenicity of Th17 cells in autoimmune diseases. *Semin. Immunopathol.* **41**, 283–297 (2019).

67. M. Maes, M. Kubera, M. Kotańska, Aberrations in the Cross-Talks Among Redox, Nuclear Factor-κB, and Wnt/β-Catenin Pathway Signaling Underpin Myalgic Encephalomyelitis and Chronic Fatigue Syndrome. *Front. Psychiatry* **13**, 822382 (2022).

68. C.-L. Han, Y.-C. Sheng, S.-Y. Wang, Y.-H. Chen, J.-H. Kang, Serum proteome profiles revealed dysregulated proteins and mechanisms associated with fibromyalgia syndrome in women. *Sci. Rep.* **10**, 12347 (2020).

69. W. C. de Vega, S. Herrera, S. D. Vernon, P. O. McGowan, Epigenetic modifications and glucocorticoid sensitivity in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *BMC Méd. Genom.* **10**, 11 (2017).

70. J. A. Macedo, J. Hesse, J. D. Turner, J. Meyer, D. H. Hellhammer, C. P. Muller, Glucocorticoid sensitivity in fibromyalgia patients: Decreased expression of corticosteroid receptors and glucocorticoid-induced leucine zipper. *Psychoneuroendocrinology* **33**, 799–809 (2008).

71. A. L. Hassett, E. Epel, D. J. Clauw, R. E. Harris, S. E. Harte, A. Kairys, S. Buyske, D. A. Williams, Pain Is Associated With Short Leukocyte Telomere Length in Women With Fibromyalgia. *The J. Pain* **13**, 959–969 (2012).

72. M. S. Rajeevan, J. Murray, L. Oakley, J.-M. S. Lin, E. R. Unger, Association of chronic fatigue syndrome with premature telomere attrition. *J Transl Med* **16**, 44 (2018).

73. Y. Zhang, T. Li, S. Preissl, M. L. Amaral, J. D. Grinstein, E. N. Farah, E. Destici, Y. Qiu, R. Hu, A. Y. Lee, S. Chee, K. Ma, Z. Ye, Q. Zhu, H. Huang, R. Fang, L. Yu, J. C. I. Belmonte, J. Wu, S. M. Evans, N. C. Chi, B. Ren, Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat. Genet.* **51**, 1380–1388 (2019).

74. K. Giménez-Orenga, E. Oltra, Human Endogenous Retrovirus as Therapeutic Targets in Neurologic Disease. *Pharm* **14**, 495 (2021).

75. P. Magnus, N. Gunnes, K. Tveito, I. J. Bakken, S. Ghaderi, C. Stoltenberg, M. Hornig, W. I. Lipkin, L. Trogstad, S. E. Håberg, 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 (2015).

76. N. Schmidt, P. Domingues, F. Golebiowski, C. Patzina, M. H. Tatham, R. T. Hay, B. G. Hale, An influenza virus-triggered SUMO switch orchestrates co-opted endogenous retroviruses to stimulate host antiviral immunity. *Proc. Natl. Acad. Sci.* **116**, 17399–17408 (2019).

77. E. W. Brenu, M. L. van Driel, D. R. Staines, K. J. Ashton, S. B. Ramos, J. Keane, N. G. Klimas, S. M. Marshall-Gradisnik, Immunological abnormalities as potential biomarkers in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. *J. Transl. Med.* **9**, 81 (2011).

78. M. Curriu, J. Carrillo, M. Massanella, J. Rigau, J. Alegre, J. Puig, A. M. Garcia-Quintana, J. Castro-Marrero, E. Negredo, B. Clotet, C. Cabrera, J. Blanco, Screening NK-, B- and T-cell phenotype and function in patients suffering from Chronic Fatigue Syndrome. *J. Transl. Med.* **11**, 68 (2013).

79. E. W. Brenu, T. K. Huth, S. L. Hardcastle, K. Fuller, M. Kaur, S. Johnston, S. B. Ramos, D. R. Staines, S. M. Marshall-Gradisnik, Role of adaptive and innate immune cells in chronic fatigue syndrome/myalgic encephalomyelitis. *Int. Immunol.* **26**, 233–242 (2014).

80. A. M. Newman, C. B. Steen, C. L. Liu, A. J. Gentles, A. A. Chaudhuri, F. Scherer, M. S. Khodadoust, M. S. Esfahani, B. A. Luca, D. Steiner, M. Diehn, A. A. Alizadeh, Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).

81. L. Wang, N. Kang, J. Zhou, Y. Guo, X. Zhang, L. Cui, D. Ba, W. He, Downregulation of CD94/NKG2A Inhibitory Receptor on Decreased γδ T Cells in Patients with Systemic Lupus Erythematosus. *Scand. J. Immunol.* **76**, 62–69 (2012).

82. N. S. Fay, E. C. Larson, J. M. Jameson, Chronic Inflammation and γδ T Cells. *Front. Immunol.* **7**, 210 (2016).

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

84. A. Henriques, L. Teixeira, L. Inês, T. Carvalheiro, A. Gonçalves, A. Martinho, M. L. Pais, J. A. P. da Silva, A. Paiva, NK cells dysfunction in systemic lupus erythematosus: relation to disease activity. *Clin. Rheumatol.* **32**, 805–813 (2013).

85. J. H. Kim, K. Park, S. B. Lee, S. Kang, J. S. Park, C. W. Ahn, J. S. Nam, Relationship between natural killer cell activity and glucose control in patients with type 2 diabetes and prediabetes. *J. Diabetes Investig.* **10**, 1223–1228 (2019).

86. N. Eaton-Fitch, S. du Preez, H. Cabanas, D. Staines, S. Marshall-Gradisnik, A systematic review of natural killer cells profile and cytotoxic function in myalgic encephalomyelitis/chronic fatigue syndrome. *Syst Rev* **8**, 279 (2019).

87. B. S. Carvalho, R. A. Irizarry, A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363–2367 (2010).

88. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).

89. K. Blighe, S. Rana, M. Lewis, EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling (available at https://github.com/kevinblighe/EnhancedVolcano).

90. A. Kassambara, F. Mundt, Factoextra: extract and Visualize the Results of Multivariate Data Analyses (available at https://rpkgs.datanovia.com/factoextra/index.html).

91. R. Kolde, A. Lizee, T. Metsalu, Pheatmap: Pretty heatmaps (available at https://github.com/raivokolde/pheatmap).

92. R. M. Kuhn, D. Haussler, W. J. Kent, The UCSC genome browser and associated tools. *Brief. Bioinform.* **14**, 144–161 (2013).

93. J. M. Bhasin, A. H. Ting, Goldmine integrates information placing genomic ranges into meaningful biological contexts. *Nucleic Acids Res.* **44**, 5550–5556 (2016).

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

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

96. M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, V. J. Carey, Software for Computing and Annotating Genomic Ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).

97. B. Gel, E. Serra, karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**, 3088–3090 (2017).

98. A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn, A. A. Alizadeh, Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).

99. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *Bmc Bioinformatics* **9**, 559–559 (2008).

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 volunteers. We also acknowledge bioMérieux and Sampled for assisting us in the analysis of samples by microarray.

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.

Author 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.

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: 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.