Title

Cell-Specific Transposable Element and Gene Expression Analysis Across Systemic Lupus Erythematosus Phenotypes

Journal

ACR Open Rheumatology, 6(11)

ISSN

2578-5745

Authors

Cutts, Zachary
Patterson, Sarah
Maliskova, Lenka
et al.

Publication Date

2024-11-01

DOI

10.1002/acr2.11713

Peer reviewed

ACR Open Rheumatology
Vol. 6, No. 11, November 2024, pp 769–779
DOI 10.1002/acr2.11713

American College of Rheumatology
Empowering Rheumatology Professionals

Cell-Specific Transposable Element and Gene Expression Analysis Across Systemic Lupus Erythematosus Phenotypes

Zachary Cutts,[1] Sarah Patterson,[1] Lenka Maliskova,[1] Kimberly E. Taylor,[1] Chun Jimmie Ye,[1] Maria Dall'Era,[1] Jinoos Yazdany,[1] Lindsey A. Criswell,[2] Gabriela K. Fragiadakis,[1] Charles Langelier,[1] John A. Capra,[1] Marina Sirota,[1*] and Cristina M. Lanata[2*]

Objective. There is an established yet unexplained link between interferon (IFN) and systemic lupus erythematosus (SLE). The expression of sequences derived from transposable elements (TEs) may contribute to SLE phenotypes, specifically production of type I IFNs and generation of autoantibodies.

Methods. We profiled cell-sorted RNA-sequencing data ($CD4^+$ T cells, $CD14^+$ monocytes, $CD19^+$ B cells, and natural killer cells) from peripheral blood mononuclear cells of 120 patients with SLE and quantified TE expression identifying 27,135 TEs. We tested for differential TE expression across 10 SLE phenotypes, including autoantibody production and disease activity.

Results. We found 731 differentially expressed (DE) TEs across all SLE phenotypes that were mostly cell specific and phenotype specific. DE TEs were enriched for specific families and open reading frames of viral genes encoded in TE sequences. Increased expression of DE TEs was associated with genes involved in antiviral activity, such as LY6E, ISG15, and TRIM22, and pathways such as IFN signaling.

Conclusion. These findings suggest that expression of TEs contributes to activation of SLE-related mechanisms in a cell-specific manner, which can impact disease diagnostics and therapeutics.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a heterogeneous chronic autoimmune disease characterized by antibodies against nucleic acids and associated proteins.[1,2] The clinical manifestations vary among different racial and ethnic groups, and the risk of developing severe manifestations is increased in Black, Asian and Pacific Islander, and Hispanic patients.[3] Transcriptomic studies have shown that SLE is also molecularly heterogeneous and that different cell types have specific gene expression profiles.[4]

Transposable element (TE)–derived sequences make up approximately 50% of the human genome. TEs are mobile genetic elements capable of changing their location within genomes. The vast majority of TE sequences in the human genome are inactive and can no longer transpose, but they still contain sequences with the potential to encode proteins and functional gene regulatory elements. A small number of specific classes of TEs are intact and active in humans, and there is some evidence of differences across human populations.[5] As a result, TEs have roles both in human health and diseases such as cancer and autoimmunity.[6–8] Recent work has suggested that two subsets of TEs, human endogenous retroviruses (HERVs) and long interspersed nuclear elements (LINEs), may play a pathogenic role in SLE.[9,10] One of many hypotheses for a mechanism by which TEs could trigger an immune response is their expression either as nucleic acids or proteins that resemble molecular patterns of exogenous viruses.[11,12] With the accumulation of TEs, an interferon (IFN) response could be triggered, which could result in a positive feedback loop that upregulates IFN-stimulated genes.[12]

Supported by the NIH (grant P30-AR-070155) and the National Center for Complementary and Integrative Health, NIH (grant K23-AT-011768).

[1]Zachary Cutts, MS, Sarah Patterson, MD, Lenka Maliskova, MS, Kimberly E. Taylor, PhD, MPH, Chun Jimmie Ye, PhD, Maria Dall'Era, MD, Jinoos Yazdany, MD, MPH, Gabriela K. Fragiadakis, PhD, Charles Langelier, MD, PhD, John A. Capra, PhD, Marina Sirota, PhD: University of California, San Francisco, San Francisco; [2]Lindsey A. Criswell, MD, MPH, DSc, Cristina M. Lanata MD: National Human Genome Research Institute, NIH, Bethesda, Maryland.

*Drs Sirota and Lanata are co-last authors and contributed equally to this work.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr2.11713).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr2.11713.

Address correspondence via email to Marina Sirota, PhD, at marina.sirota@ucsf.edu; or to Cristina Lanata, MD, at cristina.lanata@nih.gov.

Submitted for publication March 7, 2024; accepted in revised form June 4, 2024.

In previous work, over 100 locus-specific HERVs were shown to be differentially expressed (DE) in patients with SLE and correlate with lupus clinical parameters, such as the presence of double-stranded DNA (dsDNA), anti-RNP, and anti-Smith (anti-SM) antibodies.[13] Previous studies have also identified elevated expression of HERVs in SLE, implicating HERV-E clone 4-1 in peripheral blood mononuclear cells (PBMCs) and HRES1/p28 in B cells.[14,15] In addition, more recent work has characterized TEs in blood from patients with SLE compared with matched controls and found up-regulation of TEs in SLE.[8,13,16] These studies also investigated whether TEs contribute to the IFN signature observed in patients with SLE, with one study finding a positive correlation between HERV expression and the IFN signature, whereas the other did not.[8,13]

Although many studies have explored TE expression among patients with SLE and healthy patients, none have defined locus-specific TE expression in immune cell types relevant to SLE or characterized the relationship between TE expression and different lupus manifestations in a diverse patient cohort. In addition, prior work has not examined the cell-specific association of DE TEs in SLE with gene expression and gene set enrichment analysis. The goal of this study was to characterize the role of TEs in the clinical heterogeneity of SLE across multiple cell types. Here, we report on TE expression in 4 distinct cell types and 10 SLE subphenotypes, conduct family and viral gene enrichment analysis, and correlate cell-specific TE expression with SLE heterogeneity as well as cell-specific transcriptomics in a diverse cohort of patients with lupus with extensive phenotypic data.

MATERIALS AND METHODS

Cohort description and data generation. All patients in this study were from the California Lupus Epidemiology Study (CLUES). CLUES was approved by the Institutional Review Board of the University of California, San Francisco. All patients signed a written informed consent to participate in CLUES. Study procedures are described previously.[17] Clinical data collected at sampling and self-reported race (which patients chose from a fixed set of categories) was used for downstream analysis. Raw data of this study are openly available in GEO: GSE164457. All other data are available from the corresponding author upon reasonable request.

Subphenotype definitions. Disease activity was measured with the standardized SLE disease activity (SLEDAI) score.[18] A high SLEDAI score was defined as a score ≥8, whereas low was defined as <8. Besides the total SLEDAI score, we also performed analyses with specific items of the SLEDAI score, such as proteinuria and the presence of dsDNA antibody. From our own previous work, three stable clusters named mild disease, severe disease 1, and severe disease 2 were revealed from unsupervised clustering of the 11 American College of Rheumatology Classification Criteria characterized by significant differences in SLE manifestations.[17] We also performed subphenotype associations with a history of the presence of anti-SM antibody, anti-RNP antibody, and anti-dsDNA. Photosensitivity was defined as a rash or feeling sick after going out in the sun. Serologies were performed in Clinical Laboratory Improvement Amendments–certified laboratory tests and reported as abnormal or normal.

RNA-sequencing data generation. PBMCs were isolated from 120 participants with SLE. Using the EasySep protocol from STEMCELL Technologies, these cells were sorted into $CD14^+$ monocytes, B cells, $CD4^+$ T cells, and natural killer (NK) cells, for a total of 480 samples. These samples were sequenced on a HiSeq4000 PE150, gene expression data was generated using Salmon version 0.8.2 with adapter-trimmed reads, and quality control (QC) was performed as previously described.[2]

Genotyping. Genotyping for genomic DNA from peripheral blood was performed using the Affymetrix Axiom Genome-Wide LAT 1 Array. Samples with Dish QC ≥ 0.82 were retained. Using SNPolisher, single nucleotide polymorphism (SNP) genotypes were filtered for high-quality cluster differentiation and 95% call rate within batches. PLINK was used for additional QC. SNPs having an overall call rate <95% or discordant calls in duplicate samples were dropped. Samples were dropped for unexpected duplicates in identity by descent analysis or mismatched sex between genetics and self-report; for first-degree relatives, one sample was retained. All samples had at least 95% genotyping and no evidence of excess heterozygosity (maximum <2.5 SD). We tested for Hardy-Weinberg Equilibrium (HWE) and cross-batch association for batch effects using a subset of subjects that were of European ancestry and negative for dsDNA antibodies and renal disease to minimize genetic heterogeneity. SNPs were dropped if HWE $P < 1 \times 10^{-5}$ or any cross-batch association $P < 5 \times 10^{-5}$. Genetic principal components (PCs) were generated using EIGENSTRAT and used for patient stratification.

TE QC and expression quantification. TE expression quantification was performed using adapter-trimmed reads from RNA-sequencing (RNA-seq) data that were aligned with bowtie2 to hg38, allowing for 100 alignments per read, using the very sensitive local setting (-k 100 –very-sensitivie-local –score-min L,0,1.6). The resulting sorted bams with ambiguous reads were used as input into Telescope with default settings and the reference retro.hg38.v1 annotation from https://github.com/mlbendall/telescope_annotation_db/tree/master/builds to reassign reads. Because TEs are highly repetitive sequences and RNA-seq reads can align ambiguously at multiple locations in the genome (multimappers), we used Telescope, which applies an expectation-maximization algorithm to reassign multimapped

reads to a specific location.[7] Locus-specific TE read counts generated by Telescope were used for downstream analysis.

Differential TE expression analysis. TE differential expression analysis was run with DESeq2 (version 1.38.3) using counts from Telescope with outliers dropped per cell type, adjusting for age, sequencing lane, sex, genetic ancestry PCs 1–10, and medication at the time of blood draw. For data analyses, we grouped immunosuppressive medications into the following categories: biologic treatments (belimumab, abatacept, and rituximab), low-dose prednisone (<10 mg), moderate or high-dose prednisone (>10 mg), antimalarials, calcineurin inhibitors, methotrexate and leflunomide, azathioprine, mycophenolate mofetil, and cyclophosphamide. Medications were used as factors in DESeq2 as covariates. Outcomes studied included disease activity (SLEDAI score), IFN signature high/low based on gene signature developed by Kennedy et al,[19] photosensitivity (previous work has found UV light induces HERV expression),[20] proteinuria, autoantibody production (dsDNA, RNP, and SM), and disease severity (as defined by clinical clusters previously described in the same participants with SLE).[17] For stratified analyses, patients were stratified according to genetic similarity (PC1 >0.025 for Asian ancestry, <–0.025 for European ancestry), and admixed individuals were not considered for downstream stratified analysis. All *P* values from the subphenotype analysis were false discovery rate corrected using Benjamini-Hochberg. DE TEs (adjusted *P* value [padj] <0.05) for all SLE subphenotypes per cell type were used for downstream analysis. PC analyses (PCAs) of TE expression were computed with the factoextra package (version 1.0.7). Cell type PCAs were made using the variance stabilizing transformation with the DESeq2 function vst(), and variance stabilizing transformed data was visualized with plotPCA from the DESeq2 package.

Characterization of DE TEs. TE family definitions were defined using https://github.com/mlbendall/telescope_annotation_db/tree/master/builds. Locus-specific DE TEs across all SLE subphenotypes per cell type were used to calculate family enrichment. Counts per family were generated by dropping the locus from the telescope transcript name and summing the number of DE TEs according to families.tsv. LINEs were grouped into L1FLnI, L1ORF2, and L1FLI. Log odds ratio and the hypergeometric test for enrichment/depletion were calculated, and an expression threshold of four was used for filtering.

Enrichment of open reading frames (ORFs) of viral proteins derived from TE sequences was calculated using the Genome-based Endogenous Viral Element Database (gEVE). Bedtools was used to find the intersection between Hsap38.geve.v1.bed regions and Telescope annotation regions. The DE elements from the SLE subphenotype analysis that overlapped full regions of the gEVE annotation were used to calculate enrichment of ORFs of viral genes encoded in TE sequences.

TE and gene expression integration. Association of differential TE expression with gene expression was calculated using the sum of counts of DE TEs for all SLE subphenotypes per cell type. Raw counts from Telescope were normalized with DESeq2, and the counts of DE TEs for each cell type were summed to get counts of DE TEs per patient for each cell type. DESeq2 (version.1.38.3) was used with the continuous variable of summed differential counts of TEs, adjusted for age, sequencing lane, sex, genetic PCs, and medication at the time of blood draw. Gene set enrichment analysis was run with significant genes (padj < 0.05) using WebGestaltR (version 0.4.5) and the reactome database. Volcano plots were generated using Enhanced Volcano package (version 1.16.0).

RESULTS

The study consists of 120 participants with SLE from the CLUES (Figure 1A). The majority of participants were women, with an age distribution between 20 and 82 years (Table 1). To capture cell type–specific elements across major immune populations, we leveraged cell-sorted RNA-seq data ($CD4^+$ T cells, $CD14^+$ monocytes, B cells, and NK cells) from PBMCs of patients that were previously published by Andreoletti et al.[2] After standard QC measures, we quantified expression of 27,135 HERVs and LINEs (TEs) in four cell types using Telescope.[7] We found that both TE and gene expression is cell specific in patients with lupus as observed by the clear clustering in PCAs of both TE expression (Figure 1B) and gene expression (Figure 1C).

Cell-specific association between TEs and SLE phenotypes. Cell type–specific HERV expression has previously been shown in immune cells.[21] Therefore, we conducted a cell-specific comprehensive genome-wide analysis of TEs in $CD4^+$ T cells, monocytes, B cells, and NK cells, examining differential expression in SLE subphenotypes (history of anti-SM antibody, history of anti-RNP antibody, anti-dsDNA antibody at blood draw, severe disease 2 vs mild disease, severe disease 1 vs mild disease, severe disease 2 vs severe disease 1, disease activity characterized by the SLEDAI score, IFN score, proteinuria, and photosensitivity). Differential expression analysis of SLE subphenotypes identified significant (padj < 0.05) TEs in each cell type and SLE subphenotype (Supplementary Table 1).

We then explored whether these DE TEs were cell or SLE subphenotype specific. We observe very few overlaps of DE TEs across all SLE subphenotypes and cell types (Figure 2). No locus-specific TEs were shared across all cell types, whereas a small number of overlaps were observed across two to three cell types (Figure 2A).

Across subphenotypes, we also observe a small number of common DE TEs. In $CD4^+$ T cells, we identify one TE (ERVLB4_8p23.1o) that was common across five subphenotypes, including photosensitivity, anti-dsDNA antibody at time of

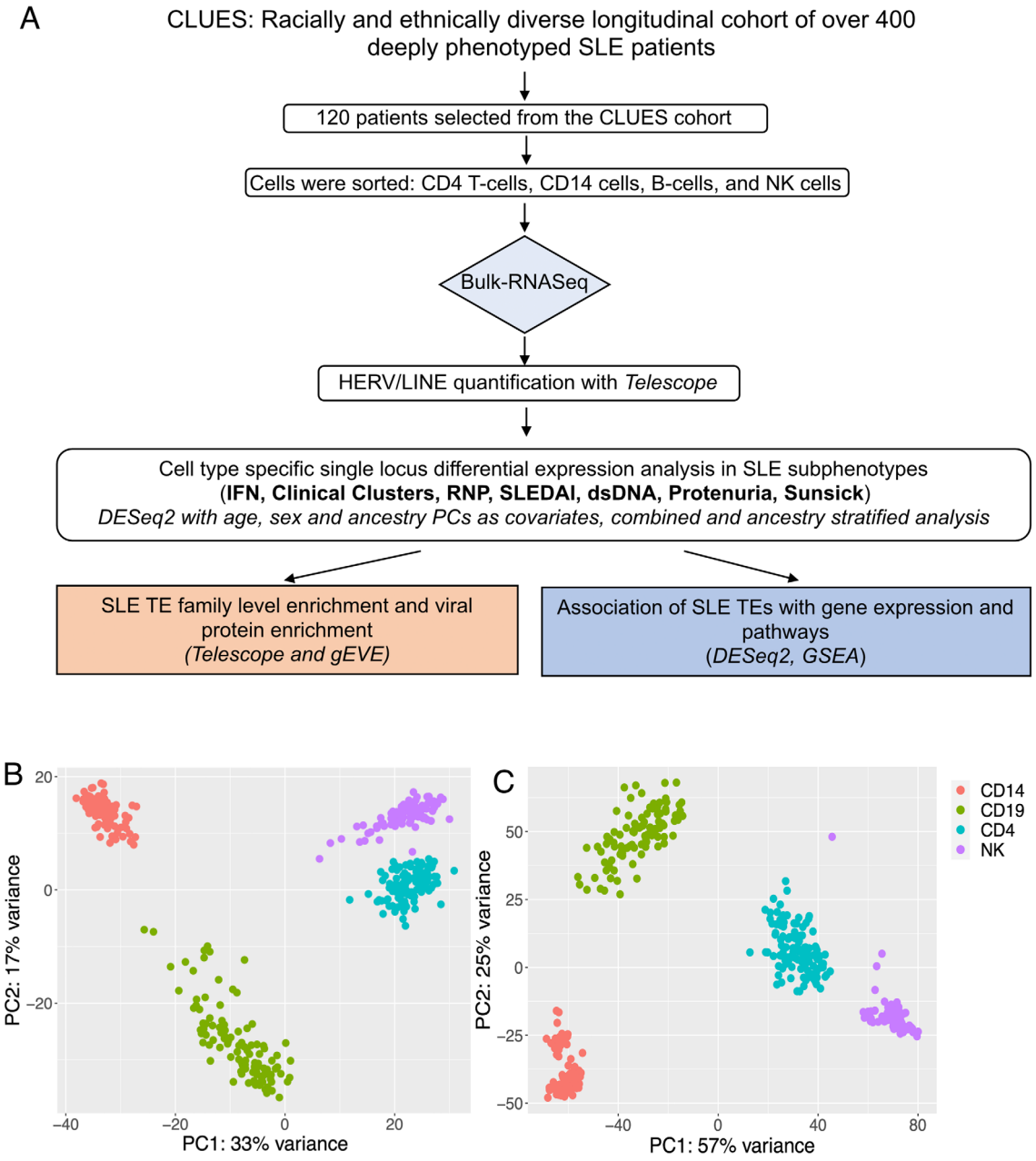

Figure 1. Study overview and comparison of TE and gene expression between cell types. (A) One hundred and twenty patients were selected for cell-sorted bulk RNA-seq analyses from the CLUES cohort. Data were QC'd, and HERVs and LINE were quantified using Telescope. DESeq2 was used for cell-specific differential expression of locus-specific TEs for SLE subphenotypes. Cell-specific differentially expressed TEs from all SLE subphenotypes were used to perform family enrichment, open reading frame of viral genes encoded in TE sequences enrichment, and association with gene expression and pathway analysis. (B) TE expression PCA plot–based visualization colored by cell type. (C) Gene expression PCA plot–based visualization colored by cell type. CLUES, California Lupus Epidemiology Study; dsDNA, double-stranded DNA; gEVE, Genome-based Endogenous Viral Element Database; HERV, human endogenous retrovirus; IFN, interferon; LINE, long interspersed nuclear element; PC, principal component; PCA, PC analysis; QC, quality control; RNA-seq, RNA-sequencing; SLE, systemic lupus erythematosus; SLEDAI, SLE disease activity; TE, transposable element.

blood draw, disease activity characterized by the SLEDAI score, history of anti-RNP antibody, and cluster severe disease 1 versus mild disease). In the monocyte analysis, we found two overlapping TEs (L1FlnI_5q35.1e and L1FlnI_Yp11.2na) common across six SLE subphenotypes (photosensitivity, anti-dsDNA antibody at the time of blood draw, disease activity characterized by the SLEDAI score, history of anti-RNP antibody, clinical cluster severe disease 1 vs mild disease, and IFN signature). In B cells, we identify one TE (HML2_8p23.1a) that is DE across eight subphenotypes (photosensitivity, anti-dsDNA antibody at the time of blood

Table 1. Cohort characteristics*

Characteristics	Overall (n = 119)	Asian enriched (n = 62)	European enriched (n = 57)
Sex, n (%)			
Female	105 (88.2)	56 (90.3)	49 (86.0)
Male	14 (11.8)	6 (9.7)	8 (14.0)
Age, y			
Mean (SD)	45.2 (13.7)	42.3 (13.9)	48.3 (12.9)
Median (Min–Max)	45.0 (20.0–83.0)	41.0 (20.0–74.0)	45.0 (27.0–83.0)
Clusters, n (%)			
Mild disease	40 (33.6)	10 (16.1)	30 (52.6)
Severe disease 1	60 (50.4)	39 (62.9)	21 (36.8)
Severe disease 2	19 (16.0)	13 (21.0)	6 (10.5)
Anti-RNP antibody, n (%)			
Abnormal	39 (32.8)	24 (38.7)	15 (26.3)
Missing	6 (5.0)	1 (1.6)	5 (8.8)
Normal	74 (62.2)	37 (59.7)	37 (64.9)
SLEDAI score, n (%)			
High	13 (10.9)	6 (9.7)	7 (12.3)
Low	106 (89.1)	56 (90.3)	50 (87.7)
Anti-SM Ab, n (%)			
Abnormal	26 (21.8)	17 (27.4)	9 (15.8)
Missing	2 (1.7)	0 (0)	2 (3.5)
Normal	91 (76.5)	45 (72.6)	46 (80.7)
High dsDNA Ab at blood draw, n (%)			
Abnormal	50 (42.0)	31 (50.0)	19 (33.3)
Normal	69 (58.0)	31 (50.0)	38 (66.7)
Proteinuria at blood draw, n (%)			
Abnormal	5 (4.2)	3 (4.8)	2 (3.5)
Normal	114 (95.8)	59 (95.2)	55 (96.5)
Photosensitivity, n (%)			
0	80 (67.2)	40 (64.5)	40 (70.2)
1	39 (32.8)	22 (35.5)	17 (29.8)

*Ab, antibody; dsDNA, double-stranded DNA; Max, maximum; Min, minimum; SLEDAI, systemic lupus erythematosus disease activity; SM, Smith.

draw, disease activity characterized by the SLEDAI score, history of anti-RNP antibody, severe disease 2 vs severe disease 1, severe disease 1 vs mild disease, IFN signature, and anti-SM antibody). In NK cells, we found one TE (LTR25_9q13a) common across four SLE subphenotypes (anti-dsDNA antibody at the time of blood draw, history of anti-RNP antibody, IFN signature, and anti-SM antibody).

Additional cell type–specific associations of TEs with SLE subphenotypes through ancestry-stratified analysis. Comparing patients based on genetic similarity revealed two strong clusters (Supplementary Figure 1), which significantly correlated with self-reported race ($r^2 = 0.938$, $P = 2.2 \times 10^{-16}$). Given the differences in SLE burden between individuals with Asian and European ancestry, we explored differential TE expression within these groups.[22] Similar to our previous analysis, we identified many DE locus-specific TEs by SLE subphenotype and cell type when stratifying by genetic ancestry groups and removing admixed individuals (Supplementary Tables 2 and 3).

We observe a larger number of DE TEs in lupus subphenotypes, especially anti-dsDNA in this stratified analysis when compared with the analysis with all patients (Supplementary Tables 2 and 3). As in the overall analyses, few locus-specific TEs were DE across all cell types or subphenotypes (Supplementary Figures 2 and 3).

In the European-enriched subgroup $CD4^+$ T cells, we found one TE (MER101_21q21.3a) common across four different subphenotypes (anti-dsDNA antibody, photosensitivity, disease activity characterized by the SLEDAI score, and anti-RNP antibody). In monocytes, we identified one TE (HERVL_Xp11.4) common across three SLE subphenotypes (photosensitivity, anti-dsDNA antibody, and IFN). In B cells, we found two TEs (ERVLB4_2q11.1b and HARLEQUIN_Yq11.223) common across five SLE subphenotypes (photosensitivity, anti-dsDNA antibody, disease activity characterized by the SLEDAI score, anti-RNP antibody, and severe disease 1 vs mild disease). In NK cells, we found one TE (L1FlnI_8q13.1d) common across four SLE subphenotypes (photosensitivity, anti-SM antibody, anti-RNP antibody, and anti-dsDNA).

In the Asian ancestry subgroup $CD4^+$ T cells, there were three common DE TEs (ERVL_17q11.1, LTR23_3q26.31, and MER41_17q23.3a) across four SLE subphenotypes (anti-dsDNA antibody, anti-RNP antibody, IFN, and anti-SM antibody). In monocytes, there were two TEs (L1FlnI_6q14.3u and MER101_6p21.1) across six subphenotypes (anti-dsDNA antibody, anti-RNP

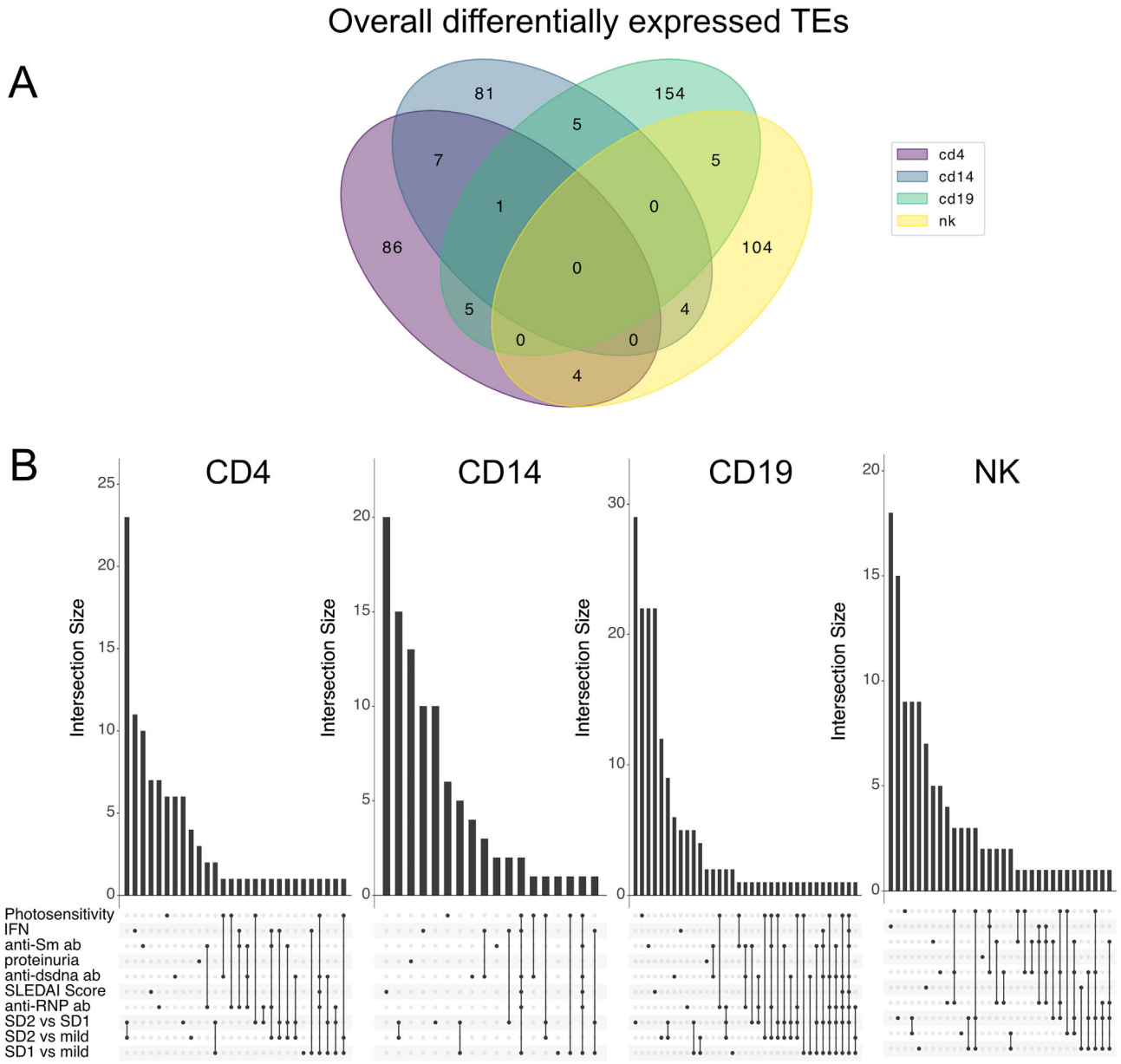

Figure 2. Identification of common differentially expressed TEs in the combined cohort shows distinct locus-specific TEs differentially expressed across cell types and SLE subphenotypes. (A) Venn diagram of overlap of DE TEs between cell types. (B) Overlap of DE TEs between lupus subphenotypes. DE, differentially expressed; dsdna, double-stranded DNA; IFN, interferon; SD, severe disease; SLE, systemic lupus erythematosus; SLEDAI, SLE disease activity; Sm, Smith; TE, transposable element.

antibody, severe disease 2 vs severe disease 1, anti-SM antibody, photosensitivity, and IFN). In B cells, there were four DE TEs (ERV316A3_4q28.3cl, ERVLE_4q32.3a, MER4B_15q21.1c, and MER4B_Xq22.3a) across four SLE subphenotypes (photosensitivity, anti-dsDNA antibody, severe disease 2 vs severe disease 1, and severe disease 1 vs mild disease). Last, in NK cells, there were four common DE TEs (ERVLB4_Xq21.31j, HUERSP2_Xq27.3a, L1FlnI_11p14.3k, and LTR25_16p12.3b) across four SLE subphenotypes (photosensitivity, anti-dsDNA antibody, anti-RNP antibody, and severe disease 2 vs severe disease 1).

Family-level enrichment of TEs and TE-derived viral gene analysis shed light on across diverse SLE subphenotypes within each unique cell type. Most families of TEs have deposited sequences at thousands of loci throughout the genome. Given their common origin, these locus-specific instances of each TE share very similar sequences and potential functional elements. Thus, expression of similar sequences from different genomic loci derived from the same family of TEs might collectively contribute to SLE. In addition, previous studies have associated specific HERV families, like HERV-K, with SLE.[16]

To test for family-level effects, we analyzed DE TEs at the family level across subphenotypes of SLE, focusing on each cell type. We discovered significant enrichment of different families among the DE TEs associated with SLE subphenotypes in each cell type (Figure 3A). In $CD4^+$ T cells, we observed enrichment of HERVH (padj = 1.35×10^{-29}), whereas in monocytes, we detected enrichment for MER61 (padj = 0.01). In B cells, we found HML2 (padj = 1.47×10^{-4}) and HERVH (padj = 1.07×10^{-30}) to be enriched, and in NK cells, ERVLB4 (padj = 1.53×10^{-6}) was enriched. We also discovered consistent depletion for L1FlnI sequences across all four cell types. In the analysis stratified by genetic similarity, notable distinctions emerged (Figure 3B and 3C).

Previous work has hypothesized that HERVs with preserved ORFs could produce proteins that could activate or depress the inflammatory cascade.[23] Therefore, we used gEVE to find ORFs of viral proteins derived from TE sequences in the locus-specific DE TEs we observed across SLE subphenotypes per cell type (Table 1). We focused our analysis on ORFs collapsed by viral genes and found enrichment of ORFs of viral proteins derived from TEs in the cell-specific DE TEs across all SLE subphenotypes (Supplementary Table 4).

IFN pathway associated with expression of DE TEs. To better understand the possible effects of differential TE expression in SLE, we asked whether the expression of host genes correlated with the expression of TEs associated with SLE subphenotypes. In the combined analysis cohort, the expression levels of many genes associated with the expression of DE TEs (Figure 4 and Supplementary Table 5). The associations of differentially up- and down-regulated genes were largely cell type specific (Figure 4E and 4F). In the $CD4^+$ T cells, we found a strong up-regulation of genes associated with the DE TEs. Some of the strongest up-regulated genes in $CD4^+$ T cells were involved in type I IFN signaling and other antiviral innate immune pathways,

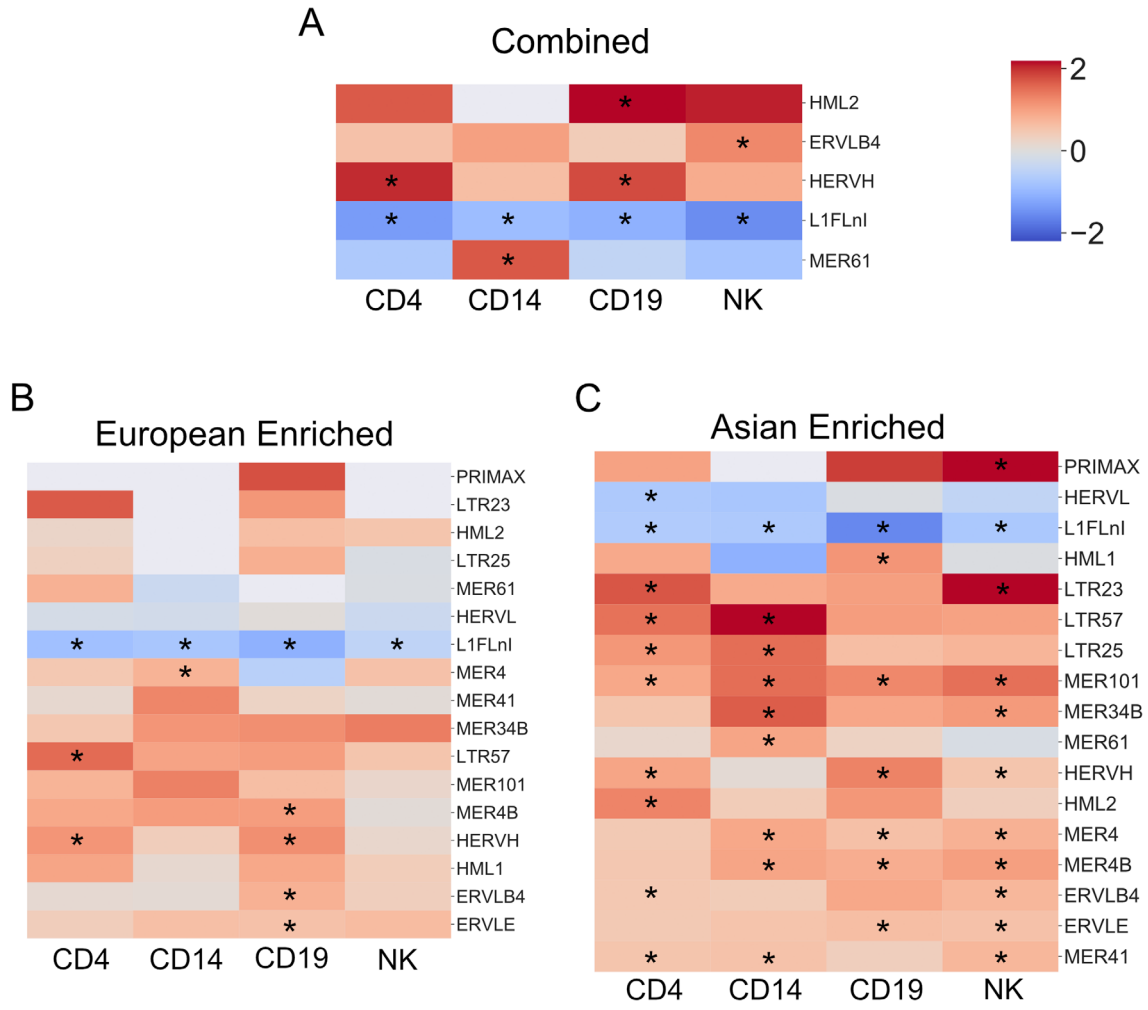

Figure 3. Log odds ratio of significant enrichment/depletion of HERV families in combined and stratified analysis. (A) Family enrichment for the combined analysis. (B) Family enrichment for the European-enriched cohort. (C) Family enrichment for the Asian-enriched cohort. Family annotation for HERV families used from https://github.com/mlbendall/telescope_annotation_db (significant families denoted by an asterisk). HERV, human endogenous retrovirus.

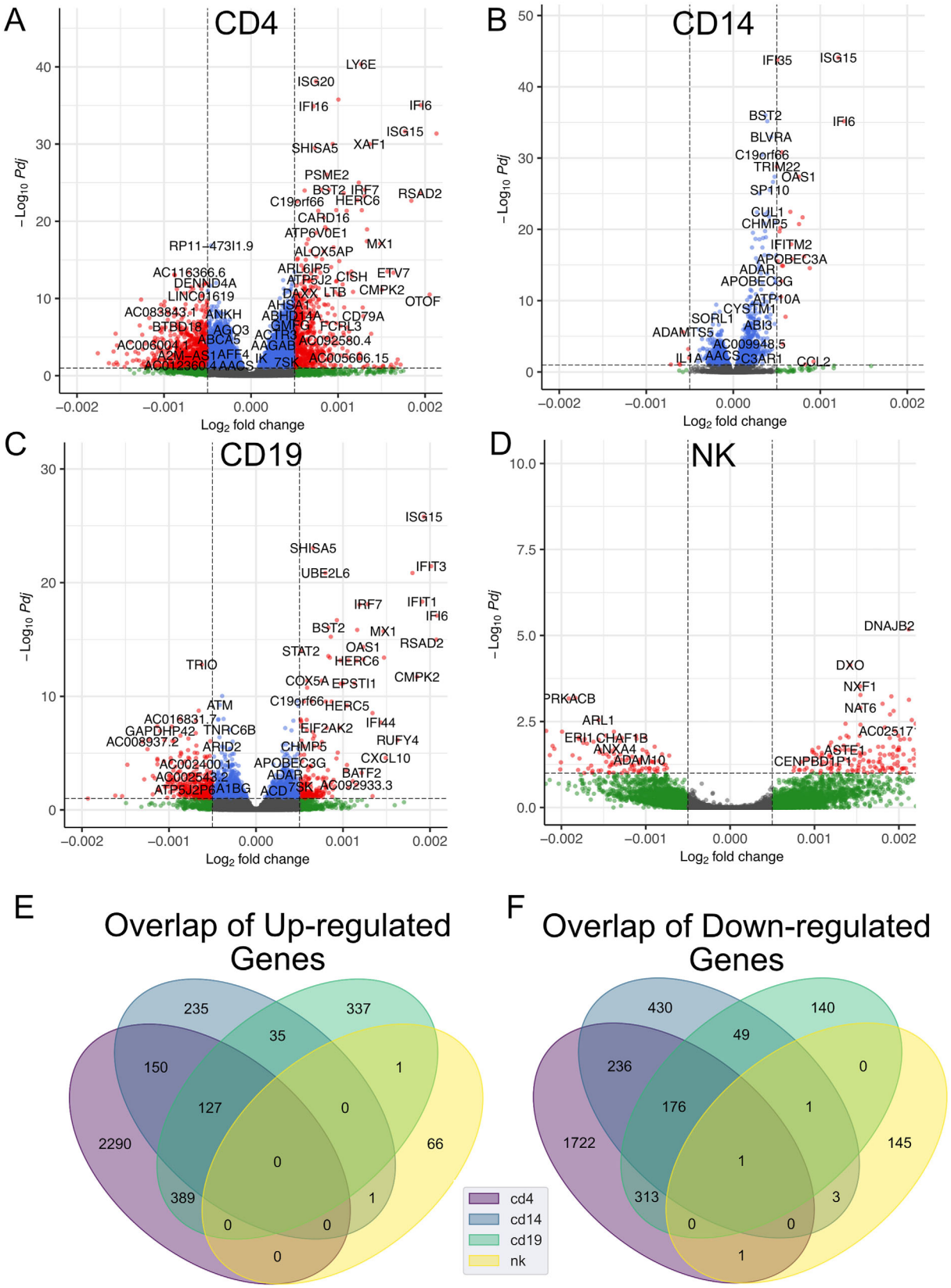

Figure 4. Cell-specific volcano plots of combined analysis differentially expressed gene associations with differential TE expression using DESeq2. (A) CD4, (B) CD14, (C) CD19, and (D) NK cells. Identification of the common DE genes (padj < 0.05) across the different cell types. (E) Up-regulated. (F) Down-regulated. DE, differentially expressed; padj, adjusted *P*-value; TE, transposable element.

such as *LY6E*, *IFI6*, *ISG15*, and *ISG20*. In the monocytes, most of the genes also appear up-regulated (Figure 4B), and the top up-regulated genes are also involved in antiviral activity, such as *ISG15*, *IFI6*, *IFI35*, *BST2*, and *TRIM22*. In CD19 cells, we also found many up-regulated genes (Figure 4C), and some of the top up-regulated genes such as *ISG15*, *IFIT3*, *IRF7*, and *BST2* play roles in response to viruses. In NK cells, we found many up- and down-regulated genes (Figure 4D); one of the top up-regulated genes we found, *NXF1*, is involved in transport of unspliced retroviral genomic RNA,[24] whereas one of the top down-regulated genes we found, *CAPZA1*, is also observed to be down-regulated in virus-infected cells.[25]

Gene set enrichment analysis revealed enrichment of pathways expressed in all cell types, except NK cells, including IFN alpha/beta signaling, IFN signaling, cytokine signaling in the immune system, and antiviral mechanism by IFN-stimulated genes (Figure 5). We also found some cell type–specific signals. For example, enriched pathways include solute carrier–mediated transmembrane transport in monocytes cells; influenza infection, viral mRNA translation, antigen presentation folding assembly, and peptide loading of class I major histocompatibility complex in CD4 cells; and G protein–coupled receptor ligand binding, host interactions with influenza factors, and NS1-mediated effects on host pathways in B cells (Supplementary Tables 6–8).

DISCUSSION

This is the first study to characterize locus and cell-specific TE expression among a deeply phenotyped SLE cohort. We found that TE differential expression is highly cell specific, with very few overlaps between immune cells. We discovered associations with several disease subphenotypes and host gene expression. This suggests that TE expression could contribute to the heterogeneity of SLE across individuals and populations.

We found associations of TE expression with production of autoantibodies against ribonucleoproteins (RNP and SM), SLE disease activity, and SLE disease severity after adjusting for potential confounders. Some of these results have been previously reported; for example, some TEs, such as HERV-E clone 4-1, have correlated with autoantibody levels (anti-U1 RNP and anti-SM nuclear antibodies) and higher IFN status in SLE.[26,13] These associations are highly cell specific and subphenotype specific, which suggests that their differential expression could contribute or be a result of different lupus subphenotypes.

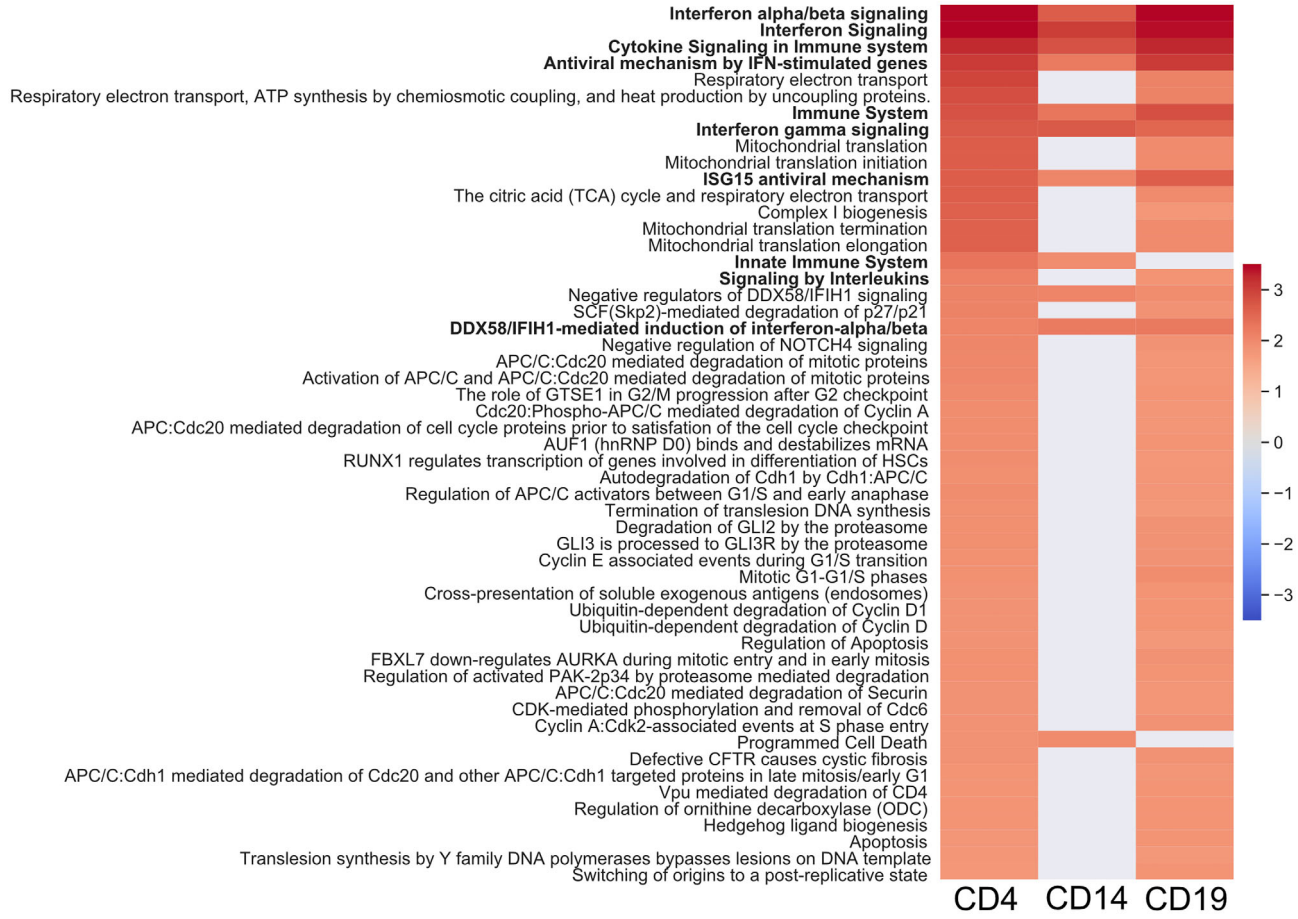

Figure 5. Significant pathways in more than one cell type from gene set enrichment analyses with webgestalt using significant genes in the combined cohort. Heatmap shows NES. NES, normalized enrichment score.

Given the differences in SLE severity and outcomes between patients who self-report as Asian or White and that TE are encoded through the germline, we explored TE differential expression patterns in individuals stratified by genetic similarity. The genetic similarity analysis showed a near complete agreement between genetic similarity–based clustering and self-reported race, as well as a small number of individuals with significant genetic admixture. Admix individuals were removed from downstream stratified analyses (Supplementary Figure 1). When the two groups, Asian enriched and European enriched, were stratified, we found more DE TEs in almost every SLE subphenotype, despite adjusting for genetic PCs 1–10. This was surprising given that the sample size for these groups was reduced from the combined analysis. Further work is needed to determine if these differences in the stratified analysis between TE expression and phenotypes are because of genetic factors or differences in environmental exposures.

Retroviral p30 gag proteins and serum reactivities to p30 gag antigen have been found in patients with SLE with proliferative glomerulonephritis.[27] Additionally, endogenous retrovirus–derived proteins such as HERV-W envelope protein and HERV-K dUTPase have been found to release antivirally active cytokines.[28] We investigated if it was possible that the DE TEs identified exert their effect as proteins because most expressed TEs do not contain full ORFs for functional viral proteins. We tested if there was significant enrichment of ORFs among the identified DE TEs and found enrichment of several TE-derived viral proteins in a cell-specific manner, such as protease, envelope (env), and reverse transcriptase. The role of TE-derived viral proteins in SLE remains to be elucidated; however, previous work has found env proteins both to trigger innate and adaptive immunity as well as possessing immunosuppressive properties.[29] Therefore, it is difficult to say whether the enrichment of ORFs of viral proteins we observe in the DE TEs is the cause of an immune response or the response acting to down-regulate an immune response.

One of the most intriguing findings was the connection between host response gene expression and the expression of TEs in relation to the SLE phenotypes. It has been hypothesized that TEs could activate the innate immune system and lead to chronic inflammation, contributing to autoimmunity.[11] Supporting this hypothesis, a study of kidney biopsies from patients with SLE with lupus nephritis found LINEs can trigger the IFN-I pathway.[30] Interestingly, in both colorectal and ovarian carcinoma cells, treatment with 5-aza-2′-deoxycytidine resulted in the accumulation of HERVs and activation of an IFN response.[31,32] An IFN response to HERVs in different cells and diseases underscores the importance of understanding how the body responds to TE expression. We found that DE TEs are associated with genes and pathways that are involved in an antiviral response, which lends support to this hypothesis. There are also cell-specific pathways of note; for instance, in $CD4^+$ T cells, we observe a correlation between DE TEs and many pathways indicating response to viral infections, which have long been associated with SLE, such as influenza infection and life cycle, viral mRNA translation, influenza viral RNA transcription, and replication. In B cells, we identify associations with inflammatory pathways, including NF-κB, which have been associated with both lupus and viral infection.[33,34] HERVs have dispersed many IFN-inducible enhances in mammalian genomes, shaping a transcriptional network underlying the IFN response, and additional studies are needed to examine the causality between TE expression and immune dysregulation to understand whether TE expression contributes to the development of SLE.[35]

Nonetheless, there exist several limitations within this study. The absence of healthy control data hampers a cell-specific comparison to a baseline TE expression profile. Future studies should include healthy individuals as well as individuals with other autoimmune diseases to better characterize the specificity of the findings. Most of the participants in this cohort exhibit low disease activity, thereby limiting our power to detect DE TEs across different levels of disease activity. Quantifying TE expression is challenging, given their repetitive nature; many tools and approaches have been used to measure their expression, each with strengths and weaknesses.[36] We selected Telescope because it quantifies locus-specific TE expression, is highly sensitive, has low Type I and Type II error rates, and has superior accuracy in comparison with other software packages for TE expression analysis.[7] We consider all TE sequences in the Telescope annotation, which include more than just full-length TE transcripts; this is justified because a TE does not have to be fully intact to influence cell physiology.[36] Finally, because this study design is cross-sectional, it is difficult to ascertain whether TE expression is a consequence rather than causal in the mechanism of SLE manifestations, and further studies are needed to understand these results.

This study aimed to investigate the impact of TE expression on SLE heterogeneity in a cell-specific manner. We discovered a significant number of locus-specific TEs and TE families whose expression is associated with specific SLE manifestations and host gene expression. In summary, our study reveals cell-specific TE expression patterns linked to disease activity, autoantibody production, and distinct disease manifestations such as lupus nephritis. TE expression is associated with expression of host genes that are relevant to SLE pathogenesis. To gain deeper insights into the causality of TEs in SLE and across its different phenotypes, future experimental work is needed in human immune cells. This future work could provide mechanistic and functional insights into the role of TEs in SLE as well as confirm the TEs identified in this study.

ACKNOWLEDGMENTS

We are grateful to all patients who participated in this study. We want to thank members of the Sirota laboratory, especially Jackie Roger, Yaqiao Li, Umair Khan, and Silvia Miramontes for technical discussions and feedback. We are also grateful to Jane Hughan, who provided feedback on the manuscript.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Drs. Cutts and Sirota had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. ZC, MS, CML, JAC, GKF.
Acquisition of data. CJY, LM, SP, MDE, LAC, JY, CML.
Analysis and interpretation of data. ZC, GKF, KET, CML, JAC, MS, CL.

REFERENCES

1. Crow MK. Pathogenesis of systemic lupus erythematosus: risks, mechanisms and therapeutic targets. Ann Rheum Dis 2023;82(8): 999–1014.
2. Andreoletti G, Lanata CM, Trupin L, et al. Transcriptomic analysis of immune cells in a multi-ethnic cohort of systemic lupus erythematosus patients identifies ethnicity- and disease-specific expression signatures. Commun Biol 2021;4(1):488.
3. Maningding E, Dall'Era M, Trupin L, et al. Racial and ethnic differences in the prevalence and time to onset of manifestations of systemic lupus erythematosus: the California Lupus Surveillance Project. Arthritis Care Res (Hoboken) 2020;72(5):622–629.
4. Nakano M, Iwasaki Y, Fujio K. Transcriptomic studies of systemic lupus erythematosus. Inflamm Regen 2021;41(1):11.
5. Kojima S, Koyama S, Ka M, et al; Biobank Japan Project Consortium. Mobile element variation contributes to population-specific genome diversification, gene regulation and disease risk. Nat Genet 2023; 55(6):939–951.
6. Shao W, Wang T. Transcript assembly improves expression quantification of transposable elements in single-cell RNA-seq data. Genome Res 2021;31(1):88–100.
7. Bendall ML, de Mulder M, Iñiguez LP, et al. Telescope: characterization of the retrotranscriptome by accurate estimation of transposable element expression. PLOS Comput Biol 2019;15:e1006453.
8. Tokuyama M, Kong Y, Song E, et al. ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. Proc Natl Acad Sci USA 2018;115(50):12565–12572.
9. Ukadike KC, Mustelin T. Implications of endogenous retroelements in the etiopathogenesis of systemic lupus erythematosus. J Clin Med 2021;10(4):856.
10. Quaglia M, Merlotti G, De Andrea M, et al. Viral infections and systemic lupus erythematosus: new players in an old story. Viruses 2021; 13(2):277.
11. Hurst TP, Magiorkinis G. Activation of the innate immune response by endogenous retroviruses. J Gen Virol 2015;96(Pt 6):1207–1218.
12. Jönsson ME, Garza R, Johansson PA, et al. Transposable elements: a common feature of neurodevelopmental and neurodegenerative disorders. Trends Genet 2020;36(8):610–623.
13. Tokuyama M, Gunn BM, Venkataraman A, et al. Antibodies against human endogenous retrovirus K102 envelope activate neutrophils in systemic lupus erythematosus. J Exp Med 2021;218(7):e20191766.
14. Fali T, Le Dantec C, Thabet Y, et al. DNA methylation modulates HRES1/p28 expression in B cells from patients with lupus. Autoimmunity 2014;47(4):265–271.
15. Ogasawara H, Hishikawa T, Sekigawa I, et al. Sequence analysis of human endogenous retrovirus clone 4-1 in systemic lupus erythematosus. Autoimmunity 2000;33(1):15–21.
16. Stearrett N, Dawson T, Rahnavard A, et al. Expression of human endogenous retroviruses in systemic lupus erythematosus: multiomic integration with gene expression. Front Immunol 2021;12: 661437.
17. Lanata CM, Paranjpe I, Nititham J, et al. A phenotypic and genomics approach in a multi-ethnic cohort to subtype systemic lupus erythematosus. Nat Commun 2019;10(1):3902.
18. Buyon JP, Petri MA, Kim MY, et al. The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic lupus erythematosus: a randomized trial. Ann Intern Med 2005;142(12 Pt 1):953–962.
19. Kennedy WP, Maciuca R, Wolslegel K, et al. Association of the interferon signature metric with serological disease manifestations but not global activity scores in multiple cohorts of patients with SLE. Lupus Sci Med 2015;2(1):e000080.
20. Min X, Zheng M, Yu Y, et al. Ultraviolet light induces HERV expression to activate RIG-I signalling pathway in keratinocytes. Exp Dermatol 2022;31(8):1165–1176.
21. Hamann MV, Adiba M, Lange UC. Confounding factors in profiling of locus-specific human endogenous retrovirus (HERV) transcript signatures in primary T cells using multi-study-derived datasets. BMC Med Genomics 2023;16(1):68.
22. Lee MH, Koh JWH, Ng CH, et al. A meta-analysis of clinical manifestations in Asian systemic lupus erythematous: the effects of ancestry, ethnicity and gender. Semin Arthritis Rheum 2022;52:151932.
23. Talotta R, Atzeni F, Laska MJ. The contribution of HERV-E clone 4-1 and other HERV-E members to the pathogenesis of rheumatic autoimmune diseases. APMIS 2020;128(5):367–377.
24. Aibara S, Katahira J, Valkov E, et al. The principal mRNA nuclear export factor NXF1:NXT1 forms a symmetric binding platform that facilitates export of retroviral CTE-RNA. Nucleic Acids Res 2015; 43(3):1883–1893.
25. Wang Y, Zhang H, Ma D, et al. Hsp70 is a potential therapeutic target for echovirus 9 infection. Front Mol Biosci 2020;7:146.
26. Piotrowski PC, Duriagin S, Jagodzinski PP. Expression of human endogenous retrovirus clone 4-1 may correlate with blood plasma concentration of anti-U1 RNP and anti-Sm nuclear antibodies. Clin Rheumatol 2005;24(6):620–624.
27. Mellors RC, Mellors JW. Type C RNA virus-specific antibody in human systemic lupus erythematosus demonstrated by enzymoimmunoassay. Proc Natl Acad Sci USA 1978;75(5):2463–2467.
28. Badarinarayan SS, Sauter D. Switching sides: how endogenous retroviruses protect us from viral infections. J Virol 2021;95:e02299–20.
29. Grandi N, Tramontano E. HERV envelope proteins: physiological role and pathogenic potential in cancer and autoimmunity. Front Microbiol 2018;9:462.
30. Mavragani CP, Sagalovskiy I, Guo Q, et al. Expression of long interspersed nuclear element 1 retroelements and induction of type i interferon in patients with systemic autoimmune disease. Arthritis Rheumatol 2016;68(11):2686–2696.
31. Roulois D, Loo Yau H, Singhania R, et al. DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. Cell 2015;162(5):961–973.
32. Chiappinelli KB, Strissel PL, Desrichard A, et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. Cell 2015;162(5):974–986.
33. Mishra RK. Involvement of NF-κB signaling pathway in the pathogenesis of systemic lupus erythematosus. Nephrol - Open J 2016;2:9–13.
34. Santoro MG, Rossi A, Amici C. NF-kappaB and virus infection: who controls whom. EMBO J 2003;22(11):2552–2560.
35. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science 2016;351(6277):1083–1087.
36. Lanciano S, Cristofari G. Measuring and interpreting transposable element expression. Nat Rev Genet 2020;21(12):721–736.