The interferon response signature links to increased expression of transposable elements in neutrophils of systemic lupus erythematosus patients

Leo J. Arteaga-Vazquez[1,2,3,*], Hugo Sepulveda [1,2,3,4,5*], Bruno Villalobos[1,2,3], Kazumasa Suzuki[1,2,3], Kenneth Kalunian[8], Ferhat Ay[1,3,6,7], Mark R. Boothby[9,10,#], Anjana Rao[1,2,3,4#]

[1] La Jolla Institute for Allergy and Immunology

[2] Department of Pharmacology, University of California, San Diego, CA, USA

[3] Moores Cancer Center, University of California, San Diego, CA, USA

[4] Sanford Consortium for Regenerative Medicine, La Jolla, CA, USA

[5] Laboratory of Transcription and Epigenetics, Institute of Biomedical Sciences, Faculty of Medicine & Faculty of Life Sciences, Universidad Andres Bello, Santiago, Chile

[6] Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA, 92093, USA

[7] Department of Pediatrics, University of California San Diego, La Jolla, CA, 92093, USA

[8] School of Medicine, University of California, San Diego, La Jolla, CA, USA

[9] Pathology-Microbiology-Immunology Dept and Rheumatology Division, Dept of Medicine, Vanderbilt University Medical Center

[10] Vanderbilt Center for Immunobiology (VCI) and Vanderbilt Institute for Infection, Inflammation, Immunology (VI4), Vanderbilt University, Nashville TN

* Equal contributions (co-first authors)

[#] Co-senior authors

Correspondence:

Leo J. Arteaga-Vazquez, larteaga@lji.org;

Hugo Sepulveda, hugo.sepulveda@unab.cl;

Bruno Villalobos, bvillalobos@lji.org;

Kazumasa Suzuki, ksuzuki@lji.org;

Kenneth Kalunian, kkalunian@health.ucsd.edu;

Ferhat Ay, ferhatay@lji.org;

Mark Boothby, mark.boothby@vumc.org (lead contact)

Anjana Rao, arao@lji.org (lead contact)

ABSTRACT

Systemic lupus erythematosus (SLE) is a spontaneous systemic auto-immune condition for which the inciting factors and genetic basis are generally unknown. Although heterogeneous in its manifestations and severity, SLE involves chronic inflammation along with sustained autoantibody production. The root causes and pathophysiology of the inflammation and breaches of tolerance are incompletely understood, but neutrophils are thought to be important elements of the pathophysiology. Type I interferons (IFN) in the bloodstream and an IFN-stimulated gene (ISG) signature in circulating leukocytes, including neutrophils, are common features in many patients. Earlier work with unfractionated peripheral blood cells of SLE patients has provided evidence of increased levels of transcripts derived from transposable elements (TEs). We sought to elucidate the relative extents of de-repression of different TE classes in neutrophils of SLE patients, and to explore the relationships between the increased ISG and TE expression as well as their correlation with the severity of the SLE disease. Published RNA-seq data obtained from neutrophils (also termed polymorphonuclear leukocytes, or PMN) were analyzed with pipelines designed to detect TEs in samples from 23 patients with SLE and healthy controls (n = 10). SLE samples were sub-divided into "IFN-positive" (IFNpos, n=12) and "IFN-negative" (IFNneg, n=11) patients, as defined in the original study based on gene expression analysis. An independent interferon-regulated gene signature was also determined for each sample. In PMN of IFNpos SLE patients, expression of ISGs and TE families were increased compared to healthy controls (HC) and IFNneg SLE patients. No TE families showed significantly increased expression in SLE patients pre-stratified as IFNneg. Increased expression of a majority of individually-mapped TEs correlated with upregulated expression of nearby genes. The majority (~80%) of upregulated TEs were present in introns of upregulated genes, and ~67% of these upregulated genes were ISGs. Of the ~20% of mapped TEs that were in the introns of genes whose expression did not change significantly, only ~28% were ISGs. Analyses of all SLE patients identified a strong correlation between expression of several specific TE families and disease activity assessed at the time of the visit. Detailed inspection of individually mapped TEs in the introns of ISGs showed a remarkable correlation of high TE expression with selective but partial intron retention as well as with annotated splicing alterations of the flanking exons in these ISGs. Neutrophils of SLE patients independently pre-classified as being IFNpos have increased expression of multiple TE classes and families, but this manifestation is not present in patients of the IFNneg stratum. The expression of only some TE families correlated with the disease activity, showing a complex relationship between TEs and SLE. The findings suggest a mechanistic relationship between the degree of IFN response and increased TE expression in a cell type centrally involved in SLE pathogenesis. The presence of highly expressed TEs in ISG introns, and the correlation of TE expression levels with partial intron retention and with splicing variations in adjacent exons, are unexpected findings that need to be further explored.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a prototypical example of a chronic autoimmune disorder, a class of systemic conditions that involve persistent auto-immune dysfunction with inflammation [1-3]. The underlying causes, pathophysiological mechanisms, and optimal treatments for most of these chronic illnesses remain to be fully elucidated. The demography, clinical features, and apparent causes for SLE vary substantially [2-4]. Nonetheless, breaches of the tolerance mechanisms that ordinarily prevent chronic and sustained production of antibodies directed against "self" or "modified self" determinants [3-6] are a central feature of SLE. These auto-antibodies can drive inflammation in various tissues.

Polymorphonuclear granulocytes (PMN), also known as neutrophils, are a prominent pathogenic cell type in SLE. PMNs undergo cell death concomitantly with a process known as NETosis [7-9] – externalisation of a chromatin meshwork that releases double-stranded DNA (dsDNA) and intranuclear antigens that elicit auto-antibodies that are a characteristic feature of SLE. There is also a strong connection between SLE and interferon (IFN)-stimulated genes (ISG) expression. Many SLE patients display a gene expression signature enriched for type I interferons (IFN-I, IFNα and IFNβ) and interferon-stimulated genes (ISGs) (reviewed in [10-14], and sera of subjects who later developed SLE had subtly higher levels of circulating IFN-I and IFN-γ as well as auto-antibodies even before the onset of clinical disease [15-17]. A variety of cell types can be sources of IFN-I, but PMNs have been suggested as

a major contributor [18]. The mechanisms responsible for the initiation of increased IFN production in most SLE patients are not yet known, and indeed the root causes may differ among SLE patients. In some instances, however, altered regulation of chromosome silencing or inadequate control of immunostimulatory RNA have been proposed as important in the pathophysiology of SLE [19-23].

In eukaryotic cells in general, and metazoans in particular, health and longevity depend on a capacity to preserve the integrity of the nuclear and mitochondrial genomes along with the epigenetic networks that stabilize cellular identities while allowing changes in gene expression [24, 25]. One challenge to genome stability is posed by the invasion of genomes by transposable elements (TEs), over millions of years of evolutionary time, transposable elements (TEs) [26-32]. After the initial invasion, TEs can expand their numbers significantly, creating large families and subfamilies of repeated TEs that include LTR-containing endogenous retroviral sequences (ERVs) as well as short and long interspersed nuclear elements (SINEs and LINEs, respectively) [26-32]. Over evolutionary time, these processes result in the generation of TEs that are nonfunctional, in part due to severe mutations in older elements, and in part due to the emergence, through evolution, of KRAB zinc finger protein (KRAB-ZFP) proteins that bind to the regulatory regions of some TEs and suppress their expression via their interaction with KAP1 [26, 31]. TEs and mutated TE-derived sequences are estimated to constitute ~40-50% of the haploid genome in humans [27, 33].

Repetitive sequences, including TE families, are sensed as threats by the non-repetitive genome due to their potential for destabilizing genomic sequences by transposition and recombination [27, 33]. In consequence, TE expression – required for transposition – is suppressed by a variety of mechanisms, among them mutations, DNA methylation, heterochromatin-associated histone modifications such as H3K9me3, and KAP1-KRAB ZFP and HUSH protein complexes [26, 27, 31, 34, 35]. Errors in these mechanisms would lead to a slow loss of heterochromatin integrity – termed heterochromatin “weakening” or “dysfunction" – which could potentially contribute to disparate diseases such as aging and cancer [35-37]. A reliable marker for heterochromatin dysfunction is increased expression of transposable elements (primarily LINEs and LTR elements) that are normally silenced in heterochromatin [35-39].

When heterochromatin integrity is compromised by decreased DNA or H3K9 methylation, TE regulatory regions (LTRs of endogenous retroviruses and the 5’ untranslated regions (5'-UTR) of LINEs) can participate in the regulation of neighboring gene transcription by acting as distal enhancers or alternative promoters, or they can modify gene structures by becoming alternative exons [26, 28-30, 38-41]. Further, the repetitive nature of TE sequences creates a potential for recombination events, genomic stress and increased mutations, and TE expression correlates with the inflammation characteristic of cancer, aging and autoimmune disease. Specifically, the micronuclei produced by chromosome mis-segregation can activate cGAS/STING pathways [42-44], and double-stranded RNA encoded by inverted-repeat Alu elements can activate innate sensor mechanisms, especially if ADAR-mediated RNA editing is compromised, culminating in MDA5 activation akin to an anti-viral response [45, 46]. The consequent release of type I interferons (IFN-I) and the increased transcription of interferon-stimulated genes (ISGs) can evoke systemic inflammation and autoimmune disease. In support of such a mechanism, GWAS studies of several autoimmune diseases have linked ADAR loss-of-function and MDA5 gain-of-function alleles to diverse autoimmune/ inflammatory conditions including SLE [47-49].

These findings raise questions about the molecular and cellular mechanisms that initiate or contribute to perpetuating the dysregulated production of type I interferons in SLE. A variety of cells can serve as sources of type I interferons when exposed to stimuli such as apoptotic cells, circulating nucleoprotein complexes derived in part from PMN, and oxidized mitochondrial DNA among others. However, the de-repression of TE transcripts offers another potential bridge connecting innate sensor mechanisms to increases in IFN-I [50-56]. With this background in mind, we set out to measure TE-derived RNA in PMN of SLE patients compared to healthy control (HC) individuals, to determine if TE de-repression is a general feature in SLE PMN. Moreover, we stratified SLE patients by whether they displayed or lacked the transcriptional signature of type I interferon responses, based on their expression of interferon-stimulated genes (ISGs); we then investigated the relationships between (i) the IFN signature and TE expression, (ii) TE expression and clinical disease in individual SLE PMN samples stratified by whether they

belonged to the IFN-positive (IFNpos) or IFN-negative (IFNneg) subset. We found that such stratification increased the sensitivity of detection of differentially expressed TE families, relative to analyses in which all SLE patients were merged without regard to IFN signatures. Stratification into IFNpos and IFNneg subsets also revealed a strong correlation of the expression levels of certain TE families with disease activity; specifically, patients with severe versus moderate or mild disease could be distinguished at least as well by quantification of the expression levels of these differentially expressed TEs as by their IFN signature. Finally, a detailed inspection of the gene locus of upregulated TEs that were present in the introns of upregulated ISGs suggested interesting connections among TE expression, increased transcription of the corresponding intron (intron retention) and annotated splicing variations in adjacent exons.

RESULTS

TEs are over-expressed in PMN of IFNpos SLE patients relative to healthy controls yet absent from the IFNneg subset

In a majority of SLE patients, the peripheral blood leukocyte pool exhibits increased levels of ISGs [10-14]. A prior analysis of RNA-seq data from SLE patients and healthy control donors identified two subgroups of SLE patients, with high ("IFNpos") and low ("IFNneg") ISGs expression, respectively [57]. In this study, polymorphonuclear leukocytes (PMNs, also known as neutrophils) in peripheral blood showed a strong upregulation of TNFSF13B (BAFF), a cytokine that supports B cell survival and can cause loss of self-tolerance when available to B cells in excess [58, 59]. Moreover, neutrophils are important mediators of SLE pathogenesis [7-9, 18] and have been reported to express LINE-1 RNAs and their encoded protein products [51]. Accordingly, we elected to perform in-depth re-analyses of the RNA-seq data from purified PMN reported in this paper [57], so as to explore more closely the TE dysregulation in SLE patients and any association with the severity of the disease.

We compared RNA-seq data from 10 healthy controls (HC) to 23 SLE patients classified by their ISG levels as IFNneg (n=11) and IFNpos (n=12) (**Fig 1A**, *top*) [57]. Notably, the overall numbers of both differentially expressed ISGs and TEs (DE TEs) were highest in PMN among the leukocyte types originally analyzed (**Fig 1B**). To weight each subject equally, we restricted our analysis to only the initial visits while also focusing on this cellular subset (**Suppl. Fig. 1A**). In a Principal Components Analysis (PCA) of the RNA-seq datasets, the 12 IFNpos SLE samples were clearly distinguished from the 10 healthy control samples and from the majority (8/11) of IFNneg samples (**Fig 1A**, *bottom*). The PCA analysis also showed that, despite clear diagnoses of SLE in all IFNneg patients, the main components of overall gene expression in 8 of 11 IFN-neg samples overlapped (commingled with) those of the healthy controls. This finding is consistent with previous studies suggesting distinct subtypes among SLE patients, and point to the existence of a subset of patients in whom the IFN signature may be muted or absent [60]. Using an extended list of IFN stimulated genes (ISGs) (**Suppl. Table 1**) ([38, 39], modified from [46]), pairwise analyses of differential ISG expression between sample sets showed the expected difference between the pool of all 23 SLE patients and healthy controls (**Fig 1C**). The data also revealed that the differences were of greater magnitude when IFNpos were compared with healthy controls (**Fig 1D, E**). In line with the PCA and the gene network stratification, ISG expression in IFNpos samples was of substantially greater magnitude than in IFNneg patients (**Fig 1F**).

The literature suggests many connections between TE expression and inflammation in cancers, autoimmune diseases and ageing (reviewed in [35, 37]), but the connections between TE and ISG expression are not yet understood. Applying the TEtranscripts pipeline (*see Methods*), which quantifies and classifies TE-related transcripts, the expression of several TE families and subfamilies was significantly increased in SLE PMN (**Fig 1G**), with LTR (endogenous retrovirus, ERV) sequences dominating the observed differentially expressed (DE) TEs. No TEs were differentially expressed when comparing patient samples stratified as IFNneg against healthy controls (**Fig 1H**). In sharp contrast, the expression of multiple TE families and subfamilies was strongly increased in IFNpos SLE patients compared to healthy controls (**Fig 1I**) or to IFNneg SLE patients (**Fig 1J**). The stratification to IFNpos increased the sensitivity of detecting increases in TE-encoded RNA five-fold when compared to the total population of all SLE patients (i.e., 55 TEs upregulated in the IFNpos to HC comparison versus 11 TEs upregulated in the total

SLE to HC comparison, **Suppl. Fig 1B**, comparing panels 3 and 1); with all TE classes (LTR, Satellite, LINE, and DNA transposons) detected (**Fig 1I; Suppl. Fig. 1B**). Hence many classes and families of TEs are derepressed in the PMN of SLE patients, but at least in this moderate-sized and diverse patient pool, the increases are confined exclusively to the SLE patients classified as IFNpos.

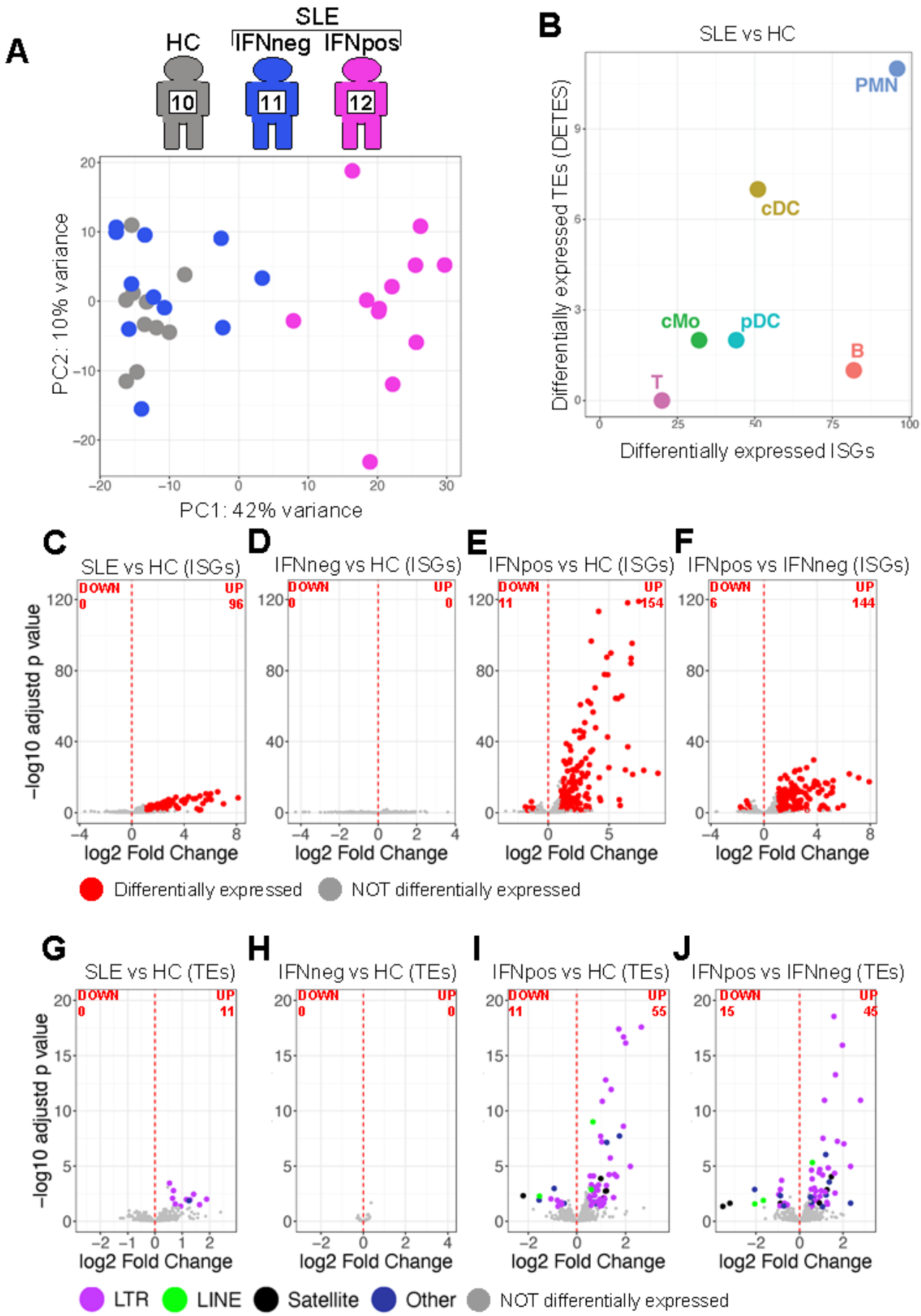

A
HC
SLE
IFNneg
IFNpos
10
11
12
PC2: 10% variance
PC1: 42% variance
B
SLE vs HC
PMN
cDC
cMo
pDC
B
T
Differentially expressed TEs (DETES)
Differentially expressed ISGs
C
SLE vs HC (ISGs)
D
IFNneg vs HC (ISGs)
E
IFNpos vs HC (ISGs)
F
IFNpos vs IFNneg (ISGs)
DOWN
UP
0
96
0
0
11
154
6
144
−log10 adjustd p value
log2 Fold Change
Differentially expressed
NOT differentially expressed
G
SLE vs HC (TEs)
H
IFNneg vs HC (TEs)
I
IFNpos vs HC (TEs)
J
IFNpos vs IFNneg (TEs)
DOWN
UP
0
11
0
0
11
55
15
45
−log10 adjustd p value
log2 Fold Change
LTR
LINE
Satellite
Other
NOT differentially expressed

Figure 1. Increased TE expression in the PMN of SLE patients is confined to the IFNpos subset.
A, *Top*, the sample numbers for patients in the subgroups (HC, healthy control; each subset of SLE). *Bottom*, A PCA plot of RNA-seq data for all PMN samples color-coded as in the homunculi (HC, gray; IFNneg, blue; IFNpos, mauve). **B,** A plot of differentially expressed TEs versus ISGs (Padj ≤ 0.05) for all cell types purified from peripheral blood, using RNA-seq data from all 23 SLE patients and 10 healthy controls (HC). T, T cells; cMo, conventional monocytes; pDC, cDC, plasmacytoid and conventional DC; B, B cells; PMN, neutrophils). **C-E,** Volcano plots of pairwise comparisons between groups, using RNA-seq data from PMN and a list of ISGs (Supplemental Table 1). Specific comparisons: **C,** All 23 SLE vs HC; **D,** IFNneg SLE samples vs HC; **E,** IFNpos SLE samples vs HC; **F,** IFNpos vs IFNneg subsets of SLE. **G-J** As in C-E, except that TE sequences were analyzed, categorized and color-coded as indicated (LTR, LINE, Satellite, Other). **G,** All SLE vs HC; **H,** IFNneg SLE vs HC; **I,** IFNpos SLE vs HC. **J,** TEs in the comparison of the IFNpos to the IFNneg subset.

Relationships between ISG and TE expression in PMN of SLE patients

The IFNpos SLE samples showed transcriptional profiles distinct from those of IFNneg and HC samples (**Fig 1A**), and a congruent result was obtained when samples were clustered based on the expression of ISGs or TEs (**Fig 2A, B**, respectively). In these clustered heatmaps, IFNpos SLE patients clearly differed from healthy control and IFNneg SLE samples; the IFNneg SLE and healthy controls samples were comingled and more similar to one another. Gene set enrichment analyses (GSEA) showed enrichment of interferon response pathways in all SLE versus healthy control PMN, as well as in IFNpos versus healthy controls or IFNneg as expected (**Fig. 2C**). Although there was no general enrichment of differentially expressed ISGs in PMN of IFNneg SLE patients' versus healthy control (**Fig. 1D**), a slight enrichment of the interferon alpha response pathway was observed (**Fig 2C**). The IFNpos signature had the major impact, as evident in comparisons of the effect sizes for IFNpos vs HC to those of the pool of all SLE patients vs HC (**Fig 2D**). These results show that the heterogeneity among SLE samples can be observed not only in the interferon signature, but also in TE expression.

The enrichment of the interferon gamma response observed in GSEA of IFNpos compared to HC or IFNneg samples suggested the involvement of T cells or NK cells, the major IFN-γ-producing immune cells. However, T cells from the IFNpos subset of SLE patients had no more *IFNG* mRNA at baseline than T cells from healthy controls, whereas those in the IFNneg subset had a modest reduction. NK cells were not examined in the prior study from which our data were drawn [57], but it would be worthwhile in future studies to compare *IFNG* mRNA and interferon-gamma protein expression in NK cells from IFNpos and IFNneg SLE patients and healthy controls.

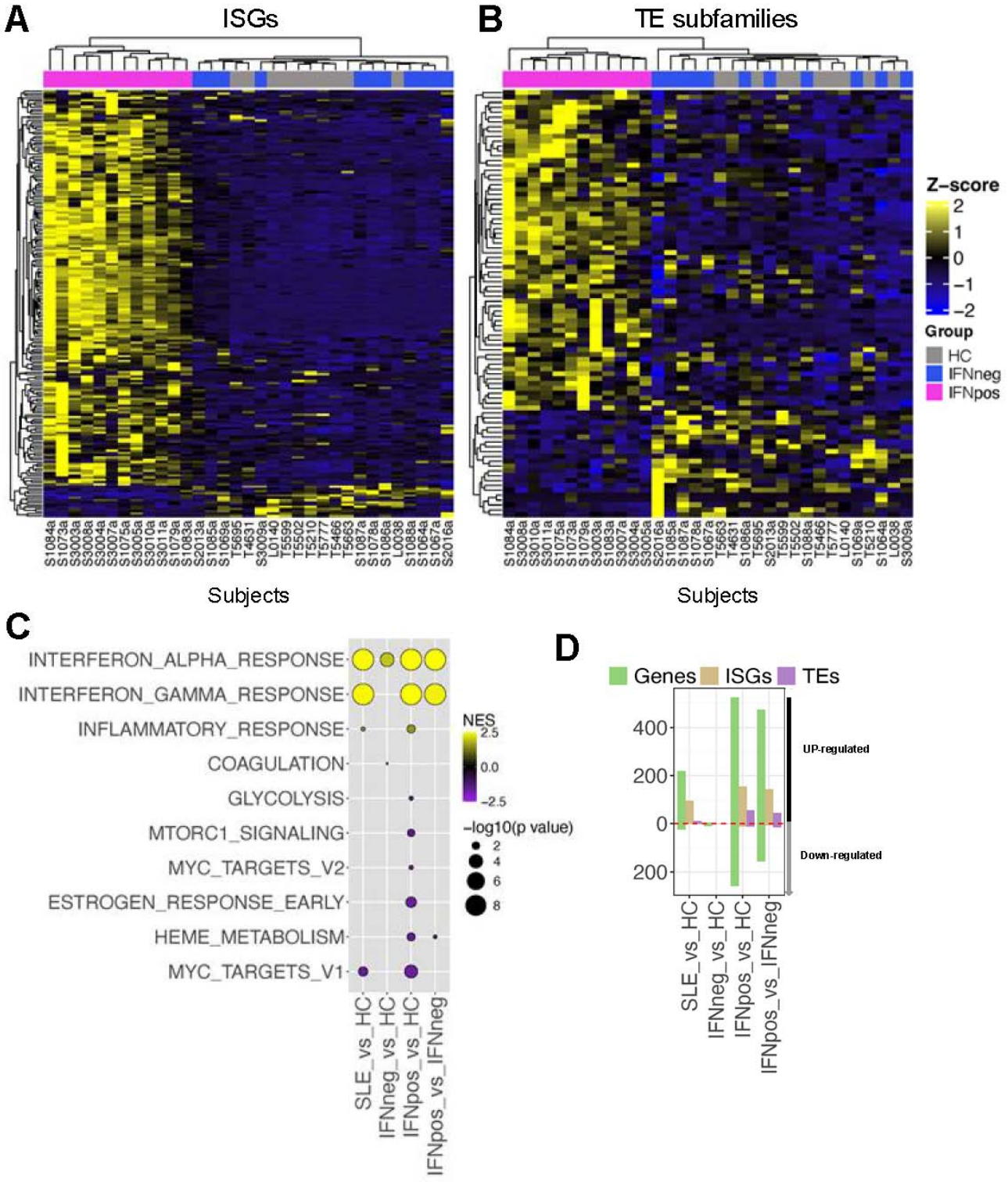


Figure 2. Distinct characteristics of the IFNneg and IFNpos subsets.

A, A self-organizing heatmap of PMN samples using differentially expressed ISGs. Each row represents an individual gene of the ISG signature, with the x-axis showing anonymized coding of patient samples ([57], see Suppl. Information). The three patient groups (HC, IFNneg SLE, pos SLE) are color coded as in Fig 1A. **B,** Heatmap of RNA-seq data with PMN samples, but showing differentially expressed TE families. **C,** Dot plot of pathway analyses applying the indicated classifications to GSEA results for each patient subgroup. **D,** Number of

genes, ISGs and TE's for which differential expression achieved Padj ≤ 0.05 in each indicated pairwise comparison, ISGs and TEs in each comparison. Note that most TEs and ISGs show predominantly increased expression.

Relationships between TE expression and clinical disease

As a chronic disease whose activity varies across time and patient visits, and which is of different overall severity in different patients, the clinical status of SLE patients can be scored either as a "SLE disease activity index" (SLEDAI), on an integer scale at the visit when blood was collected, or binned according to severity (mild, moderate, severe). SLEDAI scores of 1-5, 6-10, and 11-20 indicate mild, moderate and high disease activity at the time of the visit, evaluated based on clinical parameters such as arthritis, vasculitis, myositis, pericarditis, hematuria and so on, whereas whereas disease severity is a more subjective assessment of overall disease summed over multiple visits. These two metrics tended to correlate in the patient cohort used for our analysis (**Fig 3A**) [57]. To explore the relationship between TE expression and clinical SLE, we calculated Pearson's correlation scores and performed linear regression analyses comparing the ISG gene set variance analysis (GSVA) score for each SLE patient and their SLEDAI score or overall disease severity (higher SLEDAI scores indicate more active disease at the time of the visit). When calculated for all SLE patients, there was a weak correlation of the ISG GSVA with SLEDAI (**Fig 3B**). While "severity" is not a clearly quantifiable value, individual patients' ISG GSVA tended to increase in concert with their overall SLE severity (**Fig 3C**).

We then calculated the correlation of ISG GSVA and SLEDAI scores with normalized read counts for specific TE subfamilies. This analysis identified 17 TE subfamilies that were differentially expressed in one or more of the comparisons SLE vs HC, IFNpos vs HC, IFNpos vs IFNneg and whose expression also correlated positively and significantly with ISG GVSA and SLEDAI score (**Suppl. Fig 1C**). As examples, we show four of the top differentially expressed TE (LTR) subfamilies: (i) LTR80A:ERVL, (ii) MLT1A:ERVL-MaLR, (iii) LTR103_Mam:ERV1 and (iv) MLT1-int:ERVL-MaLR (**Fig 3D-F**). For each TE subfamily, we plotted the magnitude of TE expression (normalized read counts) for each patient's PMN against the ISG GSVA in the same sample (**Fig 3D**). The ISG GSVA and TE counts were strongly correlated, with the IFNpos subset showing both the highest TE expression levels and (as expected) the highest ISG scores.

To explore the relationship between increased TE expression and clinical SLE, we calculated Pearson's correlation scores comparing the SLEDAI score of each SLE patient with quantified TE expression (**Fig 3E**). For the four differentially expressed TE subfamilies listed above, TE counts correlated positively with SLEDAI score (**Fig 3E**) and severity of disease (**Fig 3F**). Thus, quantification of transcript levels for certain TE subfamilies correlates at least as well with SLE activity and clinical severity as do measures of ISG expression in the IFNpos subset of SLE patients.

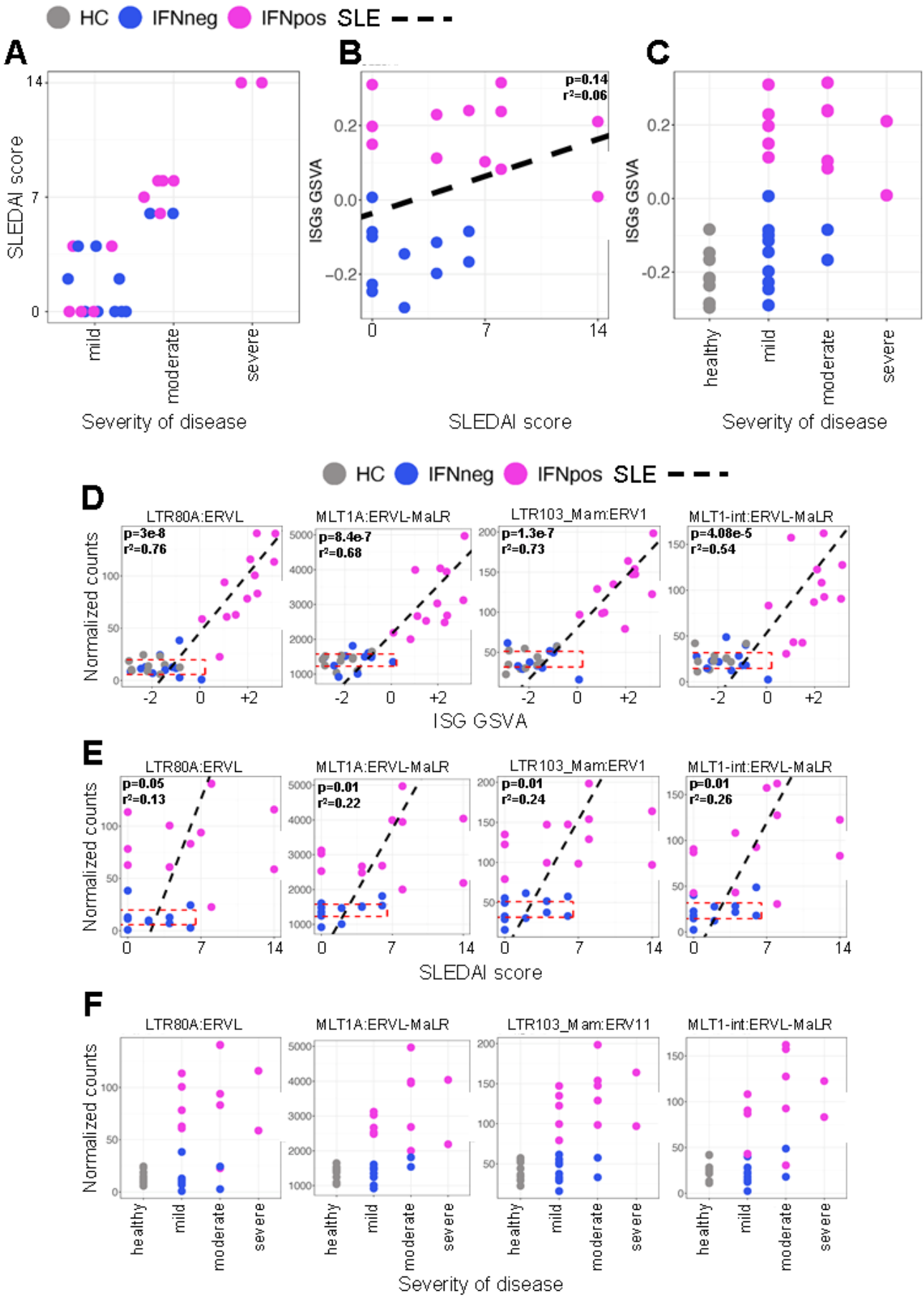

HC IFNneg IFNpos SLE
A
SLEDAI score
Severity of disease
mild
moderate
severe
B
p=0.14
r²=0.06
ISGs GSVA
SLEDAI score
C
ISGs GSVA
healthy
mild
moderate
severe
Severity of disease
D
LTR80A:ERVL
p=3e-8
r²=0.76
MLT1A:ERVL-MaLR
p=8.4e-7
r²=0.68
LTR103_Mam:ERV1
p=1.3e-7
r²=0.73
MLT1-int:ERVL-MaLR
p=4.08e-5
r²=0.54
Normalized counts
ISG GSVA
E
LTR80A:ERVL
p=0.05
r²=0.13
MLT1A:ERVL-MaLR
p=0.01
r²=0.22
LTR103_Mam:ERV1
p=0.01
r²=0.24
MLT1-int:ERVL-MaLR
p=0.01
r²=0.26
Normalized counts
SLEDAI score
F
LTR80A:ERVL
MLT1A:ERVL-MaLR
LTR103_Mam:ERV11
MLT1-int:ERVL-MaLR
Normalized counts
Severity of disease

Figure 3. Relationship between TE increases and disease activity.
A, Scatter plot of SLEDAI vs severity of disease for each of the two categories of SLE patients. **B,** For each patient sample, the ISG GSVA score in PMN is plotted in relation to the SLEDAI score, with IFNneg and IFNpos strata distinguished by color-coding. The dotted black line shows the linear regression result derived using all 23 SLE samples. **C,** Similar to B but a scatter plot of ISG GSVA scores vs severity of disease category of SLE patients and HC. **D-F,** Scatter plots focused on four specific TE families, as indicated, using data from individual patient samples. **D,** Normalized counts of the indicated TE in a subject's PMN plotted against the ISG GSVA score in the same sample. A linear regression line calculated for all samples (black dotted line) is shown, with a rectangle delimited by a red dotted line showing the 95% confidence interval for the normalized TE counts of the HC and IFNneg SLE samples with each TE family. **E,** As in D but plots of normalized counts for the indicated TE family vs SLEDAI score. Linear regression results are denoted by dotted lines as in B. **F,** As in E except the TE counts are plotted against disease severity as tabulated in the original work [57].

Assessing the relationship between expression of intronic TEs and their surrounding genes in SLE patients

Differential expression analysis using quantification of locus-specific TE copies showed enrichment of individual TEs classified as SINEs, LINES, LTRs and DNA elements in all SLE patients compared to HC, as well as in IFNpos SLE samples compared to HC and IFNneg SLE (**Suppl. Fig 1D**). We classified the genomic location of these differentially expressed TEs (DE TEs) as (a) within 2 kb upstream of a nearby gene, (b) in exons, (c) introns, (d) 3'-UTRs, (e) 5'-UTRs, and (f) intergenic regions (**Fig 4A**). Most of the DE TEs in the pool of all SLE samples (IFNneg and IFNpos) mapped to introns of differentially expressed genes (DEGs) (**Fig 4B**; **Suppl. Figs. 2, 3**). This pattern was also observed for the subgroup of IFNpos SLE samples (compared to healthy controls or IFNneg samples), but the stratification increased the sensitivity so that 4 to 7-fold more DE TEs could be assigned to genomic positions (**Fig 4C, D**). Consistent with **Fig. 1H and 2D**, there was virtually no signal for DE TEs in IFNneg SLE PMN samples compared to controls (*not shown*).

Furthermore, we found in this sample set that expression of the genes associated with upregulated TEs was generally increased (~80% of instances), and sometimes unchanged (~20% of instances), but never decreased. In no instance was the direction of change the opposite for DE TEs and the linked gene (i.e., gene UP, TE DOWN or vice versa). The majority of upregulated TEs were present in introns of genes whose expression also increased. Of this class (UP/UP in **Fig. 4C**), ~68% of the upregulated genes were ISGs. These data suggest a positive correlation (perhaps stemming from a mechanistic connection) between increased expression of TEs and increased expression of their associated genes. In over 200 instances, however, expression of a TE was increased and yet that TE mapped to the vicinity of a gene with no change in expression (**Fig 4C**). While the great majority of this set also mapped to introns, of this ~20% of mapped TEs that mapped to the introns of genes whose expression did not change (UP/NO in Fig. 4C), only ~28% were ISGs. These findings raise the possibility that a meaningful fraction of DE TEs can be regulated independent of the surrounding gene in the PMN of SLE patients. While new sequencing studies will be needed to distinguish between this possibility and alternative explanations, we conclude that the detection sensitivity for instances of "TE UP, gene expression unchanged" increased up to 10-fold by focusing the analyses on IFNpos subjects (**Fig 4B** *vs* **Fig 4C**).

The location of these TEs in intronic regions, and the finding that introns were the most common position of DE TEs (**Fig. 4; Suppl. Fig. 2**), indicated that in principle their expression might reflect transcription through the intron in which the TE was located (intron retention, IR). Indeed, intron retention has been proposed to be increased in analyses of circulating leukocytes of SLE patients [61]. To address this possibility, we quantified intron retention in the analyzed samples using the IRFinder software (**Suppl. Fig. 4**). Overall, this tool indicated that there were very few instances in which there was a substantial likelihood of intron retention (**Suppl. Fig. 4E, G**). The data suggested that, globally, intron retention in SLE PMN was not a major mechanism accounting for increased TE expression. However, a more detailed inspection of individual TEs suggested that high intronic TE expression correlated in many cases with upregulated expression of interferon-induced genes, as well as with splicing alterations in nearby exons of the gene (**Figures 5-7)**.

Fig. 5A shows several representative regions with high TE density in the *IFI44L-IFI44* locus, containing the upregulated elements HAL1B_dup2785 (intronic in *IFI44L*) and L1PBb_dup129 (3' of *IFI44*) (*red rectangles* in **Fig 5A**), as well as the L1MC5 and LTR80A elements in introns 5 and 2 of *IFI44*, respectively. Of the TEs within these regions, the intergenic L1PBb_dup129 (3' of *IFI44*) and intronic HAL1B_dup2785 elements were the most highly upregulated individual elements within their respective subfamilies, when comparing expression in IFNpos PMN compared to IFNneg or healthy controls (**Suppl. Fig. 5A**). These elements also displayed considerably higher expression in IFNpos compared to IFNneg or healthy control PMNs (see zoomed in views of **Figs. 5B, 5C**), and their increased expression was paralleled by increased expression of the enclosing *IFI44L* and *IFI44* genes (see bar graph in **Fig. 5E**. An expanded view of the *IFI44L* locus depicting additional exons and the HAL1B element is shown in **Fig. 6A**).

The finding of increased expression in IFNpos compared to IFNneg or healthy control PMNs was also true for the intronic L1MC5, LTR80A and MLT1A:ERVL-MaLR (*not shown*) elements located (with nearby TEs) (*blue rectangles* in **Fig 5A**). Zoomed-in views of the L1MC5 and LTR80A elements, adjacent TEs and the flanking exons of *IFI44*, are shown in **Figs. 5D** and **6B** respectively. The LTR80A:ERVL and MLT1A:ERVL-MaLR elements were also among the highest-expressing units of their respective subfamilies (*data not shown*), and all three TEs displayed considerably higher expression in IFNpos compared to IFNneg or healthy control PMNs.

In further illustration of this point, an MLT1A transcription unit and one of the top-expressing units of the MLT1-int family were located in intron 2 of *IFIH1*, which encodes the double-stranded (ds)RNA sensor MDA5; these TEs were highly and selectively upregulated, perhaps with adjacent TEs, in PMNs in the IFNpos subset of SLE (**Fig 7A**). Similarly, one of LTR103_Mam:ERV1's highest expressing units was located in intron 4 of *IFI6*, a "hub" gene associated with SLE [62] (**Fig 7B**), and was preferentially expressed (albeit at much lower levels than *IFI6* itself) in SLE IFNpos PMNs. We note that in the analyses with IRFinder, the intron retention ratios of introns of *IFI44*, *IFIH1* and *IFI6* had values close to zero, and comparing the results between conditions showed no significant increases between HC and either IFNpos or IFNneg (**Suppl. Fig. 4F, G** respectively). We attribute this to the fact that the averaged values for all introns did not reach statistical significance, since only certain introns of each gene showed increased expression.

Finally, detailed inspection of the exons in the vicinity of each of these upregulated TEs showed remarkable splicing variations in annotated exons adjacent to the upregulated TEs. The mechanism of this effect is unknown, but recalls observations of intron readthrough and TE "exonisation" in cancers and primary human macrophages [30, 40, 41]. In summary, our detailed analysis of TE expression in the IFNpos PMN subset emphasizes several interesting points: (i) the strong correlation of ISG upregulation with increased TE expression; (ii) the preferential enrichment of upregulated intronic TEs in introns of ISGs versus non-ISG genes; and (iii) the association of upregulated intronic TEs with splicing changes in ISG exons flanking those TEs.

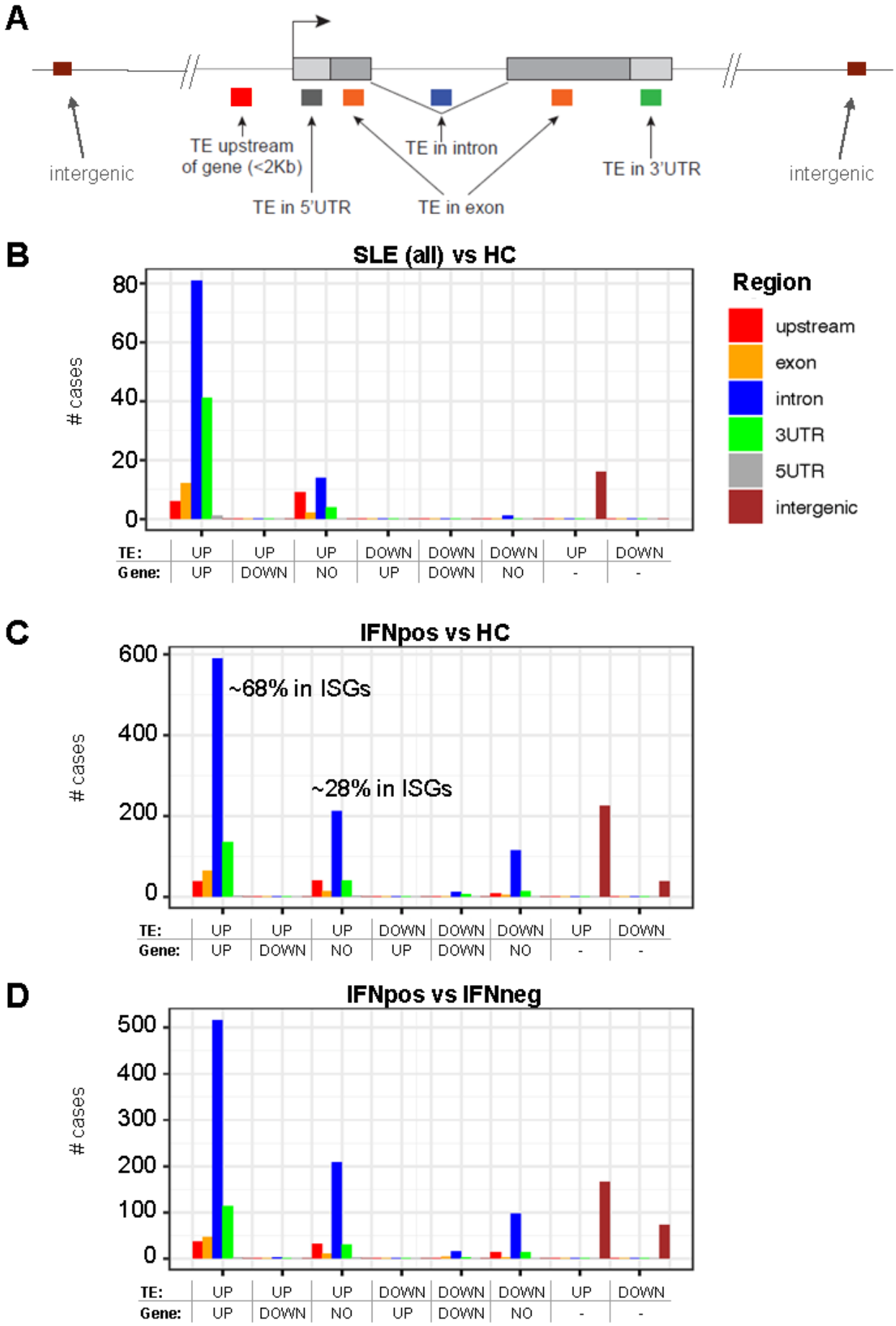

A
intergenic
TE upstream of gene (<2Kb)
TE in 5'UTR
TE in intron
TE in exon
TE in 3'UTR
intergenic
B
SLE (all) vs HC
cases
Region
upstream
exon
intron
3UTR
5UTR
intergenic
TE: UP UP UP DOWN DOWN DOWN UP DOWN
Gene: UP DOWN NO UP DOWN NO - -
C
IFNpos vs HC
~68% in ISGs
~28% in ISGs
cases
TE: UP UP UP DOWN DOWN DOWN UP DOWN
Gene: UP DOWN NO UP DOWN NO - -
D
IFNpos vs IFNneg
cases
TE: UP UP UP DOWN DOWN DOWN UP DOWN
Gene: UP DOWN NO UP DOWN NO - -

Figure 4. Genome annotations for locations of DE TEs relative to conventional genes.

A, Representation of the potential locations of a TE relative to a gene. **B-D,** Bar graph y-axes show the number of differentially expressed (DE) TEs counted for each type of genomic location denoted in A, subdivided based on whether the RNA counts for the TE increased or decreased, and then by whether expression of the linked gene increased ("UP"), decreased ("DOWN") or did not change signficantly ("NO"). **B,** Shown are the numbers of DE TEs mapping to each type of genomic location relative to identifiable genes when analyzing DE TEs identified by comparisons of all 23 SLE samples to HC. **C,** As for B but using the DE TEs identified by comparing HC samples to IFNpos SLE PMN. The inset percentages are the fraction of genes for which TE's specifically mapped to an intronic region of an ISG. **D,** As for B but using the DE TEs identified by comparing RNA-seq of IFNpos and IFNneg samples.

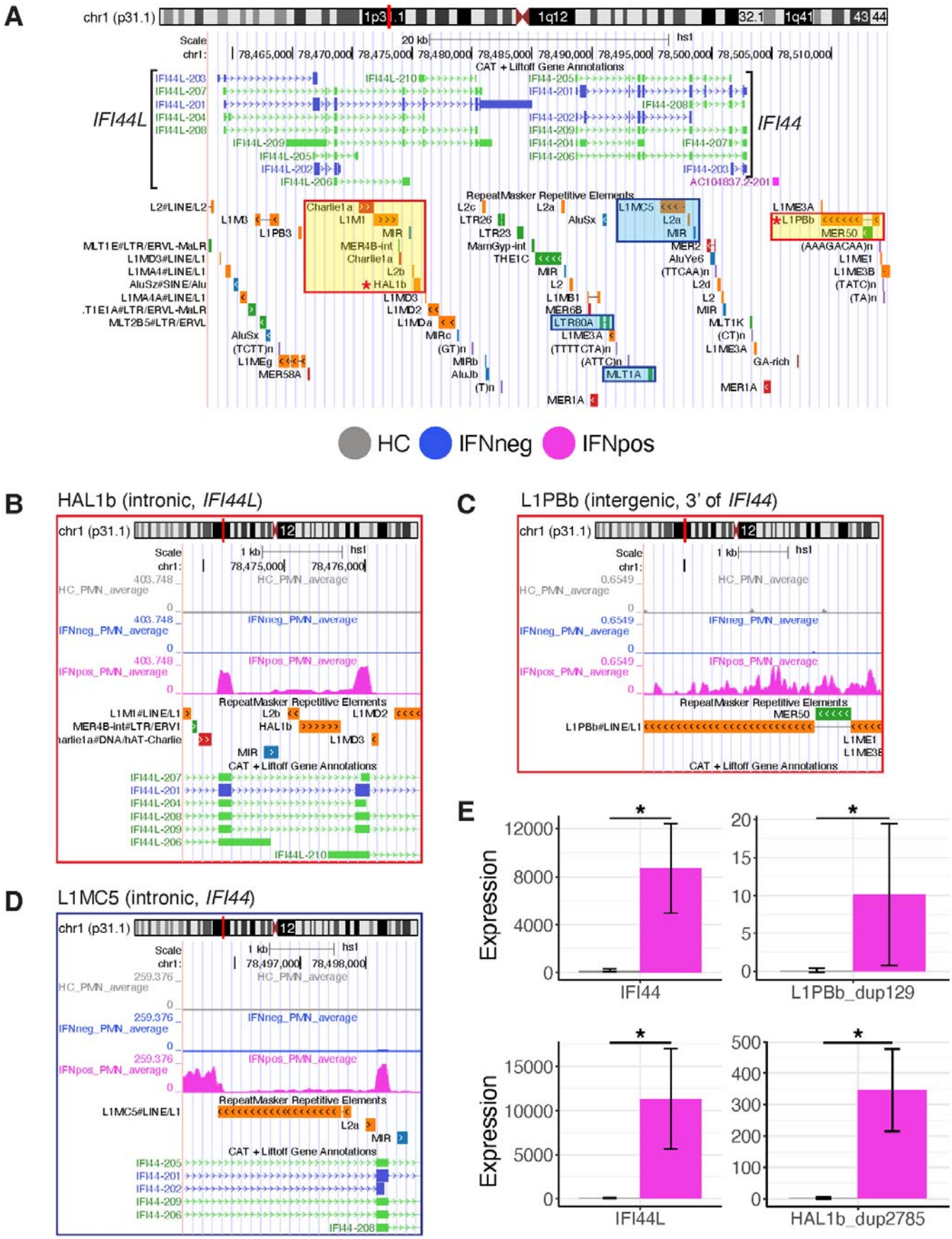


Figure 5. High intronic expression of HAL1B and L1MC5 LINE-1 elements in the *IFI44L* and *IFI44* genes

correlates with increased *IFI44L* and *IFI44* gene expression and with splicing variations in adjacent exons.

A, Genome browser view of the *IFI44L-IFI44* locus. In order, the tracks depict chromosomal location, the annotated splice isoforms of the *IFI44L* and *IFI44* genes, and all repeat/transposable elements in the *IFI44L-IFI44* locus. The high-expressing units HAL1b (panel B), L1PBb (panel C), and L1MC5 (panel D) are denoted by orange or blue boxes. The intergenic L1PBb_dup129 (3' of *IFI44*) and the intronic HAL1B_dup2785 elements (indicated by asterisks) were the most highly upregulated individual elements within their respective subfamilies, as highlighted in Suppl. Fig. 5. **B-D**, zoomed-in genome browser views showing averaged RNA-seq reads for all samples in each of the patient groups – healthy controls (HC; gray), IFNneg (blue), and IFNpos (mauve). For B and D, note the correlation of intronic transcription encompassing the TE with splicing variations in the flanking exons. **E**, Bar graph showing averaged RNA expression for the indicated genes and their corresponding intronic TEs (*top*, *IFI44* and L1MC5_dup129; *bottom*, *IFI44L* and HAL1B_dup2785). Normalized read counts from nonstranded RNA-seq +/- standard deviation are shown.

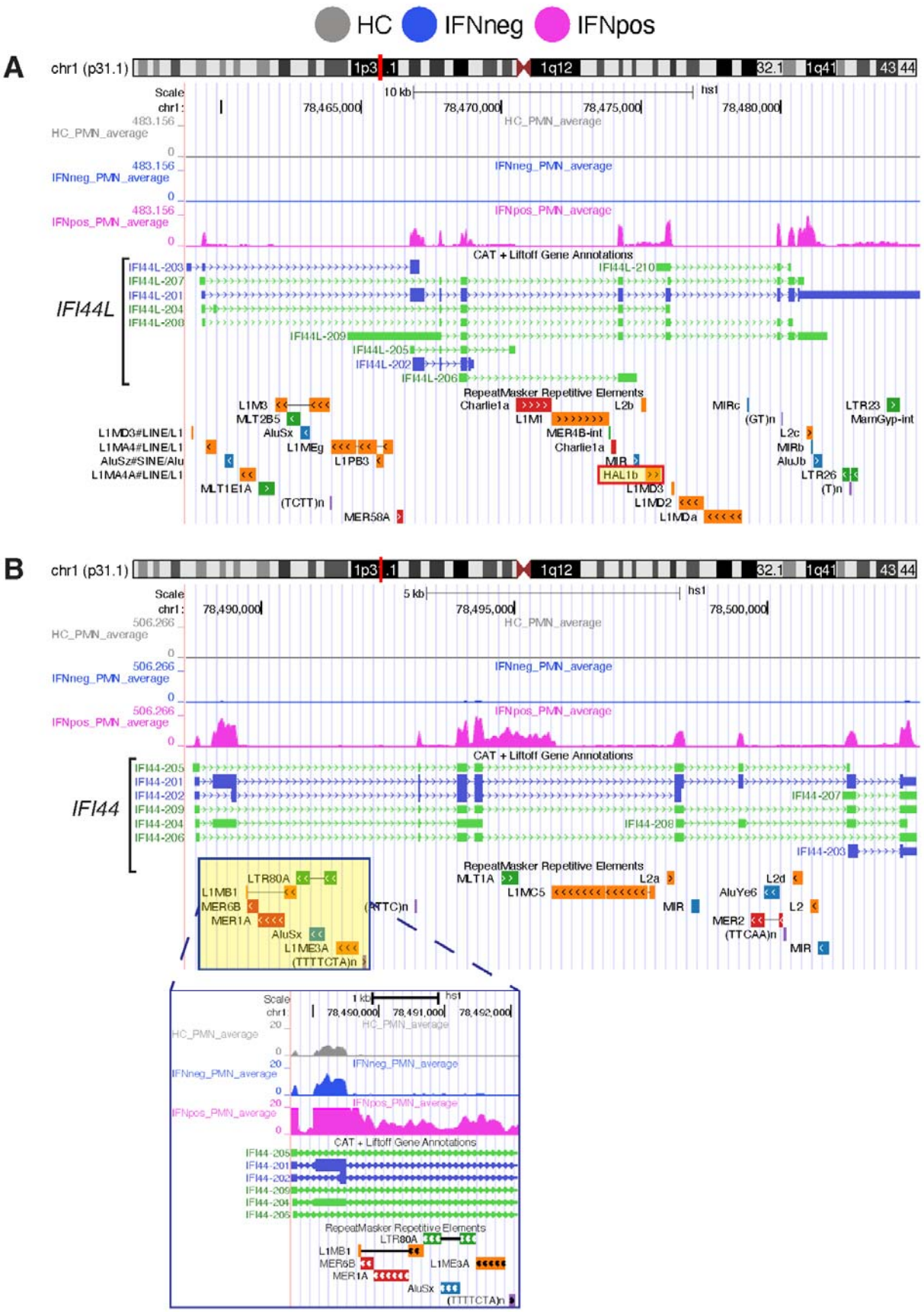


Figure 6. Expanded view of the intronic HAL1B element and LTR80A elements in the *IFI44L-IFI44* locus.

A, Expanded genome browser view of the *IFI44L-IFI44* locus showing increased expression of the same HAL1B element shown in Figure 5A, B. **B,** Expanded genome browser view of the *IFI44* gene, showing one of

LTR80A:ERVL's and one of MLT1A:ERVL-MaLR's highest expressing units located in introns 2 and 5, respectively, of IFI44. Note the selective upregulation of transcription in the introns containing these high-expressed TEs and the occurrence of annotated splice variants in the flanking exons.

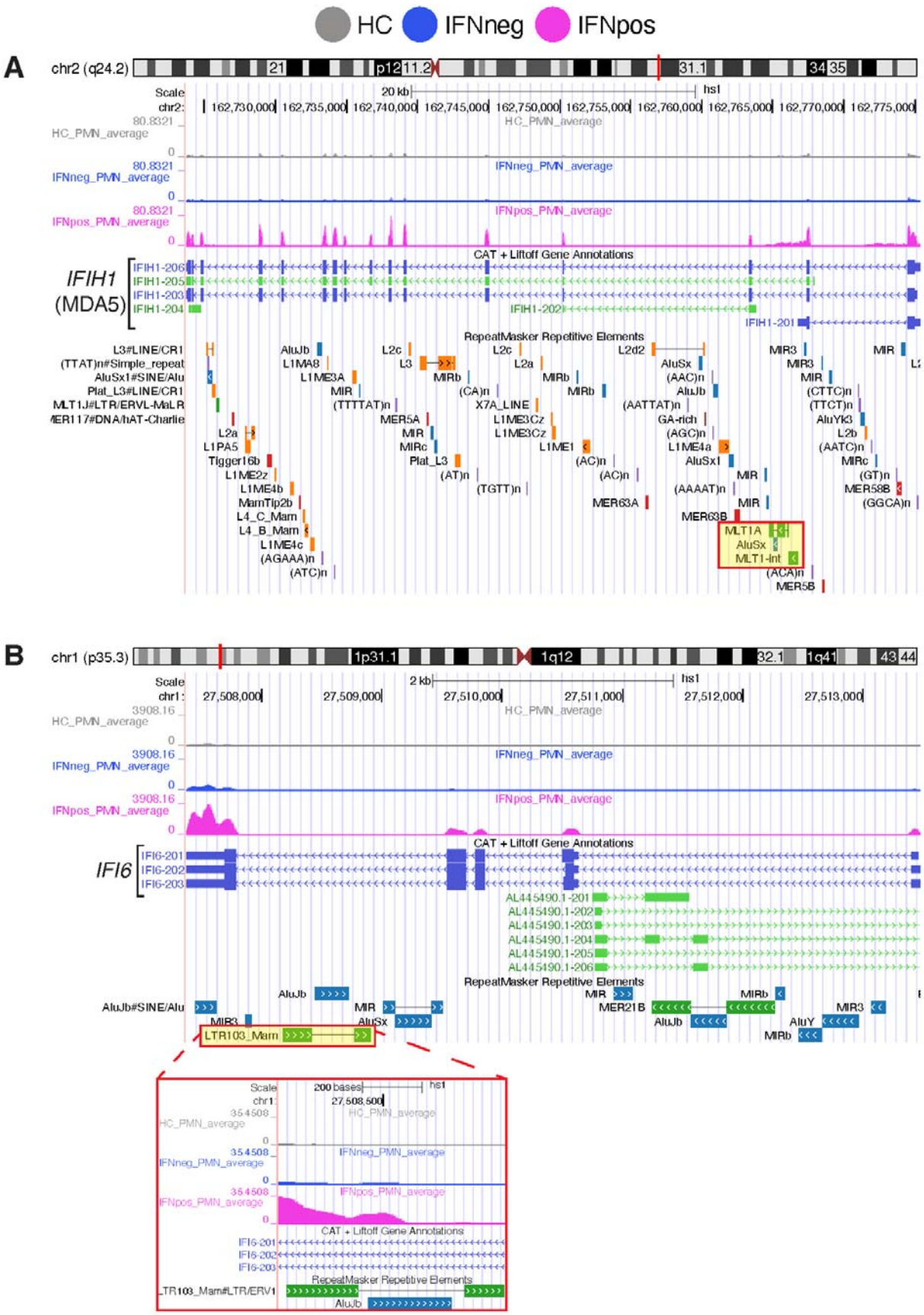


Figure 7. Genome browser views of the intronic MLT1A and MLT1-int elements in *IFIH1* locus and intronic LTR103_Mam element in the *IFI6* locus.

A, Genome browser view of the *IFIH1* locus encoding MDA5, showing the localization of one of the highest-expressing units of the MLT1A transcription unit and one of the top-expressing units of the MLT1-int family in *IFIH1* intron 2. Note the selective upregulation of transcription in the intron containing this high-expressed TE and the occurrence of annotated splice variants in the flanking exons. **B,** As in A, but showing the localization of one of LTR103_Mam:ERV1's highest expressing units in an intron of *IFI6*.

DISCUSSION

Increased expression of TE-encoded RNA can often be observed in circulating leukocytes from non-stratified collections of SLE patients [49-56, 63, 64]. Differential TE expression has been noted in $CD4^+$ T cells, $CD14^+$ monocytes, $CD19^+$ B cells, and natural killer cells from peripheral blood mononuclear cells of SLE patients [63]. Most TE classes reported as upregulated have been in the LTR class, but a subset of PMN in SLE patients express Orf1p and Orf2p proteins encoded by L1 retrotransposons [52]. We show here that within the PMN of a sample population of SLE patients, expression of TE families above the levels measured in samples from healthy controls was restricted to patients pre-classified as IFNpos based on a previously reported gene network analysis [57]. In contrast, none of the samples from IFNneg patients displayed such increases (**Figs. 1, 2**). The expression levels of several differentially expressed subfamilies of highly expressed TEs in the patient population tracked closely with ISG expression and with the clinical activity and disease severity of SLE (**Fig. 3**). Finally, individually-mapped, highly-expressed TEs tended to reside in the introns of genes (particularly ISGs) that were also highly upregulated in the IFNpos SLE population (**Fig. 4**). The expression of these TEs correlated with readthrough transcription (partial intron retention) of the corresponding introns, as well with the presence of annotated splice variations in the adjacent (flanking) exons (**Figs 5-7**).

Neutrophils (PMN) were selected as a focus for this work in part because the overall number of differentially expressed genes, and the number of those with ISG identities, were largest in this cell subset in the data set [57] on which the analyses were based (**Fig. 1B**). Abnormalities of neutrophils are widespread among SLE patients (reviewed in [8]), and neutrophil perturbations in SLE include heightened expression of multiple cytokines, including BAFF and type 1 IFN, that can promote the breaches of self-tolerance among B cells that lead to auto-Ab production [18, 58, 59, 65]. Circulating granulocytes of SLE patients include the low-density granulocyte, a variant of typical PMN [7, 8, 52, 66], suggesting that it would be worthwhile in future investigations to distinguish this PMN subset from conventional PMN. PMN also have a specific capacity to undergo 'NETosis', a form of programmed death that can be enhanced by PMN exposure to type 1 or 2 IFNs [67, 68]. An additional stimulus to NETosis that involves TE derepression may involve auto-antibodies directed against TE-encoded proteins. Immune complexes formed with circulating antibodies directed against antigens encoded by ERV-K102 have been reported to stimulate NET formation in vitro [56], with the caveat that immune complexes with other types of antigens may also induce NETosis [69]; the physiological concentrations of these ERV-K102-containing complexes are likely to be low and their impact at these presumed low concentrations remains to be established. Antibodies directed against the LINE L1-encoded ORF1p protein are common in SLE patients, but their presence does not correlate with disease activity in all patient cohorts [64]. Whatever the mechanisms of sensitization and triggers of NETosis, the ribonucleoproteins of neutrophil extracellular traps (NETs) can stimulate IFN-I production [7, 70], and the cytoplasmic and nuclear remnants of PMN NETosis provide self-antigens and stimulate the generation of auto-antibody-producing B cells that perpetuate a destructive cycle [9, 10, 66].

Our analyses of selected TE families in individual patient samples indicated that increased TE expression could be positively correlated with the presence of an IFN signature as well as with disease activity and severity in PMNs. For certain TE subfamilies, the correlation of increased TE expression with SLEDAI score and disease severity was even stronger than the correlation between SLEDAI score/ disease severity and increased ISG expression (**Fig. 3**). Since neutrophils constitute nearly half of the blood cell population, the relation of SLE disease severity to TE expression, neutrophil activation and IFN production in PMNs warrants further investigation. When ranked at the family level (**Fig. 3**), the most common differentially expressed (primarily upregulated) TEs were endogenous retrovirus (LTR/ERV)-related sequences, a finding generally consistent with prior reports in which analyses were

performed with diverse mixtures of cell types [54-56] or that did not examine PMNs [63]. However, our strategy of stratification to IFNpos patients, our focus on a single immune cell type (PMNs) and the use of the Telescope informatics tool [71] allowed us to identify differentially expressed RNA-seq reads assigned to all the varied classes of TE – HERVs, DNA transposons, LINEs, SINEs, and satellite repeats. Similar analyses are in process for other immune cell types.

Distinguishing the mechanisms that represent individual inciting events versus the diverse cellular and molecular processes that perpetuate disease is an intriguing issue in SLE research. Previous studies have suggested that increased TE expression contributes to SLE pathogenesis [55, 56], and GWAS studies have implicated ISGs in SLE [48, 49], but whether TE expression is a root cause or an effect (or both) of the abnormally heightened ISG expression observed in most SLE patients remains unclear. On the one hand, there is compelling evidence that TE expression can evoke type I IFN responses that result in ISG expression; specifically, double-stranded RNAs generated by transcription of inverted Alu repeats activate the dsRNA sensor MDA5, especially in the absence of the RNA editing enzyme ADAR1 or the presence of gain-of-function MDA5 variants associated with autoimmune disease [45, 46, 72, 73]. MDA5 forms long filaments on long stem loops of double-stranded RNA, resulting in downstream activation of the IRF3/7 and NFκB transcription factors which initiate type I interferon production and transcription of ISGs [72]. MDA5 is encoded by the *IFIH1* gene, which is one of 10 SLE "hub" genes (*IFI35, IFIT3, ISG15, OAS1, MX2, OASL, IFI6, IFITM1, IFIH1, IRF7*) identified by analysis of microarray data [62]; 9 of these (including *IFIH1*), are ISGs and the tenth, *IRF7,* encodes the interferon regulatory factor IRF7 that elicits the type I IFN response [62, 72]. GWAS studies have also implicated the Toll-like receptor TLR7 in SLE [48, 49]; TLR7 trafficked into endosomes recognizes viral and other single-stranded RNAs [74].

Whether ISG expression promotes increased TE expression is less clear. We show here that highly expressed TEs are enriched in the introns of genes that are differentially upregulated in SLE; this is particularly apparent when comparing IFNpos SLE patients versus healthy controls (GENE-UP TE-UP category **in Fig. 4**, **Suppl Fig. 3**). Additionally, more than two-thirds of the introns with highly expressed TEs are located in ISGs (**Fig. 4**), and show moderate read-through transcription (intron retention) compared to other introns in the same ISG (**Figs. 5-7**). Notably, inspection of annotated splice variants in a small number of cases suggests that the exons flanking ISG introns with increased TE expression and transcription are particularly prone to alternative splicing (**Figs. 5-7**). This phenomenon may be related to the widespread TE exonisation reported to occur in both normal and cancer cells [30, 40, 41], which in primary human macrophages was shown to have functional consequences for the innate immune response – incorporation of an Alu-derived exon generated a short isoform of the type I IFN receptor IFNAR, which suppressed ISG induction by acting as a decoy receptor for type I IFNs [41]. Long-reads whole transcriptomic sequencing will be necessary to begin probing the mechanisms underlying the relation of increased ISG expression to increased expression of intronic TEs; this effort may be complicated by the fact that alternatively spliced transcripts containing exonised TEs are often [40] – but not always [41] – expressed at much lower levels (~1%) than the canonical transcript. Follow-up studies incorporating long-read sequencing and using larger patient cohorts, or examining pre-morbid phases and progenitor cells of lupus-prone mice [75], will be required to probe early and late stages of SLE.

Dysregulation or increased expression of TEs has been observed in multiple diseases or pathological processes, particularly aging, neurodegeneration, and cancer [32, 35, 37, 76-79]. The findings raise intriguing questions about the generality of TE de-repression in inflammatory or autoimmune diseases. A number of autoimmune rheumatic diseases are distinct from SLE, but display evidence of increased type 1 or type 2 IFN levels, ISGs, and in limited literature, increased TE expression [80-83]. For example, increased expression of ERV K-102 and the LINE-1 elements have been reported in rheumatoid arthritis as in SLE [84-87]. Both type 1 and 2 IFNs are important early in the pathogenesis of type 1 diabetes mellitus (T1D) [88, 89]. We note that most of these studies are preliminary, lacking essential cell type information and stratification (e.g. IFNpos versus IFNneg). The pace of advances in sequencing – especially the capacity for long-read sequence determination – and the development of new analytic tools make it uniquely timely to investigate such questions in depth, both in SLE and across auto-immune diseases in general.

Limitations. Limitations of these analyses include the relatively small sample size and the pressing need for long-read sequencing to permit more precise determination of the structure and location of intronic TE sequences and their potential role in splicing. The threshold levels of TE-encoded RNA needed to trigger increased type I IFN production, and the extent to which increased TE expression is cell type-specific, are not known, whether for SLE, other autoimmune diseases, senescence, ageing or cancer.

METHODS

Dataset. We collected publicly available bulk RNA-seq data from Panwar, B. et al. [57]. The overall dataset contains 288 samples from 91 subjects, 65 patients with SLE and 26 healthy controls (HC), with purified polymorphonuclear neutrophils (PMN) from 33 total subjects. Samples in the original study comprised six different cell types - classical monocytes (cMo) (n = 119), conventional dendritic cells (cDC) (n = 32), plasmacytoid dendritic cells (pDC) (n = 33), T cells (n = 36), B cells (n = 32), and polymorphonuclear neutrophils (PMNs) (n = 36, of which three were samples from follow-up visits). Based on a network analysis of the expression of interferon genes and interferon stimulated genes (ISGs), the original article identified an approach that divided the SLE patients into two subgroups, one exhibiting greater interferon effect, termed 'IFNpos' [57] and a second subset designated as 'IFNneg'. This classification of the samples and terminology were used in the present study.

RNA-seq and data analyses. TrimGalore v0.6.7 [90] was used to remove adapters, reads were mapped against the T2T-CHM13v2.0 genome reference using STAR [91] v2.7.10a (--outFilterMultimapNmax 1 --alignIntronMin 20 --twopassMode Basic --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterType BySJout --outSJfilterReads Unique --outFilterMismatchNoverReadLmax 0.04) to analyze gene expression. HT-Seq [92] v2.0.4 was used to quantify gene expression with default parameters. In order to analyze transposable element expression reads were mapped against the T2T-CHM13v2.0 genome reference using STAR v2.7.10a (--outFilterMultimapNmax 100 --winAnchorMultimapNmax 100) allowing multi-mappers. Using the alignments with multi-mappers as input, we employed two programs to quantify the expression of transposable elements, TEtranscripts [93] v2.2.1 (--mode multi) was used to quantify the expression of the transposable elements at subfamily level and Telescope [94] v1.0.3 was used to quantify the expression of individual transposable elements (locus specific). Four comparisons to observe transcriptional changes between SLE samples and HC were performed: (1) all SLE sample vs HC; (2) IFNpos vs HC; (3) IFNneg vs HC; (4) IFNpos vs IFNneg using samples taken on the first visit of each patient as described in the original article. Differential expression analysis was performed in R [95] v4.2.1 using DESeq2 [96] v1.38.3. Genes with less than 10 read counts per group were filtered out before performing the analysis. Differentially expressed genes were defined using adjusted p value ≤ 0.05 and |log2 Fold Change| ≥ 1. In the case of TEs (at the individual and family levels), those with less than 1 read count in all samples were filtered out before performing the analysis. Differentially expressed TEs were defined using adjusted p value ≤ 0.05 and |log2 Fold Change| ≥ 0.5. To analyze the relationship between the expression of transposable elements and genes, we used genomic coordinates of genes and transposable elements and used bedtools [97] and R to identify the genomic locations of transposable elements relative to nearby genes. We classified the genomic location of the transposable elements as (a) 2 kb upstream of a nearby gene, (b) in exons, (c) introns, (d) 3'-UTRs, (e) 5'-UTRs, and (f) intergenic regions. Once we located the transposable elements relative to genes in the genome, we looked for the correlation between differentially expressed transposable elements and differentially expressed genes (Figure 3). IRFinder [98] v1.3.1 was used to quantify intron retention in the analyzed samples using T2T-CHM13v2.0 genome reference.

Gene set enrichment and variation analyses Gene set enrichment analysis (GSEA) was performed in R v4.2.1 using clusterProfiler [99] v4.6.2 with the hallmark gene sets from msigdbr [100] v7.5.1 with the results of the differential expression analysis as input. Pathways with p value ≤ 0.05 were considered. Gene set variation analysis (GSVA) was performed in R v4.2.1 using the gsva function from GSVA [101] v1.52.3 and personalized gene sets to quantify enrichment of interferon and ISGs in the analyzed samples.

Correlation analyses of transposable elements and clinical disease. Pearson correlations and linear regression were performed in R v4.2.1 using the cor.test function and the lm function, respectively, from the stats library using normalized counts of TEs from SLE patients and SLE disease activity index (SLEDAI) score to assess a possible correlation between TEs and severity of the disease or using normalized counts of TEs from SLE patients' PMN and the ISG GSVA scores to assess a possible correlation between the expression of TEs and the expression of ISGs. TEs with positive and significant correlation were filtered based on differential expression status, those that were differentially expressed and that correlated significantly were kept for further analyses.

Tracks. From deeptools [102] v3.5.4, multiBamSummary was used to calculate normalization factors and bamCoverage was used to generate bigWig files to explore genome coverage. Wiggletools [103] v1.2.11 and bedGraphToBigWig v4 were used to obtain average genome coverage tracks for the three conditions IFNpos, IFNneg and HC. Genome browser views were generated using the most recent update (2025) of the UCSC genome browser [104].

A

DonorID	Visit 1	Visit 2	Visit 3	Gender	Age	IFN status
S1067	■	■	■	F	63	IFNneg
S1064	■	■		F	56	IFNneg
S1083	■	■		F	50	IFNpos
L0038	■			F	33	HC
L0140	■			F	32	HC
S1069	■			F	21	IFNneg
S1073	■			F	63	IFNpos
S1075	■			F	59	IFNpos
S1078	■			F	54	IFNneg
S1079	■			F	21	IFNpos
S1084	■			F	39	IFNpos
S1085	■			F	50	IFNneg
S1086	■			F	65	IFNneg
S1087	■			F	48	IFNneg
S1088	■			F	48	IFNneg
S2013	■			F	38	IFNneg
S2016	■			F	51	IFNneg
S3003	■			F	40	IFNpos
S3004	■			F	53	IFNpos
S3005	■			F	34	IFNpos
S3007	■			F	20	IFNpos
S3008	■			M	27	IFNpos
S3009	■			F	46	IFNneg
S3010	■			F	23	IFNpos
S3011	■			F	77	IFNpos
T4631	■			F	52	HC
T5210	■			F	37	HC
T5466	■			F	24	HC
T5502	■			F	58	HC
T5599	■			F	50	HC
T5663	■			F	35	HC
T5695	■			F	36	HC
T5777	■			F	31	HC

B TE subfamilies

SLE vs HC		
TE	UP	DOWN
Satellite	0	0
SINE	0	0
LINE	0	0
DNA	1	0
LTR	10	0

IFNneg vs HC		
TE	UP	DOWN
Satellite	0	0
SINE	0	0
LINE	0	0
DNA	0	0
LTR	0	0

IFNpos vs HC		
TE	UP	DOWN
Satellite	4	1
SINE	0	0
LINE	3	1
DNA	5	3
LTR	43	6

IFNpos vs IFNneg		
TE	UP	DOWN
Satellite	3	3
SINE	0	0
LINE	1	2
DNA	8	3
LTR	33	7

D Individual TEs

SLE vs HC		
TE	UP	DOWN
Retroposon	1	0
Satellite	0	0
SINE	121	4
LINE	89	2
DNA	44	1
LTR	65	0

IFNneg vs HC		
TE	UP	DOWN
Retroposon	0	1
Satellite	0	0
SINE	0	1
LINE	0	4
DNA	1	1
LTR	1	3

IFNpos vs HC		
TE	UP	DOWN
Retroposon	3	5
Satellite	1	0
SINE	951	162
LINE	698	134
DNA	265	34
LTR	410	48

IFNpos vs IFNneg		
TE	UP	DOWN
Retroposon	1	1
Satellite	1	0
SINE	775	220
LINE	677	82
DNA	237	24
LTR	355	45

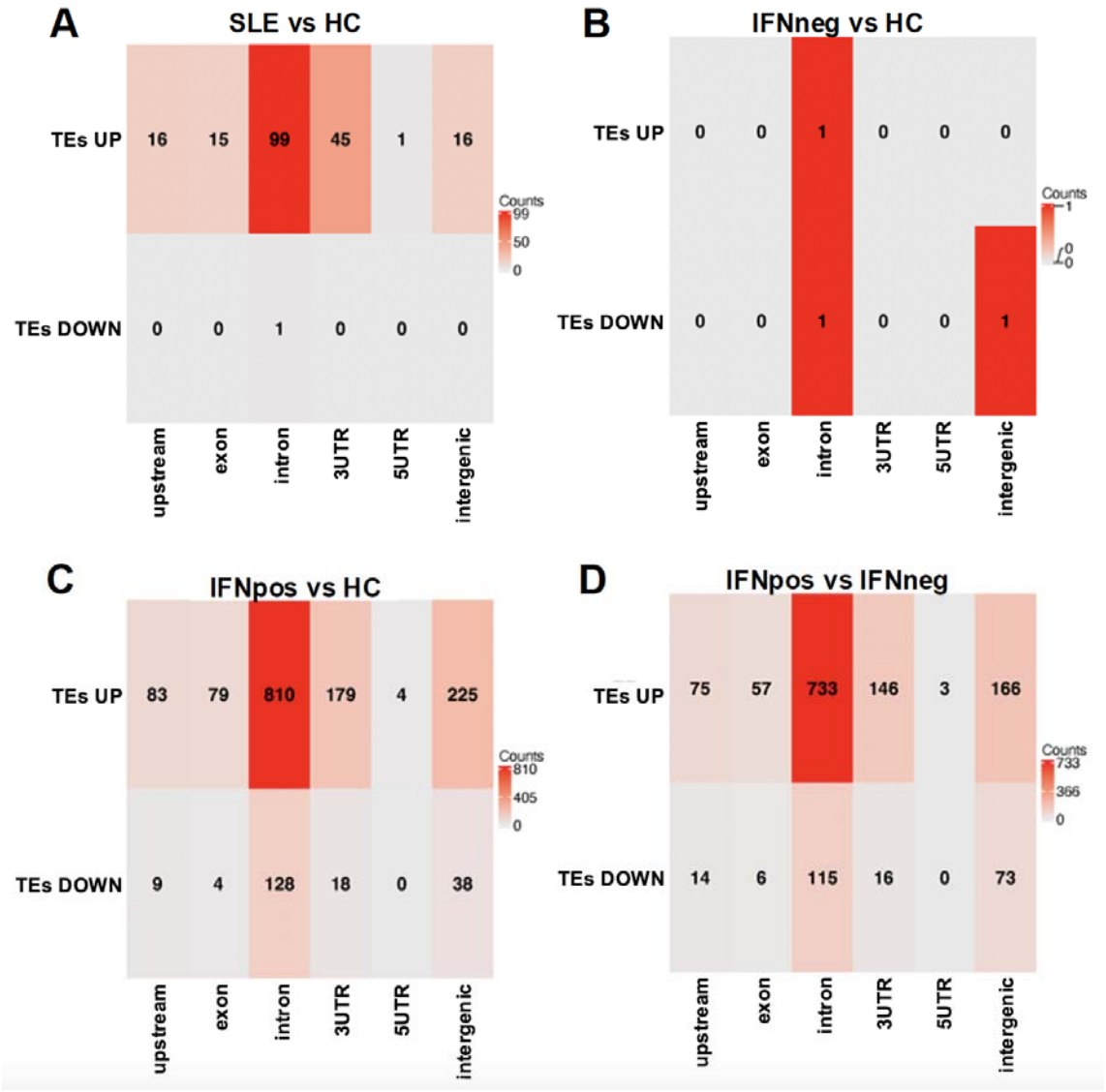
C

Correlated TEs with SLEDAI score						
TE	pvalue	r	SLE vs HC	IFNneg vs HC	IFNpos vs HC	IFNpos vs IFNneg
MamRep38:hAT-Tip100	0.0044	0.5711			✓	✓
LTR71B:ERV1	0.0054	0.5608			✓	
LTR71A:ERV1	0.0059	0.5561			✓	
MLT1-int:ERVL-MaLR	0.0079	0.5395			✓	✓
UCON37:hAT-Ac	0.0083	0.5365	✓		✓	✓
LTR103_Mam:ERV1	0.0101	0.5250	✓		✓	✓
MLT1A:ERVL-MaLR	0.0146	0.5021	✓		✓	✓
MER90a:ERV1	0.0176	0.4901				✓
HAL1b:L1	0.0227	0.4729			✓	✓
Charlie30a:hAT-Charlie	0.0256	0.4643			✓	
MER70A:ERVL	0.0262	0.4628			✓	
LTR2:ERV1	0.0262	0.4626			✓	✓
Eutr15:hAT-Tip100	0.0333	0.4450			✓	✓
UCON39:TcMar-Tigger	0.0412	0.4289				✓
LTR104_Mam:Gypsy	0.0446	0.4225			✓	✓
LTR80A:ERVL	0.0477	0.4171	✓		✓	✓
HERVK9-int:ERVK	0.0483	0.4161			✓	

Supplementary Figure 1. Tables of subjects visit, number of differentially expressed TE classes and TEs with positive and significant correlation with SLEDAI score.

A. Description of the number of subjects at each visit from whom PMN samples were collected. We confined our analysis to samples collected at the first visit. **B**. Tables describing the classes of TE subfamilies that are

differentially expressed in each comparison. **C**. Table of the differentially expressed TEs that have significant positive correlation with SLEDAI score. The check mark indicates if a given TE was found upregulated in the indicated comparison. **D**. Tables describing the classes of individual TEs that are differentially expressed in each comparison.



Supplementary Figure 2. Number of differentially expressed transposable elements found at different genomic locations.

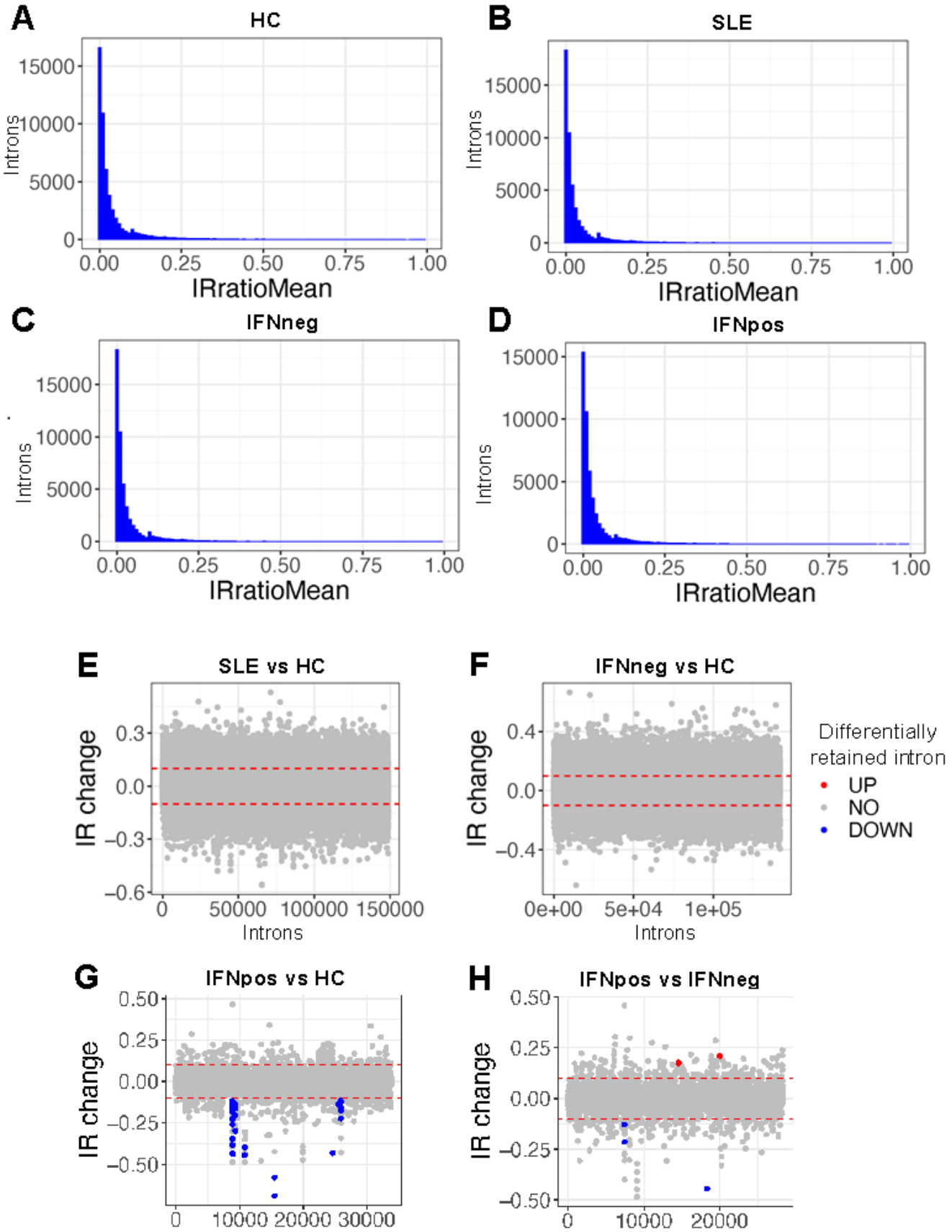
A. Number of differentially expressed transposable elements found at different genomic locations (upstream, exon, intron, 3'UTR, 5'UTR, intergenic) in the comparison SLE vs healthy controls (HC)**.** **B**. Number of differentially expressed transposable elements found in the comparison IFNneg vs HC in different genomic locations**.** **C**. Number of differentially expressed transposable elements found in the comparison IFNpos vs HC in different genomic locations. **D**. Number of differentially expressed transposable elements found in the comparison IFNpos vs IFNneg in different genomic locations.

Percentage of ISGs among total genes in each category									
SLE vs HC									
region	GENE-UP TE-UP	GENE-UP TE-DOWN	GENE-UP TE-NO	GENE-DOWN TE-UP	GENE-DOWN TE-DOWN	GENE-DOWN TE-NO	GENE-NO TE-UP	GENE-NO TE-DOWN	GENE-NO TE-NO
2kb_upstream	83.33	NA	48.15	NA	NA	0.00	0.00	NA	3.77
exon	58.33	NA	6.67	NA	NA	0.00	100.00	NA	2.37
intron	82.72	NA	49.23	NA	NA	0.00	7.14	0.00	7.47
3UTR	70.73	NA	66.67	NA	NA	0.00	75.00	NA	8.22
5UTR	100.00	NA	NA	NA	NA	NA	NA	NA	6.64
IFNpos vs HC									
region	GENE-UP TE-UP	GENE-UP TE-DOWN	GENE-UP TE-NO	GENE-DOWN TE-UP	GENE-DOWN TE-DOWN	GENE-DOWN TE-NO	GENE-NO TE-UP	GENE-NO TE-DOWN	GENE-NO TE-NO
2kb_upstream	39.47	NA	26.89	NA	NA	1.98	17.95	0.00	3.60
exon	23.81	NA	4.76	NA	0.00	3.23	7.14	0.00	2.50
intron	67.63	NA	30.81	NA	8.33	1.96	27.83	6.14	7.51
3UTR	68.89	NA	32.59	NA	40.00	1.67	35.00	23.08	7.61
5UTR	50.00	NA	27.27	NA	NA	0.00	0.00	NA	5.56
IFNpos vs IFNneg									
region	GENE-UP TE-UP	GENE-UP TE-DOWN	GENE-UP TE-NO	GENE-DOWN TE-UP	GENE-DOWN TE-DOWN	GENE-DOWN TE-NO	GENE-NO TE-UP	GENE-NO TE-DOWN	GENE-NO TE-NO
2kb_upstream	52.78	NA	31.98	NA	NA	0.00	12.50	7.14	3.56
exon	30.43	NA	5.88	NA	0.00	0.00	10.00	0.00	2.50
intron	67.31	NA	34.88	0.00	0.00	5.54	26.79	9.28	7.50
3UTR	68.42	NA	33.33	NA	0.00	5.71	44.83	0.00	7.87
5UTR	50.00	NA	20.00	NA	NA	0.00	0.00	NA	7.37

Number of ISGs/number of total genes in each category									
SLE vs HC									
region	GENE-UP TE-UP	GENE-UP TE-DOWN	GENE-UP TE-NO	GENE-DOWN TE-UP	GENE-DOWN TE-DOWN	GENE-DOWN TE-NO	GENE-NO TE-UP	GENE-NO TE-DOWN	GENE-NO TE-NO
2kb_upstream	5/6	NA	52/108	NA	NA	0/8	0/9	NA	389/10324
exon	7/12	NA	4/60	NA	NA	0/13	2/2	NA	127/5362
intron	67/81	NA	1088/2210	NA	NA	0/117	1/14	0/1	13869/185723
3UTR	29/41	NA	62/93	NA	NA	0/5	3/4	NA	496/6031
5UTR	1/1	NA	NA	NA	NA	NA	NA	NA	14/211
IFNpos vs HC									
region	GENE-UP TE-UP	GENE-UP TE-DOWN	GENE-UP TE-NO	GENE-DOWN TE-UP	GENE-DOWN TE-DOWN	GENE-DOWN TE-NO	GENE-NO TE-UP	GENE-NO TE-DOWN	GENE-NO TE-NO
2kb_upstream	15/38	NA	71/264	NA	NA	2/101	7/39	0/8	299/8314
exon	15/63	NA	5/105	NA	0/1	1/31	1/14	0/3	115/4609
intron	399/590	NA	1313/4261	NA	1/12	31/1584	59/212	7/114	11871/158050
3UTR	93/135	NA	44/135	NA	2/5	1/60	14/40	3/13	408/5360
5UTR	1/2	NA	3/11	NA	NA	0/3	0/2	NA	10/180
IFNpos vs IFNneg									
region	GENE-UP TE-UP	GENE-UP TE-DOWN	GENE-UP TE-NO	GENE-DOWN TE-UP	GENE-DOWN TE-DOWN	GENE-DOWN TE-NO	GENE-NO TE-UP	GENE-NO TE-DOWN	GENE-NO TE-NO
2kb_upstream	19/36	NA	63/197	NA	NA	0/59	4/32	1/14	292/8211
exon	14/46	NA	8/136	NA	0/4	0/36	1/10	0/2	113/4523
intron	348/517	NA	1304/3739	0/1	0/15	52/938	56/209	9/97	11591/154637
3UTR	78/114	NA	46/138	NA	0/2	2/35	13/29	0/13	416/5284
5UTR	1/2	NA	1/5	NA	NA	0/1	0/1	NA	14/190

Supplementary Figure 3. Table showing the numbers and percentages of ISGs among total genes with nearby differentially expressed (DE) TEs.

For a graphical depiction of the data see Fig. 4. *Top*, Percentages and *Bottom*, Numbers of ISGs among total genes containing DE TEs for the comparisons “All SLE vs HC”, “IFNpos vs HC”, and “IFNpos vs IFNneg”. The data are computed for TEs at each of the genomic locations shown in Fig. 4A.



Supplementary Figure 4. Computational estimation of the extents of intron retention in SLE PMN relative

to healthy control samples.

A. Mean of intron retention ratio (IRratioMean) from IRFinder of HC samples. **B**. Mean of intron retention ratio (IRratioMean) from IRFinder of SLE samples. **C**. Mean of intron retention ratio (IRratioMean) from IRFinder of IFNneg samples. **D**. Mean of intron retention ratio (IRratioMean) from IRFinder of IFNpos samples. **E**. Differential Intron retention comparing SLE against HC, significant intron retention change is indicated by color. Differentially retained intron in SLE (UP) red; differentially retained intron in HC (DOWN), blue; not differentially retained intron, grey. **F**. Differential Intron retention comparing IFNneg against HC, significant intron retention change is indicated by color. Differentially retained intron in IFNneg (UP) red; differentially retained intron in HC (DOWN), blue; not differentially retained intron, grey. **G**. Differential Intron retention comparing IFNpos against HC, significant intron retention change is indicated by color. Differentially retained intron in IFNpos (UP) red; differentially retained intron in HC (DOWN), blue; not differentially retained intron, grey. **H**. Differential Intron retention comparing IFNpos against IFNneg, significant intron retention change is indicated by color. Differentially retained intron in IFNpos (UP) red; differentially retained intron in IFNneg (DOWN), blue; not differentially retained intron, grey.

A

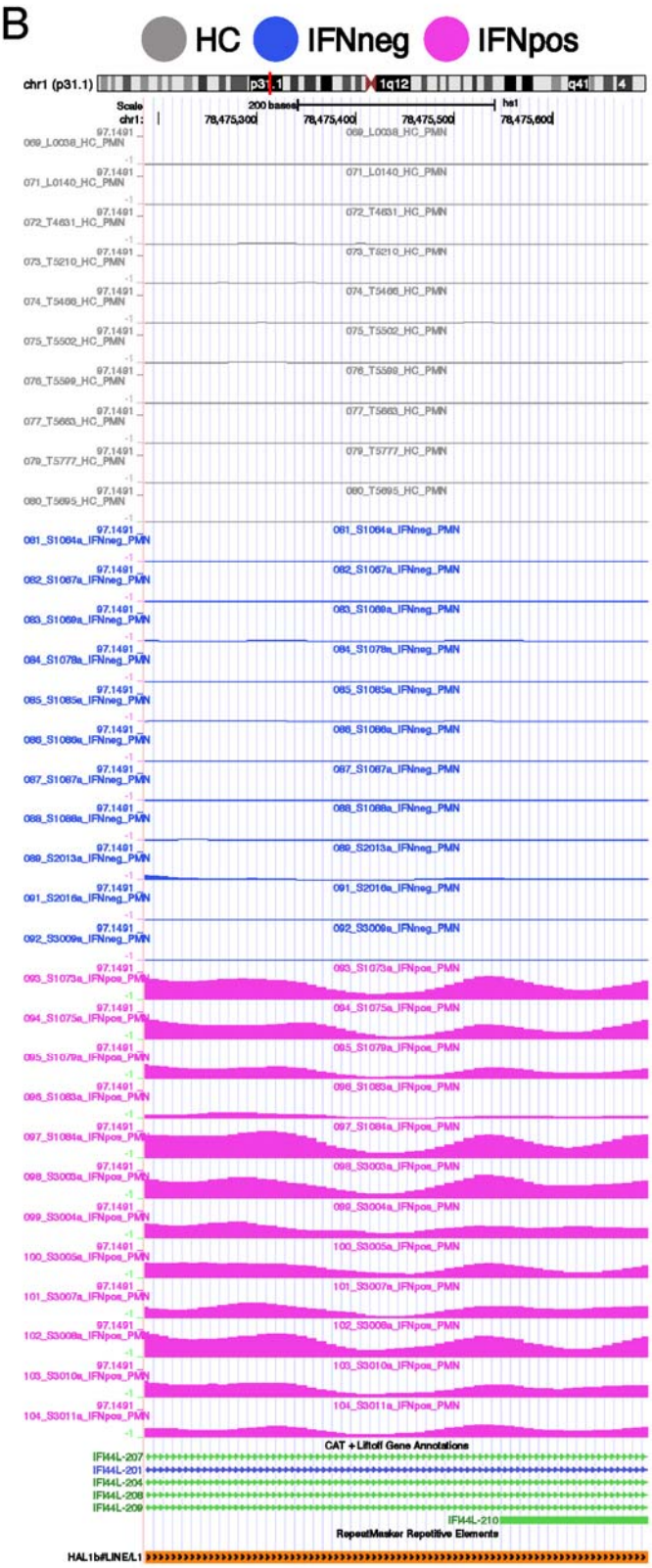
IFNpos vs HC

TE (2 elements)	log2FoldChange	padj	change
L1PBb:L1:LINE	0.953	0.027	UP
HAL1b:L1:LINE	0.66	0	UP

TE (16 total copies)	log2FoldChange	padj	change
L1PBb_dup129	5.705	0	UP
L1PBb_dup36	2.144	NA	NO
L1PBb_dup69	1.945	0.795	NO
L1PBb_dup127	0.968	0.056	NO
L1PBb_dup34	0.286	NA	NO
L1PBb_dup22	0.175	NA	NO
L1PBb_dup128	0.023	0.998	NO
L1PBb_dup38	-0.06	0.991	NO
L1PBb_dup86	-0.065	NA	NO
L1PBb_dup90	-0.594	NA	NO
L1PBb_dup44	-0.843	NA	NO
L1PBb_dup5	-0.865	NA	NO
L1PBb_dup23	-0.902	NA	NO
L1PBb_dup88	-1.387	NA	NO
L1PBb_dup45	-1.766	NA	NO
L1PBb_dup131	-2.308	0.779	NO

TE (279 total copies)	log2FoldChange	padj	change
HAL1b_dup2785	6.815	0	UP
HAL1b_dup433	4.482	0	UP
HAL1b_dup2060	1.988	0.012	UP
HAL1b_dup1477	1.88	0.018	UP
HAL1b_dup2704	1.547	0	UP
HAL1b_dup2374	3.04	NA	NO
HAL1b_dup1540	2.801	NA	NO
HAL1b_dup781	2.493	NA	NO
HAL1b_dup188	2.491	NA	NO
HAL1b_dup427	2.477	NA	NO
HAL1b_dup813	2.392	NA	NO
HAL1b_dup1570	2.284	NA	NO
HAL1b_dup806	2.127	NA	NO
HAL1b_dup310	2.093	NA	NO
HAL1b_dup1101	2.018	NA	NO
HAL1b_dup2675	1.968	NA	NO
HAL1b_dup1325	1.874	NA	NO
HAL1b_dup1452	1.863	0.539	NO

B



Supplementary Figure 5. Differential expression analysis results of L1PBb and HAL1b elements in IFNpos vs HC.

A, *Top*, L1PBb and HAL1b elements upregulated in IFNpos vs HC measured at subfamily level. *Middle*, results of the differential expression analysis of the L1PBb element's copies in IFNpos vs HC, the upregulated copy is highlighted in yellow. *Bottom*, results of the differential expression analysis of the HAL1b element's copies in

IFNpos vs HC (showing only the top expressing copies), the upregulated copies are highlighted in yellow. **B**, Genome browser view of the HAL1b_dup2785 element depicting the signal levels of all samples in the three groups, HC, IFNneg and IFNpos. Each track represents a different individual.

SUPPLEMENTARY MATERIALS

Supplementary Table 1. List of ISGs modified from ref. [46] (also used in refs. [38, 39]).

AUTHOR CONTRIBUTIONS. A.R. and M.R.B. designed and directed the study, K.K. and F.A. provided essential input, H.S. and F.A. supervised the computational analyses, L.J.A.-V., B.V. and H.S. performed the computational analyses using publicly available tools, K.S., K.K. and M.R.B. provided clinical input, and M.R.B. and A.R. wrote the manuscript with input from all authors.

ACKNOWLEDGEMENTS. We thank Q. Li (University of Pennsylvania) for sharing his curated list of ISGs, and G. J. Faulkner (University of Queensland) for critical review of our methodology and results.

FUNDING. This research was supported by NIH grants R35 CA210043, R01 CA247500 and R01 AI128589 (to A.R.) R35 GM128938 (to F.A.) and development funds of the PMI Dept, VUMC (to M.R.B.). H.S. was supported by the Pew Latin-American Fellows Program from The Pew Charitable Trusts, and by a Fellowship from the California Institute for Regenerative Medicine. K.S. received the Tullie and Rickey Families SPARK Awards for Innovations in Immunology at La Jolla Institute.

COMPETING INTERESTS. A.R. is a member of the scientific advisory board of biomodal (formerly Cambridge EpiGenetix, CEGX), Cambridge, UK. All other authors declare no competing interests.

ABBREVIATIONS. SLE, systemic lupus erythematosus; Ab, antibody; IFN interferon; type I IFN, IFN-I; ISGs, interferon signature (or stimulated) genes; TE, transposable elements; PMN, polymorphonuclear leukocytes (neutrophils); HC, healthy controls; LTR, long terminal repeat; NET, neutrophil extracellular trap; ds, double-stranded; ERV, endogenous retrovirus; SINE, short interspersed nuclear element; LINE, long interspersed nuclear element; GWAS, genome-wide association study; GSEA, gene set enrichment analysis; GSVA, gene set variance analysis; NK, natural killer; TNFSF, tumor necrosis factor(-like) secreted factor; BAFF, B cell activating factor; DE, differentially-expressed; DE gene, DEG; UTR, untranslated region; SLEDAI, SLE disease activity index; IR, intron retention.

REFERENCES

1. Tsokos GC, Lo MS, Costa Reis P, Sullivan KE. New insights into the immunopathogenesis of systemic lupus erythematosus. *Nat Rev Rheumatol*. 2016; 12: 716-730.
2. Fava A, Petri M. Systemic lupus erythematosus: Diagnosis and clinical management. *J Autoimmun.* 2019; 96: 1-13.
3. Ameer MA, Chaudhry H, Mushtaq J, Khan OS, Babar M, Hashim T, et al. 2022. An overview of systemic lupus erythematosus (SLE) pathogenesis, classification, and management. *Cureus* 2022; 14: e30330.
4. Lazar S, Kahlenberg JM. Systemic lupus erythematosus: New diagnostic and therapeutic approaches. *Annu Rev Med*. 2023; 74: 339-352.
5. Suurmond J, Atisha-Fregoso Y, Marasco E, Barlev AN, Ahmed N, Calderon SA, et al. Loss of an IgG plasma cell checkpoint in patients with lupus. *J Allergy Clin Immunol*. 2019; 143: 1586-1597.
6. Olsen NJ, Karp DR. Autoantibodies and SLE: the threshold for disease. Nat Rev Rheumatol. 2014;. 10: 181-186.

7. Garcia-Romo GS, Caielli S, Vega B, Connolly J, Allantaz F, Xu Z, Punaro M, Baisch J, Guiducci C, Coffman RL, Barrat FJ, Banchereau J, Pascual V. Netting neutrophils are major inducers of type I IFN production in pediatric systemic lupus erythematosus. *Sci Transl Med.* 2011; 3: 73ra20.
8. Tsai CY, Li KJ, Hsieh SC, Liao HT, Yu CL. What's wrong with neutrophils in lupus? *Clin Exp Rheumatol.* 2019; 37: 684-693.
9. Wang M, Ishikawa T, Lai Y, Nallapothula D, Singh RR. Diverse roles of NETosis in the pathogenesis of lupus. *Front Immunol.* 2022; 13: 895216.
10. Caielli S, Wan Z, Pascual V. Systemic lupus erythematosus pathogenesis: Interferon and beyond. *Annu Rev Immunol*. 2023; 41: 533-560.
11. Rönnblom L, Alm GV. An etiopathogenic role for the type I IFN system in SLE. *Trends Immunol*. 2001; 22: 427-431.
12. Rönnblom L, Eloranta ML, Alm GV. The type I interferon system in systemic lupus erythematosus. *Arthritis Rheum*. 2006; 54: 408-420.
13. Banchereau J, Pascual V. 2006. Type I interferon in systemic lupus erythematosus and other autoimmune diseases. *Immunity* 2006; 25: 383-392.
14. Rönnblom L, Eloranta M-L. The interferon signature in autoimmune diseases. *Curr Opin Rheumatol*. 2013; 25: 248–253.
15. Arbuckle MR, McClain MT, Rubertone MV, Scofield RH, Dennis GJ, James JA, Harley JB. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N Engl J Med.* 2003; 349: 1526-1533.
16. Lu R, Munroe ME, Guthridge JM, Bean KM, Fife DA, Chen H, Slight-Webb SR, Keith MP, Harley JB, James JA. Dysregulation of innate and adaptive serum mediators precedes systemic lupus erythematosus classification and improves prognostic accuracy of autoantibodies. *J Autoimmun*. 2016; 74: 182-193.
17. Munroe ME, Lu R, Zhao YD, Fife DA, Robertson JM, Guthridge JM, Niewold TB, Tsokos GC, Keith MP,Harley JB, James JA. Altered type II interferon precedes autoantibody accrual and elevated type I interferon activity prior to systemic lupus erythematosus. *Ann Rheum Dis.* 2016; 75: 2014-2021.
18. Nakano M, Ota M, Takeshima Y, Iwasaki Y, Hatano H, Nagafuchi Y, Itamiya T, Maeda J, Yoshida R, Yamada S, Nishiwaki A, Takahashi H, Takahashi H, Akutsu Y, Kusuda T, Suetsugu H, Liu L, Kim K, Yin X, Bang SY, Cui Y, Lee HS, Shoda H, Zhang X, Bae SC, Terao C, Yamamoto K, Okamura T, Ishigaki K, Fujio K. Distinct transcriptome architectures underlying lupus establishment and exacerbation. *Cell* 2022; 185: 3375-3389.e21.
19. Brooks W. An epigenetics-based hypothesis of autoantigen development in systemic lupus erythematosus. *Epigenomes* 2020; 4: 6.
20. Vecellio M, Wu H, Lu Q, Selmi C. The multifaceted functional role of DNA methylation in immune-mediated rheumatic diseases. *Clin Rheumatol*. 2021; 40: 459-476.
21. Yu B, Qi Y, Li R, Shi Q, Satpathy AT, Chang HY. B cell-specific XIST complex enforces X-inactivation and restrains atypical B cells. *Cell* 2021; 184: 1790-1803.e17.
22. Pyfrom S, Paneru B, Knox JJ, Cancro MP, Posso S, Buckner JH, Anguera MC. The dynamic epigenetic regulation of the inactive X chromosome in healthy human B cells is dysregulated in lupus patients. *Proc Natl Acad Sci USA* 2021; 118: e2024624118.
23. Huret C, Ferrayé L, David A, Mohamed M, Valentin N, Charlotte F, Savignac M, Goodhardt M, Guéry JC, Rougeulle C, Morey C. Altered X-chromosome inactivation predisposes to autoimmunity. *Sci Adv.* 2024; 10: eadn6537.
24. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* 2016; 17: 487-500. PMID: 27346641.
25. Newman LE, Shadel GS. Mitochondrial DNA release in Innate Immune Signaling. *Annu Rev Biochem* 2023; 92: 299-332.

26. Friedli M, Trono D. The developmental control of transposable elements and the evolution of higher species. *Annu Rev Cell Dev Biol*. 2015; 31: 429-51.

27. Kazazian HH, Moran JV. Mobile DNA in health and disease. *New Engl J Med* 2017; 377: 361-370

28. Fueyo R, Judd J, Feschotte C, Wysocka J. Roles of transposable elements in the regulation of mammalian transcription. *Nat Rev Mol Cell Biol* 2022; 23: 481-497.

29. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* 2017; 18: 71-86.

30. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, Waki K, Hornig N, Arakawa T, Takahashi H, Kawai J, Forrest AR, Suzuki H, Hayashizaki Y, Hume DA, Orlando V, Grimmond SM, Carninci P. The regulated retrotransposon transcriptome of mammalian cells. *Nat Gen* 2009; 41: 563-571. PMID: 19377475

31. Bruno M, Mahgoub M, MacFarlan TS. The arms race between KRAB-zinc finger proteins and endogenous retroelements and its impact on mammals. *Annu Rev Genet* 2019; 53: 393-416

32. Burns KH. Our conflict with transposable elements and Its Implications for human disease. *Annu Rev Pathol* 2020; 15: 51-70.

33. Hoyt SJ, Storer JM, Hartley GA, Grady PGS, Gershman A, de Lima LG, et al. From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. *Science* 2022; 376: eabk3112.

34. Seczynska M, Lehner PJ. The sound of silence: mechanisms and implications of HUSH complex function. *Trends Genet* 2023; 39: 251-267

35. Janssen A, Colmenares SU, Karpen GH. Heterochromatin: Guardian of the genome. *Annu Rev Cell Devel Biol* 2018; 34: 265-288

36. López-Moyado IF, Tsagaratou A, Delatte B, Heinz S, Benner C, Rao A. Paradoxical association of TET loss of function with genome-wide hypomethylation. *Proc Natl Acad Sci USA* 2019; 116: 16933-16942

37. López-Moyado IF*, Ko MG*, Hogan PG, Rao A. TET enzymes in the immune system: from DNA demethylation to immunotherapy, inflammation and cancer. *Annu Rev Immunol* 2024; 42: 455-488

38. Dong Z, Sepulveda H, Arteaga L, Blouin C, Fernandez J, Binder M, Chou W-C, Tien H-F, Manriquez C, Nowak R, Tien HF, Faulkner GJ, Patnaik M, Myers SA, Rao A. A mutant ASXL1-EHMT complex contributes to heterochromatin dysfunction in clonal hematopoiesis and chronic monomyelocytic leukemia. PNAS 2024; *in press*. bioRxiv preprint doi: https://doi.org/10.1101/2024.01.30.578015.

39. Sepulveda H*, Li X*, Arteaga-Vasquez LJ, López-Moyado IF, Brunelli M, Hernandez Espinosa L, Yue X, Angel JC, Brown C, Dong Z, Jansz N, Puddu F, Modat A, Scotcher J, Kennedy P, Manriquez C, Myers SA, Creed P, Crawford R, Faulkner GJ, Rao A. OGT prevents DNA demethylation and suppresses the expression of transposable elements in heterochromatin by restraining TET activity genome-wide. *Nat Struct Mol Biol* 2024; *accepted for publication*. bioRxiv preprint doi: https://doi.org/10.1101/2024.01.31.578097.

40. Arribas YA, Baudon B, Rotival M, Suárez G, Bonté PE, Casas V, Roubert A, Klein P, Bonnin E, Mchich B, Legoix P, Baulande S, Sadacca B, Diharce J, Waterfall JJ, Etchebest C, Carrascal M, Goudot C, Quintana-Murci L, Burbage M, Merlotti A, Amigorena S. Transposable element exonization generates a reservoir of evolving and functional protein isoforms. *Cell* 2024: S0092-8674(24)01328-X. *online ahead of print*.

41. Pasquesi GIM, Allen H, Ivancevic A, Barbachano-Guerrero A, Joyner O, Guo K, Simpson DM, Gapin K, Horton I, Nguyen LL, Yang Q, Warren CJ, Florea LD, Bitler BG, Santiago ML, Sawyer SL, Chuong EB. Regulation of human interferon signaling by transposon exonization. *Cell* 2024: S0092-8674(24)01333-3. *online ahead of print*.

42. Li T, Chen ZJ. The cGAS-cGAMP-STING pathway connects DNA damage to inflammation, senescence and cancer. *J Exp Med* 2018; 5: 1287-1299

43. Ablasser A, Chen ZJ. cGAS in action: expanding roles in immunity and inflammation. *Science* 2019; 363; eaat8657

44. Zhang X, Bai XC, Chen ZJ. Structures and mechanisms in the cGAS-STING innate immunity pathway. *Immunity* 2020; 53: 43-53

45. Ahmad S, Mu X, Yang F, Greenwald E, Park JW, Jacob E, Zhang C-Z, Hur S. Breaching self-tolerance to Alu duplex RNA underlies MDA5-mediated inflammation. *Cell* 2018; 172: 797-810.e13.

46. Li Q, Gloudemans MJ, Geisinger JM, Fan B, Aguet F, Sun T, Ramaswami G, Li YI, Ma JB, Pritchard JK, Montgomery SB, Li JB. RNA editing underlies genetic risk of common inflammatory diseases. *Nature* 2022; 608; 570-577

47. Bentham J, Morris DL, Graham DSC, Pinder CL, Tombleson P, Behrens TW, et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet* 2015; 47: 1457-1464

48. Crow MK. Pathogenesis of systemic lupus erythematosus: risks, mechanisms and therapeutic targets. *Ann Rheum Dis* 2023; 999-1014

49. Ghodke-Puranik Y, Olferiev M, Crow MK. Systemic lupus erythematosus: insights into pathogenesis and implications for therapy. *Nat Rev Rheumatol* 2024; 20: 635-648

50. Mustelin T, Lood C, Giltiay NV. Sources of pathogenic nucleic acids in systemic lupus erythematosus. *Front Immunol.* 2019; 10: 1028.

51. Ukadike KC, Mustelin T. Implications of endogenous retroelements in the etiopathogenesis of systemic lupus erythematosus. *J Clin Med* 2021; 10: 856

52. Ukadike KC Najjar R, Ni K, Laine A, Wang X, Bays A, Taylor MS, LaCava J, Mustelin T. Expression of L1 retrotransposons in granulocytes from patients with active systemic lupus erythematosus. *Mobile DNA* 2023; 14: 5

53. Carter V, LaCava J, Taylor MS, Liang SY, Mustelin C, Ukadike KC, Bengtsson A, Lood C, Mustelin T. High prevalence and disease correlation of autoantibodies against p40 encoded by Long Interspersed Nuclear Elements in systemic lupus erythematosus. *Arthritis Rheumatol.* 2020; 72: 89-99.

54. Treger RS, Pope SD, Kong Y, Tokuyama M, Taura M, Iwasaki A. The lupus susceptibility locus Sgp3 encodes the suppressor of endogenous retrovirus expression SNERV. *Immunity* 2019; 50: 334-347.e9

55. Tokuyama M, Kong Y, Song E, Jayewickreme T, Kang I, Iwasaki A. ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. *Proc Natl Acad Sci USA* 2018; 115: 12565-12572.

56. Tokuyama M, Gunn BM, Venkataraman A, Kong Y, Kang I, Rakib T, Townsend MJ, Costenbader KH, Alter G, Iwasaki A. Antibodies against human endogenous retrovirus K102 envelope activate neutrophils in systemic lupus erythematosus. *J Exp Med.* 2021; 218: e20191766.

57. Panwar B, Schmiedel BJ, Liang S, White B, Rodriguez E, Kalunian K, et al. Multi-cell type gene coexpression network analysis reveals coordinated interferon response and cross-cell type correlations in systemic lupus erythematosus. *Genome Res* 2021; 31: 659-676.

58. Mackay F, Woodcock SA, Lawton P, Ambrose C, Baetscher M, Schneider P, Tschopp J, Browning JL. Mice transgenic for BAFF develop lymphocytic disorders along with autoimmune manifestations. J *Exp Med* 1999; 190: 1697-1710.

59. Melchers F. Actions of BAFF in B cell maturation and its effects on the development of autoimmune disease. *Ann Rheum Dis.* 2003; 62 Suppl 2(Suppl 2): ii25-27.

60. Hubbard EL, Bachali P, Kingsmore KM, He Y, Catalina MD, Grammer AC, Lipsky PE. Analysis of transcriptomic features reveals molecular endotypes of SLE with clinical implications. *Genome Med.* 2023; 15: 84.

61. Sun X, Liu Z, Li Z, Zeng Z, Peng W, Zhu J, Zhao J, Zhu C, Zeng C, Stearrett N, Crandall KA, Bachali P, Grammer AC, Lipsky PE. Abnormalities in intron retention characterize patients with systemic lupus erythematosus. *Sci Rep.* 2023; 13: 5141.

62. Xiao L, Zhan F, Lin S. Clinical values of the identified hub genes in systemic lupus erythematosus. *Front Immunol* 2022; 13: 844025

63. Cutts Z, Patterson S, Maliskova L, Taylor KE, Ye CJ, Dall'Era M, et al. Cell-specific transposable element and gene expression analysis across systemic lupus erythematosus phenotypes. *ACR Open Rheum* 2024; 6: 769-779

64. Antiochos B, Paz M, Li J, Goldman DW, Petri M, Darrah E, Cashman K, Sanz I, Burns KH, Ardeljan D, Andrade F, Rosen A. Autoantibodies targeting LINE-1-encoded ORF1p are associated with systemic lupus erythematosus diagnosis but not disease activity. *Clin Exp Rheumatol* 2022; 40: 1636-1641

65. Domeier PP, Chodisetti SB, Schell SL, Kawasawa YI, Fasnacht MJ, Soni C, Rahman ZSM. B-cell-intrinsic type 1 interferon signaling is crucial for loss of tolerance and the development of autoreactive B cells. *Cell Rep.* 2018; 24: 406-418.

66. Liu Y, Kaplan MJ. Neutrophil dysregulation in the pathogenesis of systemic lupus erythematosus. *Rheum Dis Clin North Am*. 2021; 47: 317-333.

67. Ma Y, Wang M, Jia J, Meng J, Teng J, Zhu D, Shi H, Sun Y, Su Y, Liu H, Cheng X, Ye J, Chi H, Liu T, Chen X, Wan L, Zhou Z, Wang F, He D, Yang C, Hu Q. Enhanced type I interferon signature induces neutrophil extracellular traps enriched in mitochondrial DNA in adult-onset Still's disease. *J Autoimmun*. 2022; 127: 102793.

68. Baker T, Sharifian H, Newcombe PJ, Gavin PG, Lazarus MN, Ramaswamy M, White WI, Ferrari N, Muthas D, Tummala R, Morand EF, Furie RA, Vital EM, Chamberlain C, Platt A, Al-Mossawi H, Brohawn PZ, Csomor E. Type I interferon blockade with anifrolumab in patients with systemic lupus erythematosus modulates key immunopathological pathways in a gene expression and proteomic analysis of two phase 3 trials. *Ann Rheum Dis* 2024; 83: 1018-1027

69. Perdomo J, Leung HHL. Immune thrombosis: exploring the significance of immune complexes and NETosis. *Biology* 2023; 12: 1332

70. Blanco LP, Wang X, Carlucci PM, Torres-Ruiz JJ, Romo-Tena J, Sun HW, Hafner M, Kaplan MJ. RNA externalized by neutrophil extracellular traps promotes inflammatory pathways in endothelial cells. *Arthritis Rheumatol.* 2021; 73: 2282-2292.

71. Iñiguez LP, de Mulder Rougvie M, Stearrett N, Jones RB, Ormsby CE, Reyes-Terán G, Crandall KA, Nixon DF, Bendall ML. Transcriptomic analysis of human endogenous retroviruses in systemic lupus erythematosus. *Proc Natl Acad Sci USA* 2019; 116: 21350-21351.

72. Chen YG, Hur S. Cellular origins of dsRNA, their recognition and consequences. *Nat Rev Cell Mol Biol* 2022; 23: 286-301

73. Dias Junior AG, Sampaio NG, Rehwinkel J. A balancing act: MDA5 in antiviral immunity and autoinflammation. Trends Microbiol. 2019; 27: 75-85.

74. Lund JM, Alexopoulou L, Sato A, Karow M, Adams NC, Gale NW, Iwasaki A, Flavell RA. Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc Natl Acad Sci USA* 2004; 101: 5598-5603

75. Morel L, Croker BP, Blenman KR, Mohan C, Huang G, Gilkeson G, Wakeland EK. Genetic reconstitution of systemic lupus erythematosus immunopathology with polycongenic murine strains. *Proc Natl Acad Sci USA* 2000; 97: 6670-6675

76. Gorbunova V, Seluanov A, Mita P, McKerrow W, Fenyö D, Boeke JD, Linker SB, Gage FH, Kreiling JA, Petrashen AP, Woodham TA, Taylor JR, Helfand SL, Sedivy JM. The role of retrotransposable elements in ageing and age-associated diseases. *Nature* 2021; 596: 43-53.

77. Copley KE, Shorter J. Repetitive elements in aging and neurodegeneration. *Trends Genet.* 2023; 39: 381-400.

78. Liang Y, Qu X, Shah NM, Wang T. Towards targeting transposable elements for cancer therapy. *Nat Rev Cancer.* 2024; 24: 123-140.

79. Snowbarger J, Koganti P, Spruck C. Evolution of repetitive elements, their roles in homeostasis and human disease, and potential therapeutic applications. *Biomolecules* 2024; 14: 1250.

80. Higgs BW, Liu Z, White B, Zhu W, White WI, Morehouse C, et al. Patients with systemic lupus erythematosus, myositis, rheumatoid arthritis and scleroderma share activation of a common type I interferon pathway. *Ann Rheum Dis*. 2011; 70**:** 2029–2036.

81. Mustelin T, Ukadike KC. How retroviruses and retrotransposons in our genome may contribute to autoimmunity in rheumatological conditions. *Front Immunol.* 2020; 11**:** 593891.

82. Kuriyama Y, Shimizu A, Kanai S, Oikawa D, Motegi SI, Tokunaga F, Ishikawa O. Coordination of retrotransposons and type I interferon with distinct interferon pathways in dermatomyositis, systemic lupus erythematosus and autoimmune blistering disease. *Sci Rep*. 2021; 11: 23146.

83. Baldini C, Fulvio G, La Rocca G, Ferro F. Update on the pathophysiology and treatment of primary Sjögren syndrome. *Nat Rev Rheumatol.* 2024; 20: 473-491.

84. Khadjinova AI, Wang X, Laine A, Ukadike K, Eckert M, Stevens A, et al. Autoantibodies against the envelope proteins of endogenous retroviruses K102 and K108 in patients with systemic lupus erythematosus correlate with active disease. *Clin Exp Rheumatol.* 2022; 40: 1306-1312.

85. Wang X, Hefton A, Ni K, Ukadike KC, Bowen MA, Eckert M, Stevens A, Lood C, Mustelin T. Autoantibodies against unmodified and citrullinated Human Endogenous Retrovirus K envelope protein in patients with rheumatoid arthritis. *J Rheumatol.* 2022; 49: 26-35.

86. Neidhart M, Rethage J, Kuchen S, Künzler P, Crowl RM, Billingham ME, Gay RE, Gay S. Retrotransposable L1 elements expressed in rheumatoid arthritis synovial tissue: association with genomic DNA hypomethylation and influence on gene expression. *Arthritis Rheum.* 2000; 43: 2634-2647.

87. Mavragani CP, Sagalovskiy I, Guo Q, Nezos A, Kapsogeorgou EK, Lu P, Liang Zhou J, Kirou KA, Seshan SV, Moutsopoulos HM, Crow MK. Expression of Long Interspersed Nuclear Element 1 retroelements and induction of type I interferon in patients with systemic autoimmune disease. *Arthritis Rheumatol.* 2016; 68: 2686-2696.

88. Lombardi A, Tsomos E, Hammerstad SS, Tomer Y. Interferon alpha: The key trigger of type 1 diabetes. *J Autoimmun*. 2018; 94: 7-15.

89. Samassa F, Holtzmann C, Mallone R. Cross Talk between β Cells and Immune Cells: What We Know, What We Think We Know, and What We Should Learn. *Cold Spring Harb Perspect Med*. 2024: a041604. doi: 10.1101/cshperspect.a041604. *Online ahead of print.*

90. Krueger F, James F, Ewels P, Afyounian E, Schuster-Boeckler B. FelixKrueger/TrimGalore: v0.6.7 - DOI via Zenodo. Zenodo; 2021.

91. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; 29: 15-21.

92. Putri GH, Anders S, Pyl PT, Pimanda JE, Zanini F. Analysing high-throughput sequencing data in Python with HTSeq 2.0, *Bioinformatics* 2022; 38: 2943–2945. https://doi.org/10.1093/bioinformatics/btac166

93. Jin Y, Tam OH,Paniagua E, Hammell M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets, *Bioinformatics* 2015; 31: 3593–3599. https://doi.org/10.1093/bioinformatics/btv422

94. Bendall ML, de Mulder M, Iñiguez LP, Lecanda-Sánchez A, Pérez-Losada M, Ostrowski MA, et al. Telescope: Characterization of the retrotranscriptome by accurate estimation of transposable element expression. *PLoS Comput Biol* 2019; 15: e1006453.

95. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

96. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15: 550.

97. Aaron R. Quinlan, Ira M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics* 2010; 26: 841–842. https://doi.org/10.1093/bioinformatics/btq033

98. Middleton, R., Gao, D., Thomas, A. et al. IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol.* 2017; 18, 51.

99. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., Yu, G. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* 2021; 2(3): 100141. https://doi.org/10.1016/j.xinn.2021.100141

100. Dolgalev I. msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. R package version. 2022; 7.5.1.9001, https://igordot.github.io/msigdbr/

101. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data." *BMC Bioinformatics* 2013; 14, 7.

102. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 2016; 44: W160-5

103. Zerbino DR, Johnson N, Juettemann T, Wilder SP and Flicek PR: WiggleTools: parallel processing of large collections of genome-wide datasets for visualization and statistical analysis. *Bioinformatics* 2014; 30:1008-1009.

104. Perez P, Barber GP, Penet-Pages A, Casper J, Clawson H, Diekhans M, Fischer C, Navarro Gonzalez J, Hinrichs AS, Lee CM, Nassar LR, Raney BJ, Speir ML, van Baren MJ, Vaske CJ, Haussler D, Kent WJ, Haeussler M. The UCSC Genome Browser database: 2025 update. Nucl Acids Res 2024