Title

Integrative Precision Medicine Approach to Dissect Patient Heterogeneity in Systemic Lupus Erythematosus

Author

Cutts, Zachary

Publication Date

2024

Peer reviewed|Thesis/dissertation

Integrative Precision Medicine Approach to Dissect Patient Heterogeneity in Systemic Lupus Erythematosus

by
Zachary Cutts

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biological and Medical Informatics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Marina Sirota
925B61AB9C41499...

Marina Sirota
Chair

DocuSigned by:
John A. Capra
[illegible]AC...

John A. Capra

DocuSigned by:
Gabriela K. Fragiadakis
6FA7C49A0B43464...

Gabriela K. Fragiadakis

Committee Members

For my parents, Matthew Cutts and Ellen Wrynn

Acknowledgements

To my parents, who gave their unconditional love and support during my PhD and throughout my entire life. Especially my father, who always wanted a PhD, and could have easily completed one. Even though you have been gone for years, your memory has given me the motivation and encouragement necessary to finish this degree. Without you I would have never achieved this. To my mother, thank you for always seeing the best in me and always believing I could achieve great things. To my siblings who were always available during difficult and stressful times, thank you. I also thank my advisor, Marina Sirota, whose support has been crucial to successfully completing this PhD. In addition, I wish to thank my thesis committee, Tony Capra and Gabriela Fragiadakis, for their valuable scientific collaboration, mentorship, and belief in me and my work, which was instrumental to my success. I also wish to thank the Sirota lab. I am so grateful to have met and worked with all of you and I am lucky to call you all my friends. Also critical to my mental health, my dog Coco, who was always happy to see me regardless of how school was going. Finally, to Margot, thank you for being an amazing partner all these years.

Contributions

Basis for chapter 2:

Zachary Cutts, Sarah Patterson, Lenka Maliskova, Kimberly E. Taylor, Chun Ye, Maria Dall'Era, Jinoos Yazdany, Lindsey Criswell, Gabriela K. Fragiadakis, Charles Langelier, John A. Capra, Marina Sirota*, Cristina M Lanata*. Cell-Specific Transposable Element Gene Expression Analysis Identifies Associations with Systemic Lupus Erythematosus. *bioRxiv*. November 28, 2024. https://doi.org/10.1101/2023.11.27.567477

* Co-corresponding authors

Integrative Precision Medicine Approach to Dissect Patient Heterogeneity in Systemic Lupus Erythematosus

Zachary Cutts

Abstract

Autoimmune diseases arise from dysregulation of the immune system, leading to its attack on the body's own tissues and organs. The clinical heterogeneity of these diseases arises from several sources, such as genetic predisposition, environmental triggers, and aberrant immune responses. One emerging area of interest is the role of transposable elements (TEs) in autoimmune disease pathogenesis because these self-nucleic acids can be mistakenly detected as foreign, which can trigger a chronic immune reaction.

There is growing appreciation for the role of TEs in systemic lupus erythematosus (SLE) and studies have found differentially expressed TEs in SLE patients, which suggests a link between TE activity and disease mechanisms. Our work investigated TE expression in four immune cell types from SLE patients, revealing cell-specific and SLE subphenotype-specific differentially expressed TEs, with additional cell-type-specific TE associations in different ancestry groups. TE expression was also associated with host gene expression involved in antiviral and immune responses, supporting the hypothesis that TEs could activate the innate immune system and contribute to chronic inflammation and autoimmunity.

This study underscores the importance of TEs in SLE heterogeneity and highlights the need for further exploration of TE expression in normal immune cells and functional studies to understand their role in SLE pathogenesis. Future work to study whether antiretroviral drugs could reduce expression of TEs and mitigate SLE symptoms is warranted, given the potential involvement of TEs in autoimmune disease pathogenesis.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Contributions to Autoimmune Disease

Autoimmune diseases occur when the immune system fails to distinguish self and non-self, and they are often characterized by abnormal T and B cell reactivity[1–3]. Autoimmune diseases affect 3-5% of the population, with autoimmune thyroid disease and type I diabetes being the most common among nearly 100 autoimmune diseases[1]. The manifestations are very heterogenous, and can involve any organ system, be systemic or specific to an organ, and occur more commonly in women[3].

The core of autoimmune pathogenesis is the failure of immune tolerance, the process that normally prevents immune cells from attacking the body's own tissues. In autoimmune diseases, this breakdown leads to an inappropriate immune response against self-antigens. The exact mechanism can vary, involving both central tolerance (thymic deletion of highly self-reactive T cells) and peripheral tolerance (regulation of self-reactive cells that escape central tolerance). The breakdown of these mechanisms can be triggered by genetic defects or environmental factors, leading to the persistent activation of autoreactive T cells and the production of autoantibodies.

Autoantibodies

Common manifestations are the presence of autoantibodies, which may appear years before any sign of disease[3]. Autoantibodies are used for diagnosis and classification of autoimmune diseases; however, the link between the autoantibody and pathology is sometimes unclear. For example, in cutaneous fibrosis autoantibodies to nuclear antigens DNA topoisomerase centromere proteins are associated with diffuse skin involvement and limited skin disease, respectively.

Autoantibodies can also influence disease through antigen binding sites or immune complex formation. Through binding sites autoantibodies can alter cell function by binding with cell surface receptors. For example, in Graves' disease autoantibodies to the thyrotropin receptor activate the receptor, whereas in Myasthenia gravis autoantibodies to the acetylcholine receptor deactivate the receptor[3]. Immune complexes are formed through the binding of an autoantibody to its antigen, typically thought to occur in blood, then deposit in tissue causing inflammation and tissue damage[3].

The exact cause of autoimmunity is not well understood, but is thought to arise from an interplay of genetic, environmental, and immunological factors[4]. Rarely, autoimmunity can result from single-gene mutations, but commonly results from complex multiple gene interactions of immune cells[3].These triggers are thought to include infections, drugs, or other environmental factors like sunlight or chemicals.

Genetic Factors

More than 200 genetic loci have been found via GWAS to be associated with autoimmune disorders. Although there are some examples of monogenic autoimmune diseases, most autoimmune diseases are thought to occur due to multiple genetic factors[1]. For example, autoimmune polyendocrinopathy syndrome type 1 (APS1) is a rare aggressive autoimmune disease caused by a mutation in the *AIRE* gene. This gene affects self-antigen presentation during negative selection against self-reactive T cells in thymus[1,5]. Additionally, in immunodysregulation polyendocrinopathy enteropathy X-linked syndrome (IPEX) a defect in the *Foxp3* gene results in loss of peripheral tolerance do to mutations in CD4+CD25+ Tregs[1].

Many studies have found associations with the major histocompatibility complex (MHC). MHC encodes genes involved in antigen presentation essential for self non-self discrimination.

Additionally, human leucocyte antigens (HLAs) are gene products of MHC, linked to autoimmune disease such as type 1 diabetes, SLE, rheumatoid arthritis and psoriasis[1].

Interestingly many loci have been identified across many different autoimmune diseases suggesting a genetic commonality among autoimmune diseases[1,6]. For instance coeliac disease and rheumatoid arthritis as well as coeliac disease and type 1 diabetes share more than 50% of their loci [1,6]. Additionally, there are also overlap in genes, such as *PTPN22*, which is found in type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus, Graves' disease, and Crohns' disease. *IRF5-TNPO3* found in rheumatoid arthritis, systemic lupus erythematosus, primary biliary cholangitis, ulcerative colitis, Sjogren's syndrome. PTPN22 inhibits T-cell activation, and enhances signaling to promote type I interferon production. While *IRF5-TNPO3* is involved in failure to delete autoreactive native T cells and development of macrophage polarization[1]. Interestingly, those with autoimmune disease are more likely to develop additional autoimmune diseases, for example psoriasis patients have increased risk for developing additional autoimmune diseases, which also suggests a genetic commonality among autoimmune diseases[6].

Environmental Factors

In addition to genetic predisposition, it's thought that environmental factors contribute significantly to the onset of autoimmune diseases. Infections, particularly those caused by viruses have been implicated in the exacerbation and initiation of autoimmune responses. For example elevated type I interferon and IFN-I induced gene transcripts and proteins in blood and tissue of patients with systemic autoimmune diseases including primary Sjogren's syndrome, systemic sclerosis, dermatomyositis and SLE point to a viral trigger for these diseases, however one has not been identified[7].

Viruses in autoimmune disease

Specific viruses have been linked to autoimmune diseases, for example Epstein-Barr virus (EBV), B19V, retroviruses (RVs) and cytomegalovirus (CMV) have been linked to SLE, although the evidence for EBV is the strongest[4]. Elevated EBV viral loads have been found in peripheral B cells and in PBMCs of SLE patients and acute B19V infections can mimic SLE symptoms such as facial rash, photosensitivity for a short time[4].

Pathogens can exert strong pressure on human evolution, for example *Yersinia pestis*, which caused the Black Death and caused up to 200 million deaths, selected variants that enhanced immunity at the cost of increased susceptibility to autoimmune diseases such as Crohn's[8]. Additionally, a variant that increases risk to Crohn's disease protects against HIV progression and correlates with Human Leukocyte Antigen- C expression[9]. There seems to be a balance between the immune system needs to minimize the sensing of endogenous nucleic acids which could cause autoimmunity but also keep antiviral defense strong which relies upon nucleic acid sensing[9]. Put another way, the same pathways that protect against viral infection are thought to cause systemic autoimmunity when signaling is excessive[9].

Many hypotheses exist for how viral infections can lead to autoimmune diseases such as molecular mimicry, persistent infection, and innate immune activation. Molecular mimicry refers to the phenomenon where bacterial or viral antigens share structural similarities with self-antigens. This can mislead the immune system into attacking the body's own tissues. Molecular mimicry has long been studied for induction of autoimmune disease by viruses. For examples one of the first experiments demonstrated cross reactivity was observed in myelin proteins when a portion of hepatitis B virus polymerase was injected in rabbits[10].

Persistent viral infection can also lead to chronic inflammation and T and B cell proliferation, leading to autoantibody production[4]. Finally, innate immunity activation, where

viral DNA/RNA bind to different pattern recognition receptors which can initiate pathways leading to a type I IFN response[4].

This balance between self and non-self-recognition can be difficult for viruses, because nucleic acid sensors must discriminate between self and non-self nucleic acids[11]. Pattern recognition receptors such as toll like receptors (TLRs), RIG-I like receptors (RLRs), and cyclic GMP-AMP synthase (cGAS) are the receptors responsible for this discrimination [4,11]. Pattern recognition receptors can be either in the cytosol or transmembrane region. In the transmembrane region TLRs detect pathogen-associated molecular patterns (PAMPs), in the cytosol RIG-I-like receptors, NOD-like receptors, C-type lectin receptors, and DNA sensors recognize intracellular PAMPs[12]. Once activated pro-inflammatory molecules (cytokines, chemokines, type I interferons) are produced which create an antimicrobial environment and promote development of adaptive immunity such as clonal expansion of T cells and antibody production of B cells[12]. Interestingly TLR-7 and TLR-9 expression in B cells have also been associated with production of anti-dsDNA and anti-RNP autoantibodies, typical of SLE[4].

Another hypothesis called epitope spreading is thought to occur when following the initial immune response, the immune response broadens to target additional self-antigens not initially involved in the disease. For example in autoimmune diseases, its thought that the body responds first to an infection but a secondary response is initiated to endogenous epitopes[10,13]. Virus-epitope specific T cells become activated once a virus is recognized, however surrounding self-antigens can be released and mistakenly the T cell response to self can be initiated[10]. For example, in a mouse model for multiple sclerosis it was shown that a chronic viral infection led to a T-cell response to myelin epitopes[10].

1.2 Transposable Elements

HERVs belong to a larger class of elements called transposable elements (TEs). TEs are mobile DNA segments that can replicate independently within host genomes. TEs comprise a over 40% of human genomes[14]. TEs are classified based on their mechanism of transposition. Class I - retrotransposons which replicate through an RNA intermediate that is reverse-transcribed and integrated into the genome and Class II - DNA transposons which replicate through a cut-and paste approach where the element is physically excised from one place and inserted into another[15]. These two classes are further subdivided into subclasses, then superfamilies and families based on replication method and phylogenetics respectively[16,17].

Class I retrotransposons can be categorized into subclasses based on how they replicate and integrate, long terminal repeat (LTR)/human endogenous retroviral elements and non-LTR elements. HERV structure is similar to exogenous retroviruses, composed of two long terminal repeats (LTRs) that flank the gag, pro-pol, and env internal viral genes[12]. Gag codes structural components of matrix, capsid, and nucleocapsid, pol codes for integrase, reverse transcriptase, and RNAse, pro codes for protease, env codes for envelope and transmembrane subunits[18].

In addition to the protein-coding genes, HERVs contain a primer-binding site (PBS), which binds cellular tRNA and primes the synthesis of (-) strand DNA. The PBS sequence has been used for classification, and tRNA type for different HERV groups' PBS. For example, HERV-K for lysine tRNA, HERV-W for typtophan[12]. Additionally, HERVs are broadly divided into three classes based on similarity to exogenous viruses, class I gammaretrovirus/epsilonretrovirus-like, class II betaretrovirus-like, and class III spumaretrovirus-like.

While most HERVs are inactive through transcription and reverse transcription some still produce ssRNA, dsRNA, and cytosolic DNA that interact with TLR3, TLR7/8, and TLR9[9].

Environmental factors such as hormones, infections, and ultraviolet light can also reactivate HERV transcription[4].

HERV products can act as pathogen-associated molecular patterns (PAMPs), which trigger cellular defenses[12]. HERVs have also been found to downregulate host immunity, for example by protecting against immune activation with respect to immune tolerance of the fetus[12]. HERV sequences can also serve as regulatory elements. For example, HERV-E can act as an enhancer in parotid glands as its inserted upstream of amylase gene[12]. HERV non-coding RNAs also provide cis-regulatory elements that modulate expression of host genes such as in human embryonic stem cells, pluripotency on nuclear long ncRNAs[19]. HERVs also have a role in the interferon network, research has shown that HERV insertions dispersed IFN-inducible enhancers in mammalian genomes[12].

HERVs are thought to trigger autoimmunity through molecular mimicry or superantigens where the body reacts to viral antigens. HERV proteins have strong cross-reactivity to self-antigens, for example HTLV-1-related endogenous sequence (HRES-1), can drive formation of autoantibodies due to the similarity of U1snRNP[4]. Viral nucleic acids can directly stimulate TLRs which trigger inflammatory cascades, and by activating IFN-gamma-related and other immune-related genes[4].

Both nucleic acids and antibodies to HERV proteins may contribute to autoinflammation and act simultaneously[9]. Researchers have found increased levels of HERV nucleic acids which may be involved in an IFN I response[9]. Antibodies against ERV envelope glycoproteins have been found in Sjogren's syndrome, SLE, and juvenile SLE.

Antiviral status mediated by the innate and adaptive systems usually work to counteract exogenous viruses, and when activated eliminate these exogenous threats and shut down.

However, HERV molecules could continuously trigger immune sensors, and once IFN is produced, a positive-feedback loop could be initiated where IFN-stimulated genes are which then upregulate HERV expression.

There have been many studies that have found associations between HERVs and autoimmune diseases such as T1D, MS, and SLE. In T1D it was found that there was inhibition of insulin secretion by HERV-W env, found in 70% of sera of peripheral blood mononuclear cells of T1D patients[20]. Researchers have also found elevated levels of antibodies against HERV-K6 env protein in T1D patients[18].

In multiple sclerosis (MS) patients have also shown increased expression of HERVs compared to controls. Studies have found HERV-W, KERV-K, and HERV-H in blood, brain and CSF increased in MS patients[21]. Additional studies suggest HERV-W env contributes to axonal injury and loss of myelin and axon integrity[22]. These findings led researchers to develop an antibody to neutralize the HERV-W env protein (GNbAC1), and a clinical trial in relapsing-remitting MS that showed mixed results[23,24]. HERVs have also been found in RA, multiple studies have found upregulation of HERV-K gag mRNA expression in RA patients[25,26]. Multiple studies have also found a link between SLE and endogenous viral elements such as auto-antibodies against HERVs such as HRES-1, which may have cross reactivity with self antigens[27]. Another study found expression of HERV-E clone 4-1 correlated with autoantibodies characteristic of SLE, anti-U1 RNP and anti-Sm[28]. HERV-E clone 4-1 has been widely studied in relation to SLE and there have been many findings implicating it in SLE such as serum antibodies to HERV clone 4-1, as well as increased mRNA of HERV-E clone 4-1 gag region in SLE patients[29,30]. Additionally HERV-E clone 4-1 was found upregulated in SLE CD4+ T cells,

and correlated with SLE disease activity and negatively correlated to HERV-E LTR methylation[31].

1.3 Quantification of Transposable Elements

Recent bioinformatic tools and experimental techniques have started to address difficulties in quantification of TEs, such as TE complexity, polymorphisms in insertions and mappability[32]. Many tools have been developed based on sequencing reads and vary based on the mapping approach used (uni vs multimapping), resolution (family vs locus-specific), accounting for TE polymorphisms, and chimeric transcripts[32]. Upon initial integration into the genome, only one copy exists of the TE exists and it is fully intact. However, as the TE transposes the sequences spread throughout the genome and over time diverge due to random mutations and alterations becoming inactivated due to these mutations[33]. This also causes TEs to be highly repetitive and short read sequences can align to multiple positions, these reads are termed multimappers. Interestingly, over time older TEs can become more unique, as opposed to younger TEs which are more similar to each other, and thus more difficult to quantify at a single locus[32].

Quantification of TEs is also made difficult because of the overlap of gene transcripts and TEs, where fragments of TEs are integrated in mRNA. Over one-third of human protein-coding transcripts contain an exon derived from a TE, and three-quarter of human lncRNAs are derived from TE sequences[34,35]. Since TEs are especially abundant in intergenic and intronic regions, pervasive transcription and pre-mRNAs can make up significant portions of RNA species that contain TEs. For example it has been shown that most L1-derived RNAs are not from L1 units alone, but from co-transcription or pervasive transcription[36,37].

Conventional approaches used to measure TE-derived transcripts such as RT-qPCR have historically been used but have several limitations. These include the significant difficulty in designing primers specific for one TE family given the sequence similarity and repetitive nature of TEs. Other methods such as north blotting and reporter gene knock-in have been used to measure the size distribution of TE transcripts and autonomous transcription respectively[38,39]. Some studies have used microarrays but this wasn't widely adopted due to the same difficulties as qPCR, namely in probe design[40,41].

One solution to this problem is only using unimappers, however this greatly underestimates the signal with TE families that integrated into our genomes more recently, and reflects mappability rather than expression levels[36,42,43]. Another strategy used to address this problem is mapping against consensus sequences, which results in aggregated counts by family, for example using Repbase[44]. Finally many tools first map against the reference genome using standard tools such as Bowtie2 that allow for multimapped reads[45]. Then multimappers are reassigned, one popular strategy is statistical reassignment of these multimapped reads. Tools such as TEtranscripts, SQuIRE, and Telescope us an expectation-maximization algorithm to reassign reads[46–48]. Here I leverage advances in tools for the quantification of TEs to study their association with heterogeneity in autoimmune manifestations.

Chapter 2: Cell-Specific Transposable Element Gene Expression Analysis Identifies Associations with Systemic Lupus Erythematosus Phenotypes

2.1 Introduction

Systemic lupus erythematosus (SLE) is a heterogeneous chronic autoimmune disease characterized by antibodies against nucleic acids and associated proteins[49,50]. SLE mainly affects women, especially child-bearing age, with a relapsing-remitting course[4]. Key characteristics of SLE include impaired clearance of nucleic acids, enhanced type I Interferon (IFN) response, abnormal B cell tolerance, and production of autoantibodies[4]. Diagnosis is difficult as any 4 of 11 criteria are diagnostic of SLE, as such some consider SLE to be an umbrella term of related disorders, and the heterogeneity in clinical manifestations key to understanding the disease[11]. SLE is heterogenous disease from mild cutaneous to kidney/cardiac failure[9].

Clinical manifestations are very heterogeneous and the prevalence varies among different racial and ethnic groups, and the risk of developing severe manifestations is increased in African Americans, Asian/Pacific Islanders, and Hispanic patients[51]. Manifestations can include any organ from the central nervous system to the cardiovascular system, although mucocutaneous and musculoskeletal systems are usually the earliest signs of disease[4]. Patients that self-report as Asian experience disproportionate burden of SLE compared to White patients, including greater incidence and prevalence of SLE, higher disease activity, worse organ damage, and heightened mortality[50].

Transcriptomic studies have been undertaken to characterize the heterogeneity in SLE and have found expression of interferon (IFN)-regulated genes to be associated with disease[52,53]. About 100 SLE susceptibility loci have been identified from GWAS studies, with most explained by common variants, and IRF5 along with IRF7 and IRF8 interferon regulators are also some of

the best SLE-associated loci[54,55]. Genetic studies have shown that the heterogeneity in SLE may be due in part to the presence of pathogenic disease associated variants in non-coding regions[18,56,57].

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[58]. As a result, TEs have roles both in human health and diseases such as cancer and autoimmunity[47,59,60]. 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[4,61]. One hypothesis for how TEs could contribute to autoimmunity is that expression of TEs activates an immune response when TEs produce nucleic acids or proteins that resemble PAMPs of ancient exogenous viruses[62,63].

In previous work, over 100 locus-specific HERVs were shown to be differentially expressed in SLE patients and correlate with lupus clinical parameters such as presence of dsDNA, anti–RNP, and anti-SM antibodies[64]. In addition, HERV-E clone 4-1 mRNA expression correlated with SLEDAI score in CD4+ T cells[65]. Previous studies have identified elevated expression of HERVs in SLE, implicating HERV-E clone 4-1 in PBMCs and HRES1/p28 in B cells[31,66]. In addition, more recent work has characterized TEs in blood from SLE patients compared to matched controls, and found upregulation of TEs in SLE[60,64,67]. These studies also investigated whether TEs contribute to the IFN signature observed in SLE patients, with one

study finding a positive correlation between HERV expression and the IFN signature, while the other did not[60,64]. Another study used two independent whole blood cohorts and identified 481 locus specific HERVs differentially expressed in SLE with low overlap of these elements across datasets[67].

Long interspersed nuclear elements (LINEs) are TEs that can replicate autonomously due to their endonuclease and reverse transcriptase[68]. Accounting for ~20% of the human genome, contains two open reading frames for RNA binding protein, RT, and endonuclease, and can bind TLR7 in endosomes or retinoic acid-inducible gene (RIG-I) and melanoma differentiation-association protein 5 (MDA5), which result in IFN gene transcription[68]. It's been found that LINE-1 hypomethylation has been observed in CD4+ T cells, CD8+ T cells, and B cells in SLE patients[69]. An additional study found reduced methylation levels in neutrophils in SLE patients of LINE-1 elements, and may upregulate genes involved in the cell death[70]. Another study found hypomethylation of LINE-1, increased LINE-1 mRNA transcripts in lupus nephritis kidneys that correlated with IFN-I expression[7].

Although many studies have explored TE expression between SLE patients and controls, 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 differentially expressed (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, carry out 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 lupus patients with extensive phenotypic data.

2.2 Results

The study consists of 120 SLE participants from the California Lupus Epidemiology Study (CLUES) (**Figure 2.1**). The majority of participants were female, with an age distribution between 20 and 82 (**Table 2.1**). In order 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[50]. After standard quality control measures, we quantified expression of 27,135 HERVs and LINEs (TEs) in 4 cell-types using Telescope[47]. We found that both TE and gene expression is cell specific in lupus patients as observed by the clear clustering in principal component analyses (PCA) of both TE expression (**Figure 2.1B**) and gene expression (**Figure 2.1C**).

Association Between TEs and SLE Phenotypes is Cell Type Specific.

We carried out a comprehensive genome wide analysis of TEs in CD4+ T cells, monocytes, B cells, and NK cells in relation to 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 (adj p-value < 0.05) TEs in each cell type and SLE subphenotype (**Table 2.2**). In CD4+ T cells there were between 3 and 38 differentially expressed elements, in monocytes cells there were between 3 and 34 differentially expressed elements, in B cells there were between 2 and 57 differentially expressed elements, and in NK cells there were between 2 and 37 differentially expressed elements (**Table 2.2**).

These include two differentially expressed TEs (HML2_8p23.1b and ERVLB4_8p23.1o) in CD4+ T cells at a previously identified susceptibility locus of SLE (8p23)(21).

We then explored whether these differentially expressed TEs were cell or SLE subphenotype specific. No locus specific TEs were shared across all SLE subphenotypes across all cell types (CD4+ T cells, monocytes, B cells, and NK cells). Very few overlaps were observed across 2-3 cell types. For example, between CD4+ T cells, monocytes, and B cells L1FLnI_21q22.3a was differentially expressed.

We observe very few overlaps of TEs across all SLE subphenotypes and cell types (**Figure 2.2**). In CD4+ T cells we identify 1 TE (ERVLB4_8p23.1o) that was common across 5 subphenotypes including (photosensitivity, anti-dsDNA antibody at time of blood draw, disease activity characterized by the SLEDAI score, history of anti-RNP antibody, and cluster severe disease 1 vs mild disease). In the monocyte analysis, we found 2 overlapping TEs (L1FLnI_5q35.1e and L1FLnI_Yp11.2na) common across 6 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 Interferon signature). In B cells we identify 1 TE (HML2_8p23.1a) that is differentially expressed across 8 subphenotypes (photosensitivity, anti-dsdna antibody at the time of blood 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 1 TE common (LTR25_9q13a) across 4 SLE subphenotypes (anti-dsdna antibody at the time of blood draw, history of anti-RNP antibody, IFN signature and anti-SM antibody).

Comparing patients based on genetic similarity revealed two strong clusters (**Figure 2.8**), which significantly correlated with self-reported race (r2 0.938, p-value 2.2e-16). Given the differences in SLE burden between individuals with Asian and European ancestry, we explored differential TE expression within these groups[71]. Similar to our previous analysis, we identified many differentially expressed locus specific TEs by SLE subphenotype and cell type when stratifying by genetic ancestry groups and removing admixed individuals (**Table 2.3**).

We observe a larger number of differentially expressed TEs in lupus subphenotypes, especially anti-dsDNA in this stratified analysis when compared to the analysis with all patients (**Table 2.3 and Table 2.4**). As in the overall analyses, few locus-specific TEs were differentially expressed across cell types or subphenotypes in the stratified analyses (**Figure 2.4**).

In the Asian enriched subgroup we saw the largest number of overlaps across cell types. There were 10 TEs in common across CD4+ T cells and monocytes, 5 TEs common between monocytes, B cells, and NK cells, 20 TEs common between CD4+ T cells, monocytes, and NK cells and 5 TEs common between monocytes, B cells, and NK cells. The European enriched subgroup had fewer TEs shared across phenotypes and cell types in comparison to the Asian enriched subgroup.

In the European enriched subgroup of patients in CD4+ T cells we found 1 TE (MER101_21q21.3a) common across 4 different subphenotypes (anti-dsDNA antibody, photosensitivity, disease activity characterized by the SLEDAI score, and Anti-RNP antibody). In monocytes we identified 1 TE (HERVL_Xp11.4) common across three SLE subphenotypes (photosensitivity, anti-dsDNA antibody and IFN). In B cells we found two TE's (ERVLB4_2q11.1b, HARLEQUIN_Yq11.223) that are shared across 5 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 of patients a in CD4+ T cells there were 3 common differentially expressed TEs (ERVL_17q11.1, LTR23_3q26.31, and MER41_17q23.3a) across 4 SLE subphenotypes (anti-dsDNA antibody, anti-RNP antibody, IFN, anti-SM antibody). In monocytes cells there were 2 TEs (L1FLnI_6q14.3u, MER101_6p21.1) across 6 subphenotypes (anti-dsDNA antibody, anti-RNP antibody, severe disease 2 vs severe disease 1, Anti-SM antibody, photosensitivity, and IFN). In B cells there were 4 differentially expressed TEs (ERV316A3_4q28.3cl, ERVLE_4q32.3a, MER4B_15q21.1c, MER4B_Xq22.3a) across 4 SLE subphenotypes (photosensitivity, anti-dsDNA antibody, severe disease 2 vs severe disease 1, and severe disease 1 vs Mild Disease). Lastly, in NK cells there were 4 common differentially expressed TEs (ERVLB4_Xq21.31j, HUERSP2_Xq27.3a, L1FLnI_11p14.3k, 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 differentially expressed TEs 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. Previous studies have associated HERV families, like HERV-K, with SLE[67].

To test for family-level effects, we analyzed differentially expressed TEs at the family level across subphenotypes of SLE, focusing on each cell type. We discovered significant enrichment of different families among the differentially expressed TEs associated with SLE sub-phenotypes in each cell type (**Figure 2.5A**). In CD4+ T cells, we observed enrichment of HERVH (adj p-value 1.35E-29), while in monocytes, we detected enrichment for MER61(adj p-value 0.01). In B cells, we found HML2 (adj p-value 1.47E-04) and HERVH (adj p-value 1.07E-30) to be enriched, and in NK cells ERVLB4 (adj p-value 1.53E-06) 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. In the Asian enriched group, TEs manifested unique patterns of enrichment across different cell types. Specifically, MER101 was consistently enriched across all cell types. NK cells in this subgroup were characterized by the enrichment of PRIMAX, LTR23, MER34B, HERVH, MER4, MER4B, ERVLB4, ERVLE, and MER41. Meanwhile, B cells displayed enrichment of HML1 and HERVH, MER4, MER4B, ERVLE; and monocytes cells showed enrichment of LTR57, LTR25, MER34B, MER61, MER4, MER4B, and MER41. Enrichment in CD4+ T cells was observed for LTR23, LTR57, LTR25, HERVH, HML2, ERVLB4, and MER41; while L1FLnI demonstrated consistent depletion across all cell types.

In comparison, in the enriched European group, HERVH and LTR57 were enriched in CD4+ T cells, whereas monocytes cells displayed enrichment of MER4. MER4B, HERVH, ERVLB4, ERVLE enrichment was characteristic of B cells. As seen in the Asian enriched group, L1FLnI showed depletion across all cell types as well.

Previous work has hypothesized that HERVs with preserved open reading frames could produce proteins which could activate or depress the inflammatory cascade[72]. Therefore we used

gEVE: a genome-based endogenous viral element database to find open reading frames of viral proteins derived from HERVs in the differentially expressed TEs we observed across SLE subphenotypes. We found enrichment of open reading frames of viral proteins derived from TEs in the cell specific differentially expressed TEs across all SLE subphenotypes (**Table 2.5**).

Interferon Pathway Associated with Expression of Differentially Expressed TEs

To better understand 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 our cohort. The expression levels of many genes associated with expression of DE TEs in the combined analysis (**Table 2.6, Figure 2.6**). The associations of differentially up and down regulated genes were largely cell-type-specific (**Figure 2.6E and Figure 2.6F**). In the CD4+ T cells there is a strong upregulation of genes associated with the differentially expressed TEs. Some of the strongest up regulated genes in CD4+ T cells were involved in type I interferon signaling and other antiviral innate immune pathways such as LY6E, IFI6, ISG15, and ISG20. In the monocyte analysis, most of the genes also appear upregulated as seen in the volcano plots, and the top upregulated genes are also involved in antiviral activity such as ISG15, IFI6, IFI35, BST2, and TRIM22. In CD19 cells there are also many up-regulated genes as seen in the volcano plot. Some of the top upregulated genes are ISG15, IFIT3, IRF7, and BST2 which play roles in response to viruses. In NK cells there are many up and down regulated genes as seen in 6D, some of the top upregulated genes are NXF1, which is involved in transport of unspliced retroviral genomic RNA and CAPZA1 which was downregulated, and is also observed to be downregulated in virus infected cells[73,74].

Gene set enrichment analysis revealed that SLE associated TEs correlated with differential expression of similar pathways, such as interferon and cytokine signaling, across cell

types. Pathways expressed in all cell types except NK cells include Interferon alpha/beta signaling, Interferon Signaling, Cytokine Signaling in Immune system, and antiviral mechanism by IFN-stimulated genes (**Figure 2.7**). There are also some cell type specific signals that were identified. For example, differentially enriched pathways with TE expression include SLC-mediated transmembrane transport in monocytes cells; influenza infection, viral mRNA translation, antigen presentation: folding assembly and peptide loading of class I MHC in CD4 cells; GPCR ligand binding, host interactions with influenza factors, and NS1 mediated effects on host pathways, in B cells.

2.3 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 HERV expression could contribute to the heterogeneity of SLE across individuals and populations.

We found independent associations of TE expression with production of autoantibodies against ribonucleoproteins (RNP, 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, anti-SM nuclear antibodies) and higher interferon status in SLE[28,64]. These associations are highly cell specific, subphenotype specific, which suggests that their differential expression could contribute or be a result of different lupus subphenotypes.

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 in this cohort, as well as a small number of individuals with significant genetic admixture. Admix individuals were removed from downstream stratified analyses. When the two groups, Asian enriched and European enriched were stratified, we found more differentially expressed TEs in almost every SLE subphenotype, despite adjusting for genetic principal components 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 due to genetic factors or due to differences in environmental exposures.

Retroviral p30 gag proteins and serum reactivities to p30 gag antigen have been found in SLE patients with proliferative glomerulonephritis[75]. We investigated if it was possible that the differentially expressed TEs identified exert their effect as proteins, as most expressed TEs do not contain full open reading frames for functional viral proteins. We tested if there was significant enrichment of open reading frames among the identified differentially expressed TEs and found enrichment of several TE derived viral proteins in a cell specific manner, such as pro, env, and RT. 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[19]. Therefore it is difficult to say whether the enrichment of viral proteins we observe in the differentially expressed TEs is the cause of an immune response, or the response acting to downregulate an immune response.

One of the most intriguing findings was the connection between host response gene expression and the expression of transposable elements (TEs) in relation to the SLE phenotypes.

It has been hypothesized that TEs could activate the innate immune system, and unlike a viral infection where the immune response is curtailed, could lead to chronic inflammation and contribute to autoimmunity[62]. For example, the chronic induction of type I IFNs which could put cells into an antiviral state, shutting down metabolic processes and leading to apoptosis[62]. Supporting this, a study of kidney biopsies from SLE patients with lupus nephritis found LINE elements can trigger the IFN-I pathway[7]. We found that differentially expressed 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 differentially expressed 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-kappaB, which have been associated with both lupus and viral infection[76,77]. Additional studies are needed to examine causality between TE expression and immune dysregulation in order to understand whether TE expression contributes to the development of SLE.

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 and the total sample size is limited. While the cohort described here has individuals with Asian and European ancestry, it would be interesting if these findings hold up in a more diverse population from the US and beyond. Furthermore, 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. Finally, though we were able to account for important potential confounding factors such as medication usage, age, and gender, we cannot exclude the possibility of additional unidentified confounders

on TE expression. As 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 discover 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, further exploration into TE expression within normal immune cells might be needed. In addition, further functional studies could potentially shed light on whether TE expression contributes to the pathogenesis of SLE.

2.4 Methods

Cohort Description and Data Generation

All patients in this study were from the California Lupus Epidemiology Study (CLUES), a cohort of individuals with physician-confirmed SLE. CLUES was approved by the Institutional Review Board of the University of California, San Francisco. All participants signed a written informed consent to participate in the study. Study procedures involved an in-person research clinic visit, which included collection and review of medical records prior to the visit; history and physical examination conducted by a physician specializing in lupus; a collection of biospecimens, including peripheral blood for clinical and research purposes; and completion of a structured interview administered by an experienced research assistant. All SLE diagnoses were confirmed by study physicians based upon one of the following definitions: (a) meeting ≥4 of the 11 American College of Rheumatology (ACR) revised criteria for the classification of SLE as

defined in 1982 and updated in 1997, (b) meeting 3 of the 11 ACR criteria plus a documented rheumatologist's diagnosis of SLE, or (c) a confirmed diagnosis of lupus nephritis, defined as fulfilling the ACR renal classification criterion (>0.5 grams of proteinuria per day or 3+ protein on urine dipstick analysis) or having evidence of lupus nephritis on kidney biopsy. Based on sample availability at the time of sequencing, a total of 120 patients were profiled with bulk RNA-seq from the CLUES cohort. Clinical data collected at sampling and the self-reported race was used for downstream analyses.

Subphenotype Definitions

Disease activity was measured with the standardized disease activity score SLEDAI[78]. A high SLEDAI score was defined as a score greater than or equal to 8, while low was defined as less than 8. Besides the total SLEDAI score, we also performed analyses with specific items of the SLEDAI such as proteinuria and 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 (ACR) Classification Criteria characterized by significant differences in SLE manifestations[79]. We also performed subphenotype associations with history of presence of anti-Smith antibody (anti-Sm), anti-ribonucleoprotein antibody (anti-RNP) and anti-double stranded DNA antibody (anti dsDNA). Photosensitivity was defined as a rash or feeling sick after going out in the sun. Serologies were performed in CLIA certified labs and reported as abnormal or normal.

RNA-Seq Data Generation

Peripheral blood mononuclear cells were isolated from 120 SLE participants. Using the EasySep protocol from STEM cell technologies, these cells were sorted into CD14+ monocytes, B cells, CD4+ T cells, and NK cells, for a total of 480 samples. These samples were sequenced

on a HiSeq4000 PE150 and gene expression data was generated using Salmon v0.8.2 with adapter-trimmed reads, quality control was performed as previously described[50].

Genotyping

Genotyping for genomic DNA from peripheral blood was performed using the Affymetrix Axiom Genome-Wide LAT 1 Array. This genotyping array is composed of 817,810 SNP markers across the genome and was specifically designed to provide maximal coverage for diverse racial/ethnic populations, including West Africans, Europeans and Native Americans[80]. Samples were retained with Dish QC (DQC) ≥ 0.82. SNP genotypes were first filtered for high-quality cluster differentiation and 95% call rate within batches using SNPolisher. Additional quality control was performed using PLINK. SNPs having an overall call rate less than 95% or discordant calls in duplicate samples were dropped. Samples were dropped for unexpected duplicates in identity by descent (IBD) 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 ds-DNA antibodies and renal disease to minimize genetic heterogeneity. SNPs were dropped if HWE $p < 1e-5$ or any cross-batch association $p < 5e-5$. Genetic PCs were generated using EIGENSTRAT and used for patient stratification.

Transposable Element QC and Expression Quantification

Transposable element expression quantification was performed using adapter-trimmed reads from the 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-sensitive-local –score-

min L,0,1.6). Sorted bams 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. 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 (v.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 use at the time of blood draw. For data analyses, we grouped immunosuppressive medications into the following categories: biologic treatments (belimumab, abatacept, 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), interferon signature high/low based on gene signature developed by Kennedy et al[81], photosensitivity (previous work has found UV light induces HERV expression[82]), proteinuria, autoantibody production (dsDNA, RNP, SM), and disease severity (as defined by clinical clusters previously described in the same SLE participants)1. For stratified analyses, patients were stratified according to genetic similarity (PC1 >0.025 - Asian ancestry, <-0.025 - European ancestry) and admixed individuals were not considered for downstream stratified analysis. P-values from DESeq2 with a NA value were reassigned a value of 1, and all p values from the different cell type and subphenotype analysis were FDR-corrected using Benjamini-Hochberg. Differentially expressed TEs (padj < 0.05) for all SLE subphenotypes per cell type were used for downstream analysis. PCAs of TE expression were computed with the factoextra package (v 1.0.7). Cell type PCAs were made using the

variance stabilizing transformation with the DESeq2 function vst(), and VST-transformed data was visualized with plotPCA in the DESeq2 package.

Characterization of Differentially Expressed TEs

TE group and family definitions were taken from Telescope using families.tsv from https://github.com/mlbendall/telescope_annotation_db/tree/master/builds. Locus-specific differentially expressed TE's 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 differentially expressed TEs according to families.tsv. Line elements were grouped into L1FLnI, L1ORF2, and L1FLI. Log odds ratio and hypergeometric test for enrichment/depletion were calculated and an expression threshold of 4 was used for filtering.

Enrichment of viral proteins

Bedtools was used to find the intersection between Hsap38.geve.v1.bed regions and Telescope annotation regions. The differentially expressed elements from the SLE subphenotype analysis that overlapped full regions of the gEVE annotation were used to calculate enrichment of viral proteins.

TE and Gene Expression Integration

Association of differential TE expression with gene expression was calculated using the sum of counts of differentially expressed TEs for all SLE subphenotypes per cell type. Raw counts from Telescope were normalized with DEQseq2 and the counts of differentially expressed TEs for each cell type were summed to get counts of differentially expressed TEs per patient for each cell type. DESeq2 (v.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 us at the time of blood

draw. Gene set enrichment analysis was run with significant genes (padj < 0.05) using WebGestaltR (v 0.4.5) and the reactome database. Volcano plots were generated using Enhanced Volcano package (v 1.16.0)

Statistics

R was used to determine statistical significance of differences and a padj value of less than 0.05 was considered significant.

Study Approval

Participants were recruited from the California Lupus Epidemiology Study (CLUES). CLUES was approved by the Institutional Review Board of the University of California, San Francisco. All participants signed a written informed consent to participate in the study.

Data Availability

Raw data of this study are openly available in GEO: GSE164457. All other data are available from the corresponding author upon reasonable request.

2.5 Tables

Table 2.1 Cohort Characteristics Demographics of the subset of the CLUES cohort used for TE and gene expression analysis

	Overall (n=119)	Asian Enriched (N=62)	European Enriched (N=57)
Sex			
Female	105 (88.2%)	56 (90.3%)	49 (86.0%)
Male	14 (11.8%)	6 (9.7%)	8 (14.0%)
Age			
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			
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			
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*			
High	13 (10.9%)	6 (9.7%)	7 (12.3%)
Low	106 (89.1%)	56 (90.3%)	50 (87.7%)
Anti-Sm Ab			
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			
Abnormal	50 (42.0%)	31 (50.0%)	19 (33.3%)
Normal	69 (58.0%)	31 (50.0%)	38 (66.7%)
Proteinuria at blood draw			
Abnormal	5 (4.2%)	3 (4.8%)	2 (3.5%)
Normal	114 (95.8%)	59 (95.2%)	55 (96.5%)
Photosensitivity			
0	80 (67.2%)	40 (64.5%)	40 (70.2%)
1	39 (32.8%)	22 (35.5%)	17 (29.8%)

Table 2.2 Cell specific differentially expressed TEs across SLE subphenotypes Number of differentially expressed TEs shown below per subphenotype and cell type (adj pval < 0.05) adjusting for sex, age, sequencing lane, medications, and genetic PCs 1-10 (up-regulated/down-regulated)

	CD4	CD14	CD19	NK
Autoantibody Production				
SM + vs -	16 (7,9)	3 (0,3)	33 (6,27)	14 (6,8)
RNP + vs -	16 (6,10)	3 (2,1)	18 (11,7)	16 (7,9)
dsDNA + vs -	11 (4,7)	11 (5,6)	17 (8,9)	19 (7,12)
Disease Severity				
SD2 vs M	32 (2,30)	16 (0,16)	34 (21,13)	16 (3,13)
SD1 vs M	8 (3,5)	11 (4,7)	20 (11,9)	28 (9,19)
SD2 vs SD1	38 (3,35)	34 (4,30)	57 (38,19)	37 (2,35)
SLEDAI Score High vs Low	8 (2,6)	23 (3,20)	24 (1,23)	8 (0,8)
Clinical Manifestations				
IFN High vs Low	15 (6,9)	18 (8,10)	7 (1,6)	22 (0,22)
Proteinuria + vs -	3 (0,3)	13 (1,12)	2 (1,1)	2 (1,1)
Photosensitivity + vs -	11 (8,3)	12 (9,3)	33 (3,30)	22 (3,19)

Table 2.3 Ancestry stratified by participants defined as European enriched and clustered by genetic PC1 Number of differentially expressed TEs shown below per subphenotype and cell type (adj pval < 0.05) adjusting for sex, age, sequencing lane, medications, and genetic PCs 1-10 (up-regulated/down-regulated)

Participants defined as European Enriched by Genetic Ancestry PC1

	CD4	CD14	CD19	NK
Autoantibody Production				
SM + vs -	14 (8,6)	0 (0,0)	0 (0,0)	6 (1,5)
RNP + vs -	158 (40,118)	47 (12,35)	80 (17,63)	22 (13,19)
dsDNA + vs -	107 (26,81)	24 (9,15)	167 (44,123)	30 (6,24)
Disease Severity				
SD2 vs M	40 (15,25)	0 (0,0)	22 (16,6)	4 (2,2)
SD1 vs M	123 (45,78)	50 (21,29)	113 (29,84)	28 (9,19)
SD2 vs SD1	55 (20,35)	0 (0,0)	54 (38,16)	8 (5,3)
SLEDAI Score High vs Low	28 (17,11)	14 (13,1)	23 (11,12)	3 (2,1)
Clinical Manifestations				
IFN High vs Low	6 (4,2)	28 (11,17)	0 (0,0)	0 (0,0)
Proteinuria + vs -	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
Photosensitivity + vs -	142 (28,114)	53 (9,44)	142 (54,88)	68 (12,56)

Table 2.4 Ancestry stratified by participants defined as Asian enriched and clustered by genetic PC1 Number of differentially expressed TEs shown below per subphenotype and cell type (adj pval < 0.05) adjusting for sex, age, sequencing lane, medications, and genetic PCs 1-10 (up-regulated/down-regulated)

Participants self-reported as Asian and clustered by Genetic PC1

	CD4	CD14	CD19	NK
Autoantibody Production				
SM + vs -	102 (26,76)	62 (40,22)	105 (69,36)	92 (42,50)
RNP + vs -	193 (108,85)	85 (43,42)	123 (81,42)	203 (128,75)
dsDNA + vs -	229 (133,96)	113 (75,38)	228 (157,71)	52 (27,25)
Disease Severity				
SD2 vs M	40 (12,28)	26 (6,20)	84 (69,15)	137 (114,23)
SD1 vs M	122 (82,40)	74 (46,28)	147 (61,86)	270 (45,225)
SD2 vs SD1	201 (36,165)	115 (27,88)	235 (136,99)	162 (108,54)
SLEDAI Score High vs Low	42 (28,14)	19 (13,6)	35 (27,8)	54 (47,7)
Clinical Manifestations				
IFN High vs Low	139 (86,53)	113 (77,36)	54 (37,17)	319 (203,116)
Proteinuria + vs -	1 (1,0)	1 (0,1)	0 (0,0)	0 (0,0)
Photosensitivity + vs -	240 (48,192)	143 (38,105)	291 (56,235)	303 (29,274)

Table 2.5 Cell specific viral proteins enriched from the differentially expressed TEs
Enrichment was performed using gEVE (a genome-based endogenous viral element database) looking at a subset of viral genes, gag-structural components of matrix, capsid, and nucleocapsid, pro - protease, RT- reverse transcriptase, env - envelope-associated glycoprotein, dut- dUTPase

	CD4	CD14	CD19	NK
Combined	pro, int, RT		pro, RT	pro, env, RT
European Enriched	pro, int, gag, dut, RT		pro, int, RT	
Asian Enriched	pro, RT	pro, dut, int, RT	pro, int, RT	pro, RT

Table 2.6 Cell specific differentially expressed genes with respect to lupus associated TEs
Differential gene expression (adj p-val < 0.05) with respect to the cell specific identified locus-specific TEs that were differentially expressed across all SLE subphenotypes (Up-regulated/Down-regulated)

	CD4	CD14	CD19	NK
Combined	5405 (2449,2956)	1444 (896, 548)	1569 (680,889)	219 (151,68)
European Enriched	34 (5,29)	223 (115, 108)	70 (9,61)	288 (221,67)
Asian Enriched	18 (9,9)	17 (3,14)	16 (0,16)	24 (5,19)

2.6 Figures

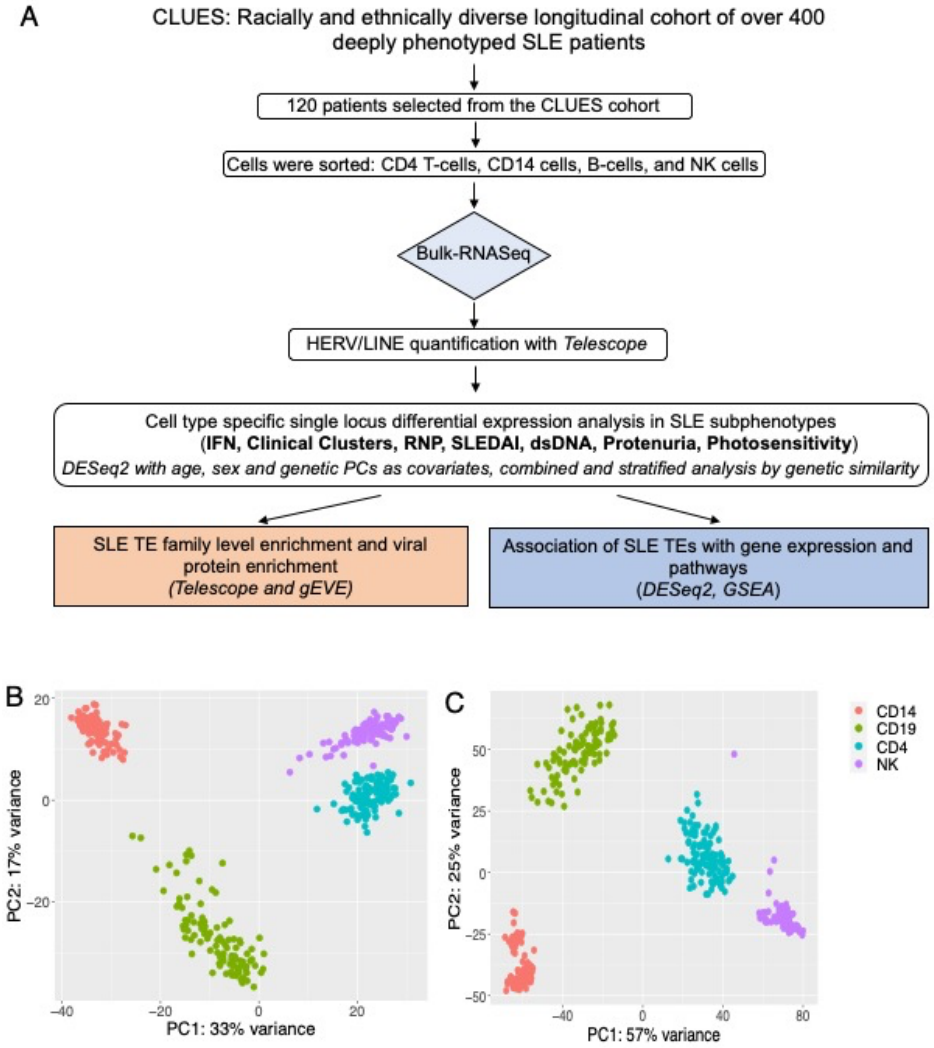

Figure 2.1 Study Overview and Comparison of Gene and TE expression between cell types
A. 120 patients selected for cell sorted bulk RNA-seq analyses from the CLUES cohort. Data were QC'ed and TEs were quantified using Telescope. DESeq2 was used for cell specific differential expression of locus specific TEs for SLE subphenotypes. Cell specific differentially expressed TEs from all SLE subphenotypes were used to perform family enrichment, TE derived viral protein enrichment, and association with gene expression and pathway analysis. B. TE expression is cell specific in lupus - PCA plot based visualization colored by cell type. C. Gene expression PCA plot based visualization colored by cell type

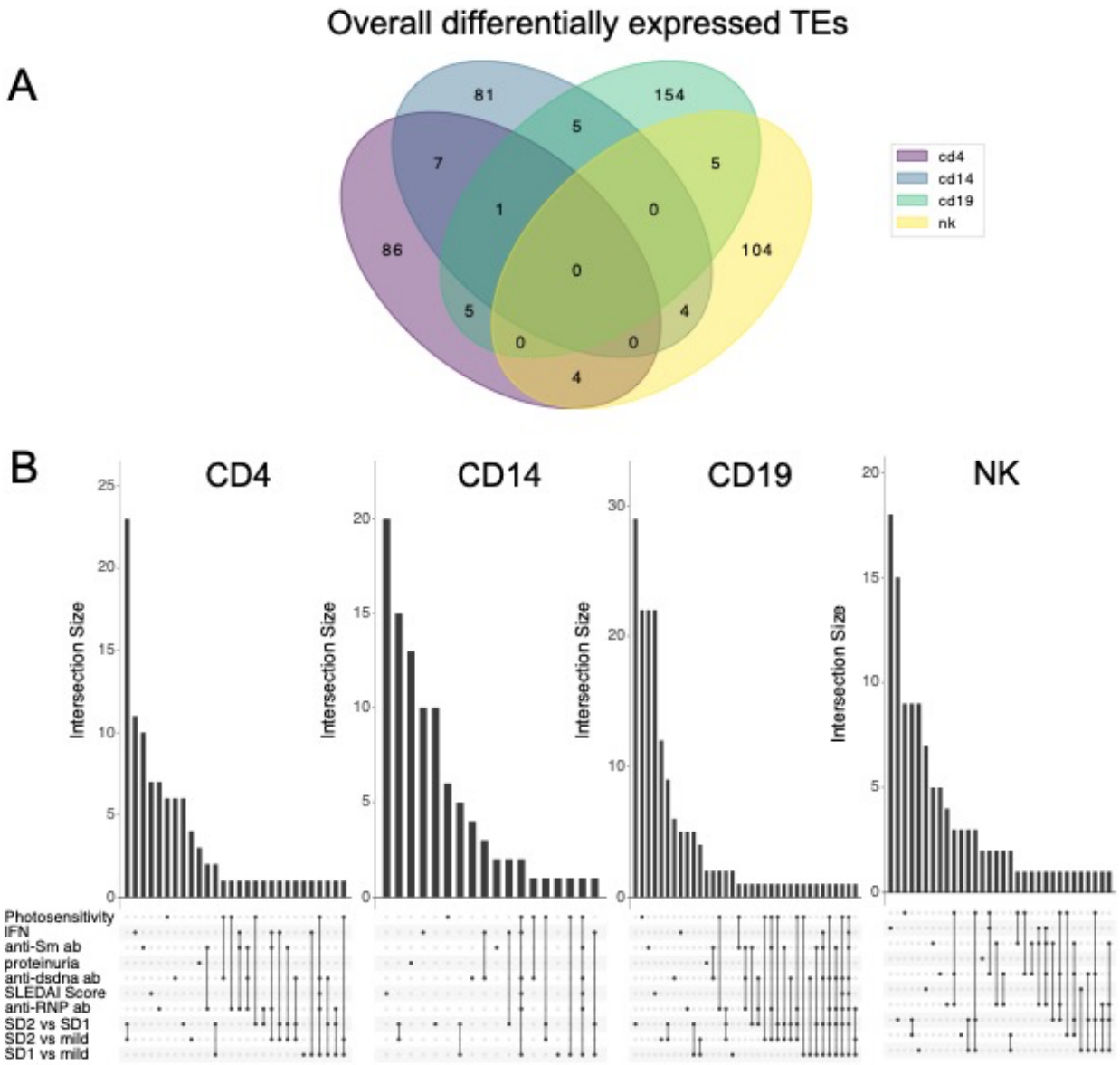

Figure 2.2 Combined cohort common differentially expressed TEs Identification of common differentially expressed TE's in the combined cohort patients shows distinct locus specific TEs differentially expressed across cell types and SLE subphenotypes. A. Venn diagram of overlap of DE TEs between cell types B. Overlap of DE TEs between lupus subphenotypes

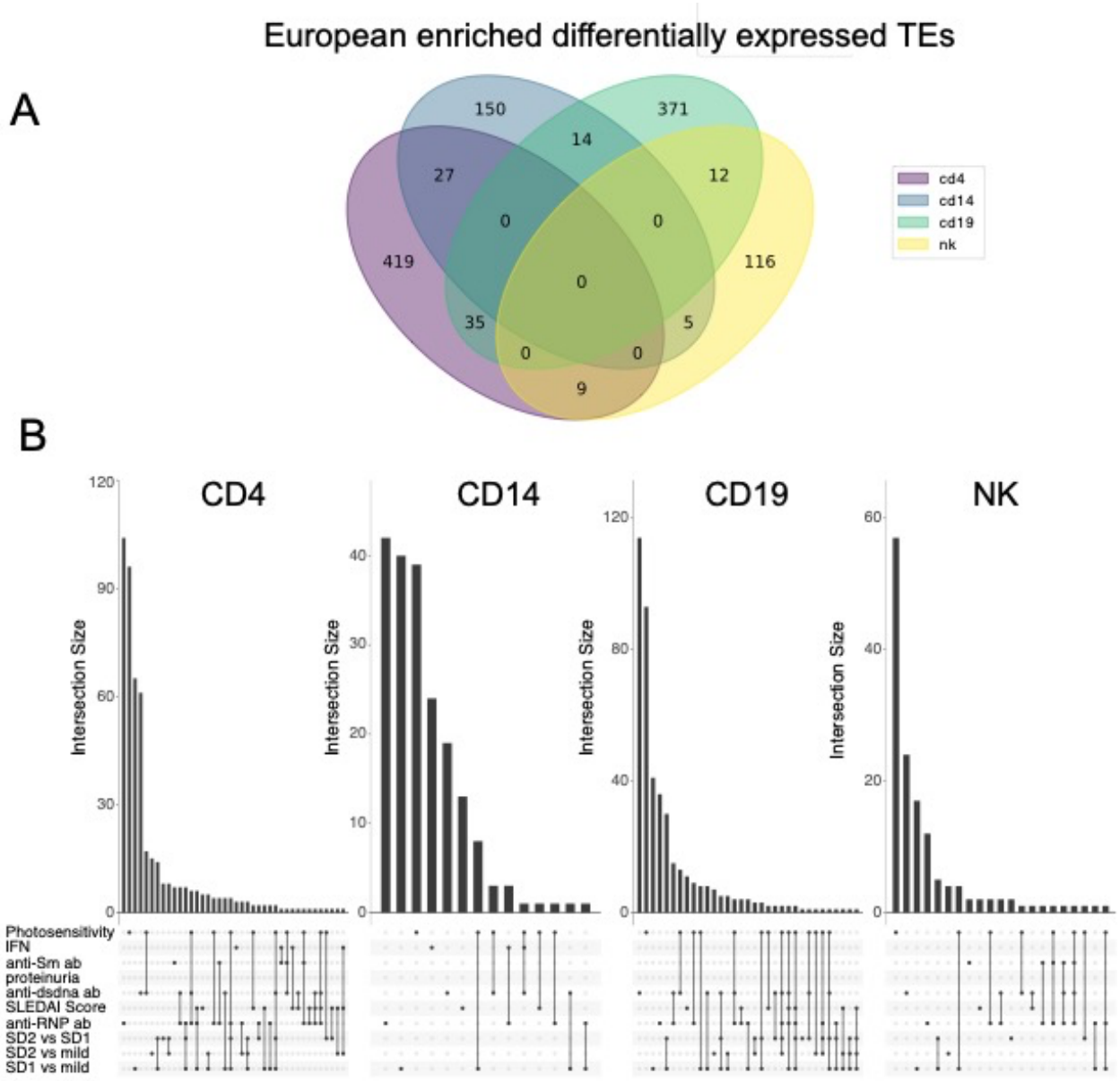

Figure 2.3 European enriched cohort common differentially expressed TEs Identification of common differentially expressed TE's in the European enriched cohort patients (ancestry stratified) shows distinct locus specific TEs differentially expressed A. across cell types and B. SLE subphenotypes

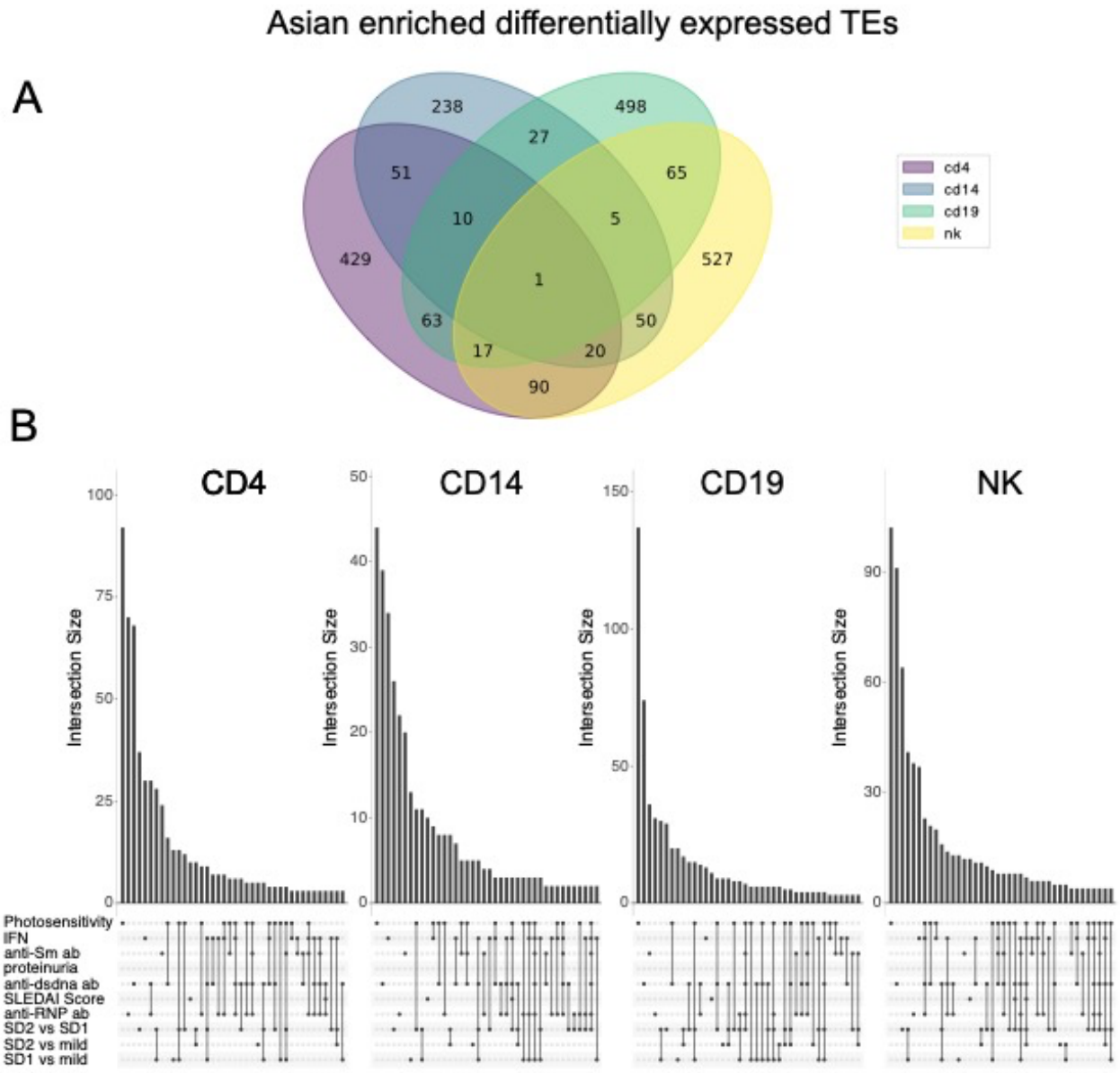

Figure 2.4 European enriched cohort common differentially expressed TEs Identification of common differentially expressed TE's in the Asian enriched cohort patients (ancestry stratified) shows distinct locus specific TEs differentially expressed A. across cell types and B. SLE subphenotypes

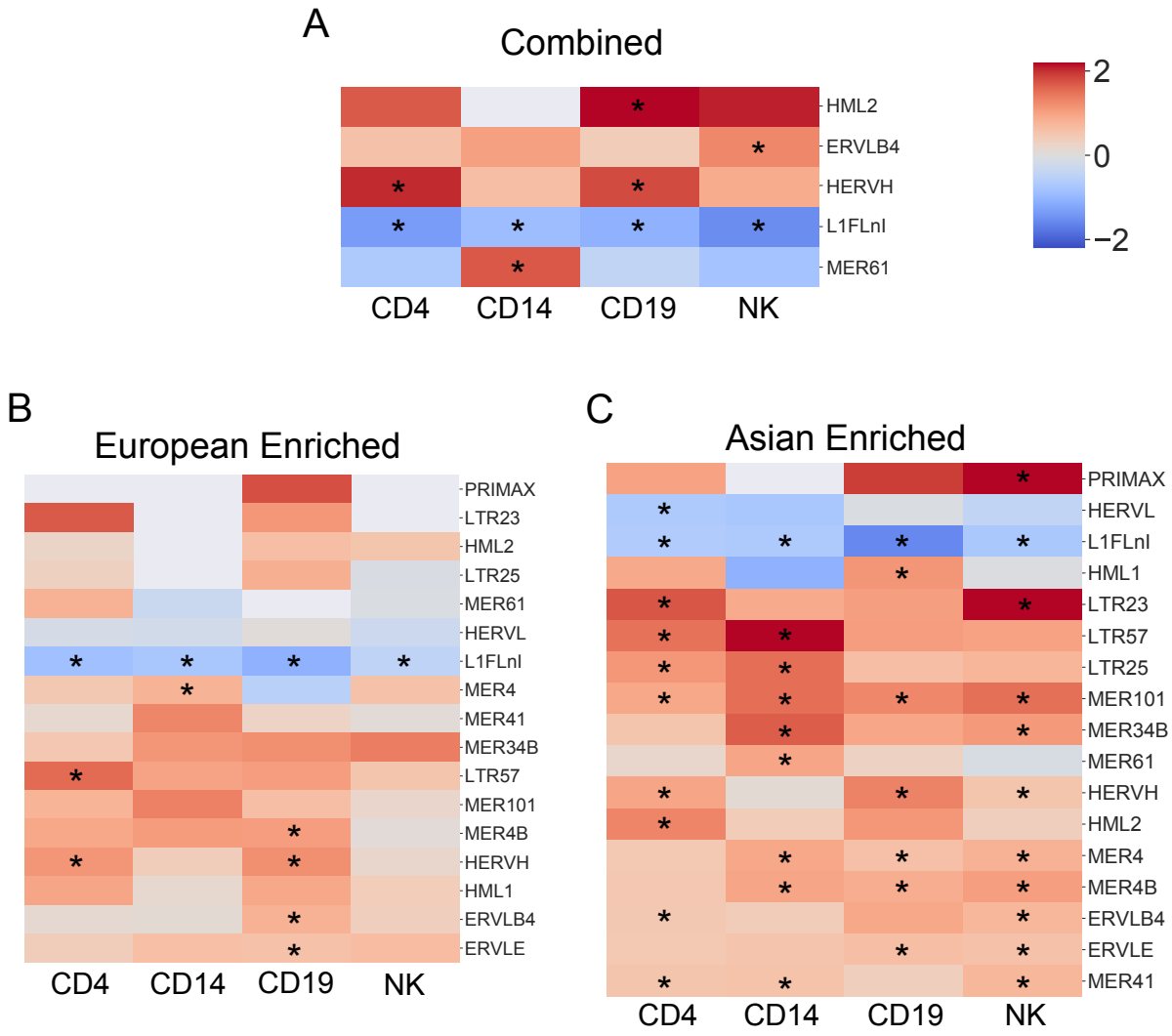

Figure 2.5 Heatmap of log odds ratio showing significant enrichment/depletion of HERV families in combined and stratified analysis Log odds ratio of significant enrichment/depletion of HERV families in combined and stratified analysis A. Family enrichment for the combined analysis B. Family enrichment 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 *)

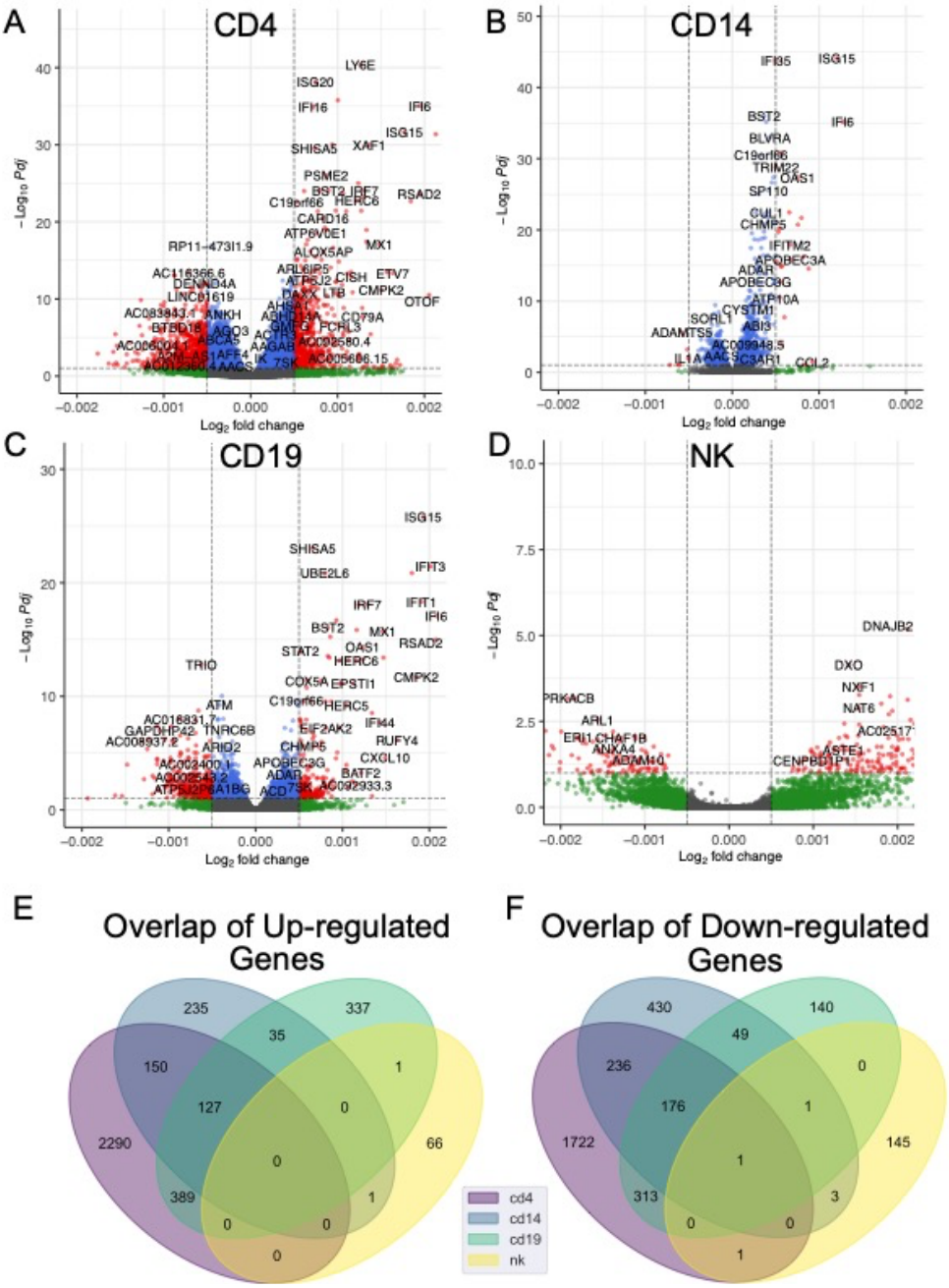

Figure 2.6 Cell specific volcano plots of combined analysis differentially expressed gene associations with differential TE expression using DESeq2 A. CD4, B. CD14, C. CD19 and D. NK. Identification of the common DE genes (padj < 0.05) across the different cell types: E. up-regulated F. down-regulated

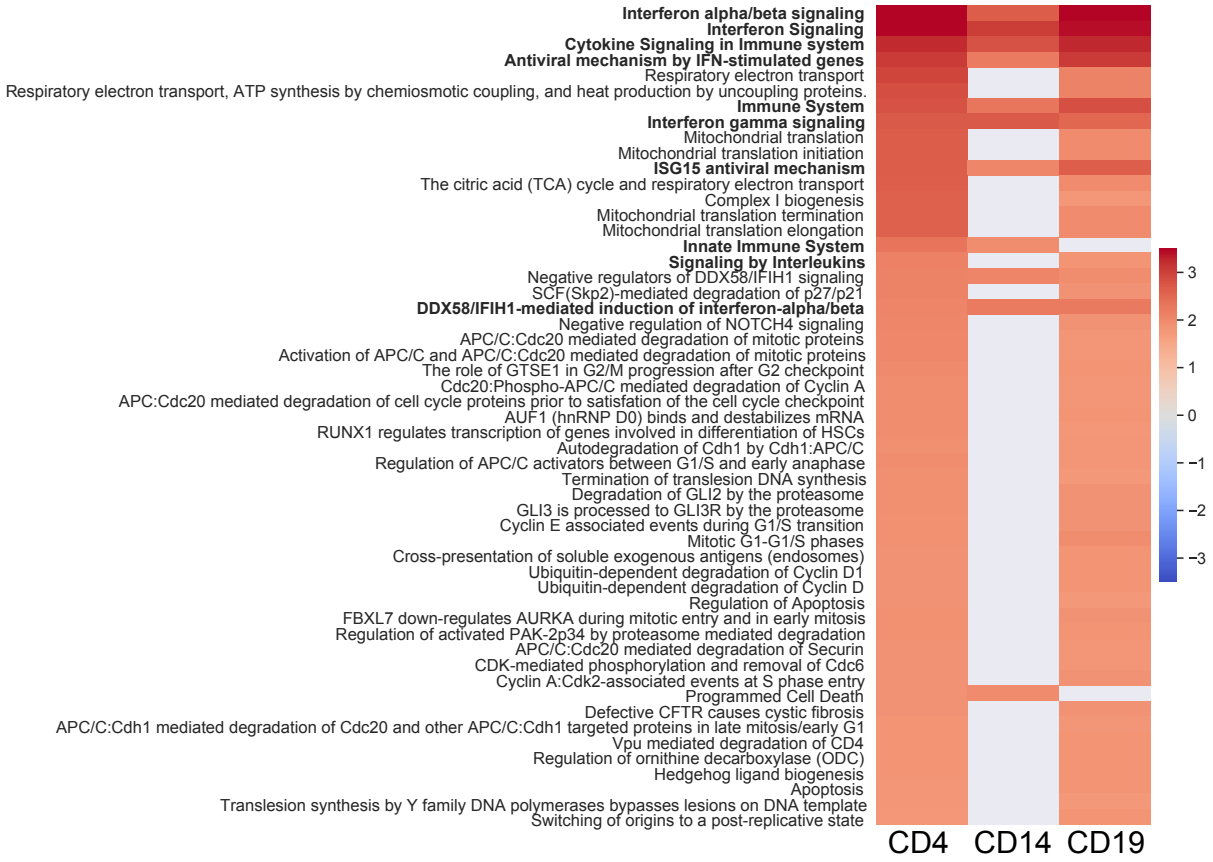

Figure 2.7 Heatmap of significantly enriched pathways Significant pathways in more than one cell type from gene set enrichment analyses with webgestalt using significant genes in the combined cohort. Heatmap shows normalized enrichment score

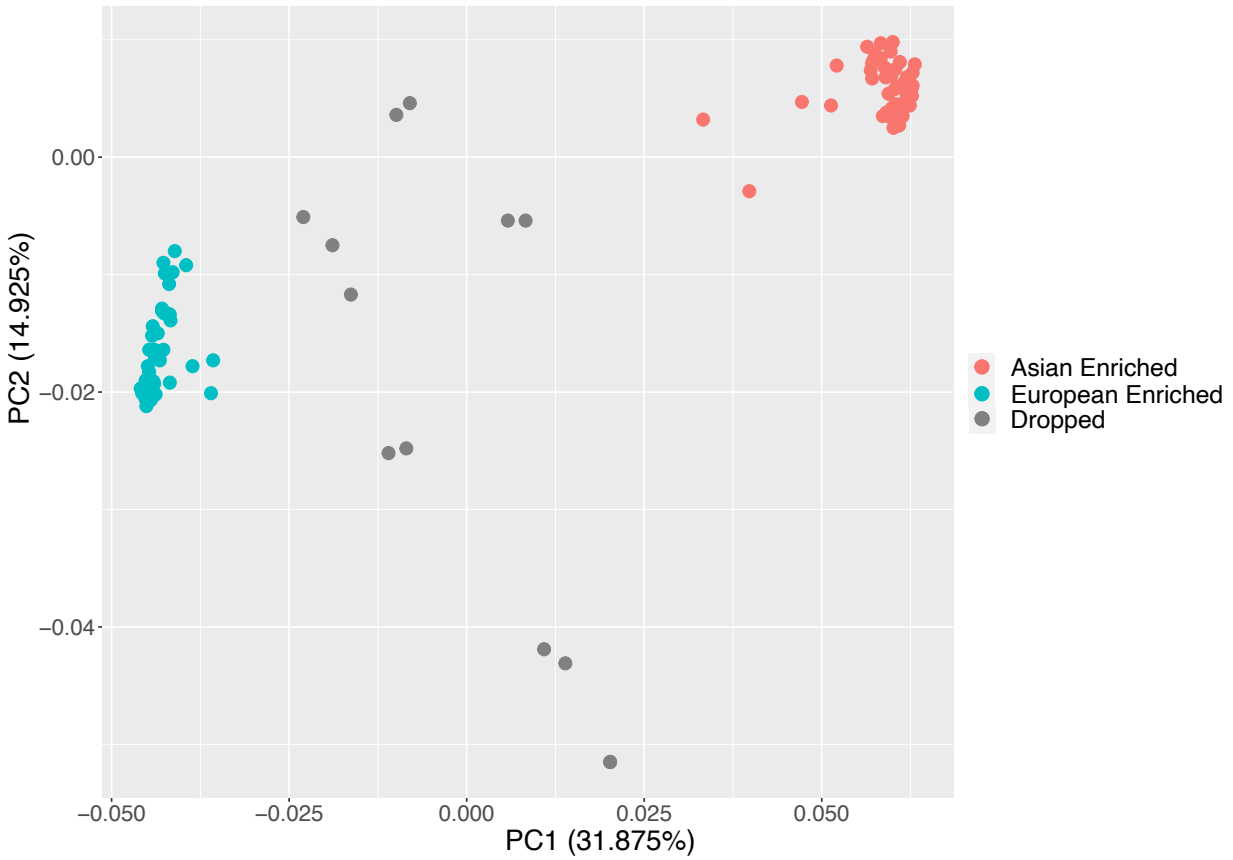

Figure 2.8 PCA plot of genotyping data across 119 patients Individuals are colored by ancestry. Individuals in gray are considered admixed and were dropped from the stratified analysis.

Chapter 3: Conclusion

Autoimmune diseases represent a significant health burden affecting millions of individuals worldwide. Understanding the underlying mechanisms driving autoimmunity is crucial for developing effective treatments and preventive strategies. Genetic and environmental factors related to immune function and viral infections have long been studied as triggers for autoimmune diseases. TEs span the divide between self and non-self which are derived from ancient infections and have emerged as intriguing candidates for triggering, modulating, and driving immune responses and contributing to disease heterogeneity.

Its thought that SLE could actually represent several different heterogeneous diseases that fall under a clinical phenotype of systemic autoimmunity, but what drives these SLE subphenotypes is unknown[83]. Therefore, in our work we investigated whether TEs might contribute to SLE heterogeneity. The phenotypes we selected fell under categories that spanned autoantibody production, disease severity, and clinical manifestations. As described previously, autoantibody production is characteristic of any autoimmune diseases. In SLE patients are often characterized and diagnosed based on the presence of autoantibodies to protein components of small nuclear ribonucleoproteins and double stranded DNA[84,85].

For over fifty years researchers have studied the link between endogenous viral proteins and reactivity with autoantibodies autoimmune diseases. These HERV proteins are hypothesized to elicit cross-reactivity through molecular mimicry between HERVs and nuclear antigens to produce antinuclear autoantibodies[28,84]. Therefore, we investigated whether there were locus-specific TEs that could be associated with autoantibody production. For patients with anti-SM, anti-RNP, and anti-dsDNA antibodies we found differentially expressed TEs across every cell type investigated. Interestingly, we found some of the highest numbers of differentially expressed TEs in CD19+ B cells, which produce antibodies. Future work could include close

examination of the differentially expressed TEs in B cells that associate with autoantibody production. This could include identifying open reading frames in those TEs and sequence homology between any HERV protein and nuclear antigens. Additionally, the CLUES cohort is deeply phenotyped, so the association of autoantibody production and molecular mimicry could be further explored with additional autoantibodies such as Anti-Ro (components of ribonucleoprotein complexes).

In addition to autoantibody production, we also investigated subphenotypes related to disease severity. Some patients with SLE only have mild symptoms, respond well to treatment, and don't develop severe organ manifestations like lupus nephritis. However, other patients can have severe rapidly worsening symptoms that become life-threatening that come and go for unknown reasons[83]. These flares in disease activity not only can cause organ damage to patients, but also the unknown trigger of these events can make living with this disease difficult for patients. Researchers have attempted to find genes responsible for the differences in disease activity, however one study comparing active to inactive disease in CD4+ T cells found zero differentially expressed genes, and only 3 in CD8+ T cells.

Researchers have also studied the correlation between HERVs and SLE disease activity and found increased HERV-E expression that correlated with disease activity[86]. Therefore we investigated whether there were locus-specific TEs that correlated with high disease activity[86]. We found mostly differentially expressed that were downregulated across all cell types in SLE patients with high disease activity, which is the opposite of what we would have expected. Upon further investigation of this signal we found that the SLEDAI cutoff of 8 to be classified a patient with high disease activity reduced our cohort to only 13 patients, whereas in the low disease activity cohort we had 113 patients. Hence, the signal we see may not be reflective of disease

activity, but rather a reflection of the number of patients we have in each cohort. Additionally, the CLUES cohort are patients who were recruited with stable disease that have responded to treatment. Therefore, we may not have enough signal to detect the locus specific HERVs associated with disease activity. Future work studying high disease activity and flares of SLE should also look at the correlation of HERVs.

In addition to disease activity, we also investigated disease severity .In previous work our group identified three lupus clinical subtypes defined mild, severe disease 1, and severe disease 2 clusters based on dimensionality reduction and clustering of American college of rheumatology (ACR) classification criteria[79]. While we found differentially expressed TEs in every comparison and cell type it is difficult to interpret. For example, when comparing severe disease 2 to mild disease in CD14 cells we found zero upregulated TEs, this is unexpected because we would expect to see a large number when comparing the most severe to least severe severity subphenotype. However, when comparing severe disease 2 to severe disease 1 in CD19 cells we found 38 upregulated TEs, which was unexpected as these two subphenotypes have overlapping ACR criteria which would make them more similar. Further work is needed to understand the patterns identified among these groups. A first step would be to look at differential expression among the individual ACR criteria before the clustering to identify which criteria are driving the differentially expressed TEs.

It is worth noting that multiple iterations of analyses were performed for the differential TE expression analysis, from only quantifying HERVs, to including different covariates. I initially only quantified HERV expression and performed differential expression analysis and found almost zero differentially expressed TEs when comparing SLE subphenotypes. However, with the strong link between SLE and both HERVs and LINEs we decided to quantify both

HERVs and LINEs. Then in our differential analysis we found thousands of differentially expressed elements in some comparisons specific to cell types, for example when comparing severe disease 2 to mild disease we found 1568 and 4100 differentially expressed TEs in CD19 and NK cells respectively.

Initially we found these results very exciting, but upon further investigation through comparisons against a null distribution, outlier samples, and including genetic ancestry PCs eliminated most of these TEs. Also, we found that the severity groups we had previously defined were characterized by race, with the severe 2 group being made up mostly of non-white, and the mild group being predominately white. This in part motivated the stratified analysis we performed.

As stated in the previous chapter, people of African, Hispanic, or Asian ancestry, compared to those of other racial or ethnic groups have increased prevalence and more serious manifestations[87]. Therefore, the stratified analysis was done to also see if there were locus-specific TEs that might contribute to heterogeneity in SLE phenotypes. Surprisingly when stratified into European enriched and Asian enriched groups by genetic ancestry PC1 there were an increased number of differentially expressed TEs in most subphenotypes and cell types. This pattern is not easily explained, and TE expression across different ancestry groups should be explored further in the context of SLE.

Next steps could include looking at TEs shared across cell types and subphenotypes and examining the overlap and different TEs between each stratified analysis and the combined cohort. Additionally, Telescope, the tool used for quantification in this analysis is reliant upon the reference used, which may not be representative for all ancestries. Additional tools for quantification should be used, especially those that take into account sequence and insertional

polymorphisms. For example, coupling whole-genome DNA sequencing and de novo assembly with RNA-seq, or long-read sequencing[32].

One of the strongest findings in our analysis was the association of the expression of cell specific TEs associated with SLE with gene expression involved in viral response. Many previous papers have looked directly at IFN and HERV or LINE expression to find associations. But in our analysis, we used an agnostic approach, only using the counts of differentially expressed TEs, we found many genes and pathways that supported the notion that the immune system interprets some TEs as viral and responds to the chronic activation of TEs as it would a viral infection.

As with the differential TE expression, it took many iterations of different analyses to come to this result. One of our initial attempts to look at the association between TE expression and gene expression used the counts of all TEs into our gene expression model, however the genes and pathways observed often had viral genes and pathways downregulated. This result was quite confusing and completely opposed to all previous work in the field. It was only when we used TEs that were differentially expressed in SLE phenotypes (i.e. those that were associated with SLE) that the pathways and genes were upregulated in response to a viral infection. While our analysis explored how TEs correlate with gene expression, it's also known that TEs act as regulators of gene expression, and future work should investigate the regulatory roles of TEs in SLE.

To investigate the role of TEs in SLE heterogeneity we examined the cell and locus-specific expression in several subphenotypes of SLE for their role in triggering autoimmune disease as nucleic acids and protein triggers and their relation to host gene expression. We quantified TE expression in four immune cell types (CD4+ T cells, monocytes, B cells, and NK

cells) from SLE patients found cell-specific and SLE subphenotype-specific differentially expressed TEs, with few overlaps across cell types or subphenotypes. Ancestry-stratified analysis revealed additional cell-type-specific TE associations in Asian and European ancestry groups. Enrichment analysis highlighted the role of specific TE families in different cell types, suggesting that TEs could contribute to the heterogeneity of SLE across individuals and populations. TE expression is associated with host gene expression involved in antiviral and immune responses, supporting the hypothesis that TEs could activate the innate immune system and contribute to chronic inflammation and autoimmunity.

This study underscores the importance of TEs in SLE heterogeneity and highlights the need for further exploration of TE expression in normal immune cells and functional studies to understand their role in SLE pathogenesis. Despite these findings, further research is needed to explore the causality between TE expression and immune dysregulation in SLE, as well as to investigate TE expression in normal immune cells and conduct functional studies to understand their role in SLE pathogenesis.

Given the potential involvement of TEs in autoimmune disease pathogenesis, targeting TE-derived pathways represents a novel therapeutic strategy for modulating immune responses and restoring immune tolerance. As this work and many others have linked retroviruses to SLE, which begs the question whether anti-retroviral drugs could help patients with SLE. Antiretroviral drugs were developed to act on different parts of the viral life-cycle, and each of these drugs alone or in combination may help to alleviate symptoms or underlying mechanisms of SLE. Some previous work has already showed that different drugs have effects on TEs such as LINEs. For example an in vitro study used stavudine, zidovudine, tenofovir disoproxil fumarate, and lamivudine, all nucleoside RT inhibitors, and found varying levels of LINE

retrotransposition[88]. In a similar study, LINEs were transfected into E. coli and nucleoside RT inhibitors were found to decrease retrotransposition and reverse transcriptase[89]. Although promising, there are some case reports of individuals developing SLE or reactivation of SLE upon highly active antiretroviral therapy, which was introduced to treat AIDS by using a combination of antiretroviral drugs[68,90]. Future work to study whether anti-retroviral drugs could reduce expression of HERVs and clinical manifestations of SLE is needed.

It remains unknown why almost all SLE patients are women, although the X chromosome houses many genes involved in immune response and appears to be a driver of autoimmune disease[87,91]. Xist is critical in silencing on of females two X chromosomes, and is thought to have originated from TEs, have sequence similarity between all its tandem repeat regions[91,92]. Xist is a non-coding RNA with high sequence similarity to HERVs, that mediates X chromosome inactivation. Spen is an Xist-binding repressor protein that recruits chromatin silencing machinery to Xist.

Research has shown that Spen loss activates ERVs in mice, that Spen binds directly to HERV RNAs, HERV RNA and Xist compete for binding of Spen. Additionally, insertion of a HERV sequence into a deficient copy of Xist rescues Spen binding to Xist which allows for local gene silencing. Future work should evaluate the role of TEs in X chromosome inactivation, including evaluation of Spen expression in female patients and correlations with HERV expression. Additionally, investigation into whether HERVs specific autoimmune disease bind Spen which could inhibit X chromosome inactivation (evaluate HERVs for A-repeat similarity to Xist). Finally, the Xist expressed in SLE patients should be analyzed for deficiencies or mutations in the A-repeat region where Spen binds and allows for X chromosome inactivation.

References

1. Wang, L., Wang, F. & Gershwin, M. E. Human autoimmune diseases: a comprehensive update. *J. Intern. Med.* **278**, 369–395 (2015).
2. Nexø, B. A. *et al.* Are human endogenous retroviruses triggers of autoimmune diseases? Unveiling associations of three diseases and viral loci. *Immunol. Res.* **64**, 55–63 (2016).
3. Pisetsky, D. S. Pathogenesis of autoimmune disease. *Nat. Rev. Nephrol.* **19**, 509–524 (2023).
4. Quaglia, M., Merlotti, G., De Andrea, M., Borgogna, C. & Cantaluppi, V. Viral Infections and Systemic Lupus Erythematosus: New Players in an Old Story. *Viruses* **13**, 277 (2021).
5. Cho, J. H. & Gregersen, P. K. Genomics and the Multifactorial Nature of Human Autoimmune Disease. *N. Engl. J. Med.* **365**, 1612–1623 (2011).
6. Zhernakova, A., Withoff, S. & Wijmenga, C. Clinical implications of shared genetics and pathogenesis in autoimmune diseases. *Nat. Rev. Endocrinol.* **9**, 646–659 (2013).
7. Mavragani, C. P. *et al.* Long interspersed nuclear element-1 retroelements are expressed in patients with systemic autoimmune disease and induce type I interferon. *Arthritis Rheumatol.* **68**, 2686–2696 (2016).
8. Klunk, J. *et al.* Evolution of immune genes is associated with the Black Death. *Nature* **611**, 312–319 (2022).
9. Vinuesa, C. G., Grenov, A. & Kassiotis, G. Innate virus-sensing pathways in B cell systemic autoimmunity. *Science* **380**, 478–484 (2023).
10. Olson, J. K., Croxford, J. L. & Miller, S. D. Virus-Induced Autoimmunity: Potential Role of Viruses in Initiation, Perpetuation, and Progression of T-Cell–Mediated Autoimmune Disease. *Viral Immunol.* **14**, 227–250 (2001).
11. Crowl, J. T., Gray, E. E., Pestal, K., Volkman, H. E. & Stetson, D. B. Intracellular Nucleic Acid Detection in Autoimmunity. *Annu. Rev. Immunol.* **35**, 313–336 (2017).

12. Grandi, N. & Tramontano, E. Human Endogenous Retroviruses Are Ancient Acquired Elements Still Shaping Innate Immune Responses. *Front. Immunol.* **9**, 2039 (2018).
13. Powell, A. M. & Black, M. M. Epitope spreading: protection from pathogens, but propagation of autoimmunity?: Epitope spreading. *Clin. Exp. Dermatol.* **26**, 427–433 (2001).
14. International Human Genome Sequencing Consortium *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
15. Wells, J. N. & Feschotte, C. A Field Guide to Eukaryotic Transposable Elements. *Annu. Rev. Genet.* **54**, 539–561 (2020).
16. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).
17. Arkhipova, I. R. Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. *Mob. DNA* **8**, 19 (2017).
18. Posso-Osorio, I., Tobón, G. J. & Cañas, C. A. Human endogenous retroviruses (HERV) and non-HERV viruses incorporated into the human genome and their role in the development of autoimmune diseases. *J. Transl. Autoimmun.* **4**, 100137 (2021).
19. Grandi, N. & Tramontano, E. HERV Envelope Proteins: Physiological Role and Pathogenic Potential in Cancer and Autoimmunity. *Front. Microbiol.* **9**, 462 (2018).
20. Levet, S. *et al.* An ancestral retroviral protein identified as a therapeutic target in type-1 diabetes. *JCI Insight* **2**, e94387 (2017).
21. Morandi, E. *et al.* The association between human endogenous retroviruses and multiple sclerosis: A systematic review and meta-analysis. *PLOS ONE* **12**, e0172415 (2017).
22. Kremer, D. *et al.* pHERV-W envelope protein fuels microglial cell-dependent damage of myelinated axons in multiple sclerosis. *Proc. Natl. Acad. Sci.* **116**, 15216–15225 (2019).

23. Kremer, D. *et al.* The neutralizing antibody GNbAC1 abrogates HERV-W envelope protein-mediated oligodendroglial maturation blockade. *Mult. Scler. J.* **21**, 1200–1203 (2015).
24. Hartung, H.-P. *et al.* Efficacy and safety of temelimab in multiple sclerosis: Results of a randomized phase 2b and extension study. *Mult. Scler. J.* **28**, 429–440 (2022).
25. Ejtehadi, H. D. The potential role of human endogenous retrovirus K10 in the pathogenesis of rheumatoid arthritis: a preliminary study. *Ann. Rheum. Dis.* **65**, 612–616 (2006).
26. Freimanis, G. *et al.* A role for human endogenous retrovirus-K (HML-2) in rheumatoid arthritis: investigating mechanisms of pathogenesis. *Clin. Exp. Immunol.* **160**, 340–347 (2010).
27. Perl, A. *et al.* Antibody reactivity to the hres-1 endogenous retroviral element identifies a subset of patients with systemic lupus erythematosus and overlap syndromes. *Arthritis Rheum.* **38**, 1660–1671 (1995).
28. Piotrowski, P. C., Duriagin, S. & Jagodzinski, P. P. 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.* **24**, 620–624 (2005).
29. Hishikawa, T. *et al.* Detection of Antibodies to a Recombinant gag Protein Derived from Human Endogenous Retrovirus Clone 4-1 in Autoimmune Diseases. *Viral Immunol.* **10**, 137–147 (1997).
30. Ogasawara, H. *et al.* Quantitative Analyses of Messenger RNA of Human Endogenous Retrovirus in Patients with Systemic Lupus Erythematosus. *J. Rheumatol.*
31. Fali, T. *et al.* DNA methylation modulates HRES1/p28 expression in B cells from patients with Lupus. *Autoimmunity* **47**, 265–271 (2014).

32. Lanciano, S. & Cristofari, G. Measuring and interpreting transposable element expression. *Nat. Rev. Genet.* **21**, 721–736 (2020).

33. Bourgeois, Y. & Boissinot, S. On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. *Genes* **10**, 419 (2019).

34. Kelley, D. & Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* **13**, R107 (2012).

35. Kapusta, A. *et al.* Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. *PLoS Genet.* **9**, e1003470 (2013).

36. Deininger, P. *et al.* A comprehensive approach to expression of L1 loci. *Nucleic Acids Res.* **45**, e31–e31 (2017).

37. Navarro, F. C. *et al.* TeXP: Deconvolving the effects of pervasive and autonomous transcription of transposable elements. *PLOS Comput. Biol.* **15**, e1007293 (2019).

38. Belancio, V. P., Roy-Engel, A. M., Pochampally, R. R. & Deininger, P. Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.* **38**, 3909–3922 (2010).

39. Skowronski, J. & Singer, M. F. Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line. *Proc. Natl. Acad. Sci.* **82**, 6050–6054 (1985).

40. Reichmann, J. *et al.* Microarray Analysis of LTR Retrotransposon Silencing Identifies Hdac1 as a Regulator of Retrotransposon Expression in Mouse Embryonic Stem Cells. *PLoS Comput. Biol.* **8**, e1002486 (2012).

41. Seifarth, W. *et al.* Comprehensive Analysis of Human Endogenous Retrovirus Transcriptional Activity in Human Tissues with a Retrovirus-Specific Microarray. *J. Virol.* **79**, 341–352 (2005).

42. Faulkner, G. J. *et al.* A rescue strategy for multimapping short sequence tags refines surveys of transcriptional activity by CAGE. *Genomics* **91**, 281–288 (2008).

43. Chung, N. *et al.* Transcriptome analyses of tumor-adjacent somatic tissues reveal genes co-expressed with transposable elements. *Mob. DNA* **10**, 39 (2019).

44. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).

45. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

46. Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593–3599 (2015).

47. Bendall, M. L. *et al.* Telescope: Characterization of the retrotranscriptome by accurate estimation of transposable element expression. *PLOS Comput. Biol.* **15**, e1006453 (2019).

48. Yang, W. R., Ardeljan, D., Pacyna, C. N., Payer, L. M. & Burns, K. H. SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Res.* **47**, e27–e27 (2019).

49. Crow, M. K. Pathogenesis of systemic lupus erythematosus: risks, mechanisms and therapeutic targets. *Ann. Rheum. Dis.* ard-2022-223741 (2023) doi:10.1136/ard-2022-223741.

50. Andreoletti, G. *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.* **4**, 488 (2021).

51. Maningding, E., Dall'Era, M., Trupin, L., Murphy, L. B. & Yazdany, J. 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.* **72**, 622–629 (2020).
52. Han, G.-M. *et al.* Analysis of gene expression profiles in human systemic lupus erythematosus using oligonucleotide microarray. *Genes Immun.* **4**, 177–186 (2003).
53. Bennett, L. *et al.* Interferon and Granulopoiesis Signatures in Systemic Lupus Erythematosus Blood. *J. Exp. Med.* **197**, 711–723 (2003).
54. Kwon, Chun, Kim, & Mak. Update on the Genetics of Systemic Lupus Erythematosus: Genome-Wide Association Studies and Beyond. *Cells* **8**, 1180 (2019).
55. Oparina, N., Martínez-Bueno, M. & Alarcón-Riquelme, M. E. An update on the genetics of systemic lupus erythematosus. *Curr. Opin. Rheumatol.* **31**, 659–668 (2019).
56. Kelly, M. *et al.* Transposable element dysregulation in systemic lupus erythematosus and regulation by histone conformation and Hsp90. *Clin. Immunol.* **197**, 6–18 (2018).
57. Mustelin, T., Lood, C. & Giltiay, N. V. Sources of Pathogenic Nucleic Acids in Systemic Lupus Erythematosus. *Front. Immunol.* **10**, 1028 (2019).
58. Kojima, S. *et al.* Mobile element variation contributes to population-specific genome diversification, gene regulation and disease risk. *Nat. Genet.* **55**, 939–951 (2023).
59. Shao, W. & Wang, T. Transcript assembly improves expression quantification of transposable elements in single-cell RNA-seq data. *Genome Res.* **31**, 88–100 (2021).
60. Tokuyama, M. *et al.* ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. *Proc. Natl. Acad. Sci.* **115**, 12565–12572 (2018).

61. Ukadike, K. C. & Mustelin, T. Implications of Endogenous Retroelements in the Etiopathogenesis of Systemic Lupus Erythematosus. *J. Clin. Med.* **10**, 856 (2021).
62. Hurst, T. P. & Magiorkinis, G. Activation of the innate immune response by endogenous retroviruses. *J. Gen. Virol.* **96**, 1207–1218 (2015).
63. Jönsson, M. E., Garza, R., Johansson, P. A. & Jakobsson, J. Transposable Elements: A Common Feature of Neurodevelopmental and Neurodegenerative Disorders. *Trends Genet.* **36**, 610–623 (2020).
64. Tokuyama, M. *et al.* Antibodies against human endogenous retrovirus K102 envelope activate neutrophils in systemic lupus erythematosus. *J. Exp. Med.* **218**, e20191766 (2021).
65. Wang, X. *et al.* Increased HERV-E clone 4–1 expression contributes to DNA hypomethylation and IL-17 release from CD4+ T cells via miR-302d/MBD2 in systemic lupus erythematosus. *Cell Commun. Signal.* **17**, 94 (2019).
66. Ogasawara, H. *et al.* Sequence Analysis of Human Endogenous Retrovirus Clone 4-1 in Systemic Lupus Erythematosus. *Autoimmunity* **33**, 15–21 (2001).
67. Stearrett, N. *et al.* Expression of Human Endogenous Retroviruses in Systemic Lupus Erythematosus: Multiomic Integration With Gene Expression. *Front. Immunol.* **12**, 661437 (2021).
68. Talotta, R., Atzeni, F. & Laska, M. J. Retroviruses in the pathogenesis of systemic lupus erythematosus: Are they potential therapeutic targets? *Autoimmunity* **53**, 177–191 (2020).
69. Nakkuntod, J., Avihingsanon, Y., Mutirangura, A. & Hirankarn, N. Hypomethylation of LINE-1 but not Alu in lymphocyte subsets of systemic lupus erythematosus patients. *Clin. Chim. Acta* **412**, 1457–1461 (2011).

70. Sukapan, P., Promnarate, P., Avihingsanon, Y., Mutirangura, A. & Hirankarn, N. Types of DNA methylation status of the interspersed repetitive sequences for LINE-1, Alu, HERV-E and HERV-K in the neutrophils from systemic lupus erythematosus patients and healthy controls. *J. Hum. Genet.* **59**, 178–188 (2014).
71. Lee, M. H. *et al.* A meta-analysis of clinical manifestations in asian systemic lupus erythematous: The effects of ancestry, ethnicity and gender. *Semin. Arthritis Rheum.* **52**, 151932 (2022).
72. Talotta, R., Atzeni, F. & Laska, M. J. The contribution of HERV-E clone 4-1 and other HERV-E members to the pathogenesis of rheumatic autoimmune diseases. *APMIS* **128**, 367–377 (2020).
73. Aibara, S., Katahira, J., Valkov, E. & Stewart, M. The principal mRNA nuclear export factor NXF1:NXT1 forms a symmetric binding platform that facilitates export of retroviral CTE-RNA. *Nucleic Acids Res.* **43**, 1883–1893 (2015).
74. Wang, Y. *et al.* Hsp70 Is a Potential Therapeutic Target for Echovirus 9 Infection. *Front. Mol. Biosci.* **7**, 146 (2020).
75. Mellors, R. C. & Mellors, J. W. Type C RNA virus-specific antibody in human systemic lupus erythematosus demonstrated by enzymoimmunoassay. *Proc. Natl. Acad. Sci.* **75**, 2463–2467 (1978).
76. Autoimmune Disease Center, The Feinstein Institute for Medical Research, 350 Community Drive, Manhasset, New York, USA & Mishra, R. K. Involvement of NF-κB Signaling Pathway in the Pathogenesis of Systemic Lupus Erythematosus. *Nephrol. - Open J.* **2**, 9–13 (2016).

77. Santoro, M. G. NEW EMBO MEMBER'S REVIEW: NF-kappaB and virus infection: who controls whom. *EMBO J.* **22**, 2552–2560 (2003).

78. Buyon, J. P. *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.* **142**, 953–962 (2005).

79. Lanata, C. M. *et al.* A phenotypic and genomics approach in a multi-ethnic cohort to subtype systemic lupus erythematosus. *Nat. Commun.* **10**, 3902 (2019).

80. Hoffmann, T. J. *et al.* Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. *Genomics* **98**, 422–430 (2011).

81. Kennedy, W. P. *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.* **2**, e000080–e000080 (2015).

82. Min, X. *et al.* Ultraviolet light induces HERV expression to activate RIG-I signalling pathway in keratinocytes. *Exp. Dermatol.* exd.14568 (2022) doi:10.1111/exd.14568.

83. Tsokos, G. C., Lo, M. S., Reis, P. C. & Sullivan, K. E. New insights into the immunopathogenesis of systemic lupus erythematosus. *Nat. Rev. Rheumatol.* **12**, 716–730 (2016).

84. Query, C. A human autoimmune protein associated with U1 RNA contains a region of homology that is cross-reactive with retroviral p30gag antigen. *Cell* **51**, 211–220 (1987).

85. Migliorini, P., Baldini, C., Rocchi, V. & Bombardieri, S. Anti-Sm and anti-RNP antibodies. *Autoimmunity* **38**, 47–54 (2005).

86. Wu, Z. *et al.* DNA methylation modulates HERV-E expression in CD4+ T cells from systemic lupus erythematosus patients. *J. Dermatol. Sci.* **77**, 110–116 (2015).

87. Tsokos, G. C. Autoimmunity and organ damage in systemic lupus erythematosus. *Nat. Immunol.* **21**, 605–614 (2020).

88. Jones, R. B. *et al.* Nucleoside Analogue Reverse Transcriptase Inhibitors Differentially Inhibit Human LINE-1 Retrotransposition. *PLoS ONE* **3**, e1547 (2008).

89. Dai, L., Huang, Q. & Boeke, J. D. Effect of reverse transcriptase inhibitors on LINE-1 and Ty1 reverse transcriptase activities and on LINE-1 retrotransposition. *BMC Biochem.* **12**, 18 (2011).

90. Drake, W. P., Byrd, V. M. & Olsen, N. J. Reactivation of Systemic Lupus Erythematosus after Initiation of Highly Active Antiretroviral Therapy for Acquired Immunodeficiency Syndrome: *JCR J. Clin. Rheumatol.* **9**, 176–180 (2003).

91. Dou, D. R. *et al.* Xist ribonucleoproteins promote female sex-biased autoimmunity. *Cell* **187**, 733-749.e16 (2024).

92. Carter, A. C. *et al.* Spen links RNA-mediated endogenous retrovirus silencing and X chromosome inactivation. *eLife* **9**, e54508 (2020).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Zachary Cutts
FE33B6FFA4B5457...
Author Signature

5/24/2024
Date