**Title:** Mapping the plasma proteomic architecture of systemic lupus erythematosus

**Authors**

Geoffrey H. D. Leung[1]

Charlotte Bottomley[1]

Norzawani Buang[1]

Robert T. Maughan[1]

Benjamin J. Whittle[1]

Boroumand Zeidaabadi[1]

Yun-Ju Huang[1,2]

Tabitha Turner-Stokes[1,3]

Marie Condon[3]

Liz Lightstone[1,3]

Tom Cairns[3]

Marina Botto*[1,3]

Matthew C. Pickering*[1,3]

James E. Peters*[1,3]

* co-senior authors

Correspondence to: j.peters@imperial.ac.uk

Affiliations

1. Department of Immunology and Inflammation, Imperial College London, London, UK
2. School of Medicine, Chang Gung University and Division of Rheumatology, Allergy and Immunology, Chang Gung Memorial Hospital, Taoyuan, Taiwan
3. Imperial Lupus Centre, Imperial College Healthcare NHS Trust, London, UK

## Key messages:

- **What is already known on this topic**

  SLE has a complex pathogenesis, involving dysregulation of both the innate and adaptive immune system. Genome-wide association studies have elucidated genetic risk factors for SLE susceptibility. Transcriptomic studies have identified molecular signatures of SLE including upregulation of genes downstream of type 1 interferon signalling. However, few studies have comprehensively surveyed the plasma proteome in SLE. The plasma proteome provides complementary information to the transcriptome. Plasma proteins include important immune effector molecules and also provide a read-out of tissue injury.

- **What this study adds**

  This study represents the most comprehensive assessment of the plasma proteome in SLE to date (7,288 protein analytes mapping to 6,595 unique protein targets). We identified dysregulated plasma proteins and pathways in SLE, protein signatures associated with disease activity, modules of proteins associated with interferon signalling and renal dysfunction, and distinct proteomic patterns correlated with anti-Smith and anti-dsDNA antibodies, even after controlling for disease activity.

- **How this study might affect research, practice or policy?**

  By identifying the proteins dysregulated in SLE and those associated with disease activity, this study provides insights into pathogenesis, identifies and highlights putative therapeutic targets. Our observation of distinct plasma proteomic signatures linked to specific autoantibodies suggests the possibility of precision medicine approaches in SLE.

## Abstract

Objectives

To characterise changes in circulating proteins associated with systemic lupus erythematosus (SLE) and identify proteins associated with clinical traits including disease activity and serological parameters.

Methods

Proteomic profiling of plasma samples from patients with SLE (n=260) and healthy volunteers (HVs) (n=86) was performed using SomaScan v4.1, an aptamer-based proteomic assay which measures 7,288 analytes. We performed differential protein abundance analysis comparing SLE and HVs using linear models in a discovery set and tested replication in a validation set. Within-cases analyses were performed to test for associations with disease activity and autoantibody status. We used network analysis to identify protein modules, which were then tested for association with clinical traits.

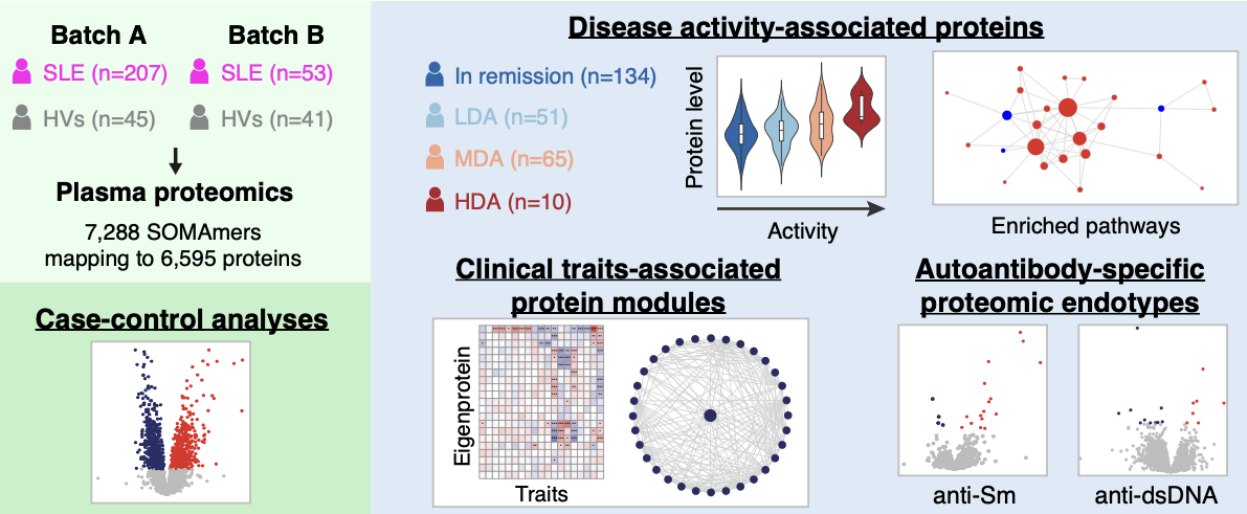
Results

215 proteins were robustly differentially abundant between SLE patients and HVs in both discovery and validation sets, including interferon-stimulated proteins. Within-cases analyses identified 421 proteins associated with disease activity. Network analysis revealed modules of correlated proteins, including an “interferon-associated” module and a “renal-related” module, each linked to distinct clinical features. The presence of anti-Sm autoantibodies was associated with elevated levels of interferon-stimulated proteins (e.g. MX1, ISG15, CXCL10), alongside decreased levels of circulating small ribonuclear proteins, independent of disease activity. Anti-dsDNA antibodies were associated with elevated levels of CD40 ligand (CD40LG) and the neutrophil protease proteinase-3. CD40LG’s association with disease activity was restricted to the dsDNA antibody positive subset.

Conclusions

Our study identified plasma protein signatures of SLE and disease activity, and revealed distinct proteomic profiles associated with the presence of specific autoantibodies.

## Graphical abstract



## Introduction

Systemic lupus erythematosus (SLE) is an autoimmune disease with complex pathogenesis involving dysregulation of both adaptive and innate immunity. A hallmark of the disease is the production of autoantibodies directed against nuclear antigens. Established pathogenic pathways include autoreactive B cells, the complement system and type 1 interferon signalling which lead to tissue inflammation and injury. Clinically, SLE is highly heterogeneous, ranging from mild to organ- or life-threatening disease, and with variable patterns of organ involvement.

The advent of genomic technologies has enabled comprehensive assessment of various molecular domains in SLE such as the genome or transcriptome. Genome-wide association studies (GWAS) have elucidated genetic risk factors for SLE [1-3], while transcriptomic studies have revealed the dysregulated gene expression programmes that occur once disease is established [4-6]. A key finding from the latter was the identification of the interferon-stimulated gene (ISG) signature, a gene expression signature indicating stimulation by interferons, particularly type 1 interferons [4, 7-9]. By contrast, few studies have investigated the proteome in SLE [10-12]. Proteins are the effector molecules of biology and the targets of most drugs. In addition, circulating proteins (measurable in plasma or serum) can provide useful and tractable biomarkers suitable for measurement in clinical practice. Thus, evaluation of proteomic changes in SLE could be valuable both for understanding pathogenesis and for translational research [13]. Advances in proteomic technologies now allow for highly multiplexed protein measurements. Here we performed plasma proteomic profiling using the SomaScan platform, measuring 7,288 protein analytes. Our data identify differentially abundant proteins in SLE vs HVs, protein signatures of active disease, and modules of correlated proteins associated with clinical traits. Moreover, we identify auto-antibody specific proteomic endotypes, with distinct proteomic patterns correlated with Smith and dsDNA auto-antibody positivity.

## Methods

Subjects and Samples

We recruited 268 patients with SLE from the Imperial Lupus Centre at the Hammersmith Hospital, London, United Kingdom (UK). All patients were anti-nuclear antibody positive and fulfilled the 1997 American College of Rheumatology criteria. We recruited 86 healthy volunteers (HVs) to provide a control group, sampled as two temporally distinct groups comprising 45 and 41 individuals, respectively. No patients had received anifrolumab since this drug is not licensed in the UK.

Human samples used in this research project were obtained from the Imperial College Healthcare Tissue and Biobank (ICHTB). ICHTB is supported by the National Institute for Health Research Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. ICHTB is approved by Wales REC3 to release human material for research (22/WA/0214). Ethical approval was provided by the ICHTB (Human Tissue Authority Licensing number 12275; Project number R14042-3A; sub-collection IMM_MB_13_001) and informed consent obtained from all participants. Peripheral blood was collected in EDTA tubes and centrifuged at 1,000G for 10 minutes at room temperature within three hours of venepuncture and stored at -70$^{o}$C without freeze-thaw before assay processing.

Patient and public involvement

Patients with SLE were involved in the conception and design of this study. The research proposal was reviewed by the NIHR Imperial BRC Community Partners group, composed of patients with immune-mediated diseases including SLE cared for at Imperial College Healthcare NHS Trust. Their feedback helped shape the study focus and ensure that research priorities aligned with unmet needs. Patients were not involved in the conduct, data analysis, or writing of the manuscript.

Proteomic measurements and quality assessment

Plasma proteome was measured using the aptamer-based SomaScan v4.1 platform (Boulder, Colorado, USA). The SomaScan v4.1 platform uses modified aptamers (SOMAmers) that bind to their targeting proteins for measurement. A total of 7,288 SOMAmers targeting 6,595 unique human proteins were measured (some proteins were assayed by multiple distinct SOMAmers; **Supplementary Table S1**). To avoid interference of anti-dsDNA autoantibodies with the SOMAmers, a validated anti-DNA antibody protocol was used [10]. This involved adding herring sperm DNA to saturate anti-dsDNA autoantibodies within each sample and prevent these antibodies from binding the SOMAmers. The platform contains buffer replicates containing no protein which are used to estimate the background noise level, calibrator replicates to control for intra-plate variations and batch effects across experiments, and quality control (QC) replicates for

signal normalisation and QC check. Multiple normalisation and calibration steps were performed using SomaLogic's standard protocol. Data normalised by adaptive normalisation by maximum likelihood was used for our analyses. Relative fluorescence units were $\log_2$-transformed before analyses. Proteomic measurements were performed in two separate batches: batch A consisted of 226 SLE and 45 healthy samples and batch B of 57 SLE and 41 HV samples.

In addition to the manufacturer's quality assessment, we performed additional steps including PCA and inspection of sample distributions via boxplots. We identified batch effects, with separation of Batch A and Batch B on PCA. **(Supplementary Figures 1A-B)**. After manual examination on the PCA plot, three samples from Batch A were removed since they were distinctly distributed from the main cohort of samples (>3 standard deviations from the means of PC1 and PC2) and also processed on the same day, suggesting technical rather than biological variation (**Supplementary Figure 1A**). Additionally, 1 sample from batch A and 4 samples from batch B were removed since they were flagged as potential poor quality samples by SomaScan's standard QC pipeline requiring extremely high (>2.5) or low (<0.4) scaling factors for normalisation. For 3 patients, serial longitudinal samples were available. In these instances, we retained the first sample for each patient and excluded the others from all downstream analyses except for the evaluation of intra-individual changes over time in disease activity-associated markers (see below). For the main analyses (excluding serial samples), post QC samples sizes were as follows: Batch A: 207 SLE, 45 HVs and Batch B: 53 SLE, 41 HVs.

<u>Protein nomenclature</u>

Due to non-standardisation of protein nomenclature (e.g. where a given protein has multiple names) and the lack of human-interpretability of Uniprot identifiers, we annotated each protein using the non-italicised gene HUGO symbol of the encoding gene. In instances where there were multiple distinct SOMAmers mapping to the same HUGO gene symbol (1,669 SOMAmers mapped to 781 unique gene symbols and 782 Uniprot identifiers), we concatenated the gene symbol with the SOMAmer identifier to preserve a unique identifier for each analyte. Importantly, in some instances, multiple SOMAmers mapping to the same gene symbol represented measurements of distinct biological entities. For example, the intact complement C3 protein and its distinct cleavage products such as C3a, C3b, and C3d, all mapped to a single HUGO gene symbol and Uniprot ID. However, in other instances there were multiple SOMAmers that are nominally directed against the same target protein. A comprehensive annotation file mapping of SOMAmers to their target proteins, Uniprot IDs and corresponding gene IDs, and gene symbols is provided in **Supplementary Table S1**.

Complement proteins

Various complement proteins and cleavage fragments were included on the SomaScan platform. To evaluate the accuracy of complement C3 and C4 quantification, we compared the levels of intact complement C3 and C4 proteins measured by SomaScan in SLE patients to their corresponding values measured in the clinical laboratory from the same blood draw. We performed this analysis separately for each batch to avoid variation related to batch.

Identification of differentially abundant proteins in SLE vs healthy samples

For each protein measured, multiple linear regression was used to analyse differences in protein abundance between SLE and HVs samples using the lm() function in R. Age and sex were adjusted for by including them as covariates. The regression model in Wilkinson notation was:

$$P \sim D + age + sex$$

where P was plasma level for a given protein and D was a binary variable disease status (SLE vs healthy).

In Batch B, we observed inflation of test statistics and an unusually high proportion of proteins that were significantly differentially abundant in the comparison of SLE vs HVs (**Supplementary Figures 2-3)**. We also observed marked separation of SLE and HV samples on PC1 (when PCA was performed limiting to Batch B; **Supplementary Figure 4**), with PC1 accounting for 42% of the variance. The most upregulated protein was PGAM1, which has been identified by Somalogic as a marker of pre-analytical variation including time to spin, suggesting that systematic non-biological variation was confounding the differences between SLE and HVs in Batch B (**Supplementary Figure 2)**. We therefore included the first principal component as a covariate to adjust for pre-analytical variation in Batch B (see **Supplementary Note**).

Statistical significance

P values were adjusted using the Benjamini-Hochberg (BH) method ($P_{BH}$) to control the false discovery rate (FDR) across multiple tests. We defined statistical significance as $P_{BH} < 0.05$ (i.e. leading to an FDR <0.05) for all analyses in this study.

Identification of proteins associated with disease activity

Disease activity was assessed using the SLE Disease Activity Index (SLEDAI)-2K [14]. Given that SLEDAI-2K scores have a highly non-normal distribution and can be sensitive to the weightings attributed to specific clinical features, prior to testing for association with proteins, we categorised disease activity into a four level ordinal variable as follows: ‘high disease activity’ (HDA) was defined as SLEDAI-2K ≥ 10; ‘moderate disease activity’ (MDA) as SLEDAI-2K between 5 and 9; ‘low disease activity’ (LDA) as SLEDAI-2K ≤ 4

but not meeting the definition for remission; 'remission' as defined by the Definition of Remission in SLE (DORIS) criteria [15], characterized by clinical components of SLEDAI-2K = 0 (i.e. not including dsDNA antibody or complement status) and a prednisolone dose ≤ 5 mg/day.

To identify proteins associated with disease activity, we performed multiple linear regression for each protein using the lm() function in R. Protein abundance was modelled as the dependent variable, with disease activity (coded as an ordinal variable: "In remission", "LDA", "MDA", "HDA"), sex, and batch as independent predictors. As a sensitivity analysis, Spearman's rank correlation was conducted between protein abundance and SLEDAI-2K scores. Prior to correlation testing, protein abundances were adjusted to account for sex and batch effects using a linear model and the resulting residuals were used in the Spearman's correlation test versus SLEDAI-2K scores.

Exploratory Longitudinal analysis

A subset of disease activity–associated proteins was selected for further investigation in a longitudinal analysis of patients (n=3) with samples taken at four or more timepoints. We first selected proteins that were: (i) potential drug targets, defined as having an associated compound with a characterized mechanism of action in Open Targets (Dataset: "Drug - mechanism of action"; data version 25.03), and (ii) were members of the pathways significantly enriched among disease activity–associated proteins. From this list we then selected the 5 proteins with the greatest effect sizes in the ordinal linear regression analysis testing for association with disease activity. These criteria resulted in 5 proteins, one of which was an interferon-stimulated protein (CXCL10). In addition, we added two interferon-stimulated proteins that were significantly associated with disease activity. Repeated measures correlation was used to assess the association between these proteins and clinically measured disease activity traits, accounting for the non-independence of observations due to repeated measurements within the same individuals [16]. This approach is more appropriate than Pearson's correlation, which assumes independent observations.

Identification of proteins associated with autoantibody status

We analysed proteomic associations with the presence of antibodies to Smith (Sm), anti-dsDNA and phospholipids. For each plasma protein, abundance (P) was regressed on antibody status (Ab; a binary variable) with batch and sex included as covariates. The regression model in Wilkinson notation was P ~ Ab + batch + sex. To identify associations with Sm antibodies independent of disease activity, we then repeated the analysis adding disease activity status as a covariate to the model.

For the dsDNA analysis, samples were classified as anti-dsDNA negative if the level was lower than 10 IUperML (the manufacturer's defined threshold) and as anti-dsDNA positive if the patient's plasma anti-dsDNA autoantibodies level was higher than 20 IUperML. Patients with anti-dsDNA levels (n=40) within the borderline range of 10-20 IUperML were excluded in this analysis. To identify proteins associated with anti-dsDNA positivity independent of disease activity, we used a modified SLEDAI-2K score excluding the anti-dsDNA component from the score to avoid over-adjustment since SLEDAI-2K includes 2 points for dsDNA antibody positivity. We re-calibrated the disease activity category thresholds between LDA, MDA and HDA based on this modified score (DORIS remission definition was unchanged since this does not include anti-dsDNA status). Specifically we normalised the modified SLEDAI-2K by the highest possible score excluding dsDNA antibody positivity (i.e., 105-2 = 103), and applied the same procedure to the unmodified SLEDAI-2K thresholds used to define activity categories described earlier. To illustrate this procedure, an individual positive for anti-dsDNA antibodies with an unmodified SLEDAI-2K score of 11 would have been defined as being in HDA (defined as SLEDAI-2K of 10 or higher) in the earlier analyses. The modified SLEDAI-2K score would be 9. The normalised score would be 9/103=0.087. The new cut-off for HDA would be 10/103=0.097. Thus this individual would be classified as in MDA for the purposes of covariate adjustment in the analysis testing for proteins associated with anti-dsDNA antibodies independent of disease activity.

As a sensitivity analysis, we tested protein abundances against anti-dsDNA titre (as a continuous variable) using all available SLE samples. To do this, protein abundances were first adjusted to account for sex, batch and disease activity (as an ordinal variable) using a linear model and the resulting residuals were used in the Spearman's correlation test versus anti-dsDNA titre.

To formally test whether CD40LG's association with disease activity was dependent on dsDNA antibody status, we fitted a regression model with an dsDNA antibody status as a binary variable (Ab) x disease activity interaction term:

CD40LG ~ Ab + disease_activity + Ab:disease_activity + batch + sex

A significant interaction term indicates that the effect of disease activity on the protein is dependent on dsDNA antibody status.

<u>Pathway enrichment analyses</u>

GeneSet Enrichment Analysis (GSEA) was performed using the fgsea R package (version 1.24.0) [17] for SLE-vs-HVs comparisons and disease activity analysis using the Human MSigDB Collection "C2: curated gene sets" (sub-collection: "Canonical

pathways") which consists of 4,023 gene sets (v2025.1) mainly sourced and curated from BioCarta, Kyoto Encyclopedia of Genes and Genomes, Pathway Interaction Database, Reactome, and WikiPathways [18]. For SLE-vs-HVs comparisons, robust-rank aggregation (RRA) [19] was used to integrate GSEA results to identify pathways that were ranked higher than expected across two batches. RRA outputs a score analogous to a P-value for each pathway term. These scores were -log10 transformed to represent the aggregated score. A higher aggregated score indicates stronger consistency.

Over-representation analyses were performed using the 'enrichPathway' function from the ReactomePA R package (v1.42.0) [20] for analysing Reactome pathways enriched in protein modules identified by WGCNA, and proteins associated with anti-Sm autoantibodies. Proteins were first mapped to corresponding Entrez IDs since this package requires Entrez ID annotation. Entrez IDs mapped to proteins in the module were tested against the background set of 6,363 Entrez IDs representing the proteins measured on SomaScan v4.1. All pathway enrichment analyses were restricted to protein sets containing 10–500 proteins to exclude very small sets vulnerable to noise and large sets representing broad and non-specific processes.

Protein network analysis

We applied the Weighted Gene Co-expression Network Analysis (WGCNA) package (v1.72-5) [21] to identify modules of correlated proteins. For this analysis we used SLE samples (n=207) from the larger Batch A. A soft thresholding power of 6 was selected based on the criterion of scale-free topology and mean connectivity. Networks were constructed and modules were identified by the adjacency and topology overlap matrices, with minimum module size set to 20. We then refined the modules by their similarity based on modular eigenprotein (representing the first principal component of each module's protein abundance profile) correlation using a dynamic tree-cut approach. Modules were merged if their eigenprotein-based dissimilarity was below a threshold of 0.25. The resulting modular eigenproteins were then used to correlate with 20 clinical traits using Spearman's rank correlation test, including 11 continuous traits: age, SLEDAI-2K, neutrophil count, lymphocyte count, haemoglobin, creatinine, urine protein-creatinine ratio (uPCR), albumin, complement C3, complement C4, and whole-blood ISG score (see below), and 9 binary traits: positivity of antibodies against cardiolipin (ACA), RNP-A, RNP-68, Sm, La, Ro52, Ro60, dsDNA, and sex. Correlations with $P_{BH} < 0.05$ were considered significant. Eigenprotein-based connectivity, or module membership, was calculated for each protein in each module by correlation between the protein abundance levels and the corresponding module's eigenprotein values. The protein with highest module membership was then defined as the hub protein for each module. The network of the proteins in the selected module was constructed using the topology overlap matrix which calculates the strength of similarity between each pair of proteins. The network was then

visualised in Cytoscape (v3.10.3) [22], with nodes representing proteins and thickness of edges representing the strength of similarity. Only the edges with a threshold value of > 0.03 were included in the visualised network.

Whole-blood interferon stimulated gene score

The whole-blood interferon-stimulated gene score was measured as described previously [23]. In brief, total RNA was isolated from whole blood samples using Tempus RNA isolation kits (Thermo Fisher Scientific) following the manufacturer's protocol. Gene expression was measured by real-time qPCR and relative gene expression of *CMPK2*, *EPSTI1* and *HERC5* were calculated by subtracting the housekeeping gene *HPRT* using the ΔCt method. The ISG score was calculated using the mean ΔCt from *CMPK2*, *EPSTI1* and *HERC5* genes *(-1).

## Results

Plasma proteomic profiles of SLE

We recruited 260 patients with SLE and 86 healthy volunteers (HVs). Patients were sampled at varied stages of disease, with a median disease duration at the time of sampling 12 years (IQR 5.5-19.6), and a range of disease activity (**Table 1**). Seventy patients (28%) had anti-double-stranded DNA (dsDNA) autoantibodies, 193 (75%) anti-Smith (Sm) autoantibodies, and 186 (72%) had anti-ribonucleoprotein (RNP) autoantibodies (**Table 1**). Reduced complement C3 (<0.70 g/L) was observed in 14% of samples and reduced complement C4 (<0.16 g/L) in 26%. One hundred and sixty eight patients (65%) had a history of previous lupus nephritis (LN) **(Supplementary Table S2)**, but only 10 patients had recent LN (defined as blood sampling within 3 months of a renal biopsy showing LN). Proteomic measurements were performed in two batches. After quality control, 260 SLE and 86 HV samples were available for analysis. Batch A ("discovery set") consisted of 207 SLE samples and 45 HV samples and batch B ("validation set") consisted of 53 SLE samples and 41 HV samples.

To identify proteomic signatures associated with SLE, we performed differential protein abundance testing of SLE (n=207) vs HVs (n=45) in batch A. For readability, where reporting the number of associations we use the term 'protein' to refer to the protein target

304 of each SOMAmer, but provide further detail of SOMAmer to protein mapping in the
305 Supplementary Tables. We identified 1,821 proteins significantly associated with SLE
306 (1,065 upregulated, 756 downregulated; FDR <0.05) (**Figure 1A**, **Supplementary Figure**
307 **5A, Supplementary Table S3**). Comparison of SLE versus HVs in batch B (53 SLE, 41
308 HVs) identified 428 differentially abundant proteins (268 upregulated, 160 downregulated;
309 FDR <0.05) **(Supplementary Figure 5B, Supplementary Table S4)**. 215 of the 1,821
310 significant associations from batch A replicated in batch B at FDR <0.05 **(Figure 1A,**
311 **Supplementary Table S5)**. The estimated effect sizes for these replicated proteins were
312 strongly correlated between the two batches (Pearson's r 0.94) (**Figure 1B**). We then
313 performed a meta-analysis of the two batches to obtain an overall estimate of the effect
314 sizes for the replicated proteins (**Figure 1C. Supplementary Table S5**). Notably, there
315 was marked upregulation of interferon-stimulated proteins (ISPs) in SLE, including ISG15,
316 MX1, STAT1, B2M, DDX58, IFIT3, CXCL11 **(Figure 1D)**. Other strongly upregulated
317 proteins included TNFRSF1B (also known as TNF-R2), VSIG4 (the receptor for
318 complement C3b and iC3b) and the complement fragment C3d. The most downregulated
319 protein was CRISPLD2, a secreted LPS-binding glycoprotein that dampens TLR4-
320 mediated cytokine responses [24].
321
322 To identify pathways dysregulated in SLE, we performed Gene Set Enrichment Analysis
323 (GSEA) in each batch and then used robust-rank aggregation to rank the pathways most
324 consistently enriched across both batches. 8 pathways were significant in both datasets,
325 including multiple terms reflecting interferon signalling, as well as SRP-mediated
326 cotranslational protein targeting, and downregulated PI3K/AKT signalling **(Figure 1E,**
327 **Supplementary Table S6).**
328
329 <u>Plasma proteomic signatures of lupus disease activity</u>
330 Samples were obtained from patients with a range of disease activities; 134 patients were
331 in clinical remission, 51 had low disease activity (LDA), 65 had moderate disease activity
332 (MDA), and 10 had high disease activity (HDA) (see **Methods** for definitions). To identify
333 proteins associated with disease activity in SLE, linear regression was performed on each
334 protein's abundance levels against disease activity (coded as ordinal variable). To

335 increase power we analysed both Batch A and B together, adjusting for batch as a
336 covariate. We identified 421 proteins significantly associated (FDR <0.05) with disease
337 activity (267 upregulated with increasing disease activity and 154 downregulated) (**Figure**
338 **2A, Supplementary Table S7**). As a sensitivity analysis, we performed Spearman's
339 correlation analysis between SLEDAI and the abundance levels of each protein. The
340 comparison between ρ obtained from the Spearman's correlation tests and the effect
341 sizes (linear gradient) obtained from the ordinal linear regression models showed
342 consistency between the two analyses (**Supplementary Figure 6**; ρ 0.73).
343
344 GSEA revealed significant enrichment of 90 pathways (**Supplementary Table S8**), the
345 most significant of which were "immunoregulatory interactions between a lymphoid and a
346 non-lymphoid cell" and "cytokine-cytokine receptor interaction". The proteins most
347 strongly associated with disease activity by p-value were CD40LG, VIP (vasoactive
348 intestinal peptide) and C1QC (all $P_{BH}$ $1.77\times10^{-6}$). Active disease was also associated with
349 other complement related proteins, including increased levels of C3d and VSIG4 (a
350 receptor for C3b [25]) and decreased levels of intact C4 and CFP (properdin) (**Figure**
351 **2B**). ISPs (MX1, B2M, CXCL10) were positively associated with disease activity, with
352 MX1 showing the greatest effect size. Levels of type I, II and III interferons were also
353 associated with active disease, with IFNL1 (a type III interferon) showing the strongest
354 effect. A caveat to the latter finding is that interferons are challenging to measure reliably
355 in plasma due to very low abundance [9]. Evaluation of connections between IFNL1 and
356 other disease activity-associated proteins revealed IFNL1 was co-correlated (|Pearson's
357 r| >0.6) with 75 other proteins (**Figure 2C, Supplementary Table S9**).
358
359 The SomaScan includes many proteins that are primarily intracellular. Their detection in
360 plasma may reflect cell death and turnover. While such proteins may be useful biomarkers
361 of disease activity, some are likely to be downstream read-outs of tissue inflammation
362 and injury rather than upstream pathogenic drivers. In contrast, plasma proteins
363 associated with disease activity that have a biological role in blood may be more
364 promising therapeutic candidates. For example, cytokines and their receptors have
365 proven to be important therapeutic targets in inflammatory diseases. We therefore

366 extracted a list of proteins annotated in the 'Cytokine–cytokine receptor interaction'
367 pathway in the KEGG database and intersected this with the list of SLE activity associated
368 proteins (**Figure 2D**). In active SLE, there was upregulation of TNF superfamily and TNF
369 receptor superfamily members (e.g., TNFSF4 (OX40L), TNFSF15 (TL1A), TNFRSF1B
370 (TNFR2), TNFRSF10B (DR5)), chemokines (e.g., CCL2, CCL27, CXCL10, CXCL14),
371 interleukins (e.g., IL6, IL11, IL24), and growth factors. In contrast, there was
372 downregulation of anti-inflammatory proteins including IL10RA and the decoy receptor
373 IL1R2.
374
375 <u>Associations with complement proteins</u>
376 Given the strong links between complement proteins and disease activity in the literature
377 and in our data, we further analysed proteomic associations with serum complement C3
378 and C4 measured in the clinical laboratory (n=260 SLE patients). Both C3 and C4 were
379 also included on the SomaScan assay. There was very strong concordance of protein
380 measurements between the SomaScan and the clinical laboratory for C4 (ρ 0.95 in Batch
381 A, 0.91 in Batch B), and moderately strong correlation for C3 (ρ 0.61 in Batch A, 0.45 in
382 Batch B) (**Supplementary Figures 7A-B**), providing validation of SomaScan
383 measurements. In addition, various complement cleavage fragments were measured;
384 correlations with intact C3 and C4 are presented in **Supplementary Figures 8-9**.
385
386 Sixty proteins were correlated with clinical laboratory measured C3 ($|\rho| > 0.3$, FDR <
387 0.05), **Supplementary Table S10**) including 21 associated with disease activity
388 (**Supplementary Figures 10A-B**). Sixty-two proteins were significantly correlated with
389 C4 (**Supplementary Table S11**), 13 of which were disease activity associated
390 (**Supplementary Figure 10C**). In total, 24 protein targets were associated with disease
391 activity and levels of C3 and/or C4 (**Supplementary Figure 10D**). Proteins associated
392 with higher disease activity and lower C3/C4 included CD40LG, interferon-stimulated
393 proteins (e.g. ISG15, IFIT3, LAG3), and C1QC, while IL36A, complement-related proteins
394 (CFP (properdin), FCN1 (ficolin) and C2) and IFNAR1 showed the opposite relationships.
395 The latter may reflect compensatory downregulation in the context of chronic interferon
396 stimulation.

397

398 Exploratory Longitudinal Analysis

399 For 3 patients, longitudinal samples taken at 4 or more timepoints were available. The
400 limited sample size precluded formal biomarker analysis, but we used these data to
401 further explore 7 selected proteins (CXCL10, CCL2, CD40LG, HAVCR2, C3d, MX1, B2M)
402 that we had identified as disease activity-associated in the cross-sectional analysis (see
403 **Methods)**. The levels of these proteins generally mirrored disease activity over time,
404 showing positive longitudinal correlations with SLEDAI-2K and anti-dsDNA, and negative
405 correlations with C3/C4 (**Supplementary Figures 11A-B**). CD40LG strongly correlated
406 with anti-dsDNA titres, in keeping with its role in B cell activation (**Supplementary Figure**
407 **11C**). Longitudinal changes in C3d were inversely correlated with clinically measured C3
408 (**Supplementary Figure 11D**), in keeping with the known biology of the complement
409 pathway activation whereby cleavage of C3 reduces the amount of intact C3 and
410 increases C3d. Several proteins returned to normal levels (defined based on HV samples)
411 as disease activity declined, though the levels of the ISPs CXCL10, MX1 and B2M failed
412 to fully normalise, suggesting persistent interferon pathway activation (**Supplementary**
413 **Figure 11A)**.

414

415 Network analysis reveals protein modules associated with clinical phenotype

416 We used weighted gene co-expression network analysis (WGCNA) to construct a protein-
417 protein correlation network, identifying 21 protein modules. We then tested for
418 associations between each module and traits including SLEDAI-2K score, ISG score
419 (measured by quantitative polymerase chain reaction, qPCR) and clinical laboratory
420 parameters. The red module, consisting of 35 proteins, was significantly correlated with
421 multiple clinical traits, with positive correlation with SLEDAI-2K score, and multiple
422 autoantibodies, and was negatively correlated with neutrophil and lymphocyte count and
423 levels of haemoglobin, albumin, and complement C3 and C4 (**Figures 3A-E,**
424 **Supplementary Figures 12A-G**). Notably, this module showed strongest correlation with
425 ISG score ($\rho$ 0.75, $P_{BH}$ $2.4\times10^{-29}$; **Figure 3B**), and pathway analysis showed enrichment
426 for interferon signalling pathways (**Figure 3F**). ISG15 (measured by SOMAmer
427 seq.14148.2) was the hub protein in this IFN-associated module (**Figure 3G,**

428 **Supplementary Table S12**). ISG15 is an ISP and the module contained multiple other
429 ISPs including (e.g. CXCL10, CXCL11, DDX58, MX1, STAT1) as well as IFNL1. While
430 red module membership was dominated by interferon-related proteins, the module also
431 contained other proteins including the complement component C1QC, the membrane
432 protein APOL2, and LAG3. These protein network correlations may suggest interplay
433 between C1Q, interferon pathways, and immune signalling. Specifically, LAG3 is a
434 negative regulator of T cell responses, thus suggesting that interferon stimulation
435 (indicated by upregulation of ISPs) may correlate with T cell inhibition.
436
437 The blue module, which showed the second highest number of associations with clinical
438 parameters, was positively correlated with creatinine and urine protein-creatinine ratio
439 (uPCR) levels and negatively correlated with albumin **(Figures 3A, 4A-B)**. This module
440 showed highest correlation with the levels of creatinine (ρ 0.52, $P_{BH}$ $7.1\times10^{-14}$; **Figure 4A**)
441 and uPCR (ρ 0.45, $P_{BH}$ $5.4\times10^{-10}$; **Figure 4B**), indicating a negative correlation with
442 glomerular filtration rate. Consistent with this, the module contained Cystatin C (CST3, a
443 marker of renal function; **Figure 4C**). In addition, the module contained Osteopontin
444 (SPP1, predominantly abundant in the kidney and associated with renal function), and
445 Parathyroid hormone (PTH, regulates calcium and phosphate balance in the kidney;
446 **Figure 4D, Supplementary Table S13**). Other renal-related proteins in the module
447 included Ephrin and Ephrin receptors, which are associated with glomerular podocytes
448 survival, angiogenic modelling and responses to renal injury [26, 27]. This blue (renal-
449 associated) module, formed by 611 proteins, was enriched for extracellular matrix-related
450 pathways, possibly reflecting renal scarring (**Figure 4E**). Moreover, the eigenprotein of
451 this module was significantly higher in patients with a history of LN, particularly those with
452 recent LN, compared with those who never had LN **(Figure 4F)**. UNC5B (Netrin receptor)
453 (measured by SOMAmer seq.7776.20) was identified as the hub protein for the renal-
454 associated module (**Supplementary Table S13**). UNC5B has been suggested to protect
455 the kidney against injury through Akt-mediated p53 inactivation [28] and to promote renal
456 proximal tubular epithelial cells proliferation and migration through the Akt and ERK
457 pathways [29].
458

459 Autoantibody-specific proteomic endotypes

460 Anti-Sm autoantibodies are highly specific for SLE and are associated with worse
461 prognosis [30, 31]. To identify a proteomic signature associated with anti-Sm
462 autoantibodies, we compared the abundance levels of each protein between anti-Sm(+)
463 (n=63) and anti-Sm(-) (n=193) patients. Eighty-three proteins were differentially
464 abundant, with 63 proteins upregulated and 20 downregulated in the anti-Sm(+) samples
465 (**Figure 5A**). We considered the possibility that disease activity might be confounded with
466 anti-Sm autoantibody positivity, since patients with Sm antibodies tend to have a worse
467 prognosis and that some of the Sm-associated proteins were associated with disease
468 activity in our previous analyses. We therefore performed multiple linear regression,
469 controlling for disease activity. This revealed 54 proteins associated with anti-Sm
470 antibodies independent of disease activity (43 proteins positively and 11 negatively
471 correlated; **Figure 5B, Supplementary Table S14**). Visualisation of abundance of these
472 proteins against anti-Sm status stratified by disease activity confirmed that the association
473 with Sm antibodies was independent of disease activity (**Figure 5C, Supplementary**
474 **Figure 13**). Pathway enrichment analysis revealed Sm-associated proteins were
475 enriched for SARS-Cov-2-related, interferon signalling, snRNP assembly, and RNA
476 processing pathway terms (**Figure 5D**). These pathway enrichment terms reflected
477 increases in ISPs and reductions of small nuclear ribonucleoproteins (snRNPs) (**Figure**
478 **5E**). ISPs increased in Sm+ SLE included ISG15 (SOMAmer seq.14148.2), MX1,
479 CXCL10, GBP1, DDX58, CXCL11, APOL2, STAT1, LGALS9, EPHB2, and LAG3. Since
480 small nuclear ribonucleoproteins (snRNPs) are the targets of anti-Sm antibodies, it is
481 possible that the reduction in circulating snRNPs reflects autoantibody-mediated
482 depletion.

483

484 We then analysed the associations between proteins and anti-dsDNA antibodies. 874
485 proteins were associated with anti-dsDNA antibody positivity (**Figure 6A**). After adjusting
486 for a modified disease activity score (which excluded dsDNA antibody status), 374
487 proteins remained associated with dsDNA antibody status (**Figures 6B-C,**
488 **Supplementary Table S15**). The most upregulated proteins included CD40LG (CD40
489 ligand), a costimulatory molecule which activates B and T cells and promotes

490 inflammatory responses [32], ARID3A (a DNA-binding protein), IL11, the neutrophil
491 protease proteinase-3 (PRTN3), and RPS7. Proteins reduced in dsDNA antibody positive
492 SLE included CFP, ITGAV|ITGB3 (the vitronectin receptor, an integrin heterodimer),
493 L1CAM, BPIFB1, and C2. Unlike for Sm autoantibody status, the association of ISPs with
494 anti-dsDNA antibody positivity became non-significant after adjustment for disease
495 activity. As a sensitivity analysis, Spearman's rank correlation was performed to test for
496 correlations between the levels of anti-dsDNA autoantibodies (as a continuous trait) and
497 each protein (**Supplementary Figure 14**). In this sensitivity analysis, CD40LG, PRTN3
498 and ARID3A were again among the strongest associations. Intriguingly, visualisation of
499 CD40LG association with disease activity stratified by antibody status showed this
500 relationship was restricted to the dsDNA antibody positive subgroup (**Figure 6D**;
501 dsDNA+: P $4.2\times10^{-3}$, Beta 0.07; dsDNA-: P 0.52, Beta -0.005). Formal interaction analysis
502 confirmed this effect (P $6.7\times10^{-5}$).
503
504 <u>Comparison with anifrolumab-induced proteomic changes in TULIP-1 trial</u>
505 Given the prominence of interferon pathways in proteomic signatures in our data, we
506 examined to what extent these signatures might be reversible with anti-type 1 interferon
507 receptor therapy through comparison with an external dataset from the study of Baker et
508 al which analysed 169 proteins in SLE pre- and post-anifrolumab in the TULIP-1 trial [33]
509 using Olink and Simoa immunoassays. 158 of these proteins were measured in our
510 SomaScan dataset (reflecting 228 SOMAmers since some proteins were measured by
511 multiple SOMAmers). Across these 158 proteins, $\log_2$ fold changes from our SLE vs HV
512 comparison were negatively correlated with anifrolumab-vs-placebo changes at week 52
513 in TULIP-1; (r -0.43; **Supplementary Figure 15A**). Similarly, protein associations with
514 disease activity in our study were inversely correlated with changes following anifrolumab
515 treatment (r -0.52; **Supplementary Figure 15B**). Of the proteins measured in both
516 studies, 30 were associated with disease activity in our study, with 13 significantly
517 modulated by anifrolumab treatment. These data underscore the centrality of the
518 interferon response in plasma proteomic signatures of SLE. However, some disease
519 activity-associated proteins were not modulated by anifrolumab, including CD40LG,

520 indicating interferon-independent disease-activity related pathways (**Supplementary**
521 **Figure 15C**).
522
523 ## Discussion
524
525 We performed wide-angled proteomic profiling in lupus, measuring 7,288 analytes
526 reflecting 6,595 unique proteins, representing the most comprehensive assessment of
527 the plasma proteome to date. Proteins differentially abundant in SLE vs healthy
528 volunteers and proteins associated with disease activity reflected known pathogenic
529 pathways in SLE, including dysregulation of adaptive immunity, complement pathways,
530 interferon and TLR signalling, validating the utility of plasma proteomics in uncovering
531 underlying disease mechanisms.
532
533 Notably, there was strong upregulation of protein products of the ISG signature in SLE vs
534 HVs, and within-case analysis revealed that this was associated with active disease. The
535 lupus ISG was first described in transcriptomic microarray studies of immune cells (whole
536 blood or PBMCs) [4, 7, 34]. Here, we demonstrate that a similar 'ISP' signature is present
537 at the protein level in the extracellular space in blood. There have been few previous
538 broad-capture plasma proteomic studies in SLE. One study used an earlier version of the
539 SomaScan to measure a smaller set of proteins (1,129) and identified 4 interferon-
540 stimulated proteins associated with SLE, consisting of: EPHB2, LAG3, CXCL13, and
541 CXCL10 [10]. In our study, all of these proteins were significantly upregulated in SLE
542 samples and with increasing disease activity, validating previous analysis. A high ISG is
543 associated with worse prognosis [9], but has not entered clinical practice. A plasma-based
544 ISP rather than a gene expression assay might prove more tractable for clinical
545 translation.
546
547 Among the most upregulated interferon-stimulated proteins in the present study were
548 ISG15, MX1, STAT1, CXCL11, B2M and LAG3. LAG3 is an inhibitory receptor expressed
549 on activated T cells. Its expression is increased in the context of chronic antigen exposure
550 (e.g. cancer and chronic viral infection) and leads to T cell exhaustion [35-37]. This may

551 represent a compensatory mechanism to limit immune-mediated tissue damage. Our
552 comparison to proteomic data from the TULIP-1 trial showed that LAG3 was
553 downregulated by anifrolumab. In addition to an interferon-driven ISP, we also directly
554 observed upregulation of interferon proteins themselves. Notably, IFNL1, a type III
555 interferon, was strongly upregulated in both comparison to HVs and the analysis of
556 disease activity. IFNL has received less attention than type 1 interferons in lupus
557 pathogenesis, but its upregulation in SLE has been previously reported [38-42].
558
559 Our SomaScan results recapitulated historical studies of the complement pathway
560 proteins beyond intact C3 and C4 (which are currently used in clinical practice as
561 biomarkers) [43]. We observed positive association of C3d and negative association of
562 CFP (properdin) with active disease, in keeping with previous reports [44-47]. C1Qc was
563 one of the most strongly associated proteins with active disease. A previous study
564 reported that C1Q negatively regulates type 1 IFN production by promoting preferential
565 binding of immune complexes (ICs) to monocytes and thus reducing IC activation of
566 pDCs, which are a major producer of type 1 IFNs [48]. However, we observed a strong
567 positive correlation between C1Qc and levels of ISPs. Our data are consistent with the
568 observation of reductions in whole-blood *C1Qc* gene expression following anifrolumab
569 therapy [33].
570
571 Disease activity-associated proteins included several members of the TNF superfamily
572 cytokines and receptors. One example was TNFSF4 (OX40L), whose plasma levels
573 increased in more active disease. The *TNFSF4* gene locus is a susceptibility locus for
574 SLE, pointing to its causal role in SLE aetiology [49]. In keeping with this, conditional
575 knock out of B cell OX40L ameliorates mouse models of SLE [50]. Our data add to these
576 lines of evidence suggesting that OX40L is a potential therapeutic target in SLE. Another
577 example was TNFSF15 (TL1A), which is a promising therapeutic target in inflammatory
578 bowel disease [51]. Upregulated TNFSF receptors included TNFRSF1B (TNF-R2), which
579 acts a receptor for both TNF and lymphotoxin alpha (LTA), and TNFRSF10B, the receptor
580 for the apoptosis-inducing cytokines TRAIL. Interleukins increased in active disease
581 included IL6, IL11 and IL24. Conversely, there were negative correlations between

582 disease activity and the anti-inflammatory proteins IL1R2 and IL10RA. IL1R2 is a decoy
583 receptor for IL-1, and reduced IL1R2 would be expected to increase pro-inflammatory IL-
584 1 signalling. IL10RA is a receptor for IL-10, an important anti-inflammatory cytokine.
585 Genetic variants in the *IL10RA* gene are associated with early-onset inflammatory bowel
586 disease [52]. Our findings raise the possibility that insufficient activation of homeostatic
587 mechanisms contribute to increased disease activity, although since our data are
588 observational, we cannot exclude the possibility that the reductions in anti-inflammatory
589 proteins is secondary to active disease rather than a contributor to it.
590
591 Co-expression network analysis identified modules of correlated proteins. Notably, two
592 modules had strong associations with clinical and laboratory parameters. One module of
593 proteins correlated with the ISG (measured by qPCR) and also had associations with
594 multiple autoantibodies. This is consistent with previous reports showing associations
595 between a high ISG and the presence of autoantibodies to Ro, U1-RNP, Sm, and dsDNA
596 [53, 54]. The hub protein in this module was ISG15. ISG15 acts as a molecular bridge
597 between the type 1 and type 2 interferon pathways. ISG15 expression is induced by type
598 1 interferons but provides a negative feedback loop to downregulate the type 1 interferon
599 response. On the other hand, it drives production of IFN-γ [55]. A previous study identified
600 ISG15-secreting plasmablasts and plasma cells in active SLE [56]. Another module of
601 proteins reflected renal function and other renal-related physiology, and was enriched for
602 ECM and fibrosis related pathways suggesting this module was capturing a read-out of
603 renal scarring.
604
605 Serological profiles in SLE correlate with clinical phenotype and prognosis [30, 31]. For
606 example, anti-dsDNA antibodies are associated with renal disease and anti-Sm
607 antibodies are associated with poor prognosis. We therefore explored whether serological
608 profiles were associated with particular proteomic signatures and identified distinct
609 proteomic profiles associated with anti-Sm and anti-dsDNA antibodies. Interestingly, even
610 after adjustment for the potential confounder of disease activity, autoantibody specific
611 proteomic profiles persisted. Anti-Sm positive SLE was associated with increased ISPs
612 and reductions in the abundance of small nuclear proteins. We hypothesise that the latter

613 may be due to depletion by anti-Sm antibodies which target epitopes on complexes
614 involving small nuclear ribonucleoproteins. We highlight that we measured proteins in
615 plasma (i.e. the extracellular space), but this may provide a proxy for intra-cellular
616 abundance. This raises the question about the implications of anti-Sm antibodies for the
617 functioning of cellular machinery. Anti-dsDNA positive SLE was associated with elevation
618 of the neutrophil protease, proteinase-3 (PRTN3). This might reflect release of neutrophil
619 contents via neutrophil extracellular traps (NETs), which are a well-established source of
620 nuclear autoantigens in SLE. It is also possible that anti-dsDNA antibodies drive further
621 NET formation and release of proteinase-3.
622
623 CD40LG showed strong associations across multiple analyses, including correlations
624 with disease activity, dsDNA antibody status, and complement C3/C4 levels. Intriguingly,
625 the association with disease activity was specific to individuals with dsDNA antibodies.
626 CD40LG/CD40 interaction drives B cell differentiation into immunoglobulin (Ig)G-
627 secreting plasma cells, making it central to antibody responses. This is consistent with
628 the associations with dsDNA antibodies (both positivity and titre) and reductions in
629 complement C3/C4 (reflecting complement consumption by immune complexes).
630 Previous studies have identified increased expression of CD40 or CD40LG in immune
631 cells and serum in a range of immune-mediated diseases, including SLE [57], and, in line
632 with our data, Goules et al found that CD40LG was elevated in the sera of SLE patients
633 and associated with dsDNA positivity and disease activity [58]. Further support for a
634 pathogenic role for the CD40 pathway in SLE comes from animal models; blocking the
635 CD40-CD40LG interaction ameliorates lupus-like disease in NZB/SWR or NZB/NZW F1
636 mice [59, 60] and also reduced dsDNA antibodies. There have been efforts to
637 therapeutically target CD40LG in SLE and other autoimmune diseases. While first
638 generation anti-CD40LG monoclonal antibodies (mAbs) were abandoned due to
639 thrombotic events [61], direct targeting of CD40 or newer generation anti-CD40LG mAbs
640 bioengineered with inert or absent Fc tails circumvents this issue. A phase 2 trial of the
641 anti-CD40 mAb BI655064 in lupus nephritis failed to meet its primary end-point, although
642 post-hoc analysis suggested potential benefits of higher doses [62]. Evaluation of
643 Frexalimab, a bioengineered anti-CD40LG mAb is ongoing (NCT05039840). Notably,

644 comparison of our data to proteomic data from the TULIP-1 study revealed that, unlike
645 many of the proteins we found associated with disease activity, CD40LG was unaffected
646 by anifrolumab. This suggests that CD40LG axis inhibition might be a complementary
647 approach to anti-IFN therapy by tackling an orthogonal pathogenic pathway. Given the
648 strong associations between CD40LG and dsDNA Ab and complement levels, our data
649 highlight that a stratified medicine approach for anti-CD40 pathway therapeutics might be
650 beneficial. This concept has parallels to the path to approval of belimumab, which proved
651 effective in SLE when used in serologically defined subsets of patients after initial trial
652 failure [63].
653
654 Our study had several limitations. Although our data were split into a discovery and
655 replication subcohorts for case versus control analysis, this was a single-centre study.
656 For within-cases analyses, we analysed both subcohorts together to boost power and
657 thus we did not have a hold-out set for validation. Patients were sampled opportunistically
658 at various points in the disease course and not at pre-defined timepoints; while this
659 enabled cross-sectional evaluation of markers of disease activity, disease duration,
660 chronic damage and past and current treatment were potential confounders. We did not
661 have sufficient longitudinal samples to formally and systematically evaluate intra-
662 individual changes in disease activity-related markers over time. Future studies using
663 inception cohorts with serial sampling are needed to validate putative biomarkers of
664 disease activity identified in the present study. Finally, interpretation of plasma proteomic
665 measurements can be challenging. In addition to canonical soluble proteins, the
666 SomaScan platform includes targets measuring intracellular and membrane-bound
667 proteins. Their detection may represent release from dead cells. Moreover, where
668 proteins exist in both membrane-bound and cleaved states, it is not always clear whether
669 plasma proteomic assays are exclusively capturing the soluble form or also protein from
670 cell membranes (for example, arising from in vivo sources such as exosomes or from
671 sample handling). Complementary cytometry data on paired samples would provide
672 additional insights.
673

674 In summary, this study provides the most comprehensive map of the plasma proteome in
675 SLE to date. Stratification by autoantibody reveals distinct proteomic endotypes even
676 after adjustment for disease activity. Our data provide insights into pathogenesis and
677 highlight potential therapeutic targets.
678
679 **Competing interests**
680 LL reports the following: Advisor / Consultant to: Alexion, Argenx, Astra Zeneca,
681 Boehringer Ingelheim, Carna Health, GSK, Incyte, Kezar, Nkarta, Novartis, Otsuka,
682 Pfizer, Roche, Sanofi, Vera, Vertex; Speaker bureaux: Astra Zeneca, GSK, Otsuka,
683 Roche; Trial data monitoring committees: Nkarta, Novartis. The other authors declare that
684 they have no known competing financial interests or personal relationships that could
685 have appeared to influence the work reported in this paper.
686
687 **Acknowledgments**
688 The study was funded by a Medical Research Foundation Fellowship to J.E.P. (MRF-
689 057-0003-RG-PETE-C0799) and by the NIHR Imperial Biomedical Research Centre
690 Immunology Theme. The views expressed are those of the authors and not necessarily
691 those of the NIHR or the Department of Health and Social Care. M.C.P. is a Wellcome
692 Trust Senior Fellow in Clinical Science (grant no. 212252/Z/18/Z). We thank Shanice
693 Lewis for technical assistance.

| | | SLE | Healthy | P value |
|---|---|---|---|---|
| **N, Samples [%]** | | 260 | 86 | |
| | Female | 229 [88.1%] | 58 [67.4%] | <0.0001 |
| | Male | 31 [11.9%] | 28 [32.6%] | |
| | | | | |
| **Median age, Years [IQR]** | | 43.9 [34.8, 55.4] | 39.5 [30.0, 49.1] | 0.001 |
| | Female | 44.0 [34.6, 55.5] | 39.6 [30.3, 48.1] | 0.002 |
| | Male | 39.9 [35.3, 51.9] | 39.4 [30.6, 50.4] | 0.271 |
| | | | | |
| **N, Disease Activity [%]** | | 260 | - | |
| | In remission | 134 [51.5%] | - | |
| | Low | 51 [19.6%] | - | |
| | Moderate | 65 [25%] | - | |
| | High | 10 [3.8%] | - | |
| | | | | |
| **Median disease duration, Years [IQR]** | | 12.0 [5.5, 19.6] | - | |
| | | | | |
| **N, Lupus nephritis [%]** | | | | |
| | Ever | 168 [64.6%] | - | |
| | Recent* | 10 [3.8%] | - | |
| | | | | |
| **Laboratory parameters** | | | | |
| | **C3 level (g/L, Median [IQR])†** | 1.03 [0.85, 1.20] | - | |
| | N, Low C3 (<0.7g/L) [%] | 36 [13.8%] | - | |
| | | | | |
| | **C4 level (g/L, Median [IQR])†** | 0.21 [0.15, 0.28] | - | |
| | N, Low C4 (<0.16g/L) [%] | 68 [26.2%] | - | |
| | | | | |
| | **N, Anti-Sm status‡ [%]** | 256 | - | |
| | Negative | 193 [75.4%] | - | |
| | Positive | 63 [24.6%] | - | |
| | | | | |
| | **N, Anti-RNP status [%]** | 260 | - | |
| | Negative | 186 [71.5%] | - | |
| | Positive | 74 [28.5%] | - | |
| | | | | |
| | **N, Anti-dsDNA status†‡ [%]** | 250 | - | |
| | Negative (0-9 IUperML) | 140 [56%] | - | |
| | Borderline (10-20 IUperML) | 40 [16%] | - | |
| | Positive (>20 IUperML) | 70 [28%] | - | |

**Table 1. Characteristics of study participants.**
P-value for differences in sex distribution was calculated using chi-squared test; P-values for differences in age were calculated using Student's t-test. *Recent nephritis defined as blood sample taken within 3 months of a renal biopsy showing lupus nephritis. †Anti-dsDNA antibodies and complement levels measured at the time of research sample. ‡Differences in N reflect individuals with missing data for autoantibody status.

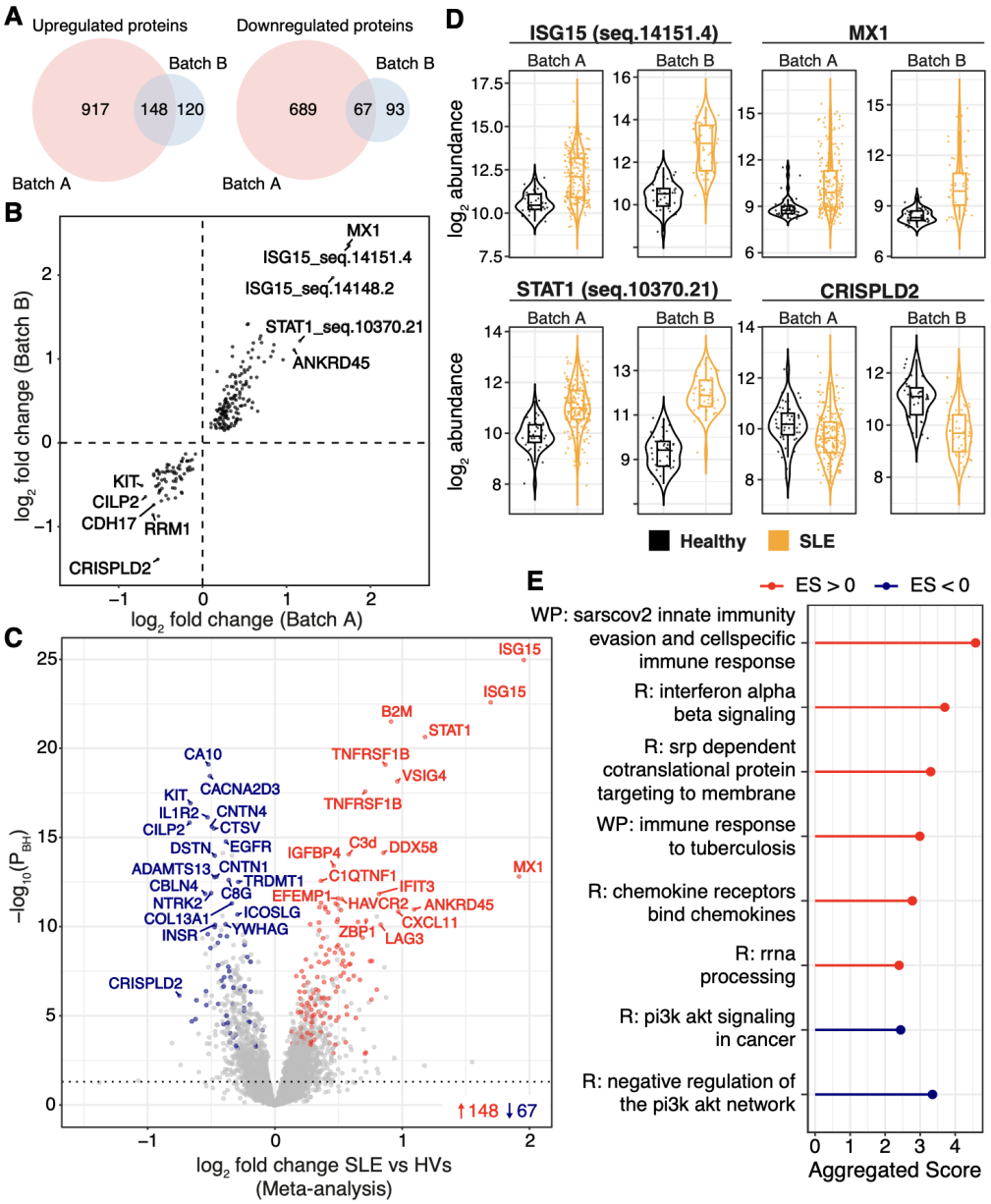



**Figure 1. Differentially abundant proteins and enriched pathways in SLE versus healthy volunteers.**

**A)** Venn diagrams of the number of significantly (FDR <0.05) upregulated (left) and downregulated (right) proteins in SLE versus HVs in batch A (n=207 SLE, n=45 HVs) and B (n=53 SLE, n=41 HVs) data. **B)** Comparison of effect size estimates (covariate-adjusted $\log_2$ fold change) for validated proteins between Batch A and Batch B. Each point represents the protein target of a SOMAmer. 215 protein targets which were significantly differentially abundant (FDR <0.05) in both batch A and batch B are shown. The top 5 upregulated and downregulated protein targets (ranked by meta-analysed effect sizes) are annotated. **C)** Volcano plot showing the results from meta-analysis of Batches A and B. Coloured points represent protein targets replicated in both batches (FDR <0.05). Red: upregulated, blue: downregulated, grey: not replicated or non-significant. **D)** Examples of most significant differentially abundant proteins. Each point represents a sample. Orange: SLE, Black: HVs. **E)** Pathways enriched (FDR <0.05) in both batches. Aggregated score was calculated using robust rank aggregation. Red: upregulated pathways. Blue: downregulated pathways. R, Reactome; WP, WikiPathways; ES, Effect size.

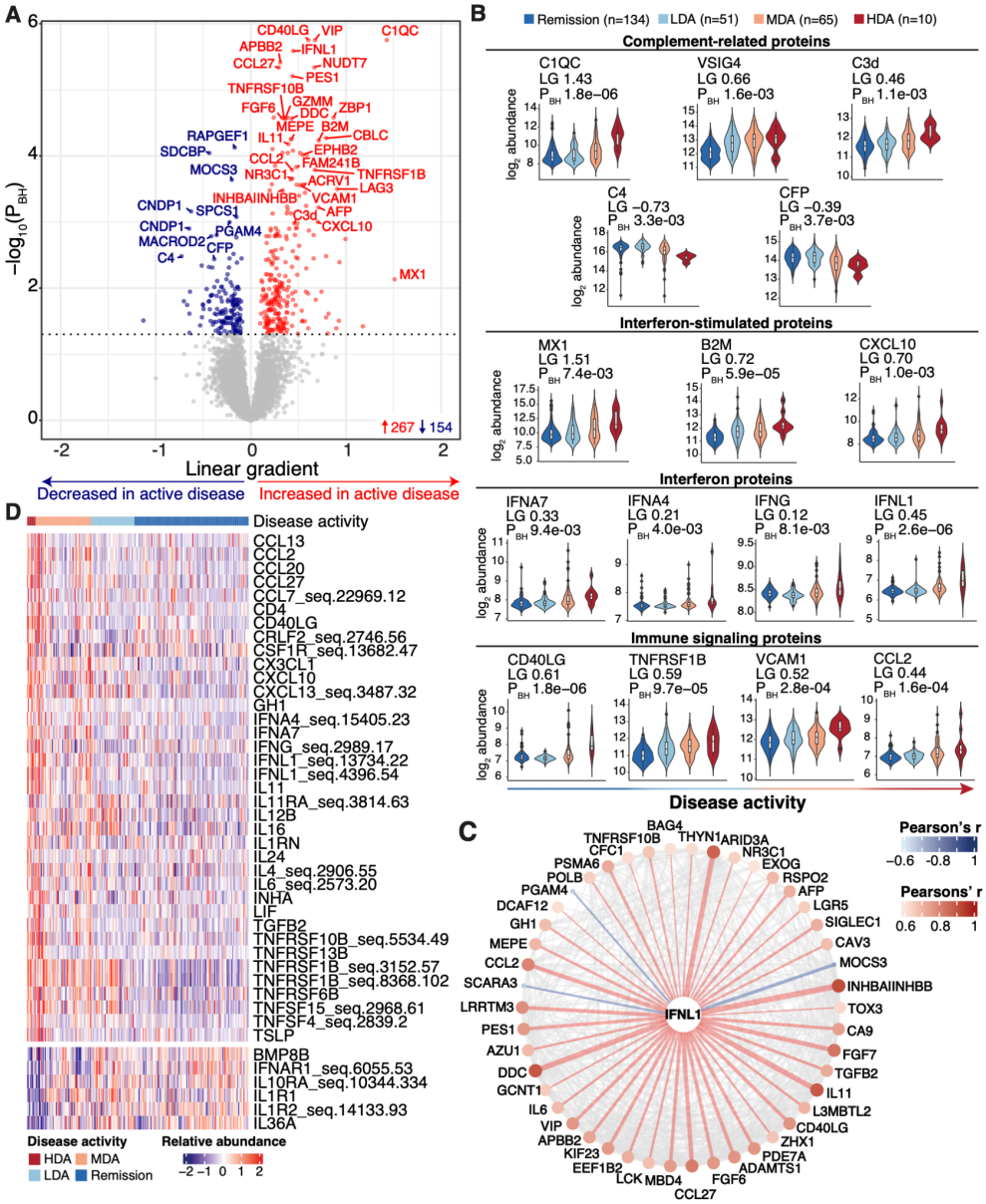



**Figure 2. Proteins associated with SLE disease activity.**
n=260 SLE samples. **A)** Volcano plot. $P_{BH}$ Benjamini-Hochberg-adjusted P-value (linear regression). Linear gradient shows the magnitude and direction of change in the protein level with increasing disease activity. Dotted line = 5% FDR. Each point represents a protein. Red: significant positive association with active disease, blue: negative association, grey: non-significant. **B)** Examples of disease-associated complement-related, interferon-pathway and immune signalling proteins. LG = Linear gradient. Unique SOMAmer identifiers for displayed proteins where the SomaScan has >1 SOMAmer are: B2M (seq.3485.28), IFNA4 (seq.15405.23), IFNG (seq.2989.17), TNFRSF1B (seq.3152.57), and VCAM1 (seq.2967.8). **C)** Disease activity-associated proteins correlated with IFNL1. For clarity of visualisation, only the proteins associated with disease activity at 1% FDR are displayed. A detailed list of proteins correlated with IFNL1 is provided in Supplementary Table S9. Edges represent pairwise protein correlations with |Pearson's r|>0.6. Node size, colour, and edge thickness indicate the strength of correlation (Pearson's r) with IFNL1. Edge colour denotes correlation direction: blue for negative and red for positive. **D)** Heatmap of significant (FDR <0.05) disease activity-associated proteins in the KEGG Cytokine-Cytokine Receptor Interaction pathway. Protein levels were adjusted for sex and batch.

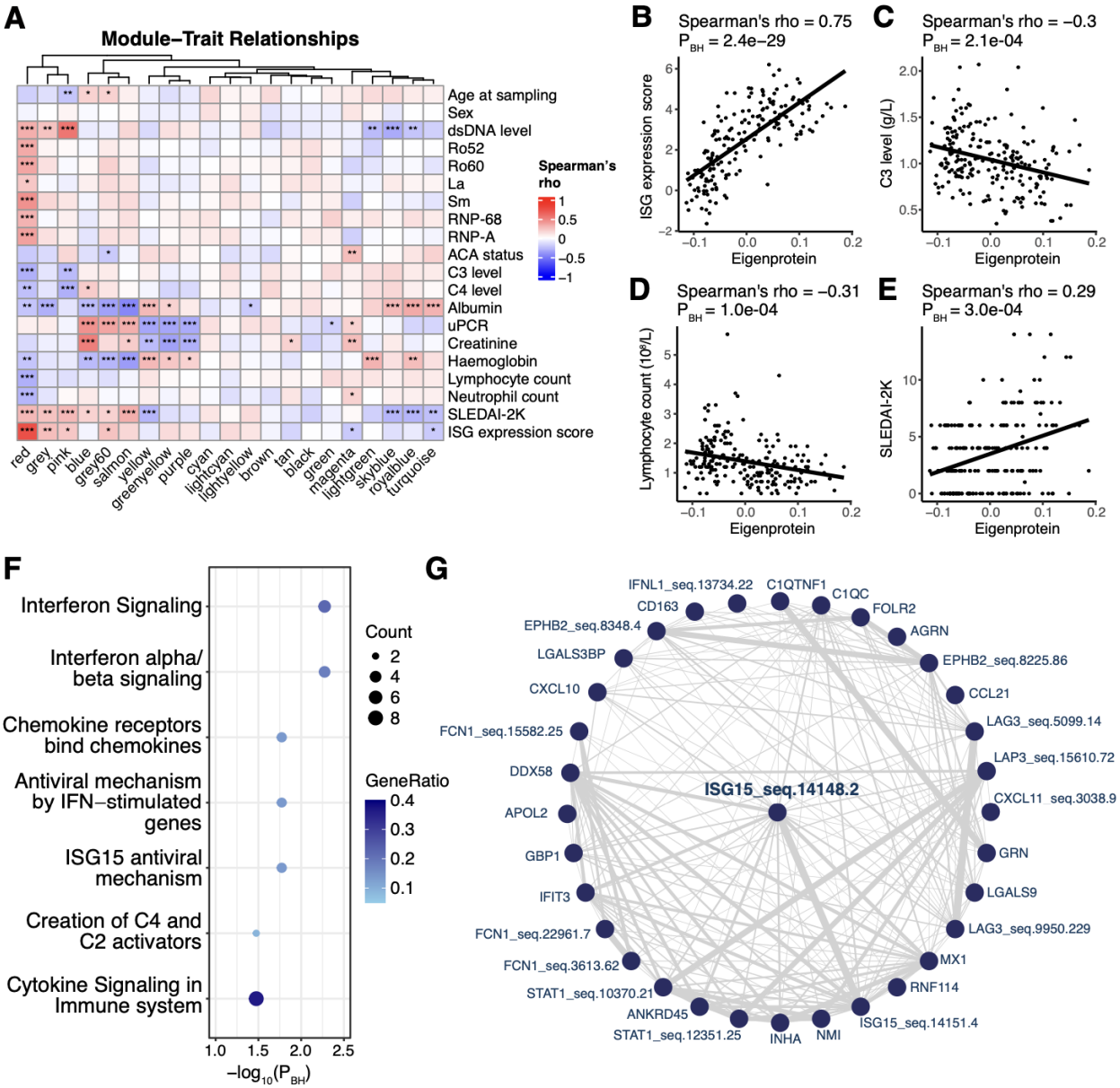



**Figure 3. Network analysis identifies protein modules correlated with clinical traits.**
**A)** Network constructed using SLE samples (n=207). Correlation between each module and clinical traits. Asterixes represent BH-adjusted P values: * < 0.05, ** < 0.01, ***< 0.001. Colour gradient represents Spearman's ρ. **B-E)** Relationships between the eigenprotein values of the red module and clinical traits: **(B)** ISG expression score, **(C)** C3 level, **(D)** lymphocyte count, **(E)** SLEDAI-2K. Each point represents a sample. Solid line indicates best fit from linear regression. **F)** Pathways enriched for proteins in the red module. **G)** Network of protein members of the red module. Edges represent the strength of similarity between each pair of proteins. ISG15 (measured by SOMAmer seq.14148.2) was the module hub protein.

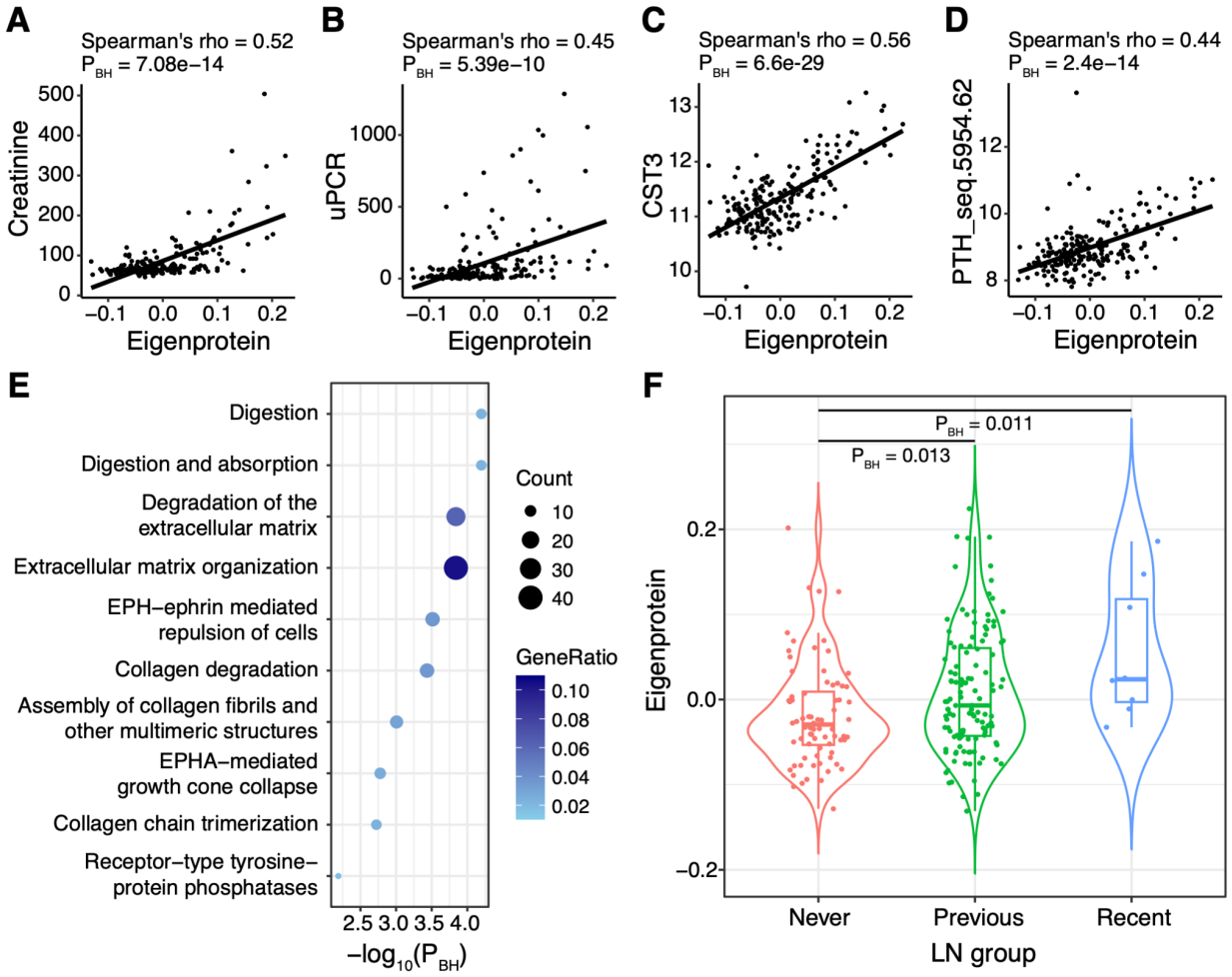



**Figure 4. Renal-associated protein module.**
**A-D)** Relationships between the eigenprotein values of the blue module and renal markers: **(A)** Creatinine, **(B)** uPCR, **(C)** cystatin C (CST3), **(D)** parathyroid hormone (PTH). Each point represents an SLE sample. Solid line represents the line of best fit from linear regression. **E)** Pathways enriched for proteins involved in the blue module. **F)** Eigenprotein values of the blue module according to lupus nephritis (LN) status. n=207 SLE samples.

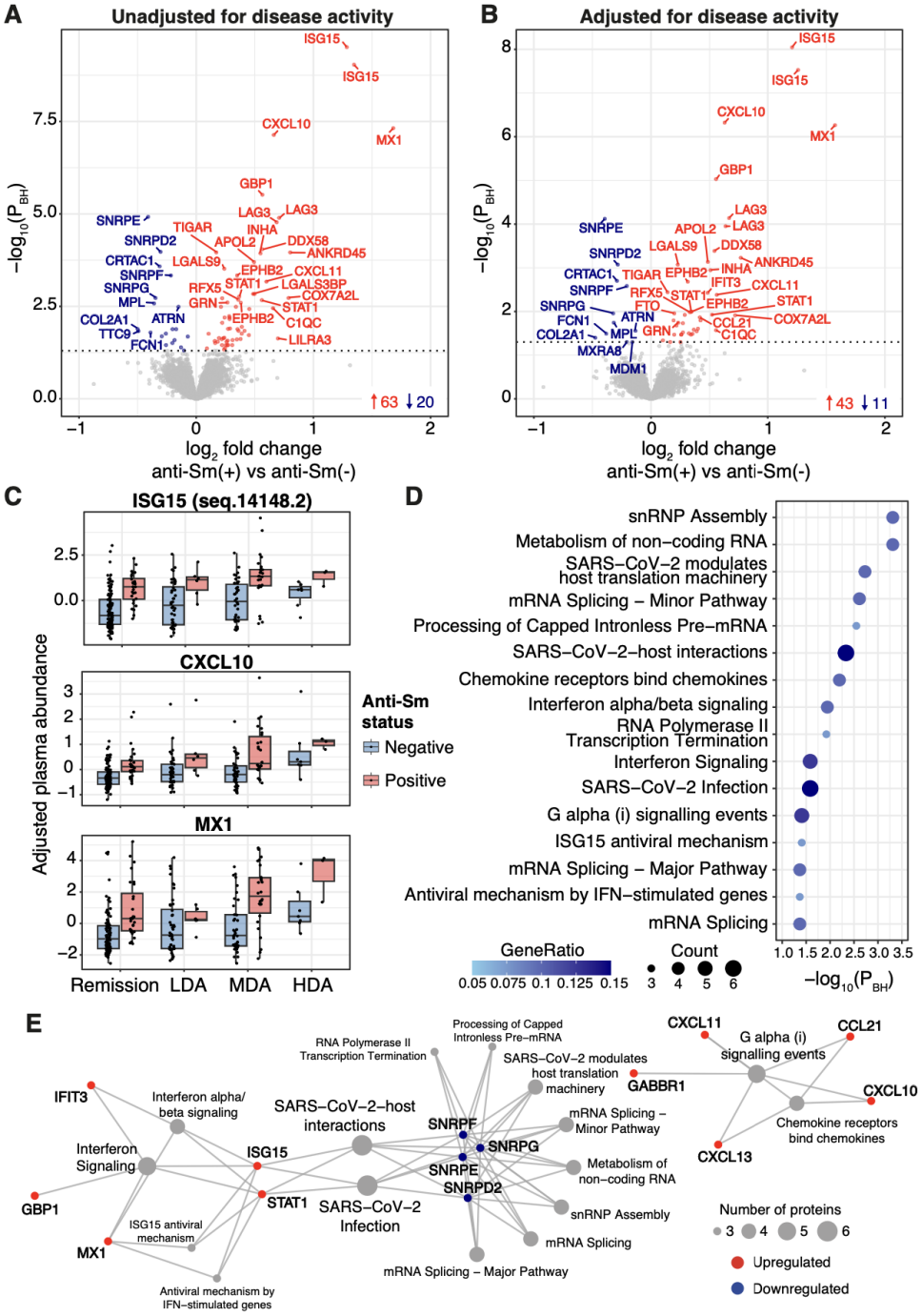



**Figure 5. Proteins associated with anti-Sm autoantibodies.**
**A)** Proteins differentially abundant in anti-Sm positive (n=63) versus anti-Sm negative SLE (n=193), without adjustment for disease activity. Each point represents a protein. Red: significantly upregulated, blue: significantly downregulated, grey: non-significant. **B)** As for **A,** but after adjustment for disease activity. **C)** Interferon-stimulated proteins are associated with Sm antibodies independent of disease activity. Displayed protein levels have been adjusted for batch and sex. **D)** Significantly enriched pathways for proteins associated with anti-Sm antibodies after adjustment for disease activity. **E)** Connections between proteins associated with anti-Sm antibodies status independent of disease activity and the enriched pathways. Node size (for pathway terms) represents the number of proteins associated with the pathway. Red: upregulated proteins. Blue: downregulated proteins.

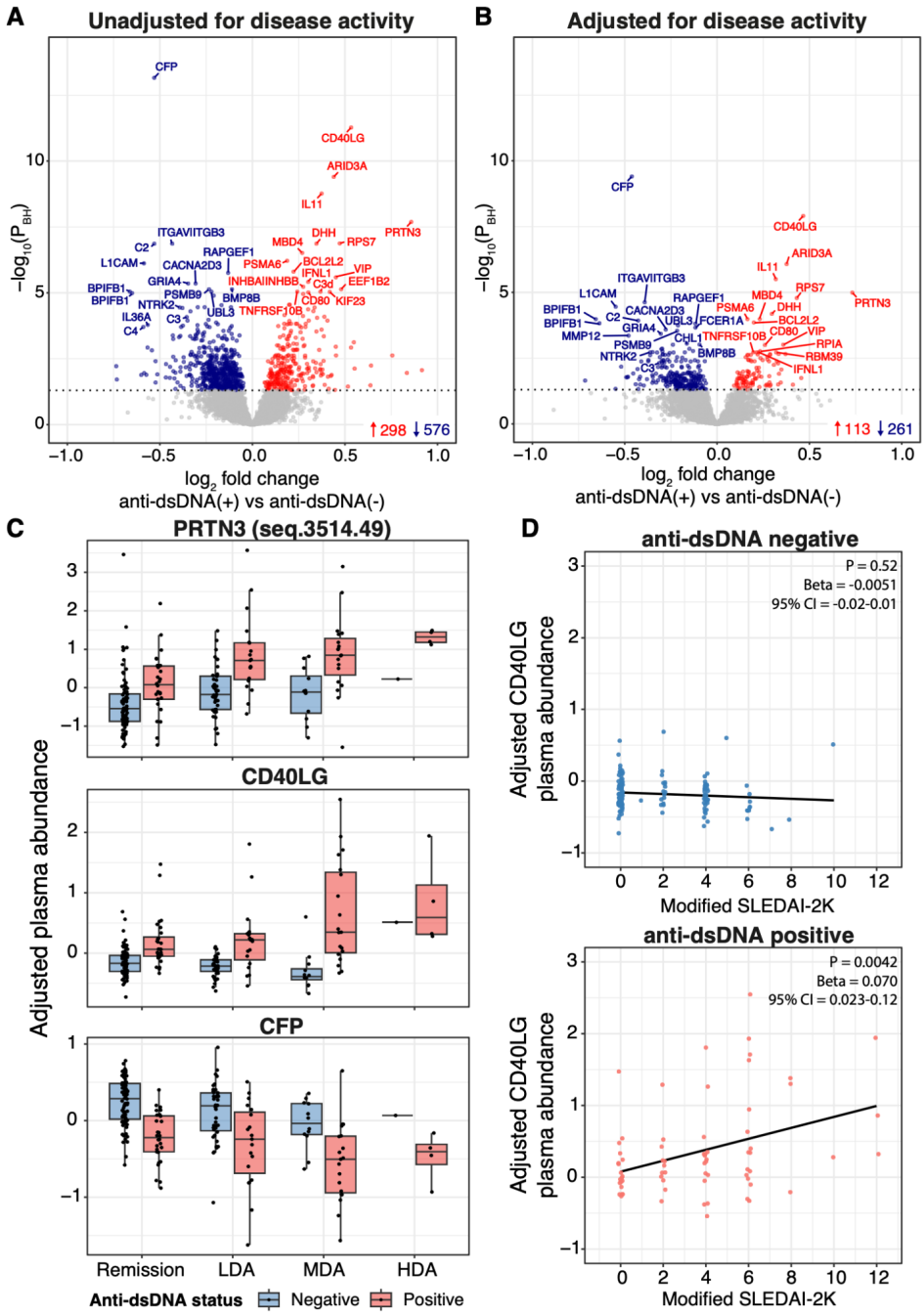



**Figure 6. Proteins associated with anti-dsDNA autoantibodies.**
**A)** Proteins differentially abundant in anti-dsDNA positive (n=70) and anti-dsDNA negative (n=140) SLE samples, without adjustment for disease activity. Each point represents a protein. Red: significantly upregulated, blue: significantly downregulated, grey: non-significant. **B)** As for **A**, but after adjustment for disease activity. **C)** Examples of proteins associated with anti-dsDNA antibodies independent of disease activity. Protein levels in ant-dsDNA antibody positive and negative samples are shown stratified by disease activity. **D)** The association of plasma CD40LG with disease activity is restricted to anti-dsDNA positive individuals. SLEDAI-2K modified to exclude anti-dsDNA status from the score. Each point represents an SLE sample. Solid line represents the line of best fit from linear regression. Protein levels in **C** and **D** have been adjusted for batch and sex.

# References


1. Bentham, J., et al., Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet*, 2015. **47**(12): p. 1457-1464.
2. Langefeld, C.D., et al., Transancestral mapping and genetic load in systemic lupus erythematosus. *Nat Commun*, 2017. **8**: p. 16021.
3. Wang, Y.F., et al., Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. *Nat Commun*, 2021. **12**(1): p. 772.
4. Baechler, E.C., et al., Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci U S A*, 2003. **100**(5): p. 2610-5.
5. Panousis, N.I., et al., Combined genetic and transcriptome analysis of patients with SLE: distinct, targetable signatures for susceptibility and severity. *Ann Rheum Dis*, 2019. **78**(8): p. 1079-1089.
6. Nakano, M., et al., Distinct transcriptome architectures underlying lupus establishment and exacerbation. *Cell*, 2022. **185**(18): p. 3375-3389 e21.
7. Bennett, L., et al., Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med*, 2003. **197**(6): p. 711-23.
8. Postal, M., et al., Type I interferon in the pathogenesis of systemic lupus erythematosus. *Curr Opin Immunol*, 2020. **67**: p. 87-94.
9. Psarras, A., M. Wittmann, and E.M. Vital, Emerging concepts of type I interferons in SLE pathogenesis and therapy. *Nat Rev Rheumatol*, 2022. **18**(10): p. 575-590.
10. Smith, M.A., et al., Using the circulating proteome to assess type I interferon activity in systemic lupus erythematosus. *Sci Rep*, 2020. **10**(1): p. 4462.
11. Su, K.Y.C., et al., Proteomic analysis identifies subgroups of patients with active systemic lupus erythematosus. *Clin Proteomics*, 2023. **20**(1): p. 29.
12. Li, Y., et al., Combined proteomics and single cell RNA-sequencing analysis to identify biomarkers of disease diagnosis and disease exacerbation for systemic lupus erythematosus. *Front Immunol*, 2022. **13**: p. 969509.
13. Piga, M. and L. Arnaud, The Main Challenges in Systemic Lupus Erythematosus: Where Do We Stand? *J Clin Med*, 2021. **10**(2).
14. Gladman, D.D., D. Ibanez, and M.B. Urowitz, Systemic lupus erythematosus disease activity index 2000. *J Rheumatol*, 2002. **29**(2): p. 288-91.
15. van Vollenhoven, R.F., et al., 2021 DORIS definition of remission in SLE: final recommendations from an international task force. *Lupus Sci Med*, 2021. **8**(1).
16. Bakdash, J.Z. and L.R. Marusich, Repeated Measures Correlation. *Front Psychol*, 2017. **8**: p. 456.
17. Korotkevich, H., et al., Fast gene set enrichment analysis. *bioRxiv*, 2021.
18. Liberzon, A., et al., The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*, 2015. **1**(6): p. 417-425.
19. Kolde, R., et al., Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*, 2012. **28**(4): p. 573-80.
20. Yu, G. and Q.Y. He, ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst*, 2016. **12**(2): p. 477-9.
21. Langfelder, P. and S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 2008. **9**: p. 559.
22. Shannon, P., et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 2003. **13**(11): p. 2498-504.
23. Buang, N., et al., Type I interferons affect the metabolic fitness of CD8(+) T cells from patients with systemic lupus erythematosus. *Nat Commun*, 2021. **12**(1): p. 1980.

24. Wang, Z.Q., et al., The novel lipopolysaccharide-binding protein CRISPLD2 is a critical serum protein to regulate endotoxin function. *J Immunol*, 2009. **183**(10): p. 6646-56.
25. Li, J., et al., VSIG4 inhibits proinflammatory macrophage activation by reprogramming mitochondrial pyruvate metabolism. *Nat Commun*, 2017. **8**(1): p. 1322.
26. Weiss, A.C. and A. Kispert, Eph/ephrin signaling in the kidney and lower urinary tract. *Pediatr Nephrol*, 2016. **31**(3): p. 359-71.
27. Huang, Z., et al., Key role for EphB2 receptor in kidney fibrosis. *Clin Sci (Lond)*, 2021. **135**(17): p. 2127-2142.
28. Ranganathan, P., et al., UNC5B receptor deletion exacerbates tissue injury in response to AKI. *J Am Soc Nephrol*, 2014. **25**(2): p. 239-49.
29. Wang, W., W.B. Reeves, and G. Ramesh, Netrin-1 increases proliferation and migration of renal proximal tubular epithelial cells via the UNC5B receptor. *Am J Physiol Renal Physiol*, 2009. **296**(4): p. F723-9.
30. Gomez, A., et al., Treat-to-target in SLE: is serology important? Results from an integrated analysis of five randomized clinical trials of belimumab. *Rheumatology (Oxford)*, 2025. **64**(6): p. 3598-3605.
31. Arroyo-Avila, M., et al., Clinical associations of anti-Smith antibodies in PROFILE: a multi-ethnic lupus cohort. *Clin Rheumatol*, 2015. **34**(7): p. 1217-23.
32. Elgueta, R., et al., Molecular mechanism and function of CD40/CD40L engagement in the immune system. *Immunol Rev*, 2009. **229**(1): p. 152-72.
33. Baker, T., et al., Type I interferon blockade with anifrolumab in patients with systemic lupus erythematosus modulates key immunopathological pathways in a gene expression and proteomic analysis of two phase 3 trials. *Ann Rheum Dis*, 2024. **83**(8): p. 1018-1027.
34. Han, G.M., et al., Analysis of gene expression profiles in human systemic lupus erythematosus using oligonucleotide microarray. *Genes Immun*, 2003. **4**(3): p. 177-86.
35. Ruffo, E., et al., Lymphocyte-activation gene 3 (LAG3): The next immune checkpoint receptor. *Semin Immunol*, 2019. **42**: p. 101305.
36. Andrews, L.P., et al., LAG-3 and PD-1 synergize on CD8(+) T cells to drive T cell exhaustion and hinder autocrine IFN-gamma-dependent anti-tumor immunity. *Cell*, 2024. **187**(16): p. 4355-4372 e22.
37. Gao, Z., et al., T-cell exhaustion in immune-mediated inflammatory diseases: New implications for immunotherapy. *Front Immunol*, 2022. **13**: p. 977394.
38. Wu, Q., et al., Interferon-lambda1 induces peripheral blood mononuclear cell-derived chemokines secretion in patients with systemic lupus erythematosus: its correlation with disease activity. *Arthritis Res Ther*, 2011. **13**(3): p. R88.
39. Oke, V., et al., IFN-lambda1 with Th17 axis cytokines and IFN-alpha define different subsets in systemic lupus erythematosus (SLE). *Arthritis Res Ther*, 2017. **19**(1): p. 139.
40. Amezcua-Guerra, L.M., et al., Type III Interferons in Systemic Lupus Erythematosus: Association Between Interferon lambda3, Disease Activity, and Anti-Ro/SSA Antibodies. *J Clin Rheumatol*, 2017. **23**(7): p. 368-375.
41. Chen, J.Y., et al., Interferon-lambda3/4 genetic variants and interferon-lambda3 serum levels are biomarkers of lupus nephritis and disease activity in Taiwanese. *Arthritis Res Ther*, 2018. **20**(1): p. 193.
42. Oke, V., et al., High levels of circulating interferons type I, type II and type III associate with distinct clinical features of active systemic lupus erythematosus. *Arthritis Res Ther*, 2019. **21**(1): p. 107.
43. Pickering, M.C. and M.J. Walport, Links between complement abnormalities and systemic lupus erythematosus. *Rheumatology (Oxford)*, 2000. **39**(2): p. 133-41.
44. Perrin, L.H., P.H. Lambert, and P.A. Miescher, Complement breakdown products in plasma from patients with systemic lupus erythematosus and patients with membranoproliferative or other glomerulonephritis. *J Clin Invest*, 1975. **56**(1): p. 165-76.

45. Schrager, M.A. and N.F. Rothfield, Clinical significance of serum properdin levels and properdin deposition in the dermal-epidermal junction in systemic lupus erythematosus. *J Clin Invest*, 1976. **57**(1): p. 212-21.
46. Rother, E., et al., Complement split product C3d as an indicator of disease activity in systemic lupus erythematosus. *Clin Rheumatol*, 1993. **12**(1): p. 31-5.
47. Kao, A.H., et al., Erythrocyte C3d and C4d for monitoring disease activity in systemic lupus erythematosus. *Arthritis Rheum*, 2010. **62**(3): p. 837-44.
48. Santer, D.M., et al., C1q deficiency leads to the defective suppression of IFN-alpha in response to nucleoprotein containing immune complexes. *J Immunol*, 2010. **185**(8): p. 4738-49.
49. Cunninghame Graham, D.S., et al., Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus. *Nat Genet*, 2008. **40**(1): p. 83-9.
50. Cortini, A., et al., B cell OX40L supports T follicular helper cell development and contributes to SLE pathogenesis. *Ann Rheum Dis*, 2017. **76**(12): p. 2095-2103.
51. Sands, B.E., et al., Phase 2 Trial of Anti-TL1A Monoclonal Antibody Tulisokibart for Ulcerative Colitis. *N Engl J Med*, 2024. **391**(12): p. 1119-1129.
52. Glocker, E.O., et al., Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N Engl J Med*, 2009. **361**(21): p. 2033-45.
53. Kirou, K.A., et al., Activation of the interferon-alpha pathway identifies a subgroup of systemic lupus erythematosus patients with distinct serologic features and active disease. *Arthritis Rheum*, 2005. **52**(5): p. 1491-503.
54. Zhuang, H., et al., Association of anti-nucleoprotein autoantibodies with upregulation of Type I interferon-inducible gene transcripts and dendritic cell maturation in systemic lupus erythematosus. *Clin Immunol*, 2005. **117**(3): p. 238-50.
55. Zhang, X., et al., Human intracellular ISG15 prevents interferon-alpha/beta over-amplification and auto-inflammation. *Nature*, 2015. **517**(7532): p. 89-93.
56. Care, M.A., et al., Network Analysis Identifies Proinflammatory Plasma Cell Polarization for Secretion of ISG15 in Human Autoimmunity. *J Immunol*, 2016. **197**(4): p. 1447-59.
57. Peters, A.L., L.L. Stunz, and G.A. Bishop, CD40 and autoimmunity: the dark side of a great activator. *Semin Immunol*, 2009. **21**(5): p. 293-300.
58. Goules, A., et al., Elevated levels of soluble CD40 ligand (sCD40L) in serum of patients with systemic autoimmune diseases. *J Autoimmun*, 2006. **26**(3): p. 165-71.
59. Mohan, C., et al., Interaction between CD40 and its ligand gp39 in the development of murine lupus nephritis. *J Immunol*, 1995. **154**(3): p. 1470-80.
60. Early, G.S., W. Zhao, and C.M. Burns, Anti-CD40 ligand antibody treatment prevents the development of lupus-like nephritis in a subset of New Zealand black x New Zealand white mice. Response correlates with the absence of an anti-antibody response. *J Immunol*, 1996. **157**(7): p. 3159-64.
61. Boumpas, D.T., et al., A short course of BG9588 (anti-CD40 ligand antibody) improves serologic activity and decreases hematuria in patients with proliferative lupus glomerulonephritis. *Arthritis Rheum*, 2003. **48**(3): p. 719-27.
62. Jayne, D.R., et al., Clinical and Biomarker Responses to BI 655064, an Antagonistic Anti-CD40 Antibody, in Patients With Active Lupus Nephritis: A Randomized, Double-Blind, Placebo-Controlled, Phase II Trial. *Arthritis Rheumatol*, 2023. **75**(11): p. 1983-1993.
63. Stohl, W. and D.M. Hilbert, The discovery and development of belimumab: the anti-BLyS-lupus connection. *Nat Biotechnol*, 2012. **30**(1): p. 69-77.