Title: Autoantibodies targeting TLR and SMAD pathways define new subgroups in Systemic Lupus Erythematosus

Authors: Myles J. Lewis,[1, **] Michael B. McAndrew,[2] Colin Wheeler,[2] Nicholas Workman,[2] Pooja Agashe,[2] Jens Koopmann,[3] Ezam Uddin,[2] David L. Morris,[4] Lu Zou,[1] Richard Stark,[2] John Anson,[2] Andrew P. Cope,[5] Timothy J. Vyse[4, *]

Affiliations:

[1]Centre for Experimental Medicine and Rheumatology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ, UK.

[2]Oxford Gene Technology, Unit 15, Oxford Industrial Park, Yarnton, Oxfordshire, OX5 1QU, UK.

[3]MedImmune, Aaron Klug Building, Granta Park, Cambridge, CB21 6GH.

[4]Department of Medical and Molecular Genetics, King's College London, London, SE1 9RT, UK.

[5]Academic Department of Rheumatology, Centre for Molecular and Cellular Biology of Inflammation, King's College London, London, SE1 9RT, UK.

[*]To whom correspondence should be addressed: timothy.vyse@kcl.ac.uk
Tel. +44 20 7188 8431
Fax +44 20 7188 2585

[**]Corresponding author: myles.lewis@qmul.ac.uk

Abstract

Objectives

The molecular targets of the vast majority of autoantibodies in systemic lupus erythematosus (SLE) are unknown. We set out to identify novel autoantibodies in SLE to improve diagnosis and identify subgroups of SLE individuals.

Methods

A baculovirus-insect cell expression system was used to create an advanced protein microarray with 1543 full-length human proteins expressed with a biotin carboxyl carrier protein (BCCP) folding tag, to enrich for correctly folded proteins. Sera from a discovery cohort of UK and US SLE individuals (n=186) and age/ethnicity matched controls (n=188) were assayed using the microarray to identify novel autoantibodies. Autoantibodies were confirmed in a validation cohort (91 SLE, 92 controls) and confounding rheumatic disease cohort (n=92).

Results

We confirmed 68 novel proteins as autoantigens in SLE and 11 previous autoantigens in both cohorts (FDR<0.05). Using hierarchical clustering and principal component analysis, we observed four subgroups of SLE individuals associated with four corresponding clusters of functionally linked autoantigens. Two clusters of novel autoantigens revealed distinctive networks of interacting proteins: SMAD2, SMAD5 and proteins linked to TGF-β signalling; and MyD88 and proteins involved in TLR signalling, apoptosis, NF-κB regulation and lymphocyte development. The autoantibody clusters were associated with different SLE manifestations (arthritis, renal and thrombocytopenia). A panel of 26 autoantibodies, derived by penalised multinomial logistic regression taking into account four SLE clusters, showed improved diagnostic accuracy compared to conventional antinuclear antibody and anti-dsDNA antibody assays.

Conclusions

These data suggest that the novel SLE autoantibody clusters may be of prognostic utility for predicting organ involvement in SLE patients and for stratifying SLE patients for specific therapies.

Keywords Systemic lupus erythematosus; autoantibodies; autoimmunity; personalised medicine; diagnostic assay; protein microarray

1. Introduction

Although first described in 1957, anti-nuclear antibodies (ANA) and anti-double-stranded DNA (dsDNA) antibody assays remain the primary diagnostic tests for systemic lupus erythematosus (SLE) [1, 2]. Following the development of assays for extractable nuclear antigens (ENA) Ro, La, Sm and U1-RNP, there have been no significant improvements in diagnostic assays for SLE for many years [3]. In contrast, the identification of citrullinated proteins as autoantigen epitopes in rheumatoid arthritis (RA) led to a marked improvement in RA diagnostic tests with the development of anti-cyclic citrullinated peptide (CCP) assays. Although numerous SLE-associated autoantibodies have been described [4], they have not significantly improved upon the diagnostic and biomarker abilities of conventional ANA, dsDNA and ENA tests, and in many cases the true molecular targets remain undefined. Initial protein microarrays used to detect autoantibodies in SLE sera were largely based on existing autoantigens, but have identified several glomerular proteins and serum factors including B cell-activating factor (BAFF) as SLE autoantigens [5-10]. Microarrays utilising large scale *de novo* synthesis of thousands of proteins have detected autoantibodies in cancer and other diseases [11, 12], but only identified a single SLE autoantigen [13]. Older protein microarrays may have failed to identify autoantibodies due to poor protein conformation caused by misfolding or lack of post-translational modification.

We used a novel protein microarray utilising 1543 distinct proteins chosen from multiple functional and disease pathways, to identify novel autoantigens in SLE. Our aim was to identify previously undiscovered autoantibodies that might act as SLE biomarkers to improve diagnostic (and potentially prognostic) performance over existing clinical assays and to determine whether subgroups of SLE patients with different autoantibody repertoires existed. Full-length human proteins bound to the microarray were expressed in a baculovirus-insect cell expression system

with a biotin carboxyl carrier protein (BCCP) folding tag. The BCCP tag enriches for correctly folded proteins, conserving protein epitopes in their native conformation, which may be necessary for high affinity antibody binding (Figure 1A) [14]. In this study, we used this newer design of protein microarray to elucidate the underlying nature of autoantigens in SLE.

2. Materials and Methods

2.1. Study population

Serum samples from SLE individuals were collected from multiple UK institutions and USA (Seralabs). Serum samples from age/ethnicity matched controls for UK individuals were obtained from the TwinsUK resource (part of the National Institute for Health Research (NIHR) BioResource) and for USA individuals from Seralabs. SLE and control samples were randomly assigned 2:1 to the Discovery cohort (186 SLE and 188 controls) and the Validation cohort (91 SLE and 92 controls). SLE patients were almost all female reflecting the sexual dimorphism of SLE, while healthy controls were exclusively female. All SLE patients fulfilled the 1997 revised American College of Rheumatology (ACR) criteria for classification of SLE. The validation cohort was compared with a Confounding/ interfering disease cohort included patients with the following conditions: systemic sclerosis (n=12), primary Sjögren's syndrome (n=6), polymyositis (n=3) and mixed connective tissue disease (n=3) sourced from USA (Seralabs), and rheumatoid arthritis (RA) (n=68) obtained from multiple UK institutions. RA patients fulfilled the 2010 ACR-EULAR (European League against Rheumatism) criteria for diagnosis of RA. Ethical approval was granted by the Independent Investigational Review Board Inc. (4/16/2008) and the UK National Research Ethics Service London (reference numbers MREC98/2/06, 06/MRE02/9 and 07/H0718/49).

2.2. Protein microarray

Protein microarray assay using the Discovery Array v3.0 protein microarray is described in the Supplementary Methods. The microarray data are available at ArrayExpress accession E-MTAB-5900.

2.3. Serum autoantibody measurement

Anti-dsDNA and ANA titres were measured in all serum samples by ELISA (QUANTA Lite cat no. 704650 & 708750; Inova Diagnostics, San Diego, USA), according to the manufacturer's instructions. ANA results were classified as negative (<20 units), moderate positive (20-60 units) or strong positive (>60 units) using thresholds determined by the manufacturer. Anti-dsDNA samples were classified as negative (<30 IU/mL), borderline (30-75 IU/mL), positive (>75 IU/mL) according to the manufacturer's protocol. ANA and dsDNA results were not based on historical case record results.

2.4. Statistical analysis

Statistical analysis, protein-protein interaction analysis, genotyping, HLA imputation and analysis, and predictive models are described in detail in the Supplementary Methods. The STARD checklist was completed and is available in the online supplement.

3. Results

3.1. Identification of novel SLE-associated autoantigens by microarray

Serum samples from a discovery cohort of 186 SLE patients and 188 age/ethnicity matched healthy controls (Table S1) were analysed for IgG autoantibody levels against 1543 correctly folded, full-length human proteins using a custom protein microarray (Oxford Gene Technology, UK) (Figure 1A). Samples were assayed by ELISA for ANA and anti-dsDNA antibodies for comparison. Normalised autoantibody levels were compared between SLE individuals and healthy controls in the discovery cohort, using a linear regression model adjusting for age, gender, ethnicity and country. A total of 226 autoantibodies, which were increased in the SLE individuals compared to controls in the discovery cohort at FDR-corrected $P<0.05$, were investigated in a validation cohort of 91 SLE individuals and 92 age/ethnicity matched controls. Demographics for the discovery and validation cohorts are shown in Table S1. Of 226 autoantigens observed in the discovery cohort, a total of 79 autoantibodies were also significantly increased in SLE individuals in the validation cohort at $FDR<0.05$ (Figure 1C, Table S2). The well-known SLE autoantigens TROVE2 (Ro60) and SSB (La) showed the most significant difference between SLE and control groups in both cohorts. The array validated a further nine previously reported SLE autoantigens (Figure 1D, Table S2). A total of 68 novel autoantigens were validated by the microarray, with the most statistically significant four novel autoantibodies shown in Figure 1E.

A post-validation meta-analysis was performed using a regression model adjusting for age, gender, ethnicity and country. Suggestive evidence at $FDR_{meta}<0.01$ was found for a further 41 autoantibodies (Table S3), of which 38 were novel. Nine of the validated autoantigens have been shown to be implicated in SLE pathogenesis through immunological studies, but were not

previously known to be autoantigens: CREB1, ZAP70, VAV1, PPP2CB, IRF4, IRF5, EGR2, PPP2R5A and LYN [15-19], while TEK (Tie2 receptor) was identified in the meta-analysis [20]. Five novel autoantigens are the products of SLE susceptibility genes: *IRF5*, *LYN*, *PIK3C3*, *NFKBIA* and *DNAJA1* [21-25]. In summary, 26 of 120 autoantigens (79 validated and 41 identified in the meta-analysis) have a previously identified link to SLE, either as known autoantibody targets or directly implicated in SLE pathogenesis.

In a secondary analysis of the discovery cohort, autoantibodies from the array were ranked by positivity in SLE patients, defined as autoantibody levels >2 SD of the control population, and tested for statistical significance using Fisher's exact test, corrected for multiple testing. Autoantibodies with FDR-corrected $P<0.05$ were analysed for positivity in the validation cohort. A total of 60 autoantibodies showed a significant increase in antibody positivity in both discovery and validation cohorts (Figure 2A). The most prevalent autoantibodies were the known SLE autoantigens Ro60 (overall prevalence 37.5%), SSB/La (35.4%), HNRNPA2B1 (29.6%) and PSME3/Ki (23.8%). The most prevalent novel autoantibodies were LIN28A (22.4%), IGF2BP3 (21.7%) and HNRNPUL1 (21.3%). SLE patients tended to be simultaneously positive for multiple autoantibodies in contrast to the control group (Kruskal-Wallis test with post-hoc Nemenyi test, $P<2\times10^{-16}$) and confounding disease group ($P=6.7\times10^{-9}$), with some individuals producing antibodies against over 60 antigens (Figure 2B).

3.2. SLE autoantibodies cluster into four distinct subgroups

Since we observed that groups of autoantibodies showed strong cross-correlation, we performed unsupervised hierarchical clustering of autoantibody levels in SLE individuals in the discovery cohort and compared with clustering of the validation cohort. In both the discovery and

validation cohorts, SLE individuals clustered into four subgroups, designated: SLE 1a, 1b, 2 and 3 (Figure 3A). Each subgroup was associated with four distinct clusters of autoantibodies (Clusters 1a, 1b, 2 and 3). We have applied this nomenclature based on functional characterization of the autoantigen clusters (see below). SLE subgroup 1a individuals were characterised by being strongly anti-Ro60 and anti-La positive. SLE subgroup 2 showed the broadest range of autoantibody positivity, SLE subgroup 3 were mainly positive for cluster 3 antibodies, while SLE group 1b showed a mixed pattern. The existence of these groupings is borne out in a condensed subspace heatmap of the SLE subgroups which also shows the striking similarity in antibody patterns across the four patient clusters between discovery and validation cohorts (Figure S1).

Cross-correlation of the autoantigens (Figure 3B) revealed strong internal correlation within each antibody cluster, confirming existence of four autoantibody clusters, with certain autoantigens (e.g. RQCD1, SUB1) showing a tendency to inverse correlation with autoantibodies from other clusters. The existence of the autoantibody subgroups of response was confirmed by clustering of all four SLE subgroups when data were re-analysed with inclusion of controls (n=280) using principal component analysis (PCA) (Figure 4A, Movie S1 and Figure S2). PCA showed delineation between SLE patient clusters 1a, 2 and 3 on PC2 and PC3, with PC1 aiding delineation between control and SLE individuals as well as SLE cluster 1b. Component loadings plots showed clear separation of the four subgroups of autoantigens (Figure 4B, Movie S2 and Figure S2).

3.3. Autoantibody cluster-defined SLE subgroups show different disease characteristics

To probe whether the autoantibody clusters were linked to differential SLE phenotype, we compared ANA and dsDNA antibody levels in the four SLE subgroups. SLE group 1a, whose individuals are strongly Ro60 and La positive, showed very high levels of ANA positivity ($P<0.01$, Kruskal-Wallis test), while SLE group 2 showed the highest levels of anti-dsDNA antibodies ($P<0.01$) (Figure 4C). Groups 1b and 3 showed significantly lower levels of both ANA and dsDNA antibodies, consistent with these groups being distinct entities. Furthermore, analysis of ANA negative individuals (at the time of the assay) showed that these were almost exclusively in subgroups 1b and 3, which constituted 90% of ANA negative individuals ($P=1.3\times10^{-9}$, χ^2 test) (Figure 4D). Similarly, subgroups 1b and 3 made up 69% of dsDNA negative SLE individuals ($P=8.4\times10^{-5}$). ANA and dsDNA antibody levels were measured by ELISA on the same serum samples as the microarray assay, and were not based on patient clinical records. This suggests that the novel autoantibodies from cluster 1b and 3 are particularly important for diagnosing SLE patients with intermittently negative ANA/dsDNA antibodies for whom existing diagnostic tests may be unreliable.

SLE subphenotype clinical data available on 184 UK SLE individuals was analysed for trends in autoantibody positivity across autoantibody clusters. While some subphenotype characteristics such as skin rash were similar across all four clusters, autoantibody clusters showed specificity for presence or absence of arthritis ($P=0.00063$ for interaction between cluster and subphenotype by two-way ANOVA), pulmonary ($P=0.0059$) and neurological involvement ($P=0.038$) (Figure 4E), suggesting that there are subphenotype differences between the autoantibody clusters. Group 2 autoantibody positivity was higher in the presence of arthritis, while group 1A was lower. Renal involvement was associated with IGF2BP3 positivity

(P=0.015) and higher cluster 1B positivity. CARHSP1 positivity showed particularly strong association with the presence of thrombocytopenia (P=0.00028). TWIST2 positivity was associated with oral ulceration (P=0.016).

Medication usage data was available on 234 SLE individuals from both UK and US cohorts. No difference was seen for the majority of immunosuppressive medications between SLE clusters, however, SLE3 individuals showed lower prednisolone usage (P=0.0044, Fisher's exact test) and higher warfarin usage (P=0.0078). This raises the possibility that the cluster 3 autoantibodies might be associated with anti-phospholipid syndrome.

3.4. SLE autoantigen clusters associate with functional protein-protein interaction networks

To examine whether the clusters of autoantigens identified associated with different SLE subgroups showed themes of molecular or functional categorisation, each cluster of autoantigens was investigated using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database, and cross-referenced against Ingenuity Pathway Analysis and the PANTHER (protein annotation through evolutionary relationship) classification system. The 79 validated autoantigens were insufficient for meaningful pathway analysis, so were supplemented with 41 putative autoantigens identified by post-validation meta-analysis. Protein interaction networks identified using STRING were discernible in cluster 2 and cluster 3 autoantigens. In cluster 2, the largest network of interacting proteins was centred around SMAD2, SMAD5 and included proteins associated with TGF-β, Wnt and bone morphogenic protein (BMP) signalling such as PPP2CB, ID2, TWIST2, CSNK2A1 and CSNK2A2 (Figure 5A). In cluster 3, a protein network of genes involved in toll-like receptor (TLR) signalling and NF-κB activation including MYD88,

BIRC3 (cIAP-2), NFKBIA (IκBα), MAP3K7 (TAK1) and MAP3K14 (NIK) was observed, interlinked with genes involved in apoptosis regulation such as BIRC3 (cIAP-2), ANXA1 (Annexin A1), CASP9 (caspase-9), ZMYND11 and BCL2A1 (Figure 5B). The cluster 3 network also incorporated key proteins involved in lymphocyte differentiation such as VAV1, EGR2, ZAP70 and SH2B1. STRING identified TGFBR1 (TGF-β receptor 1) and RELA (NF-κB p65) as predicted functional nodes for clusters 2 and 3 respectively (prediction score 0.999). The functional themes of the autoantigen clusters are summarised in Figure 5C.

3.5. Improved diagnostic accuracy of expanded autoantibody panels

Elastic net regularised logistic regression was employed as a variable selection method to identify an optimal autoantibody panel for SLE diagnosis. 10-fold cross-validation (using the discovery cohort) was used to select L1-L2 parameter α and shrinkage parameter λ (Figure S4). The optimal penalized binomial logistic regression model (α=0.7, λ_{1se}=0.00764), employing 17 autoantibodies (Table S4), was tested on the Validation cohort using Receiver Operating Characteristic (ROC) curve analysis (Figure 6A). The performance of autoantibody models at discriminating SLE individuals from a non-SLE group including both healthy controls and confounding group individuals (mostly RA), to mimic the real-world situation of a typical rheumatology clinic. Low level ANA positivity is commonly observed in healthy individuals [26], other autoimmune diseases or pregnant individuals. Thus in clinical practice ANA performs more poorly, since it is significantly less specific than dsDNA antibodies at lower titres. The elastic net binomial regression model showed improved sensitivity (59.3%) at high specificity (90%) (Figure 6A,B), compared to standard ANA (37.0%) and dsDNA antibody (38.6%) assays and combined ANA+dsDNA regression model (47.8%). However, this model did not reflect the

different patient clusters as well as other autoantibodies (Figure S5). We hypothesized that a biomarker model, which exploited the distinct clustering of autoantibodies in SLE individuals, could be superior to binomial regression models. First, we used elastic net regularized multinomial logistic regression for variable selection to narrow the autoantibodies to a set of 26 autoantibodies which optimally identified the four SLE clusters in the discovery cohort (Figure 6C). This reduced set of 26 autoantibodies was trained using penalised mixture discriminant analysis (MDA) [27, 28] to enable separation of clustered data, specifying one control cluster and four SLE clusters. The MDA model showed superior diagnostic classification compared to the binomial elastic net regression model, with a sensitivity of 67.0% at specificity 90%. Addition of ANA and dsDNA antibodies to the MDA model did not improve prediction (Figure S6). The improvement in the MDA model compared to the elastic net binomial regression model is likely to be due to the non-linear decision boundary (Figure S7), which delineates four separate SLE clusters from healthy controls, in both discovery and validation cohorts. The panel of 26 autoantibodies was able to delineate different patterns of subphenotype and organ involvement (Figure S3).

4. Discussion

Using a novel protein microarray design optimised to enhance presentation of correctly folded proteins, we have identified 68 proteins as novel autoantigens in SLE, and confirmed 11 known SLE autoantibody targets. Post-validation meta-analysis found suggestive evidence for a further 41 autoantigens. The design of microarray used in our study found a large number of novel autoantigens in stark contrast to previous proteomic approaches to autoantigen discovery which have only identified a handful of new autoantigens [13, 29]. A striking feature of the novel SLE

autoantigens found in our study is that many are clearly identifiable as important immune system regulators, in multiple cases already implicated in SLE pathogenesis. Thirteen of the 106 novel autoantigens have been directly implicated in SLE pathogenesis or genetic susceptibility. This helps to confirm the validity of this new protein microarray approach for identification of novel autoantibodies.

Unsupervised hierarchical clustering of the novel autoantigens revealed four SLE subgroups present in both the discovery and validation cohorts, associated with four clusters of autoantigens (Figure 3A). The clustering designation of both the SLE subgroups and autoantigen clusters was strongly supported by principal component analysis (Figure 4A,B). The most well-known autoantigens form cluster 1a, which includes TROVE2 (Ro60), SSB (La), the proteasome subunit PSME3 (Ki, PA28 gamma) and SMN1, which complexes with Sm and U1-RNP autoantigens as part of the spliceosome. Cluster 1b includes known autoantigens HNRNPA2B1, PABPC1 and HMGB2. Autoantigens in clusters 1a and 1b are distinguished by a functional theme of involvement in RNA processing including mRNA decay (RQCD1), mRNA splicing (SMN1), mRNA editing (APOBEC3G, PABPC1, IGF2BP3), nucleocytoplasmic RNA transport (Ro60, SSB/La and HNRNPA2B1) and microRNA binding (LIN28A). Other cluster 1b antigens are involved in chromatin remodelling and DNA binding. Comparison with ANA and dsDNA antibody levels showed that group 1a were strongly ANA positive and group 2 were strongly dsDNA positive. Group 1b and 3 showed lower levels of ANA and dsDNA antibodies and 90% of the ANA negative individuals were from SLE1b and SLE3. It is important to note that ANA and dsDNA positivity was measured on the same samples as used for the microarray assay, to allow direct comparison. Although ANA positivity is typically quoted as >95% in SLE individuals at or around the time of diagnosis, ANA autoantibodies can become negative over

time in >20% of individuals [30]. In a large multicentre study of 446 SLE individuals, of whom 98% were historically ANA positive, only 71% were ANA positive at the time of study entry [31]. It is well established that low titres of ANA can be detected in healthy individuals [26]. Thus, cluster 1b and 3 autoantigens are of major clinical importance for aiding diagnosis of SLE individuals who are ANA of low titre or intermittently negative and/or anti-dsDNA antibody negative.

Specific autoantibody clusters were associated with significant differences in subphenotype. The presence of arthritis was associated with lower cluster 1A autoantibody positivity and higher levels of cluster 2 autoantibodies such as PRKRA, consistent with the importance of Wnt signalling in synovial biology. Renal involvement was associated with higher cluster 1B autoantibodies, specifically IGF2BP3. Another 1B autoantibody HNRNPA2B1 has been previously associated with lupus nephritis [32], but showed less strong association than IGF2BP3 in our cohort. Certain novel autoantibodies were associated with certain SLE manifestations: CARHSP1 positivity showed strong association with thrombocytopenia and TWIST2 was associated with oral ulceration. There were differences in pulmonary and neurological involvement between autoantibody clusters, however due to the heterogeneous nature of pulmonary and neurological manifestations, these associations require validation in cohorts with more detailed clinical information and granularity. Overall, the novel autoantibodies have potential prognostic utility for predicting specific disease manifestations and organ involvement in SLE.

We used the STRING database to analyse the autoantigen clusters for protein-protein interactions (Figure 5). Two key themes emerged: cluster 2 autoantigens centred around SMAD2 and SMAD5 were linked to TGF-β/Wnt/BMP signalling; cluster 3 autoantigens were implicated

in TLR/NF-kB signalling, apoptosis regulation, and B and T lymphocyte development. The SLE2 subgroup associated with highest positivity for cluster 2 autoantigens showed the highest levels of arthritis and Raynaud's, consistent with TGF-β pathway involvement. Excess TLR7 activity is linked to development of SLE [33], and we identified a distinct subgroup of SLE patients (SLE3) associated with autoantibodies against the TLR adaptor MYD88, NF-κB signalling proteins TAK1 and MAP3K14 (NIK), and apoptosis regulators BIRC3 (cIAP-2) and ANXA1 (annexin A1) [34]. The demonstration that anti-Ro and anti-La antibodies in SLE sera bound apoptotic cell blebs [35] led to the 'waste disposal hypothesis', which proposed that defective clearance of dying cells is the source of autoantigen exposure [36]. Annexin A1 is released by apoptotic neutrophils and promotes phagocytosis of apoptotic neutrophils by macrophages [37]. TLR7 is upregulated in SLE neutrophils and primes neutrophils for production of neutrophil extracellular traps (NETs), which have been proposed as a source of antigen for anti-dsDNA antibody formation [38]. It is conceivable that some cluster 3 antigens may originate from neutrophils undergoing NETosis or apoptosis.

SLE3 cluster autoantigens also included transcription factors and adaptors important for regulating lymphocyte development including ZAP70, EGR2, CREB1 and VAV1. Egr-2 controls T cell self-tolerance and *Egr2* deficient mice develop lupus-like autoimmune disease [18]. ZAP70 and VAV1, which strongly clustered together, are both recruited to the immunological synapse following T cell receptor stimulation, and may reside in membrane microdomains leading to inclusion in secreted exosomes [39]. Excess type I interferon activity plays an important role in SLE pathogenesis, and several notable interferon pathway genes (IRF4, IRF5) were identified as autoantigens.

The clustering of antigens into functional groups hints at different underlying pathogenic mechanisms defining SLE subgroups. If the new classes of autoantibodies represent different underlying pathogenic mechanisms, this would have important clinical ramifications, with the prediction that the SLE subgroups defined by this study might require different treatment strategies. For example, patients with NF-κB and B cell differentiation genes as antigens may be a subgroup which are more likely to respond to B cell therapies (e.g. Rituximab, Belimumab), while patients with TGF-β/Wnt signalling pathways may be at risk of fibrotic manifestations overlapping with systemic sclerosis, and might respond to non-B cell specific therapies (e.g. cyclophosphamide).

This study has a number of limitations including: the single timepoint for sample collection; lack of clinical information on disease activity at the time of sample collection; incomplete information on organ manifestations (e.g. pulmonary and neurological) and anti-phospholipid syndrome status; and absence of HEp-2 ANA assay as a comparator. ANA ELISA was employed in this study since it is less prone to operator-dependent subjectivity than the standard HEp-2 ANA fluorescence assay, however HEp-2 ANA is more sensitive than ELISA. Thus, future studies to further investigate which of these novel autoantibodies are useful for prognosis, therapeutic stratification or monitoring disease activity alongside anti-dsDNA antibodies, will necessitate longitudinal, prospective studies to collect serial samples alongside more detailed clinical information on type of organ involvement and neurological features. HEp-2 ANA assay should also be included in the comparison. Since some autoantibodies, such as cardiolipin antibodies [40], can be triggered by acute infections, sera from an infectious diseases cohort should be compared with the SLE cohort. Following the identification of TGF-β pathway autoantigens, the confounding disease cohort should include a larger cohort of other connective

tissue diseases including a large systemic sclerosis cohort with detailed clinical information on systemic sclerosis type (limited or diffuse) and patterns of organ involvement (interstitial lung disease, Raynaud's manifestations etc).

We identified a 17-autoantibody autoantibody biomarker panel which showed improved sensitivity and specificity for diagnosis of SLE in comparison to standard ANA and anti-dsDNA assays (Figure 6A,B). However, this binomial model, which was trained for simple discrimination of SLE patients from controls, was outperformed by a multinomial regression 26-autoantibody model trained to discriminate four clusters of SLE individuals by penalised mixture discriminant analysis (MDA). The MDA model, by accounting for clustering of SLE individuals, showed enhanced diagnostic performance in the validation cohort compared to conventional ANA and dsDNA assays. The use of a repertoire of autoantibodies for SLE diagnosis has parallels with the peptide libraries employed by anti-CCP diagnostic assays for RA, and the 26-autoantibody biomarker panel demonstrates comparable sensitivity/specificity for SLE diagnosis to anti-CCP assays in RA [41].

In summary, using improved protein microarray technology with attention to optimal protein folding and synthesis, we have identified a large number of novel SLE autoantigens. Our study suggests that SLE can be subgrouped by molecular signature through four distinct autoantibody patterns. We propose that each SLE subgroup may have diverse pathogenic and/or genetic mechanisms underlying the differential autoantigen response.

Declaration of interest

MBM, CW, NW, PA, JK, EU, RS, JA are or were full-time employees of Oxford Gene Technology.

Funding

T.J.V. was awarded funding to from the George Koukis Foundation and an Arthritis Research UK Special Strategic Award. The study received support from the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and the Biomedical Research Centre based at Guy's & St. Thomas' National Health Service (NHS) Foundation Trust, in partnership with King's College London. The TwinsUK study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013).

Acknowledgements

The authors thank the volunteers who participated in this study. We thank Su Wang for statistical advice (predictive models).

References

[1] Friou GJ. Clinical application of a test for lupus globulin-nucleohistone interaction using fluorescent antibody. Yale J Biol Med, 1958;31**:**40-7.

[2] Miescher P, Strassle R. New serological methods for the detection of the L.E. factor. Vox Sang, 1957;2**:**283-7.

[3] Isenberg DA, Manson JJ, Ehrenstein MR, Rahman A. Fifty years of anti-ds DNA antibodies: are we approaching journey's end? Rheumatology (Oxford), 2007;46**:**1052-6.

[4] Yaniv G, Twig G, Shor DB, Furer A, Sherer Y, Mozes O *et al.* A volcanic explosion of autoantibodies in systemic lupus erythematosus: a diversity of 180 different antibodies found in SLE patients. Autoimmunity reviews, 2015;14**:**75-9.

[5] Robinson WH, DiGennaro C, Hueber W, Haab BB, Kamachi M, Dean EJ *et al.* Autoantigen microarrays for multiplex characterization of autoantibody responses. Nat Med, 2002;8**:**295-301.

[6] Li QZ, Xie C, Wu T, Mackay M, Aranow C, Putterman C *et al.* Identification of autoantibody clusters that best predict lupus disease activity using glomerular proteome arrays. J Clin Invest, 2005;115**:**3428-39.

[7] Li QZ, Zhou J, Wandstrat AE, Carr-Johnson F, Branch V, Karp DR *et al.* Protein array autoantibody profiles for insights into systemic lupus erythematosus and incomplete lupus syndromes. Clin Exp Immunol, 2007;147**:**60-70.

[8] Chong BF, Tseng LC, Lee T, Vasquez R, Li QZ, Zhang S *et al.* IgG and IgM autoantibody differences in discoid and systemic lupus patients. J Invest Dermatol, 2012;132**:**2770-9.

[9] Papp K, Vegh P, Hobor R, Szittner Z, Voko Z, Podani J *et al.* Immune complex signatures of patients with active and inactive SLE revealed by multiplex protein binding analysis on antigen microarrays. PloS one, 2012;7**:**e44824.

[10] Price JV, Haddon DJ, Kemmer D, Delepine G, Mandelbaum G, Jarrell JA *et al.* Protein microarray analysis reveals BAFF-binding autoantibodies in systemic lupus erythematosus. J Clin Invest, 2013;123**:**5135-45.

[11] Desmetz C, Mange A, Maudelonde T, Solassol J. Autoantibody signatures: progress and perspectives for early cancer detection. Journal of cellular and molecular medicine, 2011;15**:**2013-24.

[12] Abel L, Kutschki S, Turewicz M, Eisenacher M, Stoutjesdijk J, Meyer HE *et al.* Autoimmune profiling with protein microarrays in clinical applications. Biochim Biophys Acta, 2014;1844**:**977-87.

[13] Huang W, Hu C, Zeng H, Li P, Guo L, Zeng X *et al.* Novel systemic lupus erythematosus autoantigens identified by human protein microarray technology. Biochem Biophys Res Commun, 2012;418**:**241-6.

[14] Boutell JM, Hart DJ, Godber BL, Kozlowski RZ, Blackburn JM. Functional protein microarrays for parallel characterisation of p53 mutants. Proteomics, 2004;4**:**1950-8.

[15] Katsiari CG, Kyttaris VC, Juang YT, Tsokos GC. Protein phosphatase 2A is a negative regulator of IL-2 production in patients with systemic lupus erythematosus. J Clin Invest, 2005;115**:**3193-204.

[16] Krishnan S, Juang YT, Chowdhury B, Magilavy A, Fisher CU, Nguyen H *et al.* Differential expression and molecular associations of Syk in systemic lupus erythematosus T cells. J Immunol, 2008;181**:**8145-52.

[17] Biswas PS, Gupta S, Stirzaker RA, Kumar V, Jessberger R, Lu TT *et al.* Dual regulation of IRF4 function in T and B cells is required for the coordination of T-B cell interactions and the prevention of autoimmunity. J Exp Med, 2012;209**:**581-96.
[18] Zhu B, Symonds AL, Martin JE, Kioussis D, Wraith DC, Li S *et al.* Early growth response gene 2 (Egr-2) controls the self-tolerance of T cells and prevents the development of lupuslike autoimmune disease. J Exp Med, 2008;205**:**2295-307.
[19] Silver KL, Crockford TL, Bouriez-Jones T, Milling S, Lambe T, Cornall RJ. MyD88-dependent autoimmune disease in Lyn-deficient mice. Eur J Immunol, 2007;37**:**2734-43.
[20] Kumpers P, David S, Haubitz M, Hellpap J, Horn R, Brocker V *et al.* The Tie2 receptor antagonist angiopoietin 2 facilitates vascular inflammation in systemic lupus erythematosus. Ann Rheum Dis, 2009;68**:**1638-43.
[21] Graham RR, Kyogoku C, Sigurdsson S, Vlasova IA, Davies LR, Baechler EC *et al.* Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. Proc Natl Acad Sci U S A, 2007;104**:**6758-63.
[22] Kariuki SN, Franek BS, Mikolaitis RA, Utset TO, Jolly M, Skol AD *et al.* Promoter variant of PIK3C3 is associated with autoimmunity against Ro and Sm epitopes in African-American lupus patients. Journal of biomedicine & biotechnology, 2010;2010**:**826434.
[23] Lu R, Vidal GS, Kelly JA, Delgado-Vega AM, Howard XK, Macwana SR *et al.* Genetic associations of LYN with systemic lupus erythematosus. Genes and immunity, 2009;10**:**397-403.
[24] Li Y, Cheng H, Zuo XB, Sheng YJ, Zhou FS, Tang XF *et al.* Association analyses identifying two common susceptibility loci shared by psoriasis and systemic lupus erythematosus in the Chinese Han population. J Med Genet, 2013;50**:**812-8.
[25] Ramos PS, Williams AH, Ziegler JT, Comeau ME, Guy RT, Lessard CJ *et al.* Genetic analyses of interferon pathway-related genes reveal multiple new loci associated with systemic lupus erythematosus. Arthritis Rheum, 2011;63**:**2049-57.
[26] Tan EM, Feltkamp TE, Smolen JS, Butcher B, Dawkins R, Fritzler MJ *et al.* Range of antinuclear antibodies in "healthy" individuals. Arthritis Rheum, 1997;40**:**1601-11.
[27] Hastie T, Buja A, Tibshirani R. Penalized discriminant analysis. Annals of Statistics, 1995;23**:**73-102.
[28] Hastie T, Tibshirani R. Discriminant Analysis by Gaussian Mixtures. J R Statist Soc B, 1996;58**:**155-76.
[29] Katsumata Y, Kawaguchi Y, Baba S, Hattori S, Tahara K, Ito K *et al.* Identification of three new autoantibodies associated with systemic lupus erythematosus using two proteomic approaches. Molecular & cellular proteomics : MCP, 2011;10**:**M110 005330.
[30] Sjowall C, Sturm M, Dahle C, Bengtsson AA, Jonsen A, Sturfelt G *et al.* Abnormal antinuclear antibody titers are less common than generally assumed in established cases of systemic lupus erythematosus. J Rheumatol, 2008;35**:**1994-2000.
[31] Wallace DJ, Stohl W, Furie RA, Lisse JR, McKay JD, Merrill JT *et al.* A phase II, randomized, double-blind, placebo-controlled, dose-ranging study of belimumab in patients with active systemic lupus erythematosus. Arthritis Rheum, 2009;61**:**1168-78.
[32] Schett G, Dumortier H, Hoefler E, Muller S, Steiner G. B cell epitopes of the heterogeneous nuclear ribonucleoprotein A2: identification of a new specific antibody marker for active lupus disease. Ann Rheum Dis, 2009;68**:**729-35.

[33] Pisitkun P, Deane JA, Difilippantonio MJ, Tarasenko T, Satterthwaite AB, Bolland S. Autoreactive B cell responses to RNA-related antigens due to TLR7 gene duplication. Science, 2006;312**:**1669-72.
[34] Silke J, Brink R. Regulation of TNFRSF and innate immune signalling complexes by TRAFs and cIAPs. Cell Death Differ, 2010;17**:**35-45.
[35] Casciola-Rosen LA, Anhalt G, Rosen A. Autoantigens targeted in systemic lupus erythematosus are clustered in two populations of surface structures on apoptotic keratinocytes. J Exp Med, 1994;179**:**1317-30.
[36] Walport MJ. Complement. Second of two parts. N Engl J Med, 2001;344**:**1140-4.
[37] Perretti M, D'Acquisto F. Annexin A1 and glucocorticoids as effectors of the resolution of inflammation. Nat Rev Immunol, 2009;9**:**62-70.
[38] Garcia-Romo GS, Caielli S, Vega B, Connolly J, Allantaz F, Xu Z *et al.* Netting neutrophils are major inducers of type I IFN production in pediatric systemic lupus erythematosus. Science translational medicine, 2011;3**:**73ra20.
[39] Wang GJ, Liu Y, Qin A, Shah SV, Deng ZB, Xiang X *et al.* Thymus exosomes-like particles induce regulatory T cells. J Immunol, 2008;181**:**5242-8.
[40] Adebajo AO, Charles P, Maini RN, Hazleman BL. Autoantibodies in malaria, tuberculosis and hepatitis B in a west African population. Clin Exp Immunol, 1993;92**:**73-6.
[41] Taylor P, Gartemann J, Hsieh J, Creeden J. A systematic review of serum biomarkers anti-cyclic citrullinated Peptide and rheumatoid factor as tests for rheumatoid arthritis. Autoimmune diseases, 2011;2011**:**815038.

Figure legends

Figure 1. Novel autoantigens identified by protein microarray in Systemic Lupus Erythematosus (SLE)

(**A**) Novel protein microarray technology used BCCP folding tag to improve protein folding conformation of array-bound proteins**.** (**B**) Volcano plot of autoantigens in the Discovery cohort displaying each microarray autoantigen as a single point with P value on the y-axis versus $\log_2$ fold change in antibody levels between SLE and matched controls on the x-axis. P values were calculated using a linear regression model adjusting for cohort, sex, age and ethnicity. Blue points signify FDR-corrected $P_{train}<0.01$. (**C**) Volcano plot of autoantigens validated in the validation cohort. Red points show autoantigens validated in both cohorts (FDR-corrected P_{train} and $P_{test}<0.01$), blue points show autoantigens found in Discovery cohort but not replicated in Validation cohort. Red points show autoantigens validated in both cohorts, blue points show autoantigens with $FDR_{meta}<0.01$. (**D** & **E**) Tukey boxplots of median normalised IgG binding data showing IgG autoantibody reactivity against specific antigens on the protein microarray in the discovery cohort (Control1, n=188; SLE1, n=186) and the validation cohort (Control2, n=92; SLE2, n=91). (**D**) Top four previously identified autoantigens confirming validation of lesser known antigens PABPC1 and HMGB2. (**E**) Top four novel autoantigens identified by microarray. Box plots show median, upper and lower quartiles, with whiskers denoting maximal and minimal data within 1.5 × interquartile range (IQR). Dark blue dots represent antibody positivity defined as >2 SD of control population. Confounding group includes individuals with rheumatoid arthritis, Sjogren's syndrome and other connective tissue diseases.

Figure 2. Hierarchy of autoantibody positivity in SLE individuals

(**A**) Autoantigens ranked by positivity in SLE patients in both Discovery and Validation cohorts. P values were calculated by Fisher test with FDR correction for multiple testing. FDR-corrected $P<0.05$ in both discovery and validation cohorts was considered significant. (**B**) Distribution histogram showing total number of positive autoantibodies for each individual showing that sera from SLE patients can recognise over 60 discrete autoantigens.

Figure 3. Hierarchical clustering identifies four SLE autoantibody subgroups

(**A**) Heatmap of unsupervised hierarchical clustering of 79 validated autoantibody levels in SLE individuals from the discovery cohort (n=186) and validation cohort (n=91), using correlation as distance metric and Ward's clustering method. Rows were clustered based on the discovery cohort. Autoantibody levels were Z score normalised against control population mean and SD, with Z scores >2 corresponding to positive autoantibody levels. Autoantibodies cluster into four major clusters, with four matching clusters SLE1a, SLE1b, SLE2 and SLE3 identified in SLE individuals and the four patient clusters were observed in both the discovery and validation cohorts. (**B**) Correlation plot of Pearson r values shows significant cross-correlation of autoantibodies within each cluster.

Figure 4. Autoantibody cluster-defined SLE subgroups show different disease characteristics

(**A**) Principal component analysis of 79 autoantibody levels in SLE individuals and controls. Principal component (PC) scores for PC1-3 showing clusters of SLE individuals identified by hierarchical clustering. Ellipsoids show 95% confidence intervals. (**B**) PC loading scores for

PC2, PC3 and PC4 showing clustering of autoantigens identified by hierarchical clustering. (**C**) Anti-nuclear and anti-double-stranded DNA antibody results according to SLE cluster groups SLE1a, SLE1b, SLE2 and SLE3. ANA and dsDNA sera levels were measured by ELISA, and are not based on patient clinical records. Statistical analysis by Kruskal-Wallis test. (**D**) Comparison of SLE subgroups among ANA and dsDNA negative individuals, showing that ANA negative individuals are predominantly from subgroups SLE1b and SLE3. (**E**) Subphenotype comparison of autoantibody clusters. Heatmap represents subphenotype fold change for mean autoantibody levels for each autoantibody cluster. P values calculated for interaction between autoantibody cluster and subphenotype by two-way ANOVA. *P<0.05 for pairwise comparisons for presence/absence of subphenotype. (**F**) Positivity of individual autoantibodies across subphenotypes. (**G**) Differential usage of medications in SLE clusters. Statistical analysis in **D**, **F**, **G** by Fisher's exact test.

Figure 5. SLE autoantigen clusters demonstrate functional protein-protein interaction networks

(**A** & **B**) Protein-protein interaction networks were derived from STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database and plotted using Cytoscape for (**A**) cluster 2 and (**B**) cluster 3 autoantigens. Line thickness represents strength of interaction confidence. Predicted nodes are shown in orange. (**C**) Phylogenetic tree of autoantigens summarising key protein functions for autoantigens in each cluster.

Figure 6. Improved diagnostic performance of 26-autoantibody biomarker panel

(**A-B**) Diagnostic biomarker panels were derived by elastic net penalised logistic regression using binomial model (control, SLE) or multinomial model (control, 4 SLE clusters) for variable selection. The optimal model was the 26-autoantibody panel selected by multinomial elastic net logistic regression and trained using penalised mixture discriminant analysis (MDA) employing one control cluster and four SLE clusters. All models were trained exclusively on the Discovery cohort and tested on the Validation test cohort. (**A**) Receiver operating characteristic (ROC) curves for Validation cohort are shown for MDA and elastic net derived biomarker panels compared with ANA, dsDNA and combined ANA + dsDNA model. (**B**) Table of area under curve (AUC), sensitivity and specificity of biomarker panels (MDA, binomial elastic net regression) compared to ANA, dsDNA and ANA + dsDNA model. (**C**) Heatmap of Z scores of mean autoantibody levels in each SLE subgroup for 26-autoantibody biomarker panel. Statistical analysis by one-way ANOVA with FDR correction.

Supplementary Material

Movie S1. 3D plot of principal component scores for SLE individuals and controls

Three-dimensional plot of principal component scores 1-3 for SLE individuals and controls (grey spheres). SLE individuals are coloured according to subgroup SLE1a (red), SLE1b (purple), SLE2 (green), SLE3 (blue) as determined by hierarchical clustering.

Movie S2. 3D plot of principal component loading scores for SLE autoantigens

Three-dimensional plot of principal component loading scores for SLE autoantigens, along axes PC2, PC3 and PC4. Autoantigens are coloured according to antigen cluster 1a (red), cluster 1b (purple), cluster 2 (green), cluster 3 (blue) as determined by hierarchical clustering.

Figure 1

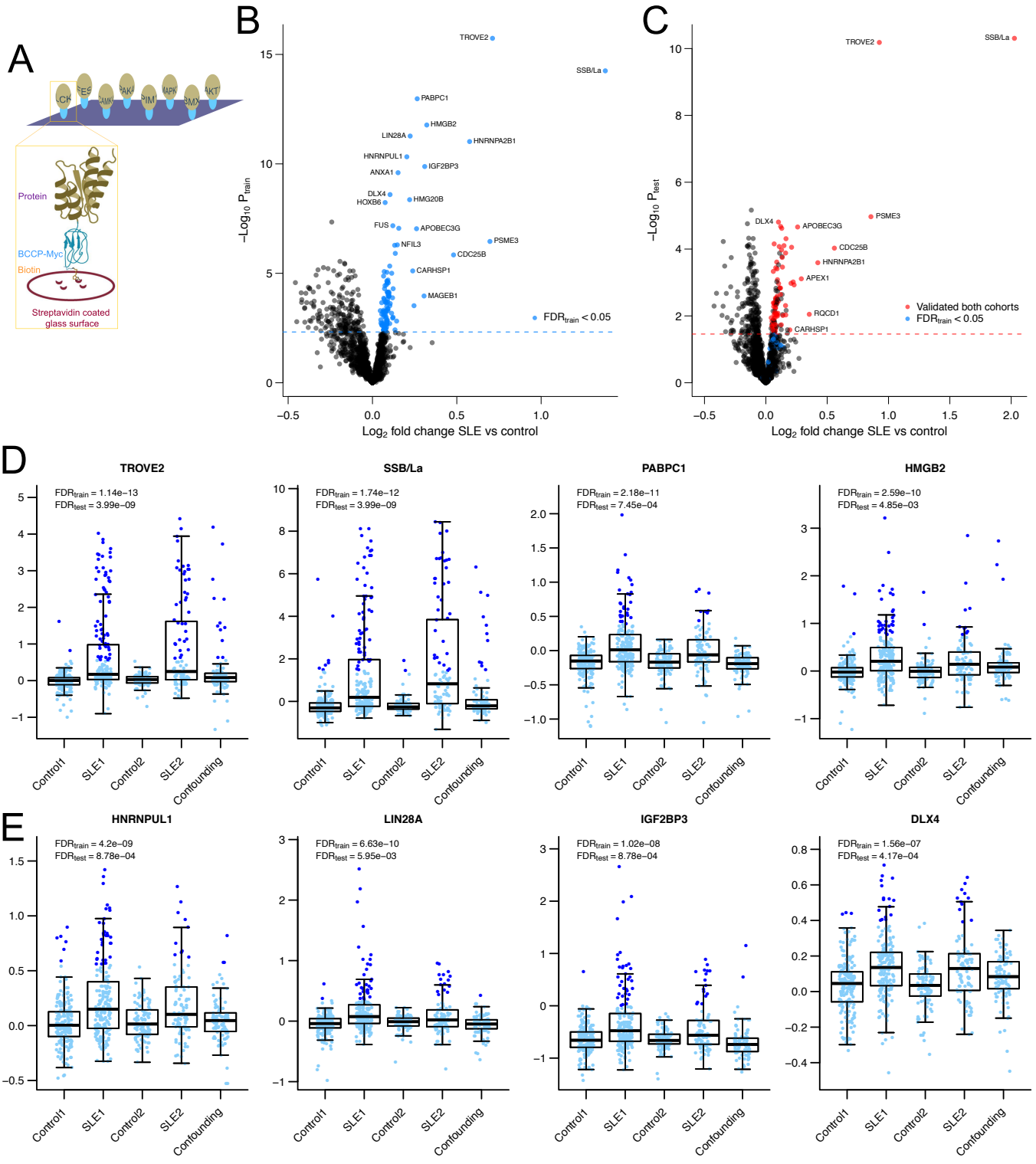

A
Protein
BCCP-Myc
Biotin
Streptavidin coated glass surface
B
TROVE2
SSB/La
PABPC1
HMGB2
LIN28A
HNRNPA2B1
HNRNPUL1
IGF2BP3
ANXA1
DLX4
HOXB6
HMG20B
FUS
APOBEC3G
NFIL3
PSME3
CDC25B
CARHSP1
MAGEB1
FDRtrain < 0.05
−Log10 Ptrain
Log2 fold change SLE vs control
C
TROVE2
SSB/La
DLX4
APOBEC3G
PSME3
CDC25B
HNRNPA2B1
APEX1
RQCD1
CARHSP1
Validated both cohorts
FDRtrain < 0.05
−Log10 Ptest
Log2 fold change SLE vs control
D
TROVE2
FDRtrain = 1.14e−13
FDRtest = 3.99e−09
SSB/La
FDRtrain = 1.74e−12
FDRtest = 3.99e−09
PABPC1
FDRtrain = 2.18e−11
FDRtest = 7.45e−04
HMGB2
FDRtrain = 2.59e−10
FDRtest = 4.85e−03
Control1
SLE1
Control2
SLE2
Confounding
E
HNRNPUL1
FDRtrain = 4.2e−09
FDRtest = 8.78e−04
LIN28A
FDRtrain = 6.63e−10
FDRtest = 5.95e−03
IGF2BP3
FDRtrain = 1.02e−08
FDRtest = 8.78e−04
DLX4
FDRtrain = 1.56e−07
FDRtest = 4.17e−04
Control1
SLE1
Control2
SLE2
Confounding

Figure 2

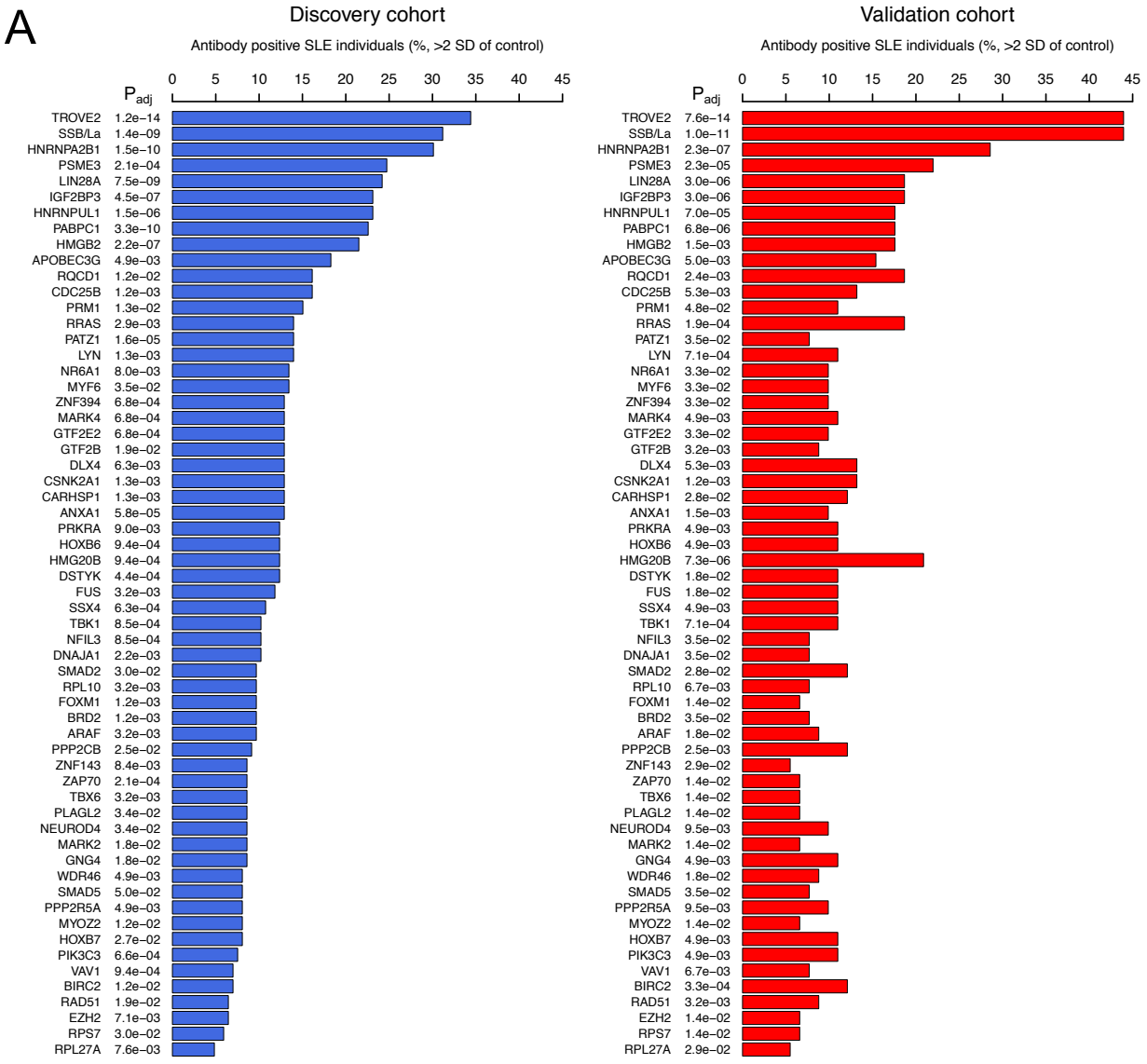

A
Discovery cohort
Antibody positive SLE individuals (%, >2 SD of control)
Validation cohort
Antibody positive SLE individuals (%, >2 SD of control)
0 5 10 15 20 25 30 35 40 45
P adj
TROVE2 1.2e−14
SSB/La 1.4e−09
HNRNPA2B1 1.5e−10
PSME3 2.1e−04
LIN28A 7.5e−09
IGF2BP3 4.5e−07
HNRNPUL1 1.5e−06
PABPC1 3.3e−10
HMGB2 2.2e−07
APOBEC3G 4.9e−03
RQCD1 1.2e−02
CDC25B 1.2e−03
PRM1 1.3e−02
RRAS 2.9e−03
PATZ1 1.6e−05
LYN 1.3e−03
NR6A1 8.0e−03
MYF6 3.5e−02
ZNF394 6.8e−04
MARK4 6.8e−04
GTF2E2 6.8e−04
GTF2B 1.9e−02
DLX4 6.3e−03
CSNK2A1 1.3e−03
CARHSP1 1.3e−03
ANXA1 5.8e−05
PRKRA 9.0e−03
HOXB6 9.4e−04
HMG20B 9.4e−04
DSTYK 4.4e−04
FUS 3.2e−03
SSX4 6.3e−04
TBK1 8.5e−04
NFIL3 8.5e−04
DNAJA1 2.2e−03
SMAD2 3.0e−02
RPL10 3.2e−03
FOXM1 1.2e−03
BRD2 1.2e−03
ARAF 3.2e−03
PPP2CB 2.5e−02
ZNF143 8.4e−03
ZAP70 2.1e−04
TBX6 3.2e−03
PLAGL2 3.4e−02
NEUROD4 3.4e−02
MARK2 1.8e−02
GNG4 1.8e−02
WDR46 4.9e−03
SMAD5 5.0e−02
PPP2R5A 4.9e−03
MYOZ2 1.2e−02
HOXB7 2.7e−02
PIK3C3 6.6e−04
VAV1 9.4e−04
BIRC2 1.2e−02
RAD51 1.9e−02
EZH2 7.1e−03
RPS7 3.0e−02
RPL27A 7.6e−03
TROVE2 7.6e−14
SSB/La 1.0e−11
HNRNPA2B1 2.3e−07
PSME3 2.3e−05
LIN28A 3.0e−06
IGF2BP3 3.0e−06
HNRNPUL1 7.0e−05
PABPC1 6.8e−06
HMGB2 1.5e−03
APOBEC3G 5.0e−03
RQCD1 2.4e−03
CDC25B 5.3e−03
PRM1 4.8e−02
RRAS 1.9e−04
PATZ1 3.5e−02
LYN 7.1e−04
NR6A1 3.3e−02
MYF6 3.3e−02
ZNF394 3.3e−02
MARK4 4.9e−03
GTF2E2 3.3e−02
GTF2B 3.2e−03
DLX4 5.3e−03
CSNK2A1 1.2e−03
CARHSP1 2.8e−02
ANXA1 1.5e−03
PRKRA 4.9e−03
HOXB6 4.9e−03
HMG20B 7.3e−06
DSTYK 1.8e−02
FUS 1.8e−02
SSX4 4.9e−03
TBK1 7.1e−04
NFIL3 3.5e−02
DNAJA1 3.5e−02
SMAD2 2.8e−02
RPL10 6.7e−03
FOXM1 1.4e−02
BRD2 3.5e−02
ARAF 1.8e−02
PPP2CB 2.5e−03
ZNF143 2.9e−02
ZAP70 1.4e−02
TBX6 1.4e−02
PLAGL2 1.4e−02
NEUROD4 9.5e−03
MARK2 1.4e−02
GNG4 4.9e−03
WDR46 1.8e−02
SMAD5 3.5e−02
PPP2R5A 9.5e−03
MYOZ2 1.4e−02
HOXB7 4.9e−03
PIK3C3 4.9e−03
VAV1 6.7e−03
BIRC2 3.3e−04
RAD51 3.2e−03
EZH2 1.4e−02
RPS7 1.4e−02
RPL27A 2.9e−02
B
Control
Confounding
SLE
Proportion of individuals (%)
0 20 40 60 80 100
0−2 3−5 6−9 10−19 20−29 30−39 40−49 50−59 60−69
Number of positive autoantibodies per individual

Figure 3

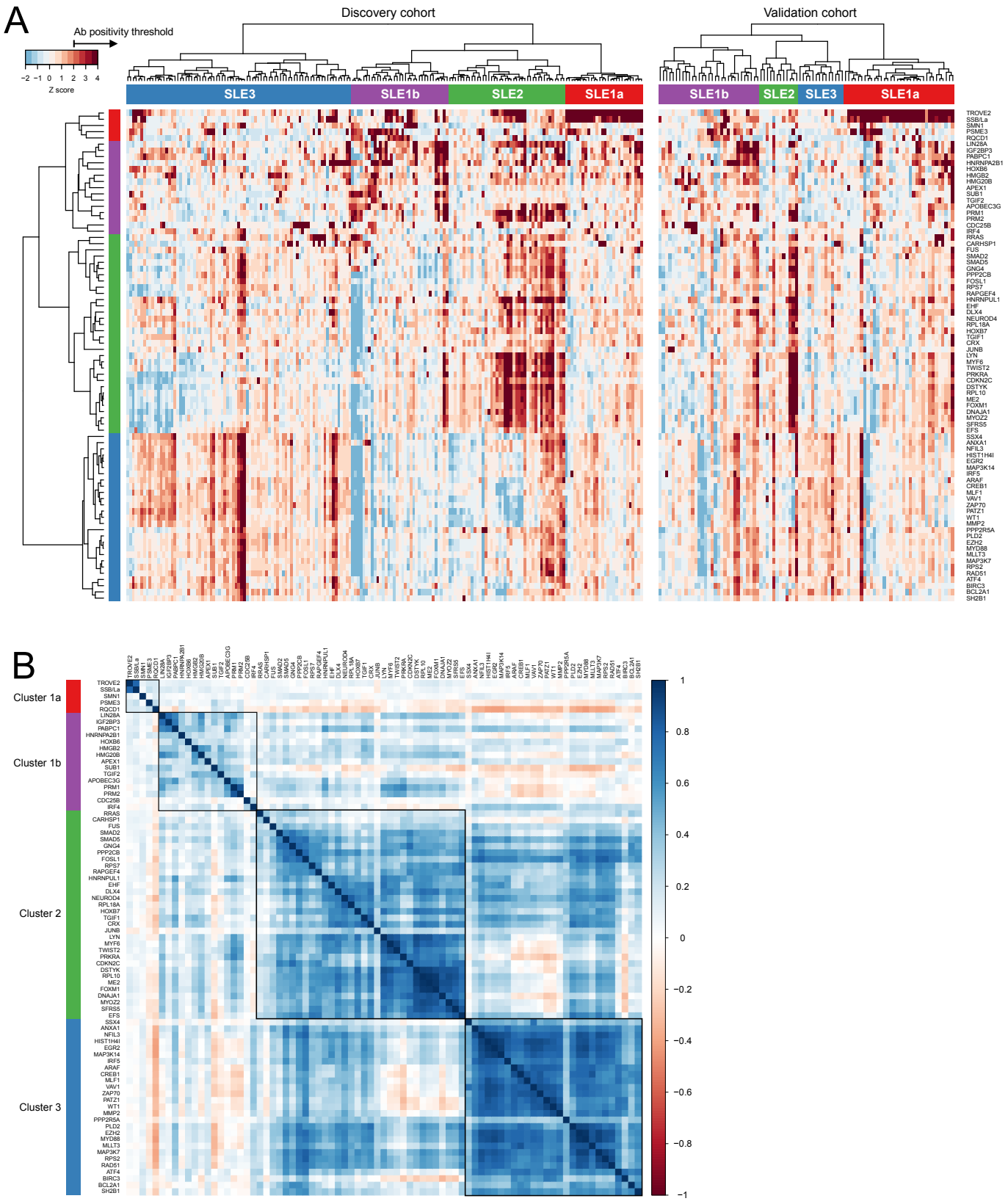

A
Ab positivity threshold
−2 −1 0 1 2 3 4
Z score
Discovery cohort
Validation cohort
SLE3
SLE1b
SLE2
SLE1a
SLE1b
SLE2
SLE3
SLE1a
B
Cluster 1a
Cluster 1b
Cluster 2
Cluster 3
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

Figure 4

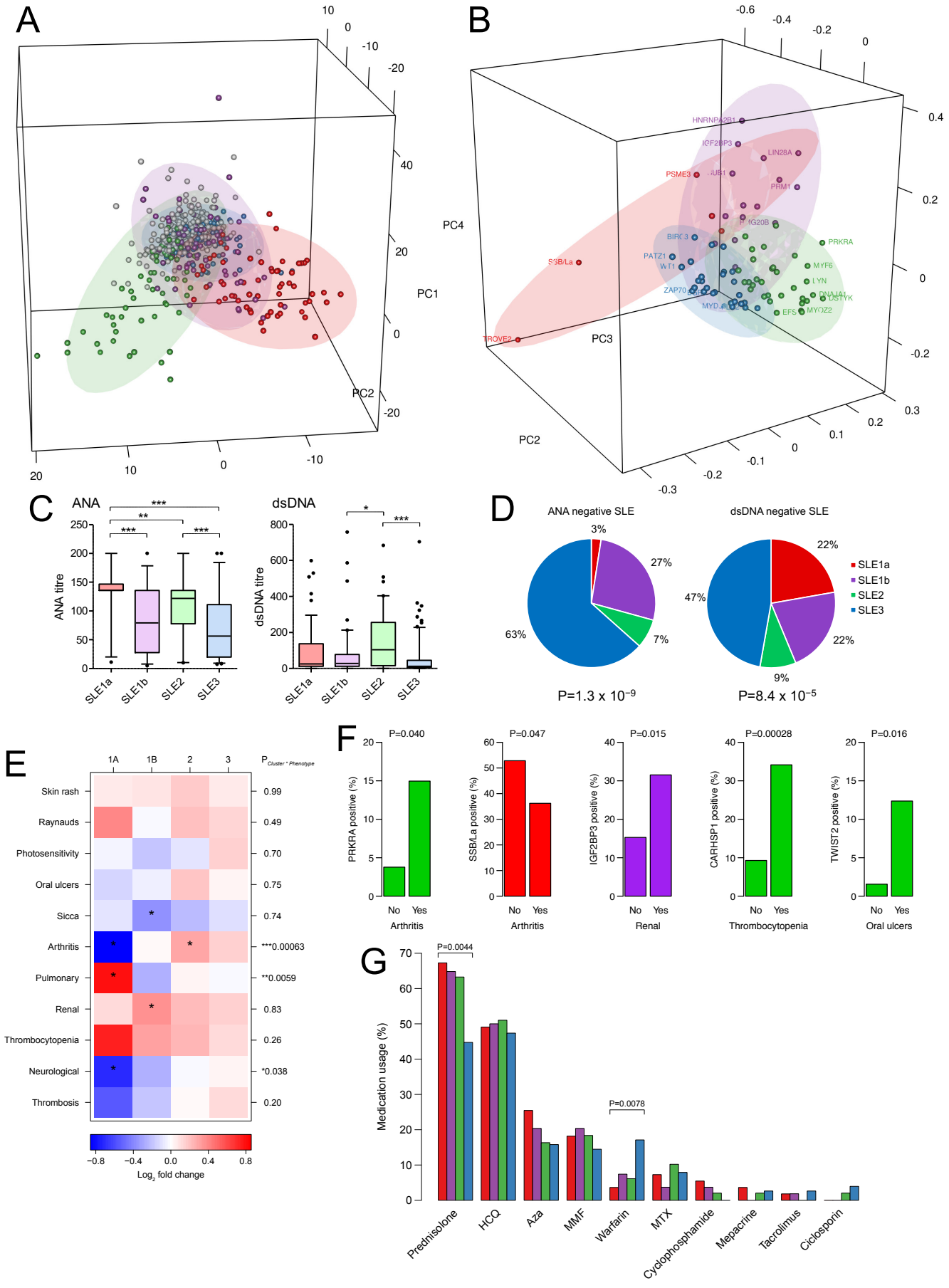

A
PC1
PC2
B
PC4
PC3
PC2
C
ANA
ANA titre
SLE1a
SLE1b
SLE2
SLE3
dsDNA
dsDNA titre
D
ANA negative SLE
3%
27%
7%
63%
P=1.3 x 10−9
dsDNA negative SLE
22%
47%
22%
9%
P=8.4 x 10−5
SLE1a
SLE1b
SLE2
SLE3
E
1A
1B
2
3
Skin rash
Raynauds
Photosensitivity
Oral ulcers
Sicca
Arthritis
Pulmonary
Renal
Thrombocytopenia
Neurological
Thrombosis
0.99
0.49
0.70
0.75
0.74
***0.00063
**0.0059
0.83
0.26
*0.038
0.20
Log2 fold change
F
PRKRA positive (%)
P=0.040
Arthritis
SSB/La positive (%)
P=0.047
Arthritis
IGF2BP3 positive (%)
P=0.015
Renal
CARHSP1 positive (%)
P=0.00028
Thrombocytopenia
TWIST2 positive (%)
P=0.016
Oral ulcers
No
Yes
G
Medication usage (%)
P=0.0044
P=0.0078
Prednisolone
HCQ
Aza
MMF
Warfarin
MTX
Cyclophosphamide
Mepacrine
Tacrolimus
Ciclosporin

Figure 5

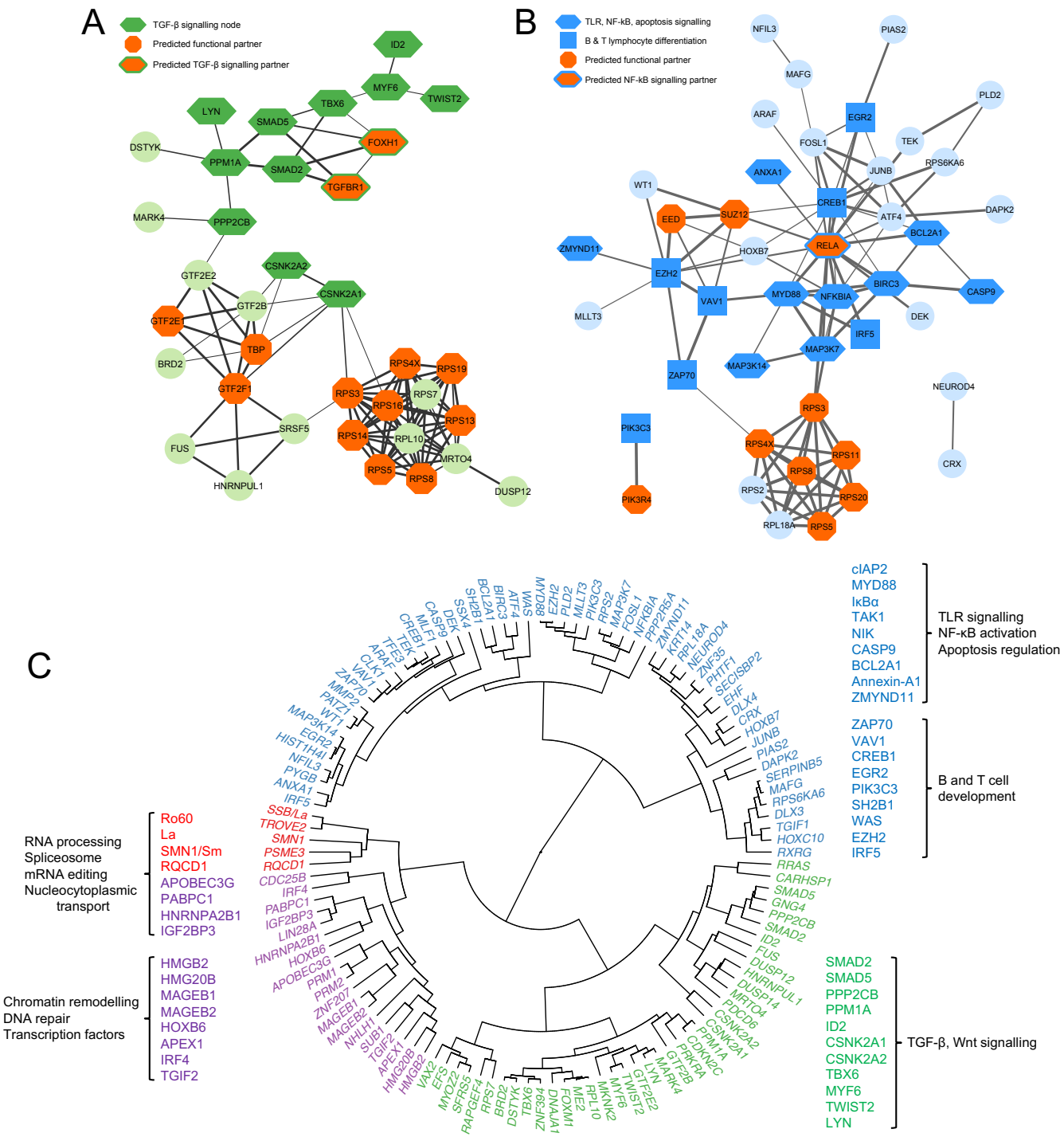
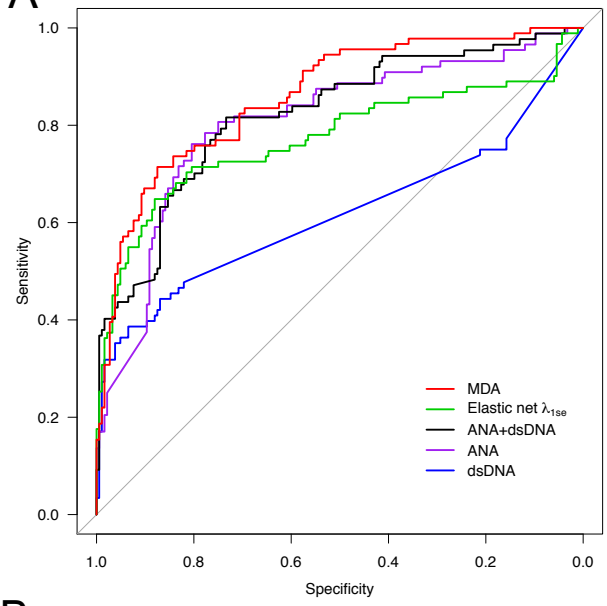

A
TGF-β signalling node
Predicted functional partner
Predicted TGF-β signalling partner
ID2
MYF6
TWIST2
TBX6
LYN
SMAD5
FOXH1
DSTYK
PPM1A
SMAD2
TGFBR1
MARK4
PPP2CB
GTF2E2
CSNK2A2
CSNK2A1
GTF2B
GTF2E1
TBP
BRD2
GTF2F1
RPS4X
RPS19
RPS3
RPS7
RPS16
RPS13
RPS14
RPL10
SRSF5
FUS
MRTO4
RPS5
RPS8
HNRNPUL1
DUSP12
B
TLR, NF-kB, apoptosis signalling
B & T lymphocyte differentiation
Predicted functional partner
Predicted NF-kB signalling partner
NFIL3
PIAS2
MAFG
PLD2
ARAF
EGR2
FOSL1
TEK
ANXA1
JUNB
RPS6KA6
WT1
CREB1
EED
SUZ12
ATF4
DAPK2
BCL2A1
ZMYND11
HOXB7
RELA
EZH2
BIRC3
CASP9
VAV1
MYD88
NFKBIA
MLLT3
DEK
IRF5
MAP3K7
MAP3K14
ZAP70
NEUROD4
RPS3
PIK3C3
RPS4X
RPS11
RPS8
RPS2
RPS20
CRX
PIK3R4
RPL18A
RPS5
C
cIAP2
MYD88
IκBα
TAK1
NIK
CASP9
BCL2A1
Annexin-A1
ZMYND11
TLR signalling
NF-κB activation
Apoptosis regulation
ZAP70
VAV1
CREB1
EGR2
PIK3C3
SH2B1
WAS
EZH2
IRF5
B and T cell development
SMAD2
SMAD5
PPP2CB
PPM1A
ID2
CSNK2A1
CSNK2A2
TBX6
MYF6
TWIST2
LYN
TGF-β, Wnt signalling
Ro60
La
SMN1/Sm
RQCD1
APOBEC3G
PABPC1
HNRNPA2B1
IGF2BP3
RNA processing
Spliceosome
mRNA editing
Nucleocytoplasmic transport
HMGB2
HMG20B
MAGEB1
MAGEB2
HOXB6
APEX1
IRF4
TGIF2
Chromatin remodelling
DNA repair
Transcription factors

Figure 6

A



B

	MDA (26 Ab model)	Elastic net (17 Ab model)	ANA + dsDNA	ANA	dsDNA
AUC	0.860	0.773	0.822	0.806	0.618
AUC 95% CI	0.813 - 0.907	0.704 - 0.842	0.766 - 0.878	0.748 - 0.865	0.538 - 0.697
Sensitivity (specificity 85%)	71.4%	65.9%	65.5%	67.0%	44.3%
Sensitivity (specificity 90%)	67.0%	59.3%	47.8%	37.0%	38.6%
Sensitivity (specificity 95%)	56.0%	50.5%	43.7%	29.3%	36.4%

C