Int J Clin Exp Pathol 2017;10(10):10681-10694
www.ijcep.com /ISSN:1936-2625/IJCEP0061261

Original Article

Serum proteomics study reveals candidate biomarkers for systemic lupus erythematosus

Lijun Zhong[2*], Jiao Liu[2*], Juntuo Zhou[3], Lin Sun[1], Changhong Li[1], Xinyi Li[1], Rui Liu[1], Jinxia Zhao[1], Bin Yang[2], Xiangyuan Liu[1], Xiaoli Deng[1]

[1]Department of Rheumatolgoy and Immunology, Peking University Third Hospital, Beijing, China; [2]Medical and Health Analytical Center, Peking University Health Science Center, Beijing, China; [3]Department of Pathology, School of Basic Medical Sciences, Peking University Health Science Center, Beijing, China. [*]Equal contributors.

Received July 11, 2017; Accepted September 29, 2017; Epub October 1, 2017; Published October 15, 2017

Abstract: Systemic lupus erythematosus (SLE) is an autoimmune disease which is characterized by the presence of autoantibodies. It will be helpful if specific serum biomarkers can be used for monitoring the disease activity as well as differentiating SLE from other diseases. For this purpose, we used a label free-based two dimensional liquid chromatography mass spectrometry platform to analyze serum samples from SLE patients in active or inactivestage. Significant differences were found for 42 serum proteins implicated in pathways including complement and coagulation cascades. Further gene set enrichment analysis revealed that gene sets including formation of fibrin clot, ECM glycoproteins and innate immune system were highly correlated with the SLE disease activity. To further assess the validity of these findings, thrombospondin-4 was selected for subsequent ELISA assays. We also explored the autoantibody of three candidate biomarkers in larger cohorts including SLE, Rheumatoid arthritis, Sjogrensyndrome patients and normal controls. Our findings provided valuable information on the proteomic changes in the serum of different SLE disease activity. Serum properdin, collectin-11 and thrombospondin-4 were valuable in monitoring the disease activity of SLE, and the autoantibodies to them may be valuable in differentiating SLE from other diseases for clinical diagnosis in the future.

Keywords: Biomarkers, systemic lupus erythematosus, label-free proteomics, bioinformatics

Introduction

Systemic lupus erythematosus (SLE) is a complex systemic autoimmune disease that can produce a lot of autoantibodies. For SLE, the molecular diagnostics are limited and pathogenesis is not clearly understood, because the disease characteristic is heterogeneous and the disease activity changes with time and therapy [1, 2]. Specific biomarkers that can differentiate SLE from other diseases and can monitor the change of disease activity are very important for both clinic diagnosis and mechanism study. Traditional biomarkers such as anti-dsDNA antibody and anti-Sm antibody are relative specific but not sensitive enough to differentiate SLE from other disease and there are some drawbacks for them to monitor the change of SLE disease activity [2, 3]. Therefore, the identification and characterization of more specific molecular and cellular targets in SLE target tissue and biomarkers of early-onset and effective response to treatment of SLE complications is meaningful and necessary [4].

Proteomics, the collective study of all expressed proteins in biological samples, can reveal information on not only the independent parts (protein expressions) but also the inter play of protein complexes and signaling pathways [5-7]. A proteomics study can be achieved using high-throughput experimental platforms such as liquid chromatography coupled tandem mass spectrometers (LC-MS/MS) [8-10]. These platforms can measure the abundance of proteins in different biological conditions, proved to be a powerful tool for biomarker study [11]. While isotopic labelling of proteins can achieve more accurate quantitative measurements [12] label-free quantitative proteomics is also popular because it is easily accessible and required less sample preparation [13].

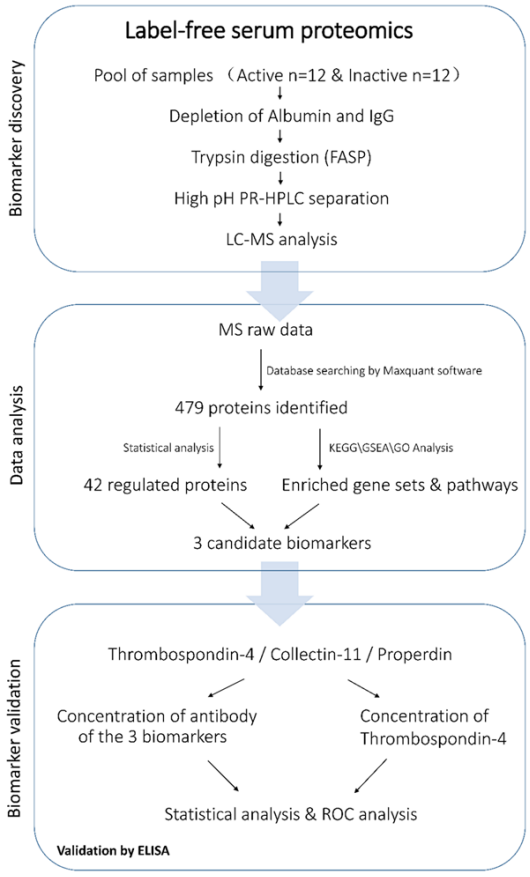

Figure 1. Schematic diagram of the overall study design. It can be divided into two main stages. 1) The biomarker discovery stage consisted of protein expression analysis and bioinformatic analysis. 2) The biomarker validation stage, which was divided into two parts: one was to verify the target protein thrombospondin-4 found in the biomarker discovery stage by ELISA; another was to make clear by antibody microarray whether the antibodies against the three target proteins (Thrombospondin-4, collectin-11, and properdin) was specific to SLE and whether they were correlated with the disease activity.

In recent years, proteomics technologies and applications have facilitated the biomarker discoveries and mechanism studies of SLE [4]. Proteomics study can provide diagnostic information for SLE by measurement of immune cell profiles and activity [14], identification of specific autoantibodies [15], and identification of changes in protein expression profiles in bodily fluid such as urine, blood [16], and cerebral spinal fluid [17]. Among these studies, varieties of assay methods were used, including MALDI-TOF, 2D-PAGE and 2D-LC-MS [4]. In the present study, we performed a label-free LC-MS workflow to investigate the protein profiling differences in serum between active and inactive stage of SLE. Assisted by following bioinformatic analysis and ELISA validation, we have found novel candidate serum biomarkers for monitoring the disease activity of SLE, as well as differentiating SLE from other diseases.

Materials and methods

Patients

From 2008 to 2015, patients from Peking University Third Hospital, who fulfilled the 1997 SLE classification criteria revised by ACR, were recruited in our study. This protocol was approved by the Ethics Committee at Peking University Third Hospital and informed consent was obtained from each patient. Three control groups were also established: one was RA (Rheumatoid arthritis, RA) group,one was SS (Primary Sjogrensyndrome, SS) group, the other was healthy control group (NC), which was consisted of healthy volunteers. Two different sample cohorts (SLE active and inactive) were used for the biomarker discovery stage, and five different sample cohorts (SLE active, SLE inactive, RA, SS and healthy control) were used for the validation stage. People in all the groups were well matched in age and gender. Detailed information of the enrolled patients and the criteria used for defining the disease stage and sample groups can be found in the Supplementary Methods section.

Shotgun analysis sample preparation

The whole workflow is shown in **Figure 1** total of 12 SLE patients in active stage and 12 in inactive stage were used in the discovery stage of this study. Serum samples in the same phenotypic group were pooled for proteomic analysis. We created three serum pools for each disease stage group, each pool contained equal amounts of serum from 4 subjects. Each pooled serum sample was subjected to albumin and IgG depletion using Aurum Serum Protein Mini Kit (Bio-Rad) according to the manufacturer's protocol. The flow-through fractions (low abundance proteins) were collected for trypsin digestion. Protein samples were digested according to the manufacturer's protocol for filter-aided sample preparation (FASP) [18]. The protein to enzyme ratio was 50:1. Samples were incubated overnight at 37°C and released peptides were collected by centrifugation.

High pH reverse phase chromatography was performed using the Dionex Ultimate 3000 Micro Binary HPLC Pump system [19]. The mobile phase used were: buffer A (20 mM ammonium formate in water, pH 10) and buffer B (20 mM ammonium formate in 80% acetonitrile, pH 10). Digested peptides mixture were loaded onto a 2.1 mm × 150 mm Waters BEH130 C-18 column containing 3.5 μm particles. Peptides were eluted at a flow rate of 230 μL/min with a gradient of 5% buffer B for 5 min, 5% to 15% buffer B for 15 min, 15% to 25% buffer B for 10 min, 25% to 55% buffer B for 10 min, and 55% to 95% buffer B for 5 min. The system was then maintained in 95% buffer B for 5 min before equilibrating with 5% buffer B for 10 min prior to the next injection. Elution was monitored by measuring the absorbance at 214 nm, and fractions were collected every 2 min. Fractions containing eluted peptides were collected into 15 fractions based on peptide density, and then were vacuum-dried before nano-ESI-LC-MS/MS analysis.

LC-MS/MS analyses

The MS analysis experiments were performed on a nano-flow HPLC system (Easy-nLC II, Thermo Fisher Scientific, USA) connected to a LTQ-OrbitrapVelos Pro mass spectrometer equipped with a Nanospray Flex Ion Source (Thermo Fisher Scientific, USA). Peptide mixtures (5 μL) were injected at a flow rate of 5 μL/min onto a pre-column (Easy-column C18-A1, 100 μm I.D. × 20 mm, 5 μm, Thermo Fisher Scientific). Chromatographic separation was performed on a reversed phase C18 column (Easy-column C18-A2, 75 μm I.D. × 100 mm, 3μm, Thermo Fisher Scientific) at a flow rate of 300 nL/min with a 60 min gradient of 2% to 40% acetonitrile in 0.1% formic acid. The electrospray voltage was maintained at 2.2 kV, and the capillary temperature was set at 250°C. The LTQ-Orbitrap was operated in data-dependent mode to simultaneously measure full scan MS spectra (m/z 350-2000) in the Orbitrap with a mass resolution of 60,000 at m/z 400. After full-scan survey, the 15 most abundant ions detected in the full-MS scan were measured in the LTQ-Orbitrap using collision-induced dissociation (CID). Each group had triple biological replicates.

Protein identification and quantitation

The data analysis was performed with MaxQuant software [20] (version 1.4.1.2, http://www.maxquant.org/). For protein identification, the MS/MS data were submitted to the Uniprot human protein database (release 3.43, 72, 340 sequences) using the Andromeda search engine with the following settings: trypsin cleavage; fixed modification of carbamidomethylation of cysteine; variable modifications of methionine oxidation; a maximum of two missed cleavages; and false discovery rate was calculated by decoy database searching. Other parameters were set as default. The results were imported into Microsoft excel for further analysis. Label-free quantitation (LFQ) was also performed in MaxQuant, the minimum ratio count for LFQ was set to 2, and the match-between-runs option was enabled. Unsupervised hierarchical clustering (Pearson linkages), heat map generation and scatter plot were carried out with the MetaboAnalyst 3.0 Web service (http://www.metaboanalyst.ca/).

Bioinformatic analysis

Pathway enrichment analysis was performed using DAVID [21], and identified proteins were mapped to the coagulation and complement cascade pathway using the pathway mapping tools of KEGG (http://www.kegg.jp/kegg/). The BiNGO plugin [22] in the Cytoscape environment [23] was used to retrieve the Gene Ontology Consortium (GOC, http://geneontology.org/) in terms of molecular function, biological process and cellular component. The statistical test used was Hypergeometric test, and the false discovery rate (FDR) associated with multiple testing was corrected using the Benjamini-Hochberg method and an FDR-corrected p value < 0.05 was considered significant.

We used the gene set enrichment analysis (GSEA) method for functional enrichment analysis [24]. Proteins with more than two unique peptides identified in all six samples were defined as qualified proteins, and used for GSEA analysis. The GSEA was performed using java GSEA (gsea2-2.1.0. jar from http://www.broadinstitute.org/gsea/downloads.jsp). The phenotypes of analyzed data were given to two classes, A (Active) and B (Inactive). All curated canonical pathways (C2) in curated molecular signature database (MSigDB, v4.0) were selected as the gene sets. The permutation type was set to gene set, and other settings were set as default. A normal p value < 0.05 and FDR q value < 0.05 was considered as a significantly enriched

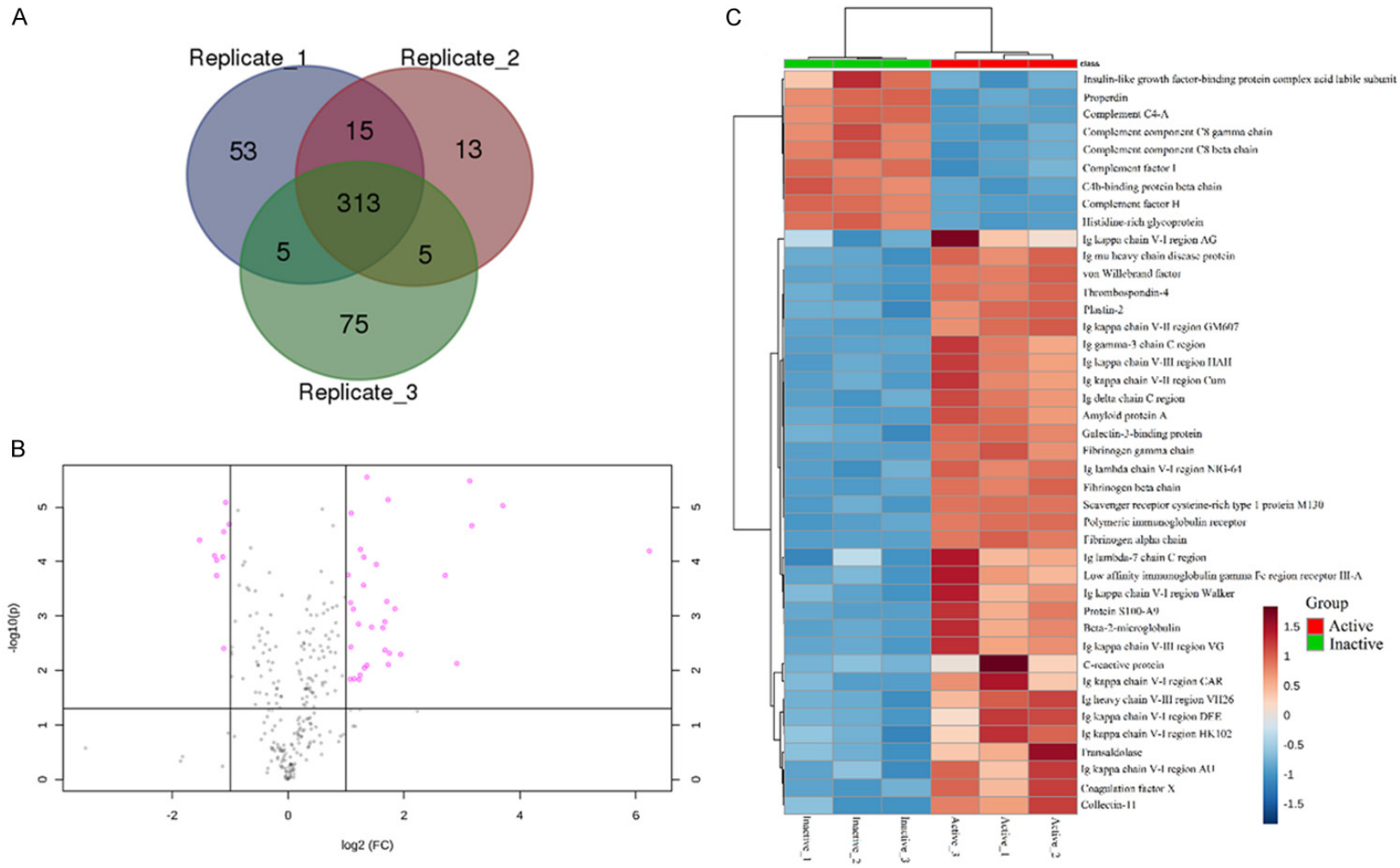

Figure 2. Protein expression analysis by LC-MS. A. Venn diagrams representing the number of proteins identified in each of the three biological replicates. B. Scatter plots constructed with $\log_2$FC (x-axis) and $\log_{10}$ *p*-value (y-axis). Pink points represent significantly dysregulated proteins with FC > 2 and *p*-value < 0.05. C. Hierarchical clustering analysis of the 42 significantly differentially expressed proteins in the proteomic analysis. 33 proteins were up-regulated (red) and 9 were down-regulated (blue) in active stage.

pathway according to GSEA documentation. The significantly enriched pathways, expression data, and all curated canonical pathways were subsequently subjected to Cytoscape (version 3.2.1) and interpreted by the Enrichment Map plugin according to user manual. The representative pathways were obtained using an overlap coefficient cutoff > 0.5.

Protein level analysis by ELISA and antibody microarray

Levels of antibodies to properdin, collectin-11 and thrombospondin-4 in serum samples from all the five groups were measured using antibody microarray. Polystyrene micro well plates (Maxisorp, Nunc, Roskilde, Denmark) were coated with anti-COLEC11 antibody (ab91483, Abcam), anti-Properdin antibody (ab25850, Abcam) or anti-THBS4 antibody (ab76861, Abcam) at 1:1000 dilution (100 µl/well). After overnight incubation at 4°C, the coated wells were washed three times and left to block with 5% milk for 2 hours at room temperature. The calibrator, controls and samples were diluted in 2% milk and incubated for 2 hours. After three washes, HRP-conjugated Streptavidin anti human IgG antibody diluted to 1/5000 in washing buffer containing 2% milk was added to the wells and incubated for 1 hour at room temperature. The wells were washed three times. TMB was added for 15 min, the color development was stopped with 1 M H_2SO_4. Optical density was measured at 490-650 nm using Vmax Kinetic Microplate Reader and the data were processed using SoftMax Pro software (Molecular Devices, Wokingham, United Kingdom). The samples were diluted to 1/100 and the calibrator and samples were run in triplicates unless otherwise stated.

Statistical analysis

For biomarker discovery stage. All statistical analyses were performed using Preseus and Graphpad prism. For the discovery stage, a 2-fold change and Student's t-test and a *p* value of 0.05 were used as combined thresholds to define biologically dysregulated proteins. For GSEA analysis, a normal *p* value of 0.05 and FDR q value of 0.25 was used as cutoff to define significantly enriched gene sets.

For biomarker validation stage. For the result of ELISA, the concentration of thrombospondin-4 was expressed as the mean ± S.E.M, and for the results of antibody arrays, relative quantification was reported as mean ± S.E.M. Statistical significance was determined by one-way ANOVA and post hoc Tukey's test, *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001; NS, not significant. The diagnostic performance of the protein biomarkers were estimated using the apparent area under the receiver operating characteristic curve (ROC) with its 95% confidence interval (CI) [25]. Results were represented as a histogram or ROC using GraphPad 6.0.

Results

Study design and workflow flowchart

The workflow of the present study is shown in **Figure 1.** It can be divided into two main stages. In the biomarker discovery stage, it consisted of protein expression analysis and bioinformatics analysis. From the result of proteomics study, we found 42 significantly differentially expressed proteins (candidate biomarkers) between active and inactive SLE. Further combining the bioinformatics analysis, we chose three proteins for the next validation stage. In the validation stage, the concentration of thrombospondin-4 in serum was analyzed using ELISA, and the protein level of anti-collectin-11, anti-thrombospondin-4 and anti-properdin in serum were analyzed by antibody microarray.

Protein expression analysis

For biomarker discovery stage, altogether 24 patients (12 active and 12 inactive) were recruited and the disease duration was from 1 month to 35 years. There was no significant difference between the age of the patients in the two groups (33±10 vs 29±10, P=0.36). The detailed data of the two groups are illustrated in Table S1. The Venn diagrams showed the number of proteins identified in each of the three biological replicates (**Figure 2A**). A total of 479 proteins were identified in all three biological replicates using the shotgun method, and 276 of these proteins have been quantified by label-free algorithm (at least 2 unique peptides of a protein have been repeatedly measured in all six samples). To select the proteins that were differentially expressed in active stage, we used the following criteria: fold change higher than 2 and *p*-value of less than 0.05 (using the Student's t-test, pink points in **Figure 2B**). As a result, we found a total of 42 significantly dif-

Table 1. Details of the 42 significantly dysregulated proteins revealed by LC-MS

Uniprot ID	Gene name	Protein name	Log2 Ratio (Active/inactive)	-Lg t-test *p* value
P27918	CFP	Properdin	-1.52901	4.30401
P05156	CFI	Complement factor I	-1.27587	3.45964
P07360	C8G	Complement component C8 gamma chain	-1.23329	3.85219
P07358	C8B	Complement component C8 beta chain	-1.23228	3.89911
P20851	C4BPB	C4b-binding protein beta chain	-1.12639	4.21013
P0C0L4	C4A	Complement C4-A	-1.1101	4.8848
P35858	IGFALS	Insulin-like growth factor-binding protein complex acid labile subunit	-1.10345	2.72782
P08603	CFH	Complement factor H	-1.07566	5.39542
P04196	HRG	Histidine-rich glycoprotein	-1.01697	4.87906
P13796	LCP1	Plastin-2	1.03747	3.36114
P01593		Lg kappa chain V-I region AG	1.06913	1.48691
P01594		Lg kappa chain V-I region AU	1.08419	2.51964
P04275	VWF	Von Willebrand factor	1.08821	4.98184
P01860	IGHG3	Lg gamma-3 chain C region	1.12164	3.7318
P37837	TALDO1	Transaldolase	1.12927	2.20462
P61769	B2M	Beta-2-microglobulin	1.20682	3.47086
A0M8Q6	IGLC7	Lg lambda-7 chain C region	1.26382	1.79461
P01702		Lg lambda chain V-I region NIG-64	1.25278	3.80361
P01880	IGHD	Lg delta chain C region	1.30188	3.75883
P04220		Lg mu heavy chain disease protein	1.30983	3.75921
P01597		Lg kappa chain V-I region DEE	1.29854	2.41149
P01602	IGKV1-5	Lg kappa chain V-I region HK102	1.37157	2.10286
P01833	PIGR	Polymeric immunoglobulin receptor	1.36094	4.91676
P00742	F10	Coagulation factor X	1.43658	3.20172
Q08380	LGALS3BP	Galectin-3-binding protein	1.5376	3.14521
P06702	S100A9	Protein S100-A9	1.02215	2.52182
P01764		Lg heavy chain V-III region VH26	1.64034	2.89607
P04433		Lg kappa chain V-III region VG	1.66446	3.52739
P04431		Lg kappa chain V-I region Walker	1.67393	2.71896
P18135		Lg kappa chain V-III region HAH	1.70047	3.80851
F5GZZ9	CD163	Scavenger receptor cysteine-rich type 1 protein M130	1.73306	4.08511
P08637	FCGR3A	Low affinity immunoglobulin gamma Fc region receptor III-A	1.7483	2.75632
Q9BWP8	COLEC11	Collectin-11	1.2424	2.1414
P01614		Lg kappa chain V-II region Cum	1.84169	3.69246
P06309		Lg kappa chain V-II region GM607	1.36063	3.67862
P02741	CRP	C-reactive protein	1.57288	1.31637
P0DJI8	SAA1	Amyloid protein A	2.72141	3.99998
P01596		Lg kappa chain V-I region CAR	2.35734	1.97073
P02671	FGA	Fibrinogen alpha chain	3.13721	5.7251
P35443	THBS4	Thrombospondin-4	3.29386	2.76728
P02675	FGB	Fibrinogen beta chain	3.787	3.41729
P02679	FGG	Fibrinogen gamma chain	6.23518	5.82024

ferentially expressed proteins, among which 33 were up-regulated and 9 were down-regulated in active stage. Hierarchical clustering analysis was performed to visualize the 42 significantly differentially expressed proteins (**Figure 2C**). The details of the 42 differentially expressed proteins, including protein ID (Uniprot), gene name, protein name, Log2 ratio active/inactive, and -Log10 *p*-value, are listed in **Table 1**.

Bioinformatic analysis

The 42 significantly dysregulated proteins were interrogated and mapped to KEGG pathways

Table 2. Significantly enriched KEGG pathways by DAVID

Term	Count	%	*P*-Value	Benjamini
Complement and coagulation cascades	11	3.4	3.1E-16	3.7E-15
Systemic lupus erythematosus	4	1.2	0.0027	0.015
Prion diseases	2	0.6	0.099	0.32

(**Table 2**). The first ranked pathway was the complement and coagulation cascade, a total of 37 detected serum proteins were mapped to the pathway, with 11 demonstrating a significant regulation between active and inactive stage of SLE (**Figure 3**). To further extend our knowledge about the change of serum proteins between disease stages, gene ontology (GO) analysis was performed to reveal the molecular function, biological process and cellular component associated with the 42 significantly dysregulated proteins. As a result, the significantly dysregulated proteins are highly correlated with protein binding and scavenger receptor activity in terms of molecular function (**Figure 4A**), participating the processes of inflammatory response, compliment activation and coagulation (**Figure 4B**), and mainly exist at extracellular region and platelet alpha granule lumen (**Figure 4C**). Details of the GO analysis results are shown in Table S4.

At last, we applied gene set enrichment analysis using the profiling result of all the 276 quantified serum proteins to detect more biology-driven gene sets without biases toward significantly different expressed genes. The results showed that 4 gene sets in phenotype A (active) and 8 gene sets in phenotype B (inactive) were significantly enriched at FDR < 25%. Formation of fibrin clot clotting cascade, ECM glycoproteins, core extracellular matrix were enriched in active stage while complement cascade and innate immune system were enriched in inactive stage at both NOM *p*-value < 0.01 and FDR < 25% (**Figure 5A**, **Table 3**). A network of the gene sets was constructed using Enrichment map plug in to visualize the significantly enriched gene sets and their relation with each other. The enrichment plot (profile of the running ES Score & positions of gene set members on the rank ordered list) of the 5 significantly enriched gene sets were showed in **Figure 5B**. The details of the 5 significantly enriched gene sets, including name, size, NES, nom *p*-value, FDR q-value, core enrichment genes are listed in **Table 3**. Considering all the analysis results mentioned above, we chose the following threeproteins for further validation and exploration: properdin, collectin-11 and thrombospondin-4.

Biomarker validation and autoantibody quantification

For the quantification of serum thrombospondin-4, patients belonged to three groups were recruited. There was no significant difference between the age of the patients in the groups. The detailed data of the patients and groups is illustrated in Table S2. There was significant difference in the serum level of thrombospondin-4 between patients in SLE active group and inactive group ($P < 0.0001$, **Figure 6A**). The AUC of the ROC analysis was 0.8622 (**Figure 6B**).

Figure 7A-C shows the relative quantification results of the antibodies against thrombospondin-4, collectin-11 and properdin. The difference of all the three autoantibodies between the SLE group (including both active group and inactive group) and the other two groups (SS and NC) were statistically significant. The difference of none of the serum level of autoantibodies between the two SLE groups was statistically significant ($P > 0.05$). In details, serum level of anti-thrombospondin-4 was significantly higher in SLE group than that in SS group and NC group ($P < 0.01$, $P < 0.0001$, respectively), but was not different from that in RA group ($P > 0.05$, **Figure 7A**). Serum level of anti-collectin-11 was significantly higher in SLE group than that in RA, SS and NC groups ($P < 0.0001$, $P < 0.0001$, $P < 0.05$, respectively, **Figure 7B**). Serum level of anti-properdin was significantly higher in SLE group than that in SS group and NC group ($P=0.08$, $P < 0.0001$, respectively), but was not different from that in RA group ($P > 0.05$, **Figure 7C**). The ROC curve of SLE and NC group are shown in **Figure 7D-F**. Furthermore, receiver-operating characteristic (ROC) curve analysis revealed that the protein levels of anti-thrombospondin-4, anti-collectin-11 and anti-properdin yielded an area under the curve (AUC) of 0.8031, 0.7368, and 0.8910, separately.

Discussion

The large-scale study of biological systems by mass-spectrometry based shotgun proteomics can provide deep insights into protein abundance and their expression patterns, which

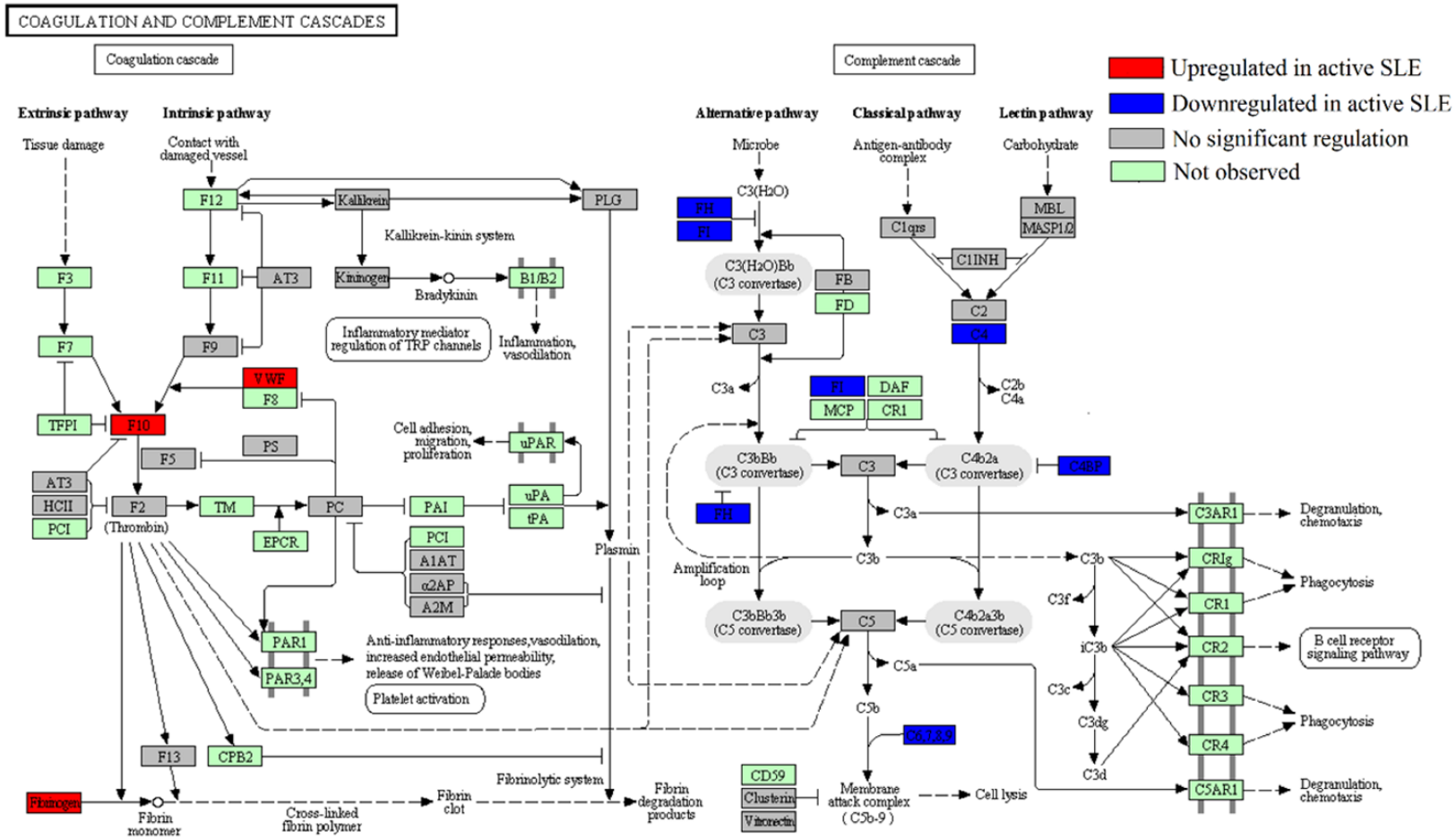

Figure 3. Pathway of coagulation and coagulation cascades appear inversely regulated in different SLE disease activity. The 42 significant dysregulated proteins were interrogated by DAVID and mapped to KEGG pathways. The most significant pathway was the coagulation and complement cascade ($P=3.1e^{-16}$, Benjamini $=3.7e^{-15}$). A total of 37 detected serum proteins mapped to the pathway, with 11 demonstrating a significant ($P < 0.05$) regulation between disease activity. The colors of the nodes represent protein levels in active SLE stage revealed by LC-MS (Red, up-regulated; blue, down-regulated; gray, detected with no significant change; green, not detected).

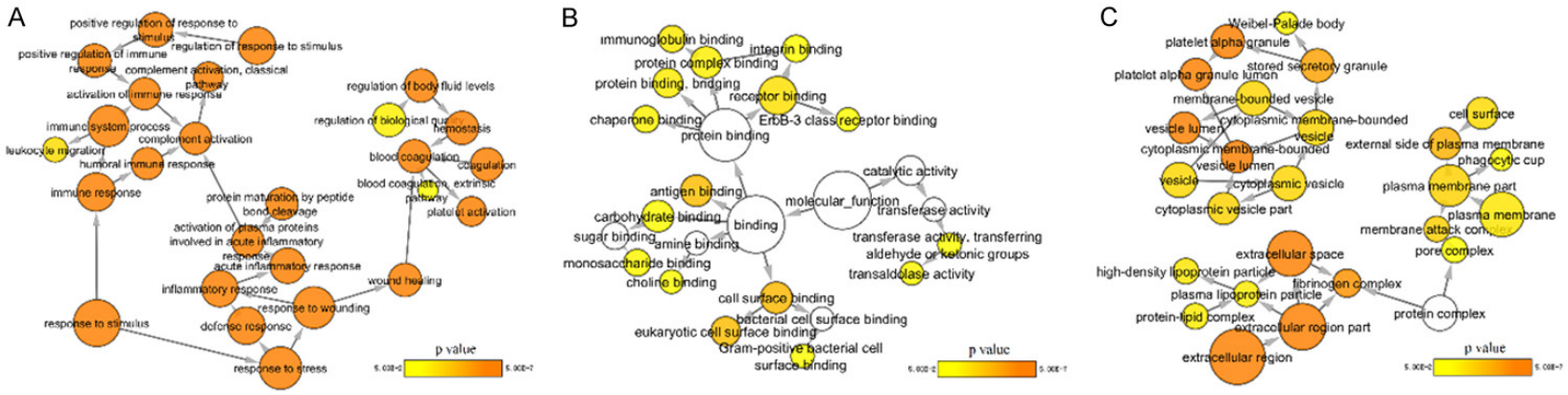

Figure 4. Gene ontology annotation of the 42 significantly dysregulated proteins. Gene ontology annotation was performed by BINGO plugin and visualized in Cytoscape. The color of the node represent the significant *p* value from high (Yellow, 0.05) to low (Orange, 5e-7). A: Gene ontology enrich result in terms of molecular function. B: Gene ontology enrich result in terms of biological processes. C: Gene ontology enrich result in terms of cellular components.

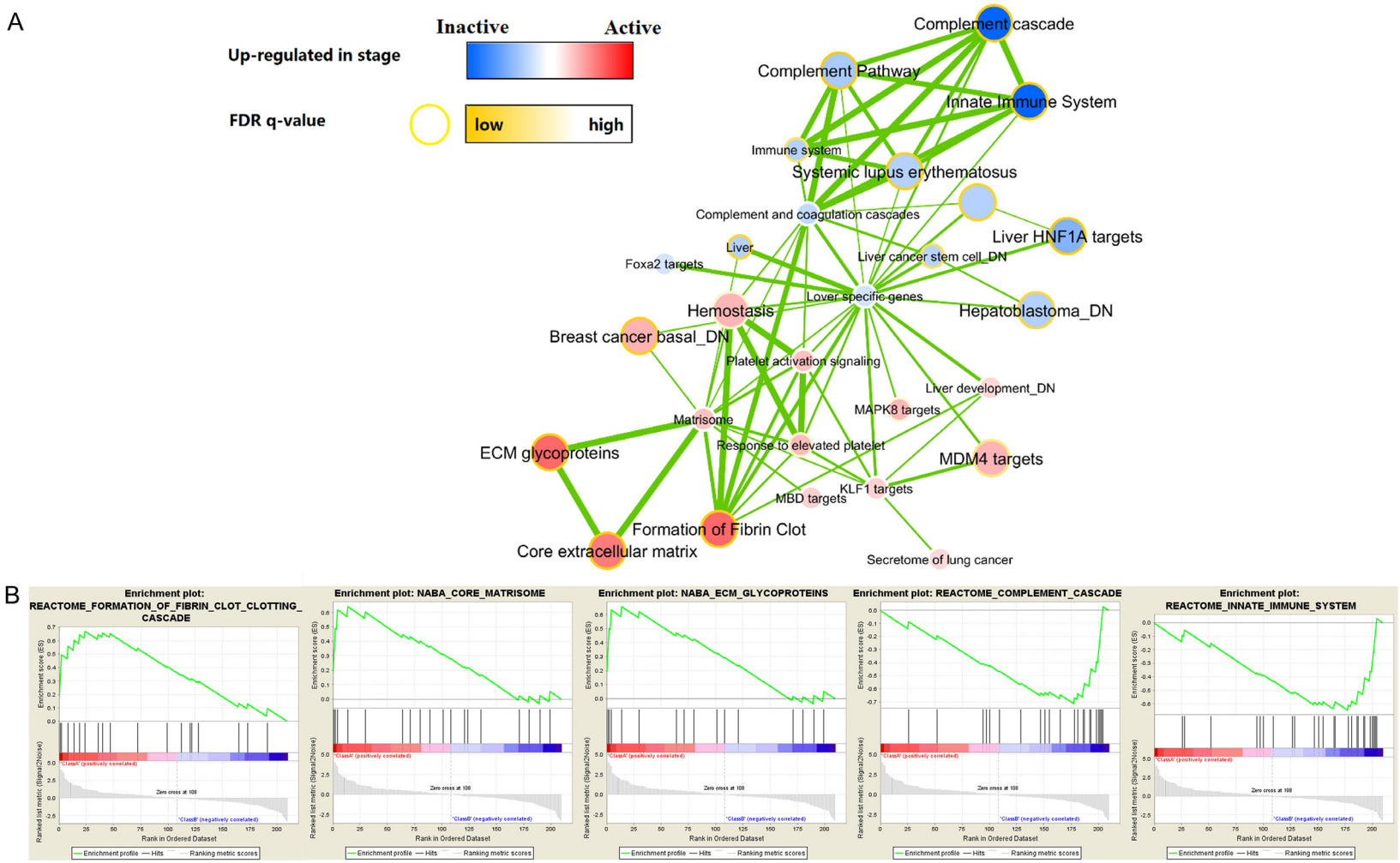

Figure 5. Gene Set Enrichment Analysis (GSEA) of the proteomic result. A. A network of the gene sets constructed using Enrichment map plugin for visualizing the significantly enriched gene sets and their relation with each other. B. The Enrichment plot (profile of the running ES Score & positions of gene set members on the rank ordered list) of the 5 significantly enriched gene sets.

Table 3. Details of the enriched gene sets of the GSEA analysis

Enriched in class	Name	Size	NES	NOM *p*-value	FDR q-value	Core enrichment genes
Active	REACTOME_FORMATION_OF_FIBRIN_CLOT	19	1.7631	0.0051	0.1405	FGG, FGA, FGB, F10, VWF, F9, PROC
Active	NABA_ECM_GLYCOPROTEINS	16	1.7153	0.0052	0.1225	FGG, FGA, FGB, THBS4, VWF
Active	NABA_CORE_MATRISOME	19	1.7066	0.0064	0.0909	FGG, FGA, FGB, THBS4, VWF
Inactive	REACTOME_COMPLEMENT_CASCADE	23	-2.092	0	0.0035	MBL2, C4BPA, C6, C3, C5, C8A, CFH, C4A, C4BPB, C8B, C8G, CFI
Inactive	REACTOME_INNATE_IMMUNE_SYSTEM	26	-2.0001	0	0.005	PROS1, LBP, MBL2, C4BPA, C6, C3, C5, C8A, CFH, C4A, C4BPB, C8B, C8G, CFI

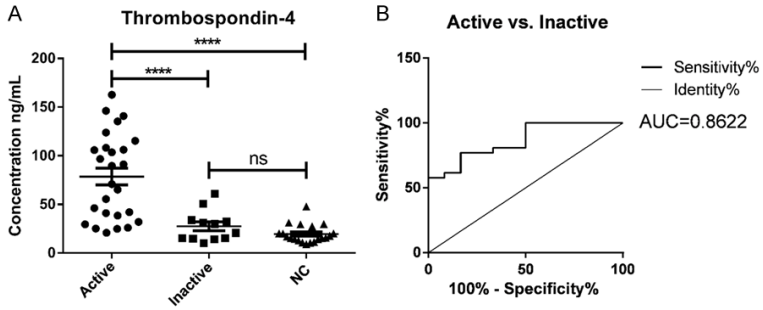

Figure 6. The serum level of thrombospondin-4 analyzed by ELISA. A: Serum level of thrombospondin-4 in three groups (SLE active, SLE inactive, and normal control). There was significant difference in the serum level of thrombospondin-4 between SLE active group and inactive group (****P < 0.0001) but no difference between SLE inactive group and normal control group (ns). B: Receiver-operating characteristic (ROC) curve analysis revealed that the protein level of thrombospondin-4 yielded an area under the curve (AUC) of 0.8622.

may carry much important biological information. Lots of studies have been performed to investigate biomarkers for SLE using a variety of methods. Christopher et al. performed sequencing and profiling of autoantibodies of SLE using mass spectrometry [26], while Brad et al. used ELISA to study plasma, urine, and renal expression of adiponectin in human SLE [27]. To the best of our knowledge, few label-free LC-MS based study of serum biomarker screening in different stage of SLE (Active vs inactive) has been reported currently. The results of our quantitative proteomics revealed large valuable information about SLE, including 479 identified proteins and 42 candidate biomarkers (**Table 1**). We used the sample pooling design for the proteomic study to reduce the analysis time as well as the influence of individual variation on selecting proper candidate biomarkers in the discovery stage. And our further ELISA experiments (**Figure 6**) in the validation stage using a larger sample size of individual samples fully confirmed the results in discovery stage, which illustrates the good performance and credibility of our proteomic study.

The following bioinformatic analysis revealed much information about the pathways, gene ontology categories and gene sets those were highly correlated with the SLE disease stage. The pathway of coagulation cascade was significantly up-regulated in active stage of SLE while complement cascade was significantly down-regulated (**Figure 3**, **Table 2**), which was consisted with the GO result (**Figure 4**; Table S4) and GSEA result (**Figure 5**; **Table 3**). The differently regulated proteins mainly have the molecular function of protein binding, and localized at extracellular regions (**Figure 4**, Table S4). These findings inspire us that the coagulation and complement relevant proteins in the extracellular regions may play important roles when the SLE disease activity changes.

Based on the protein profiling and bioinformatic analysis, we finally focused on three serum proteins and their antibodies for the validation stage: properdin, collectin-11 and thrombospondin-4. The collectins are a group of innate immune proteins structurally characterized by their content of a carbohydrate recognition domain and a collagen-like region [28]. Collectin-11 is the more recently described member of this group [29]. Collectin-11 is a secreted-type collectin and it is a soluble protein found in the serum at a mean concentration of 284 ng/ml, and it exists in complex with MASPs (mannan-binding lectin (MBL)-associated serine proteases, MASPs.). Collectin-11 is also involved in the lectin activation. It binds to microorganisms and apoptotic cells, and its binding to microorganisms leads to complement activation via MASPs in vitro [30]. Colle-

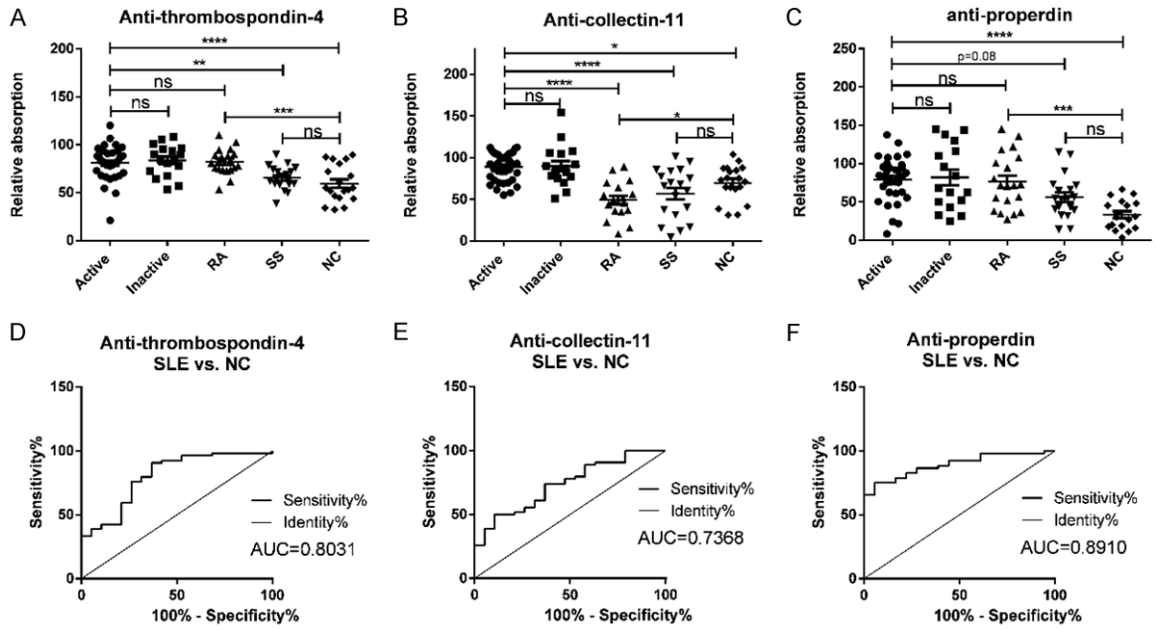

Figure 7. The serum level of antibodies against the three proteins. Grouped scatter plot reported as mean ± S.E.M. and ROC curve were used to present the relative quantification result of the three autoantibodies in groups of SLE active, SLE inactive, RA (Rheumatoid arthritis), SS (Primary sjogren syndrome) and NC (Normal control). Statistical significance was determined by one-way ANOVA and post hoc Tukey's test, *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001; NS, not significant. A: Serum level of anti-thrombospondin-4 in all the five groups. B: Serum level of anti-collectin-11 in all the five groups. C: Serum level of anti-properdin in all the five groups. D: ROC curve constructed with serum level of anti-thrombospondin-4 between SLE (Both active and inactive stage) and NC. E: ROC curve constructed with serum level of anti-collectin-11 between SLE (Both active and inactive stage) and NC. F: ROC curve constructed with serum level of anti-properdin between SLE (Both active and inactive stage) and NC.

ctin-11 is expressed primarily by cells in the adrenal gland, kidney, and liver.

Properdin is a plasma glycoprotein of the complement system. It is the only known positive regulator of the complement cascade [31]. Based on recent studies [32], the role of properdin in alternative pathway complement activation should be viewed as a stabilizer of preformed C3bBb convertase on the cell surface, as well as a platform to recruit and assemble new C3bBb complexes. Properdin was one factor of alternative complement activation pathway so the decrease of it during disease activity indicated that alternative activation pathway played an important role in the pathogenesis of SLE. This result was consistent with the results of Sato et al. [33]. The result of their study showed that the glomerular deposition relevant complement component, especially properdin, may be an index of the histological activity of lupus nephritis. It indicated that maybe properdin was involved in the pathogenesis of SLE.

There are three pathways for the activation of complement: classical, alternative and lectin. Collectin-11 and properdin belong to lectin and alternative pathway respectively. Our results indicated that both the alternative and the lectin pathways were also involved in the complement activation of SLE. Whereas the change of serum properdin was similar to that of complement 3 and complement 4, the change of serum collectin-11 was on the contrary. This was in accordance with the previous investigations by A. Troldborg [29], which showed that patients with the highest disease activity have higher levels of collectin-11. It is possible that different pathways of complement activation played different roles in the pathogenesis of SLE. Complement activation may act as a double-edged sword, being highly important in preventing SLE and exacerbating it once the disease has been established.

The concentration of thrombospondin-4 in active SLE patients was significantly higher than those in remission. Thrombospondin-4 is a secreted multi-domain glycoprotein of the extracellular matrix belonging to a family of at least five thrombospondins [34]. Studies have provided little information about the physiological

functions of thrombospondin-4. Thrombospondin-4 has been shown to stimulate the proliferation of erythroid cells, hematopoietic precursors (CD34-positive cells), skin fibroblasts and kidney epithelial cells. However, the protein also has anti-proliferatory effects, for example in endothelial cells. Other proposed functions include a supportive role in myoblast adhesion and interactions with other extracellular matrix proteins, such as certain collagens, laminin a, fibronectin and matrilin. Some studies also found that thrombospondin-4 has proangiogenic effect [35]. There were no previous studies investigating the correlation between thrombospondin-4 and SLE. Given these information and our results, we hypothesized that thrombospondin-4 plays a role in the pathogenesis of SLE, however more studies are needed to clarify its mechanism.

In summary, our present study revealed valuable information about the differences in serum protein profile between active and inactive stage of SLE, and indicated that serum levels of thrombospondin-4 were positively correlated with the disease activity of SLE and they might be valuable in the monitoring of the disease activity of SLE. What's more, the antibodies of collectin-11, thrombospondin-4 and properdin serum were also confirmed to be able to distinguish SLE from healthy controls, and the combination of these proteins and their antibodies would help us to identify SLE from other autoimmune disease as well as evaluate the disease activity of SLE. These candidate biomarkers are potential for the diagnostic usage in clinical assay, even though studies including larger number of samples are still needed to verify the results in the future.

Acknowledgements

This study was supported by a Project of the National Natural Science Foundation of China (81501390).

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Xiaoli Deng, Department of Rheumatology and Immunology, Peking University Third Hospital, 49 North Garden Road, Haidian District, Beijing 100191, China. Tel: +86-10-82266919; E-mail: dxiaoli0910@hotmail.com

References

[1] Font J, Cervera R. 1982 revised criteria for classification of systemic lupus erythematosus--ten years later. Lupus 1993; 2: 339-41; discussion 43.

[2] Molino C, Fabbian F, Longhini C. Clinical approach to lupus nephritis: recent advances. Eur J Intern Med 2009; 20: 447-53.

[3] Rahman A, Isenberg DA. Systemic lupus erythematosus. N Engl J Med 2008; 358: 929-39.

[4] Korte EA, Gaffney PM, Powell DW. Contributions of mass spectrometry-based proteomics to defining cellular mechanisms and diagnostic markers for systemic lupus erythematosus. Arthritis Res Ther 2012; 14: 204.

[5] Wu X, Hasan MA, Chen JY. Pathway and network analysis in proteomics. J Theor Biol 2014; 362: 44-52.

[6] Paul D, Kumar A, Gajbhiye A, Santra MK, Srikanth R. Mass spectrometry-based proteomics in molecular diagnostics: discovery of cancer biomarkers using tissue culture. Biomed Res Int 2013; 2013: 783131.

[7] Luber CA, Cox J, Lauterbach H, Fancke B, Selbach M, Tschopp J, Akira S, Wiegand M, Hochrein H, O'Keeffe M, Mann M. Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. Immunity 2010; 32: 279-89.

[8] Ono M, Shitashige M, Honda K, Isobe T, Kuwabara H, Matsuzuki H, Hirohashi S, Yamada T. Label-free quantitative proteomics using large peptide data sets generated by nanoflow liquid chromatography and mass spectrometry. Mol Cell Proteomics 2006; 5: 1338-47.

[9] Altelaar AF, Munoz J, Heck AJ. Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet 2013; 14: 35-48.

[10] Kingsmore SF. Multiplexed protein measurement: technologies and applications of protein and antibody arrays. Nat Rev Drug Discov 2006; 5: 310-20.

[11] Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003; 422: 198-207.

[12] Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol 2005; 1: 252-62.

[13] Neilson KA, Ali NA, Muralidharan S, Mirzaei M, Mariani M, Assadourian G, Lee A, van Sluyter SC, Haynes PA. Less label, more free: approaches in label-free quantitative mass spectrometry. Proteomics 2011; 11: 535-53.

[14] Dai Y, Hu C, Huang Y, Huang H, Liu J, Lv T. A proteomic study of peripheral blood mononuclear cells in systemic lupus erythematosus. Lupus 2008; 17: 799-804.

[15] Seko Y, Matsumoto A, Fukuda T, Imai Y, Fujimura T, Taka H, Mineki R, Murayama K, Hirata Y, Nagai R. A case of neonatal lupus erythematosus presenting delayed dilated cardiomyopathy with circulating autoantibody to annexin A6. Int Heart J 2007; 48: 407-15.
[16] Suzuki M, Wiers K, Brooks EB, Greis KD, Haines K, Klein-Gitelman MS, Olson J, Onel K, O'Neil KM, Silverman ED, Tucker L, Ying J, Devarajan P, Brunner HI. Initial validation of a novel protein biomarker panel for active pediatric lupus nephritis. Pediatr Res 2009; 65: 530-6.
[17] Katsumata Y, Kawaguchi Y, Baba S, Hattori S, Tahara K, Ito K, Iwasaki T, Yamaguchi N, Oyama M, Kozuka-Hata H, Hattori H, Nagata K, Yamanaka H, Hara M. Identification of three new autoantibodies associated with systemic lupus erythematosus using two proteomic approaches. Mol Cell Proteomics 2011; 10: M110.005330.
[18] Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods 2009; 6: 359-62.
[19] Yang F, Shen Y, Camp DG 2nd, Smith RD. High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. Expert Rev Proteomics 2012; 9: 129-34.
[20] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 2008; 26: 1367-72.
[21] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4: 44-57.
[22] Maere S, Heymans K, Kuiper M. BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 2005; 21: 3448-9.
[23] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-504.
[24] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102: 15545-50.
[25] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143: 29-36.
[26] Tipton CM, Fucile CF, Darce J, Chida A, Ichikawa T, Gregoretti I, Schieferl S, Hom J, Jenks S, Feldman RJ, Mehr R, Wei C, Lee FE, Cheung WC, Rosenberg AF, Sanz I. Diversity, cellular origin and autoreactivity of antibody-secreting cell population expansions in acute systemic lupus erythematosus. Nat Immunol 2015; 16: 755-65.
[27] Rovin BH, Song H, Hebert LA, Nadasdy T, Nadasdy G, Birmingham DJ, Yung Yu C, Nagaraja HN. Plasma, urine, and renal expression of adiponectin in human systemic lupus erythematosus. Kidney Int 2005; 68: 1825-33.
[28] Ohtani K, Suzuki Y, Wakamiya N. Biological functions of the novel collectins CL-L1, CL-K1, and CL-P1. J Biomed Biotechnol 2012; 2012: 493945.
[29] Troldborg A, Thiel S, Jensen L, Hansen S, Laska MJ, Deleuran B, Jensenius JC, Stengaard-Pedersen K. Collectin liver 1 and collectin kidney 1 and other complement-associated pattern recognition molecules in systemic lupus erythematosus. Clin Exp Immunol 2015; 182: 132-8.
[30] Selman L, Hansen S. Structure and function of collectin liver 1 (CL-L1) and collectin 11 (CL-11, CL-K1). Immunobiology 2012; 217: 851-63.
[31] Lesher AM, Nilsson B, Song WC. Properdin in complement activation and tissue injury. Mol Immunol 2013; 56: 191-8.
[32] Hourcade DE. The role of properdin in the assembly of the alternative pathway C3 convertases of complement. J Biol Chem 2006; 281: 2128-32.
[33] Sato N, Ohsawa I, Nagamachi S, Ishii M, Kusaba G, Inoshita H, Toki A, Horikoshi S, Ohi H, Matsushita M, Tomino Y. Significance of glomerular activation of the alternative pathway and lectin pathway in lupus nephritis. Lupus 2011; 20: 1378-86.
[34] Frolova EG, Drazba J, Krukovets I, Kostenko V, Blech L, Harry C, Vasanji A, Drumm C, Sul P, Jenniskens GJ, Plow EF, Stenina-Adognravi O. Control of organization and function of muscle and tendon by thrombospondin-4. Matrix Biol 2014; 37: 35-48.
[35] Muppala S, Frolova E, Xiao R, Krukovets I, Yoon S, Hoppe G, Vasanji A, Plow E, Stenina-Adognravi O. Proangiogenic properties of thrombospondin-4. Arterioscler Thromb Vasc Biol 2015; 35: 1975-86.

Table S1. Characteristics of SLE patients in the two groups for proteomics study

Group	Active stage (n=12)	Inactive stage (n=12)
Demographic characteristics		
Age (years)	33±10 (20-52)	29±10 (20-52)
Gender		
M	1	1
F	11	11
Disease duration	1 month-35 years	1 month-30 years
Clinical and lab characteristics		
SLEDAI-2000	13±5 (8-30)	2±2 (0-4)
Anti-dsDNA (%)	6 (50.0%)	0
Anti-Sm (%)	5 (41.7%)	4 (33.3%)
Anti-nRNP (%)	7 (58.3%)	5 (41.7%)
Anti-SSA (%)	8 (66.7%)	8 (66.7%)
Anti-SSB (%)	3 (25.0%)	1 (8.3%)
Anti-Jo-1 (%)	0	0
Anti-SCl-70 (%)	0	0
Anti-rRNP (%)	5 (41.7%)	5 (41.7%)
ACL (%)	1 (8.3%)	8 (8.3%)
Organs involved		
Kidney (%)		
Joint and muscle (%)	4 (33.3%)	
Gastrointestine (%)	7 (58.3%)	
Cytopenia (%)	1 (8.3%)	
Lung (%)	4 (33.3%)	
Nervous system (%)	1 (8.3%)	
Heart (%)	2 (16.7%)	
Complications	0	
APS (%)	1 (8.3%)	

Table S2. Characteristics of SLE patients in the three groups for ELISA validation

Group	Active stage (n=26)	Inactive stage (n=12)	Normal control (n=20)
Demographic characteritics			
Age (years)	37±14 (20-77)	29±7 (20-48)	38±11 (22-59)
Gender			
M	2	1	4
F	24	11	16
Disease duration	1 month-30 years	1 month-29 years	
Clinical and lab characteristics			
SLEDAI-2000	14±5 (9-28)	2±1 (0-3)	
Anti-dsDNA (%)	12 (46.2%)	0	
Anti-Sm (%)	10 (38.4%)	5 (41.7%)	
Anti-nRNP (%)	10 (38.4%)	6 (50.0%)	
Anti-SSA (%)	15 (57.7%)	9 (75%)	
Anti-SSB (%)	4 (15.4%)	2 (16.7%)	
Anti-Jo-1 (%)	0	0	
Anti-SCl-70 (%)	0	0	
Anti-rRNP (%)	8 (30.7%)	5 (41.7%)	
ACL (%)	2 (7.7%)	2 (7.7%)	
Organs involved			
Kidney (%)			
Joint and muscle (%)	11 (42.3%)		
Gastrointestine (%)	12 (46.2%)		
Cytopenia (%)	2 (7.7%)		
Lung (%)	7 (27.0%)		
Nervous system (%)	2 (7.7%)		
Heart (%)	4 (15.4%)		
Complications	0		
APS (%)	2 (7.7%)		

Table S3. Characteristics of the patients in the five groups for the antibody microarray

Group	SLE active (n=37)	SLE remission (n=17)	RA (n=20)	SS (n=20)	Healthy control (n=20)
Age (years)	37±14 (20-77)	29±7 (20-48)	37±15 (17-63)	38±11 (25-58)	35±13 (17-56)
Gender					
M	2	1	3	1	3
F	35	16	17	19	17
Disease duration	1 month-35 years	1 month-29 years	1 month-24 years	3 month-18 years	

Table S4. Details of the results of the gene ontology analysis

Biological process

GO-ID	Description	p-val	corr p...	cluster freq	total freq	genes
9611	response to wounding	1.7540E-19	7.8578E-17	18/28 64.2%	541/14305 3.7%	CRP FGB FGA CD163 CFH F10 VWF CFI FGG C4BPB CFP C8B C4A C8G SAA1 SAA2 HRG S100A9
2526	acute inflammatory response	6.3975E-18	1.4330E-15	11/28 39.2%	90/14305 0.6%	CRP C4A CD163 C8G CFH CFI SAA1 SAA2 C4BPB CFP C8B
6956	complement activation	5.2919E-15	7.3546E-13	8/28 28.5%	40/14305 0.2%	CRP C4A C8G CFH CFI C4BPB CFP C8B
2541	activation of plasma proteins involved in acute inflammatory response	6.5666E-15	7.3546E-13	8/28 28.5%	41/14305 0.2%	CRP C4A C8G CFH CFI C4BPB CFP C8B
2376	immune system process	8.7124E-14	7.8063E-12	17/28 60.7%	949/14305 6.6%	CRP CFH CFI IGKV1-5 C4BPB CFP C8B IGHG3 C4A FCGR3A C8G SAA1 IGHD SAA2 LCP1 B2M S100A9
6954	inflammatory response	2.4324E-13	1.3752E-11	12/28 42.8%	316/14305 2.2%	CRP C4A CD163 C8G CFH CFI SAA1 SAA2 C4BPB CFP C8B S100A9
7596	blood coagulation	2.4557E-13	1.3752E-11	9/28 32.1%	104/14305 0.7%	FGB FGA F10 VWF FGG SAA1 SAA2 C4BPB HRG
50817	coagulation	2.4557E-13	1.3752E-11	9/28 32.1%	104/14305 0.7%	FGB FGA F10 VWF FGG SAA1 SAA2 C4BPB HRG
7599	hemostasis	4.1204E-13	2.0511E-11	9/28 32.1%	110/14305 0.7%	FGB FGA F10 VWF FGG SAA1 SAA2 C4BPB HRG
6959	humoral immune response	1.3826E-12	6.1942E-11	8/28 28.5%	77/14305 0.5%	CRP C4A C8G CFH CFI C4BPB CFP C8B
6955	immune response	1.5984E-12	6.5100E-11	14/28 50.0%	619/14305 4.3%	CRP CFH CFI IGKV1-5 C4BPB CFP C8B IGHG3 C4A FCGR3A C8G IGHD LCP1 B2M
50896	response to stimulus	1.7600E-12	6.5706E-11	25/28 89.2%	3633/14305 25.3%	LGALS3BP CRP CFH CFI C4BPB CFP C8B IGHG3 C4A FCGR3A C8G B2M FGB FGA CD163 F10 VWF
51605	protein maturation by peptide bond cleavage	2.1029E-12	7.2470E-11	8/28 28.5%	81/14305 0.5%	CRP C4A C8G CFH CFI C4BPB CFP C8B
48583	regulation of response to stimulus	4.2777E-12	1.3688E-10	13/28 46.4%	525/14305 3.6%	CRP CFH CFI C4BPB CFP C8B THBS4 C4A C8G SAA1 SAA2 HRG B2M
50778	positive regulation of immune response	5.4888E-12	1.6351E-10	9/28 32.1%	146/14305 1.0%	CRP C4A C8G CFH CFI C4BPB CFP C8B B2M
50878	regulation of body fluid levels	5.8397E-12	1.6351E-10	9/28 32.1%	147/14305 1.0%	FGB FGA F10 VWF FGG SAA1 SAA2 C4BPB HRG
2253	activation of immune response	8.5086E-12	2.2423E-10	8/28 28.5%	96/14305 0.6%	CRP C4A C8G CFH CFI C4BPB CFP C8B
6950	response to stress	1.2287E-11	3.0581E-10	19/28 67.8%	1773/14305 12.3%	LGALS3BP CRP FGB FGA CD163 CFH F10 VWF CFI FGG C4BPB CFP C8B C4A C8G SAA1 SAA2 HRG
6958	complement activation, classical pathway	1.4603E-11	3.4431E-10	6/28 21.4%	29/14305 0.2%	CRP C4A C8G CFI C4BPB C8B
16485	protein processing	1.7691E-11	3.9628E-10	8/28 28.5%	105/14305 0.7%	CRP C4A C8G CFH CFI C4BPB CFP C8B
2455	humoral immune response mediated by circulating immunoglobulin	2.2574E-11	4.5969E-10	6/28 21.4%	31/14305 0.2%	CRP C4A C8G CFI C4BPB C8B
30168	platelet activation	2.2574E-11	4.5969E-10	6/28 21.4%	31/14305 0.2%	FGB FGA VWF FGG SAA1 SAA2
6952	defense response	3.5268E-11	6.8697E-10	13/28 46.4%	621/14305 4.3%	LGALS3BP CRP CD163 CFH CFI C4BPB CFP C8B C4A C8G SAA1 SAA2 S100A9
51604	protein maturation	3.7063E-11	6.9185E-10	8/28 28.5%	115/14305 0.8%	CRP C4A C8G CFH CFI C4BPB CFP C8B
2684	positive regulation of immune system process	3.9320E-11	7.0461E-10	10/28 35.7%	265/14305 1.8%	CRP C4A C8G CFH CFI C4BPB CFP C8B B2M THBS4
48584	positive regulation of response to stimulus	4.3937E-11	7.5707E-10	10/28 35.7%	268/14305 1.8%	CRP C4A C8G CFH CFI C4BPB CFP C8B B2M THBS4
6957	complement activation, alternative pathway	5.8377E-11	9.6862E-10	5/28 17.8%	15/14305 0.1%	C4A C8G CFH CFP C8B
42060	wound healing	8.9477E-11	1.4316E-9	9/28 32.1%	199/14305 1.3%	FGB FGA F10 VWF FGG SAA1 SAA2 C4BPB HRG
2252	immune effector process	9.3852E-11	1.4499E-9	8/28 28.5%	129/14305 0.9%	CRP C4A C8G CFH CFI C4BPB CFP C8B
50776	regulation of immune response	3.9364E-10	5.8783E-9	9/28 32.1%	235/14305 1.6%	CRP C4A C8G CFH CFI C4BPB CFP C8B B2M
16064	immunoglobulin mediated immune response	4.7513E-10	6.8664E-9	6/28 21.4%	50/14305 0.3%	CRP C4A C8G CFI C4BPB C8B
19724	B cell mediated immunity	5.3777E-10	7.5287E-9	6/28 21.4%	51/14305 0.3%	CRP C4A C8G CFI C4BPB C8B
2449	lymphocyte mediated immunity	1.9969E-9	2.7110E-8	6/28 21.4%	63/14305 0.4%	CRP C4A C8G CFI C4BPB C8B
2460	adaptive immune response based on somatic recombination	3.8180E-9	4.9054E-8	6/28 21.4%	70/14305 0.4%	CRP C4A C8G CFI C4BPB C8B
2682	regulation of immune system process	3.8324E-9	4.9054E-8	10/28 35.7%	424/14305 2.9%	CRP C4A C8G CFH CFI C4BPB CFP C8B B2M THBS4
2250	adaptive immune response	4.1650E-9	5.1831E-8	6/28 21.4%	71/14305 0.4%	CRP C4A C8G CFI C4BPB C8B
2443	leukocyte mediated immunity	1.0804E-8	1.3081E-7	6/28 21.4%	83/14305 0.5%	CRP C4A C8G CFI C4BPB C8B

Cellular component

GO-ID	Description	p-val	corr p-val	cluster freq	total freq	genes
5576	extracellular region	1.1314E-23	1.1087E-21	28/30 93.3%	2027/16374 12.3%	LGALS3BP CRP COLEC11 PIGR CFH CFI C4BPB CFP C8B THBS4 IGHG3 C4A FCGR3A C8G B2M FGB FGA CD163 F10 VWF FGG
5615	extracellular space	2.4556E-14	1.2032E-12	16/30 53.3%	748/16374 4.5%	LGALS3BP CRP FGB FGA CFH CFI FGG CFP C8B THBS4 C4A SAA1 SAA2 IGFALS HRG B2M
44421	extracellular region part	8.9568E-14	2.9259E-12	17/30 56.6%	986/16374 6.0%	LGALS3BP CRP FGB FGA CFH VWF CFI FGG CFP C8B THBS4 C4A SAA1 SAA2 IGFALS HRG B2M
31091	platelet alpha granule	3.7786E-10	9.2576E-9	6/30 20.0%	51/16374 0.3%	FGB FGA VWF FGG HRG THBS4
31093	platelet alpha granule lumen	4.5426E-9	8.6052E-8	5/30 16.6%	35/16374 0.2%	FGB FGA FGG HRG THBS4
60205	cytoplasmic membrane-bounded vesicle lumen	5.2685E-9	8.6052E-8	5/30 16.6%	36/16374 0.2%	FGB FGA FGG HRG THBS4
31983	vesicle lumen	6.9969E-9	9.7956E-8	5/30 16.6%	38/16374 0.2%	FGB FGA FGG HRG THBS4
5577	fibrinogen complex	1.9329E-7	2.3678E-6	3/30 10.0%	7/16374 0.0%	FGB FGA FGG
30141	stored secretory granule	9.3810E-7	1.0215E-5	6/30 20.0%	186/16374 1.1%	FGB FGA VWF FGG HRG THBS4
9897	external side of plasma membrane	4.5898E-6	4.4980E-5	5/30 16.6%	137/16374 0.8%	FGB FGA F10 VWF FGG
5579	membrane attack complex	6.7761E-5	6.0369E-4	2/30 6.6%	7/16374 0.0%	C8G C8B
44459	plasma membrane part	1.0996E-4	8.9798E-4	12/30 40.0%	1999/16374 12.2%	FGB PIGR FGA CD163 C8G F10 VWF FGG LCP1 C8B HRG B2M
44433	cytoplasmic vesicle part	1.1943E-4	9.0032E-4	5/30 16.6%	270/16374 1.6%	FGB FGA FGG HRG THBS4
16023	cytoplasmic membrane-bounded vesicle	1.3388E-4	9.3717E-4	7/30 23.3%	647/16374 3.9%	FGB PIGR FGA VWF FGG HRG THBS4
31988	membrane-bounded vesicle	1.5875E-4	1.0371E-3	7/30 23.3%	665/16374 4.0%	FGB PIGR FGA VWF FGG HRG THBS4
31410	cytoplasmic vesicle	1.9066E-4	1.1678E-3	7/30 23.3%	685/16374 4.1%	FGB PIGR FGA VWF FGG HRG THBS4
31982	vesicle	2.4604E-4	1.4184E-3	7/30 23.3%	714/16374 4.3%	FGB PIGR FGA VWF FGG HRG THBS4
9986	cell surface	3.4846E-4	1.8972E-3	5/30 16.6%	340/16374 2.0%	FGB FGA F10 VWF FGG
34364	high-density lipoprotein particle	9.4836E-4	4.8916E-3	2/30 6.6%	25/16374 0.1%	SAA1 SAA2
5886	plasma membrane	1.0038E-3	4.9189E-3	15/30 50.0%	3732/16374 22.7%	FGB PIGR FGA CD163 F10 VWF FGG C8B FCGR3A C8G IGHD LCP1 HRG B2M S100A9
34358	plasma lipoprotein particle	1.8596E-3	8.2839E-3	2/30 6.6%	35/16374 0.2%	SAA1 SAA2
32994	protein-lipid complex	1.8596E-3	8.2839E-3	2/30 6.6%	35/16374 0.2%	SAA1 SAA2
33093	Weibel-Palade body	3.6611E-3	1.5599E-2	1/30 3.3%	2/16374 0.0%	VWF
1891	phagocytic cup	5.4868E-3	2.2404E-2	1/30 3.3%	3/16374 0.0%	LCP1
46930	pore complex	9.3848E-3	3.6788E-2	2/30 6.6%	80/16374 0.4%	C8G C8B

Molecular function

GO-ID	Description	p-val	corr p-val	cluster freq	total freq	genes
43498	cell surface binding	8.3656E-7	8.3656E-5	4/28 14.2%	41/15440 0.2%	CRP FGB FGA FGG
43499	eukaryotic cell surface binding	2.3951E-6	1.1975E-4	3/28 10.7%	15/15440 0.0%	FGB FGA FGG
3823	antigen binding	3.6763E-6	1.2254E-4	4/28 14.2%	59/15440 0.3%	IGHG3 IGLC7 IGKV1-5 IGHD
5044	scavenger receptor activity	7.2022E-5	1.8005E-3	3/28 10.7%	45/15440 0.2%	LGALS3BP CD163 CFI
5102	receptor binding	1.6801E-4	3.3601E-3	8/28 28.5%	923/15440 5.9%	FGB FGA VWF FGG SAA1 SAA2 HRG THBS4
19865	immunoglobulin binding	4.2411E-4	7.0685E-3	2/28 7.1%	17/15440 0.1%	FCGR3A VWF
30674	protein binding, bridging	6.4109E-4	9.1584E-3	3/28 10.7%	94/15440 0.6%	FGB FGA FGG
32403	protein complex binding	8.4260E-4	1.0532E-2	4/28 14.2%	238/15440 1.5%	FCGR3A VWF C8B THBS4
51087	chaperone binding	1.3369E-3	1.4855E-2	2/28 7.1%	30/15440 0.1%	FGB VWF
4801	transaldolase activity	1.8135E-3	1.8135E-2	1/28 3.5%	1/15440 0.0%	TALDO1
48029	monosaccharide binding	2.7351E-3	2.4865E-2	2/28 7.1%	43/15440 0.2%	COLEC11 TALDO1
30246	carbohydrate binding	3.9761E-3	3.3134E-2	4/28 14.2%	364/15440 2.3%	COLEC11 TALDO1 HRG THBS4
33265	choline binding	5.4309E-3	4.0194E-2	1/28 3.5%	3/15440 0.0%	CRP
5178	integrin binding	5.7847E-3	4.0194E-2	2/28 7.1%	63/15440 0.4%	VWF THBS4
30297	transmembrane receptor protein tyrosine kinase activator activity	7.2349E-3	4.0194E-2	1/28 3.5%	4/15440 0.0%	HRG
16744	transferase activity, transferring aldehyde or ketonic groups	7.2349E-3	4.0194E-2	1/28 3.5%	4/15440 0.0%	TALDO1
43125	ErbB-3 class receptor binding	7.2349E-3	4.0194E-2	1/28 3.5%	4/15440 0.0%	HRG
51637	Gram-positive bacterial cell surface binding	7.2349E-3	4.0194E-2	1/28 3.5%	4/15440 0.0%	CRP

Supplementary methods

Materials

Ammonium bicarbonate, sodium deoxycholate, iodoacetamide, and dithiothreitol were purchased from Sigma (St. Louis, MO, USA). Tris-(2-carboxyethyl) phosphine was acquired from Thermo Scientific (Rockford, Il, USA). Modified sequencing-grade trypsin was obtained from Promega (Madison, WI, USA). All mobile phases and solutions were prepared with HPLC grade solvents (i.e. water, acetonitrile, methanol, and formic acid) from Sigma Aldrich. All other reagents were purchased from commercial suppliers with standard biochemical quality.

Patients

For biomarker discovery phase. From 2008 to 2015, patients from Peking University Third Hospital, who fulfilled the 1997 SLE classification criteria revised by ACR, were recruited in our study. This protocol was approved by the Ethics Committee at Peking University Third Hospital and informed consent was obtained from each patient. SLEDAI2000 was used to measure the disease activity for all the patients. These enrolled SLE patients were divided into two groups according to the SLEDAI2000 score: active group (SLEDAI2000 > 9) and inactive group (SLEDAI2000 ≤4). The patients in the two groups were well matched in age and gender (Table S1).

For biomarker validation for thrombospondin-4. From 2008 to 2015, the same protocols used in biomarker discovery phase were used to recruit patients and assess the disease activity. The SLE patients group of active and inactive stage were established, and a healthy control group, which was consisted of healthy volunteers, was also involved. People in all the three groups were well matched in age and gender (Table S2).

For antibodies analysis against thrombospondin-4, collectin-11, and properdin. From 2008 to 2015, another cohort of SLE patients (fulfilled the 1997 revised classification criteria for SLE) were enrolled and informed consent was obtained from each patient. These enrolled SLE patients were divided into two groups according to the SLEDAI2000 score: active group (SLEDAI2000 > 9) and inactive group (SLEDAI2000 ≤4). Three control groups were also established: one was RA (Rheumatoid arthritis, RA) group, which was consisted of patients with RA (fulfilled the 1987 classification criteria for RA); one was SS (Primary sjogren syndrome, SS) group, which was consisted with SS (fulfilled the 2002 classification criteria for SS); the other was healthy control group, which was consisted of healthy volunteers. People in all the five groups were well matched in age and gender (Table S3).