Proteomic Insights into C3dg as Biomarker of Systemic Lupus Erythematosus [†]

Anne Margrethe Trolborg [2,‡]; Thomas Bouet Guldbæk Poulsen [1]; Jacob Skallerup Andersen [1]; Mikkel Eggert Thomsen [1], Christopher Aboo [1] and Allan Stensballe [1,‡]

[1] Department of Health Science and Technology; Aalborg University, Denmark; tbgp@hst.aau.dk (T.B.G.P.), mht@et.aau.dk (M.E.T.), as@hst.aau.dk (A.S.)

[2] Advanced center for Human Genetics, SK Institute of Medical Sciences (SKIMS), Srinagar-19001; amt@biomed.au.dk

[3] Immunology and Molecular Medicine, SK Institute of Medical Sciences (SKIMS), Srinagar-190011

[‡] Authors with equal contribution

***Correspondence: as@hst.aau.dk; Tel.: +45-61608786

[†] Presented at the 1st International Electronic Conference on Biomedicine, 1–26 March 2021; Available online: https://ecb2021.sciforum.net.

Abstract: Systemic Lupus Erythematosus (SLE) is an autoimmune disease which presents clinical manifestations in different organs and presenting autoantibodies targeting its own body. The pathogenesis of SLE is not yet fully understood. However, there is no proper biomarker to diagnose SLE and to measure disease activity. Plasma samples from the four different SLE patient groups (low vs. high SLEDAI; low vs. high C3gd) were selected based on clinical scores from the SLE patients (*n* = 40). Plasma samples were analyzed by quantitative tandem mass spectrometry. Autoantigen profiles were determined by 1536plex Immunome array. Quantification of cfDNA as marker for neutrophil degranulation was determined by dsDNA Assay.

Keywords: SLE, biomarkers, cytokines, proteomics, cell free DNA, immunome array

1. Introduction

Systemic lupus erythematosus (SLE) is a systemic chronic autoimmune multisystem disease, which has a variable presentation, unpredictable disease course and prognosis [1,2]. The clinical presentation of SLE is very heterogeneous and multi-system complications include neurological, cutaneous, mucosal, lung, heart kidney, gastrointestinal, vascular, haematological, and musculoskeletal [3]. Mortality of patients with SLE have improved during the past 50 years, there have been a slow down in improvements in the 1980s and 1990s [4], and only one recent approvement of new drug in recent years [5]. Measuring disease activity of SLE patients is difficult, and few serological markers are measured, but does not necessary reflect the clinical state of the patient [3]

The pathogenesis of SLE has yet to be fully established, but defective clearance of apoptotic cells is a key driver of serological findings. Clearance deficiency is recognized initially by a rise of danger associated molecular patterns (DAMPS), e.g., heat shock proteins, toll-like receptor ligands, but also extracellular DNA, metabolites (e.g., monosodium urate crystals) and cytokines (IL-1β, IL-18, and IL-6) [6]. The continued deficiency leads to formation of a secondary necrotic core of uncleared cell debris, which leads to immune complex formation, exposed autoantigens with associated rise in autoantigens (ANA, ANCA, anti-dsDNA), and consequently Th1 and inflammasome driven inflammation [7,8]. In addition, the complement is implicated in fixing immune complexes [9], and complement to be involved in the pathogenesis of SLE patients [10,11].

Based on the molecular patterns of SLE it is conceivable, that thorough patient or disease subtype specific investigation of extracellular DNA, autoantibodies and complement system processing could lead to biomarkers that can distinguish and characterize the heterogeneity between patients, which was the aim of this study. We have previously shown that polymyalgia patients express more extracellular DNA that rheumatoid arthritis (RA) patients [12], that patients with RA have very distinct autoantigen profiles [13], and that patients can be grouped based on their serum proteome [12,14,15].

We have previously proposed that proteins involving clearance of dying cells could be implicated in the pathogenesis of SLE [2], and indeed the results in this study highlight this. Recent understanding of SLE has led to improved management recommendations of SLE [1], but a personalized approach similar to other rheumatic diseases [16] is still to come, and more biomarker studies are needed to characterize this disease [17].

There is no proper clinical tool available to precisely identify the SLE and measure the disease activity at the current time, hence the interest in finding SLE biomarkers for such purpose. Biomarkers are genetic, biological, biochemical, molecular or genetic alteration which can give indication on the disease activity or any abnormality and have measured and evaluated qualitative and quantitatively at the laboratory or the clinical scenario [6,18]. Biomarkers can be almost everything, e.g., autoantibodies, complement factors, and cytokines. Current biomarkers include the immune system and inflammatory components including acute phase autoantibodies, chemokines and complement proteins. The majorly studied antibodies are ANA, ANuA, anti-dsDNA antibodies. Chemokines like CXCL proteins, which are found on the surface of lymphocytes, are also being studied. With the complement factors, C1q, C3 and C4 have shown promising results. Cytokines such as IL-6 and IL-10 have been regarded as the potential SLE biomarkers. However, all of them have not been approved yet and the reliability, sensitivity and specificity of their clinical use is still in discussion [19,20]. Recently, the use of autoantigen reactive protein arrays have applied for personalized diagnostics of autoimmune diseases including RA and SLE [13,21]. When treating and monitoring the progression in patients then companion diagnostics (CDx) have been developed as a biomedical strategy to provide essential information regarding the effective use of drugs and biological products often on the individual patient. CDx helps the physicians to decide if the benefits from a therapeutic product outweigh possible side effects [22]. A recent study evaluate C3 fragments for improved patient diagnostics [23]. In this study we further investigate the implications of underlying proteomic signatures of clinical classifications associated with patient sub-typing for further disease insight and improved diagnostics for personalized treatment.

2. Material and methods

2.1. Study subjects

The cross-sectional cohort study has obtained approval from the ethical committee (Case number: 1-10-72-214-13) and informed consent has been received from the participating patients. The inclusion criteria include that patients are over 18 years old and fulfilled at least four out of 11 classification criteria of the ACR's revised version of 1992. Patients with active treatment for acute or chronic infection or active treatment for the cancer were excluded.

Blood samples from SLE patients were taken by the medical laboratory technicians working in Aarhus University Hospital. The blood samples were taken during the routine blood test. The samples were then collected and processed within an hour of collection by the same technician, who collected the samples. The serum, EDTA-plasma and blood for purification of blood cells were taken. Within 1 h, the EDTA-tubes were then centrifuged at 2000g for 10 min, followed by immediate freezing of the plasma at 80°C (aliquots of 0.5 mL). The plasma samples were allocated in different SLE groups partly based on the SLEDAI scores and partly on the measurements of C3dg levels. These samples were randomly selected from the cohort study after the allocation in the different groups.

Serum samples were obtained from 40 SLE patients grouped into four distinct groups: SLEDAI (high/low), and C3dg (high/low). Based on previous experiments we simulated data to determine the statistical power versus the limit to detect a certain fold change.

2.2. Circulating cfDNA levels in serum

cfDNA was measured for all serum samples in duplicates with the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies) according to the manufacturers and kit supplied reference DNA as previous done (M. K. Meyer, M. Andersen, T. V. Stausbo, K. J. Elbæk, G. N. Andersen). Group wise non-parametric (Mann-Whitney) tests were applied to determine statistically significant differences between groups.

2.3. Proteomics – Mass Spectrometry Analysis

A modified FASP protein digestion for plasma with trypsin was performed, with phase inversion surfactant removal essentially according to Nguyen et al. [23]. For each biological replicate sample a total of 100 µg protein was transferred to individual YM-10kDa spinfilters (Millipore, Billerica, MA, USA) and buffer exchanged to 5%SDC in 50mM triethylammonium bicarbonate (TEAB) by centrifugation. All centrifugation steps were performed at 14,000 g for 15 min at 4°C. The proteins were then subjected to alkylation with 12 mM tris(2-carboxyethyl)phosphine (Thermo Scientific, Waltham, MA, USA) for 30 min at 37°C, and reduction with 50 mM chloroacetamide (Sigma-Aldrich, St. Louis, MO, USA) for 20 min at 37°C in the dark. The reducing and alkylating agents were dissolved in 120 mM SDC in 50 mM TEAB, pH 8.5, and centrifuged after each step. In preparation for digestion, 100 µL digestion buffer (0.5% in 50 mM TEAB) was added to the spinfilter and centrifuged. A 1:50 (w/w) trypsin:protein ratio dissolved in 50 µL digestion buffer was added to the spinfilter, and the samples were digested overnight at 37°C. The flow-through containing the tryptic peptides were recovered by centrifugation followed by a phase separation performed with 3:1 (v/v) ethyl acetate:sample, acidified by addition of formic acid (FA) to a final concentration of 0.5%. Total phase separation was achieved by 1 min vortexing followed by centrifugation. The aqueous phase was collected and vacuum centrifuged overnight and stored at -80°C until time of analysis.

2.4. Proteomics – Mass Spectrometry Analysis.

The loaded samples amounts were normalized using A280 on a NanoDrop 1000 (Thermo Scientific, Wilmington, DE, USA), and 5 µg total peptide material was analyzed per UPLC-MS analysis.

The samples were analyzed using a UPLC-nanoESI HCD MS/MS setup with an RSLC nanopump module. The system was coupled online with an emitter for nanospray ionization (New objective picotip 360-20-10) to a QExactive mass spectrometer Plus (Thermo Scientific, Waltham, USA). The peptide material was loaded onto a 2 cm trapping reversed phase Acclaim PepMap RSLC C18 column (Dionex) and separated using an analytical 50 cm reversed phase Acclaim PepMap RSLC C18 column (Dionex). Both columns were kept at 40°C. The sample was eluted with a gradient of 96% solvent A (0.1% FA) and 4% solvent B (0.1% FA in ACN), which was increased to 8% solvent B on a 5 min ramp gradient and subsequently to 30% solvent B in 35 min ramp gradient, at a constant flow rate of 300 nL/min. The mass spectrometer was operated in positive mode (m/z 375-1400), selecting up to 12 precursor ions with a mass window of m/z 1.6 based on highest intensity for HCD fragmenting, at a normalized collision energy of 27. Selected precursors were dynamically excluded for fragmentation for 30 sec.

3. Results:

To examine the clinical data multiple clinical assessment parameters in the diagnosis of the four SLE subgroups (High activity, Low activity, High C3dg, and Low C3dg) were evauated statistically. An experimental strategy for discovery based proteome analysis of plasma samples by mass

spectrometry was performed. The scope of this proteomics based analysis is to determine the immunological and pathophysiological reason for the four distinct patient groups.
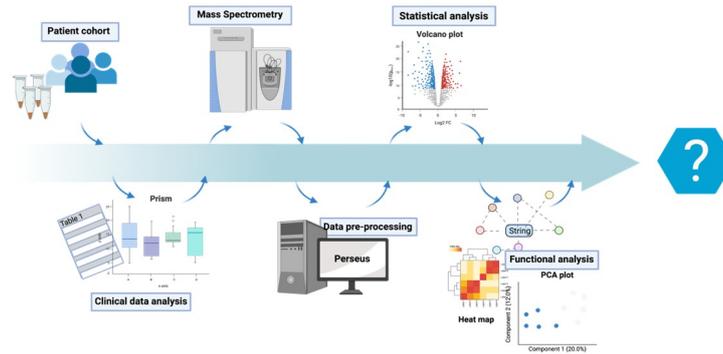

Figure 1. Flowchart visualizing the data processing of the patient cohort including the clinical information and proteomics data. Created with BioRender.com.

3.1. Overview of Clinical Characteristics of the Patient Cohort

The clinical data were analyzed with the purpose of generating an overview of the clinical information of both disease groups and their subgroups as well as the total setup. A summary table of clinical characteristics was created and presented as an amount (*n*)-(%) if the parameter was a "yes/no" question. However, if the parameter was a measurable variable it was presented with a mean and interquartile range (IQR). To illustrate the distribution of the clinical information within the four subgroups and across the disease groups (Table 1). The data generated from the graphs were then analyzed to find statistically significant differences within the four SLE subgroups.

Table 1. Overview of the clinical information within the four SLE subgroups based on the ACR criteria. The data are presented as numbers (percentage (%)) or medians (interquartile range (IQR)).

Table 1

Variables	Disease activity			C3dg levels			Total
	Total n=21	High n=11	Low n=10	Total n=19	High n=9	Low n=10	n=40
Sex - n (%)							
Male	5 (23.8%)	2 (18.8%)	3 (30.0%)	4 (21.0%)	1 (11.1%)	3 (30.0%)	9 (22.5%)
Female	16 (76.2%)	9 (81.2%)	7 (70.0%)	15 (79,0%)	8 (88.9%)	7 (70.0%)	31 (77.5%)
Age at diagnosis							
Median (IQR)	37.0 (27.0-43.0)	27.0 (20.5-41.5)	43.0 (35.5-54.5)	30.0 (22.0-41.0)	26.0 (22.0-49.0)	31.0 (24.8-33.0)	32.0 (22.8-44.5)
Years of disease							
Median (IQR)	6.0 (3.0-12.0)	8.0 (4.0-11.5)	5.0 (3.3-10.5)	10.0 (4.5-17.5)	15.0 (7.0-25.0)	6.5 (4.0-13.8)	8.0 (4.0-15.0)
SLICC							
Median (IQR)	0.0 (0.0-10.0)	0.0 (0.0-1.0)	0.5 (0.0-1.8)	0.0 (0.0-1.5)	1.0 (1.0-3.0)	0.0 (0.0-0.0)	0.0 (0.0-1.3)
ACR - n (%)							
ACR 1	17 (81.0%)	11 (100.0%)	6 (60.0%)	13 (68.4%)	4 (44.4%)	9 (90.0%)	30 (75.0%)
ACR 2	2 (9.5%)	0 (0.0%)	2 (20.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (5.0%)
ACR 3	12 (57.1%)	7 (63.6%)	5 (50.0%)	13 (68.4%)	6 (66.7%)	7 (70.0%)	25 (62.5%)
ACR 4	9 (42.9%)	7 (63.6%)	2 (20.0%)	7 (36.8%)	2 (22.2%)	5 (50.0%)	16 (40.0%)
ACR 5	17 (81.1%)	9 (81.8%)	8 (80.0%)	19 (100.0%)	9 (100.0%)	10 (100.0%)	36 (90.0%)
ACR 6	6(28.6%)	2 (18.2%)	4 (40.0%)	7 (36.8%)	4 (44.4%)	3 (30.0%)	13 (32.5%)
ACR 7	7 (33.3%)	5 (45.5%)	2 (20.0%)	5 (26.3%)	2 (22.2%)	3 (30.0%)	12 (30.0%)
ACR 8	1 (4.8%)	0 (0.0%)	1 (10.0%)	5 (26.3%)	4 (44.4%)	1 (10.0%)	6 (15.5%)
ACR 9	20 (95.2%)	10 (90.9%)	10 (100.0%)	15 (78.9%)	5 (55.5%)	10 (100.0%)	35 (87.5%)
ACR 10	19 (90.5%)	10 (90.9%)	9 (90.0%)	17 (89.5%)	7 (77.7%)	10 (100.0%)	36 (90.0%)
ACR 11	21 (100.0%)	11 (100.0%)	10 (100.0%)	17 (89.5%)	7 (77.7%)	10 (100.0%)	38 (95.0%)
Total of ACR criteria at diagnosis - n (%)							
4	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (10.5%)	2 (22.2%)	0 (0.0%)	2 (5.0%)
5	6 (28.6%)	2 (18.2%)	4 (40.0%)	2 (10.5%)	2 (22.2%)	0 (0.0%)	8 (20.0%)
6	7 (33.3%)	3 (27.3%)	4 (40.0%)	8 (42.1%)	4 (44.4%)	4 (40.0%)	15 (37.5%)
7	5 (23.8%)	4 (36.4%)	1 (10.0%)	4 (21.1%)	0 (0.0%)	4 (40.0%)	9 (22.5%)
8	3 (14.3%)	2 (18.2%)	1 (10.0%)	3 (15.8%)	1 (11.1%)	2 (20.0%)	6 (15.0%)
SLEDAI							
Median (IQR)	2.0 (0.0-10.0)	10.0 (10.0-11.5)	0.0 (0.0-0.0)	4.0 (0.5-5.5)	2.0 (1.7-2.0)	4.9 (4.0-7.0)	3.0 (0.0-9.3)
Biomarker - Median (IQR)							
Anti-ds-DNA-ab *10^3 IU/L	4.0 (1.0-22.0)	22.0 (1.0-59.5)	1.5 (1.0-7.8)	12.0 (1.5-21.1)	2.0 (0.0-13.0)	19.5 (5.8-156.5)	8.5 (1.0-22.5)
C3 level g/L	1.1 (0.7-1.1)	0.8 (0.6-1.2)	1.1 (1.0-1.1)	1.0 (0.8-1.1)	1.1 (1.0-1.3)	0.8 (0.7-1.0)	1.0 (0.8-1.1)
C4 level g/L	0.2 (0.1-0.2)	0.1 (0.1-0.2)	0.2 (0.2-0.2)	0.2 (0.1-0.2)	0.2 (0.2-0.2)	0.1 (0.0-1.0)	0.2 (0.1-0.2)
C1q µmol/L	0.5 (0.4-0.7)	0.5 (0.4-0.7)	0.5 (0.4-0.8)	0.4 (0.3-0.5)	0.4 (0.4-0.5)	0.3 (0.1-0.5)	0.5 (0.3-0.6)
CRP	3.1 (1.2-5.8)	3.2 (1.6-6.7)	2.0 (1.3-4.2)	2.7 (2.2-8.2)	7.6 (2.2-8.2)	2.4 (1.0-4.9)	2.9 (1.1-7.6)
ANA at inclusion - n (%)							
Yes	18 (85.7%)	11 (100.0%)	7 (70.0%)	17 (89.5%)	7 (77.8%)	10 (100.0%)	36 (90.0%)
APLS - n (%)							
Yes	3 (14.3%)	2 (18.2%)	1 (10.0%)	3 (15.8%)	1 (11.1%)	2 (20.0%)	6 (15.0%)
Proteinurea at inclusion - n (%)							
Yes	2 (20.0%)	3 (27.3%)	5 (23.8%)	5 (26.3%)	2 (10.5%)	3 (30.0%)	10 (25.0%)

As it appears from Table 1, the patients within the High activity subgroup were diagnosed with SLE at an earlier age (27 years (20.5–41.5)) compared to the Low activity subgroup (43 years (35.5–54.5)). The biomarker anti-ds-DNA is more elevated in the High activity subgroup with a median value at 22.0*10^3 IU/L compared to Low activity (1.5*10^3 IU/L). The opposite tendency appears from the C3dg disease group, where Low C3dg has a higher amount of anti-ds-DNA (19.5*10^3 IU/L) compared to High C3dg (2.0*10^3 IU/L). For the complement factors, C3, C4, and C1q the concentrations in all four SLE subgroups are within their respective reference values. However, for the hemogram the lymphocyte count was below the reference values whereas High activity had the lowest count (0.8*10^9/L) compared to Low activity (1.1*10^9/L). For the C3dg patient group, Low C3dg had a slightly lower concentration (1.0*10^9/L) than High C3dg (1.1*10^9/L). No significant differences were observed between the subgroups in regard to the lymphocyte count.

3.2. Clinical Manifestations, Medication and Disease Activity Assessment

To further examine the patient cohort, certain parameters from the clinical dataset are presented in the following sections and are used to investigate if High and Low C3dg can be correlated to the biological presentation of SLE patients.

By examining figure 2A, it is worth noting that there is a significant difference between males and females (p < 0.001), where females account for an overall of 77,5%. Additionally, the sex distribution

within the four subgroups is visualized in figure 2B, where the median indicates that the duration of the disease seems to be higher in High activity compared to Low activity, as well as High C3dg compared to Low C3dg. However, none of these differences, both in females and males, was found to be significant. The number of patients and the variation between the four SLE subgroups within each ACR criterion is illustrated in Figure 2C. It is generally seen that the number of patients within each ACR criteria differs among the SLE subgroups. For ACR 1 there is a significantly higher number of patients in High activity compared to Low activity (p = 0.035). Moreover, ACR 9 is significantly higher in Low C3dg compared to High C3dg (p = 0.033). By inspecting the graph, it can be seen that the ACR criteria across, regardless of the subgroups, differ in frequency. An example is the ACR criteria 6 and 8, which are less common compared to criteria 10 and 11 in this patient cohort. When evaluating the total number of ACR criteria fulfilled at diagnosis in figure 2D, there is no significant difference in any of the SLE subgroups. Aside from the clinical manifestations, the SLEDAI score within the four SLE subgroups is visualized in figure 2E and indicates the disease activity of the patients.

It is observed in figure 2E, that the SLEDAI score within the clinical dataset varies from 0 to 14 and shows a significantly higher score in the subgroup High activity compared to Low activity (p < 0.001). Moreover, a significantly higher score is seen in Low C3dg compared to High C3dg (p= 0.028). SLE has several liquid biomarkers used as diagnostic criteria, where C3, C4, and ANA are the most approved. ANAs vary over time, depending on the disease activity, where C3 and C4 are continuously low in SLE patients. These factors are illustrated with the distribution in the four subgroups below.

In the SLE patient cohort, nine different treatment options were available. The distribution of the different treatments supplied to the patients in the four groups is visualized in figure 3F. The treatment within the four subgroups differs, however, no significant difference between High activity and Low activity as well as between High C3dg and Low C3dg is found in any of the treatment options. Still, it is seen that most patients receive the treatment chloroquine and glucocorticosteroids. As visualized in figure 2G, the amount of the complement factor C3 is generally higher than C4 in all of the subgroups. For both of the components, C3 and C4 respectively, it applies that they are significantly higher in High C3dg compared to Low C3dg (p = 0.002; p = 0.007). However, no significant difference was observed in either C3 or C4, between High activity and Low activity. When inspecting C1q, it is seen that neither High/Low activity nor High/Low C3dg is significant. Another biomarker seen in the four SLE subgroups is the Anti-ds-DNA-ab, which is visualized in figure 2H. A significant difference between the High C3dg and Low C3dg (p = 0.035) is observed, however, no significant difference is observed between High activity and Low activity.
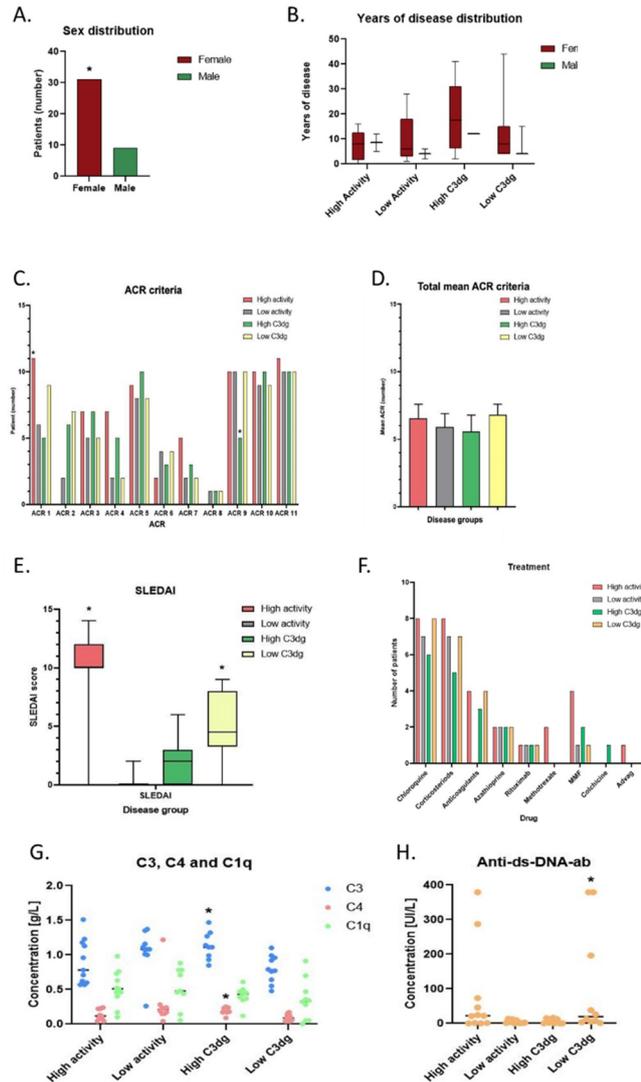
Figure 3. Assessment of the clinical data from the cohort of SLE patients * indicates a p < .05. (A.) The distribution of sex within each SLE subgroup and the disease duration. A) The distribution of females and males in the patient cohort. B) The distribution of the disease duration divided into females and males for each SLE subgroup. (C.) Clinical manifestations of the SLE subgroups according to ACR criteria fulfilled at diagnosis. (D.) The variation within the individual ACR criteria. B) The total number of ACR criteria within each subgroup, presented as mean; (E.) distribution of the SLEDAI scores within the four SLE subgroups. (F.) Nine different treatment options supplied to the four SLE subgroups in the patient cohort.

3.3. Analysis MS-based Proteomics of SLE Patient Plasma

A biological insight into the differences observed in the clinical data, a statistical and functional analysis of the obtained proteomic data comparing the proteomic profiles of each disease groups; High/Low activity and the High/Low C3dg, respectively. The distribution of the identified statistically significant proteins between the activity patient subgroup were compared statistically using t-test and visualized in figure 4A .

It is observed that the majority of the proteins identified are upregulated in High activity compared to Low activity. In total 31 proteins in the Activity disease group fulfilled the statistical. Based on gene ontology (GO) annotations, 21 proteins were found to have a possible relation to the immune system, which is illustrated in table 2. A total of 16 proteins showed to have a FC (Log2) value above 0, and are thereby upregulated in the High activity group, compared to the Low activity group. The

highest fold change is observed in the immunoglobulin kappa variable 6–21 with a 3.85 fold increase in the High activity subgroup (p = .023). Conversely, five proteins are observed to be downregulated. Among these, the most downregulated protein in the High activity group compared to Low activity is Ig delta chain C region with a 0.26 fold decrease (p =.044). As well as for the Activity patient group, a scatterplot of the significant proteins identified between the High and Low C3dg subgroups was performed and visualized in figure 3B.
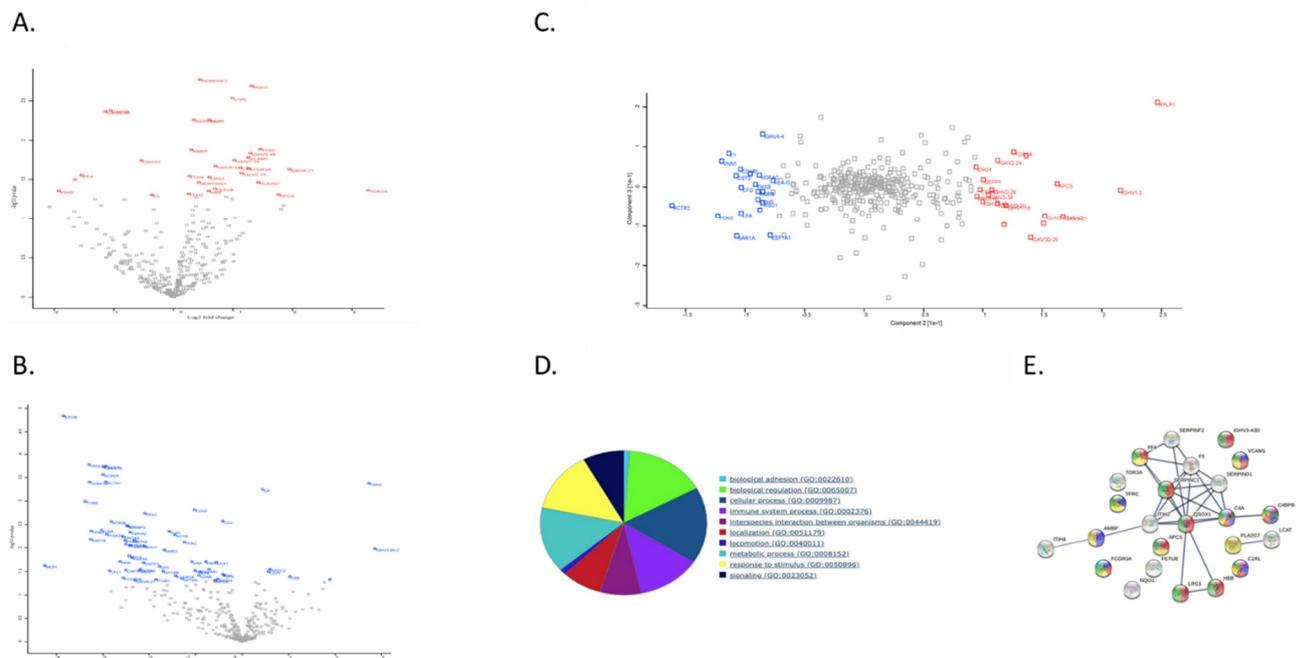

Figure 4. Proteomic investigation of SLE subtypes. (A) statistically significant proteins, that are differentially expressed in High activity compared to Low activity. The significant proteins are marked with red with their respective gene tags (p < 0.05). (B.) The statistically significant proteins that are differentially expressed in High C3dg compared to Low C3dg. (C) PCA plot of the proteins responsible for the separation between High and Low activity. Proteins highlighted in blue are responsible for the Low activity cluster, whereas the proteins highlighted in red are responsible for the High activity cluster.

In this patient group, it is observed that the majority of the identified proteins that were differentially expressed are downregulated in High C3dg compared to Low C3dg. In total 63 proteins were identified as differentially regulated however, 32 of these are included in table 3, based on selected GO terms relevant to the immune system. In High C3dg, a total of 4 proteins were upregulated compared to Low C3dg. Among these proteins, the immunoglobulin heavy variable 1–69D exhibited the highest fold change with a 7.22 increase (p = 0.010). As mentioned, most of the differentially expressed proteins in C3dg were downregulated and involved a total of 28 proteins. Among these proteins, the tubulin beta chain had the most considerable downregulation with a 0.10 fold change (p = .001).

The identified proteins from the statistical analysis were further analyzed to investigate the biological function the proteins were associated with. A pie plot was made to illustrate the distribution of the biological functions of the proteins, as seen in figure 3D . It is visualized in the pie chart that the disease group High/Low activity has a lot of different biological functions associated with the identified proteins. The majority of the proteins are related to the following functions: biological regulation, cellular process, metabolic process, response to stimulus, and immune system process.

To explore the protein-protein interactions from the identified proteins a STRING analysis was performed. In the analysis, the KEGG pathway and some GO terms showed a relation to the immune system, as seen in figure 3E. By examining the STRING network, it is possible to observe a large network consisting of 13 protein-protein interactions with high confidence. Among these proteins, Complement C4-A (C4A) is found to be involved in several GO-terms associations and linked to multiple proteins within this network. Furthermore, C4A and Low-affinity immunoglobulin gamma Fc region receptor III-A (FCGR3A) are shown to be associated with SLE after imputation of the KEGG pathways.

To analyze the statistically significant proteins and the resemblance between the patient subgroups, different unsupervised PCA plots were made to illustrate clusters based on their similarities. Moreover, a STRING analysis and pie chart for both of the patients groups were performed to examine if the biological functions of the significant proteins were connected. By examining the STRING network, it is possible to observe a large network consisting of 13 protein-protein interactions with high confidence. Among these proteins, Complement C4-A (C4A) is found to be involved in several GO-terms associations and linked to multiple proteins within this network. Furthermore, C4A and Low-affinity immunoglobulin gamma Fc region receptor III-A (FCGR3A) are shown to be associated with SLE after imputation of the KEGG pathways.

Table 2. Relevant proteins for the pathogenesis of SLE identified in plasma samples from the Activity disease group. The proteins are presented with their gene name, protein name, Uniprot accession number, and p-value. The FC (Log2) value corresponds to the student's T-test difference value.

Table 2

Gene	Protein Name	Uniprot Accession	p-value	FC (Log2)	FC
IGKV6-21	Immunoglobulin kappa variable 6-21	A0A0C4DH24	.023	1.95	3.85
APCS	Serum amyloid P-component	P02743	.049	1.76	3.40
TFRC	Transferrin receptor protein 1	P02786	.013	1.47	2.76
PLA2G7	Platelet-activating factor acetylhydrolase	Q13093	.034	1.43	2.69
IGHV3-49	Immunoglobulin heavy variable 3-49	A0A0A0MS15	.015	1.30	2.46
FCGR3A	Low affinity immunoglobulin gamma Fc region receptor III-A	P08637	.022	1.26	2.40
VCAM1	Vascular cell adhesion protein 1	P19320	.017	1.26	2.40
IGLV2-18	Immunoglobulin lambda variable 2-18	A0A075B6J9	.026	1.14	2.21
C4A	Complement C4-A	P0C0L4	.022	1.13	2.18
IGHV1-24	Immunoglobulin heavy variable 1-24	A0A0C4DH33	.018	1.02	2.03
C1RL	Complement C1r subcomponent-like protein	Q9NZP8	.003	0.98	1.98
IGHV5-51	Immunoglobulin heavy variable 5-51	A0A0C4DH38	.021	0.71	1.63
LRG1	Leucine-rich alpha-2-glycoprotein	P02750	.029	0.61	1.53
HBB	Hemoglobin subunit beta	P68871	.044	0.59	1.50
SERPINC1	Antithrombin-III	P01008	.005	0.35	1.28
AMBP	Protein AMBP	P02760	.013	0.30	1.23
QSOX1	Sulfhydryl oxidase 1	O00391	.018	-0.55	0.69
C4BPB	C4b-binding protein beta chain	P20851	.004	-1.06	0.48
PF4	Platelet factor 4	P02776	.028	-1.54	0.34
IGHV3-43D	Immunoglobulin heavy variable 3-43D	P0DP04	.032	-1.65	0.32
IGHD	Ig delta chain C region	P01880	.044	-1.92	0.26

Table 3: Relevant proteins for the pathogenesis of SLE identified in plasma samples from the C3dg patient group. The proteins are presented with their gene name, protein name, Uniprot accession number, and p-value. The FC

(Log2) value corresponds to the student's T-test difference value.

Table 3

Gene	Protein Name	Uniprot Accession	p-value	FC (Log2)	FC
IGHV1-69D	Immunoglobulin heavy variable 1-69D	A0A0B4J2H0	.010	2.85	7.22
IGLV1-44	Ig lambda chain V-I region VOR	P01699	.046	1.89	3.71
C4B	Complement C4-B	P0C0L5	.041	1.02	2.03
LCP1	Plastin-2	P13796	.032	0.58	1.49
VTN	Vitronectin	P04004	.038	-0.38	0.77
CLU	Clusterin	P10909	.003	-0.41	0.75
FN1	Fibronectin	P02751	.036	-0.43	0.74
IGHG3	Ig gamma-3 chain C region	P01860	.046	-0.53	0.69
IGKV1-17	Ig kappa chain V-I region Gal	P01599	.028	-0.72	0.61
PF4	Platelet factor 4	P02776	.033	-0.91	0.53
FCN3	Ficolin-3	O75636	.002	-1.02	0.49
TFRC	Transferrin receptor protein 1	P02786	.007	-1.24	0.42
GP1BA	Platelet glycoprotein Ib alpha chain	P07359	.045	-1.29	0.41
FCGR3A	Low affinity immunoglobulin gamma Fc region receptor III-A	P08637	.039	-1.44	0.37
ACTB	Actin, cytoplasmic 1	P60709	.005	-1.45	0.37
NME2	Nucleoside diphosphate kinase B	P22392	.011	-1.68	0.31
ITGB3	Integrin beta-3	P05106	.047	-1.85	0.28
GSTP1	Glutathione S-transferase P	P09211	.021	-2.12	0.23
ACTN1	Alpha-actinin-1	P12814	.028	-2.20	0.22
IGKV6-21	Immunoglobulin kappa variable 6-21	A0A0C4DH24	.049	-2.31	0.20
MYH9	Myosin-9	P35579	.007	-2.32	0.20
MYL9	Myosin regulatory light polypeptide 9	P24844	.015	-2.44	0.18
HSPA8	Heat shock cognate 71 kDa protein	P11142	.016	-2.36	0.19
VCL	Vinculin	P18206	.003	-2.48	0.18
CAP1	Adenylyl cyclase-associated protein 1	Q01518	.029	-2.50	0.18
PPBP	Platelet basic protein	P02775	.007	-2.52	0.17
PRDX6	Peroxiredoxin-6	P30041	.006	-2.56	0.17
FCGR3B	Low affinity immunoglobulin gamma Fc region receptor III-B	O75015	.044	-2.57	0.17
PKM	Pyruvate kinase PKM	P14618	.019	-2.64	0.16
ITGA2B	Integrin alpha-IIb	P08514	.005	-2.92	0.13
ACTG1	Actin, cytoplasmic 2	P63261	< .001	-2.93	0.13
ALDOA	Fructose-bisphosphate aldolase A	P04075	< .001	-2.99	0.13
RAP1B	Ras-related protein Rap-1b-like protein	A6NIZ1	.006	-3.26	0.10
TUBB	Tubulin beta chain	P07437	.001	-3.38	0.10
STOM	Erythrocyte band 7 integral membrane protein	P27105	< .001	-3.85	0.07

4. Discussion (condensed).

In the process of observing potential differences in the clinical characteristics of the patient subgroups High/Low activity and High/Low C3dg, respectively, several parameters showed significant differences.

When inspecting the individual ACR criteria in table 1 and figure 2C (individual ACR), the prevalence of ACR 1 (malar rash) was significantly more present in the High activity subgroup, compared to the Low activity subgroup (p = .035). The link between skin lesions and disease activity remains unclear in literature but may be due to the fact that the presence of any rash is a part of the SLEDAI score [24]. Therefore, malar rash, in particular, may be of importance in discriminating High disease activity from Low disease activity according to the present study. When comparing the disease group C3dg, a tendency showed that malar rash was more frequently observed in the Low C3dg subgroup (90.0%) compared to the High C3dg subgroup (44.4%), however, this difference was not significant.

As opposed to the ACR criteria involving malar rash and hematological disorder, a significant difference was observed between both the disease activity subgroups and C3dg subgroups regarding the SLEDAI scores, as seen in figure 2G. For the disease activity subgroups, High activity scored a significantly higher value than Low activity (p < .001), which was expected as the patients were allocated based on this scoring system. For the C3dg subgroups, a higher SLEDAI score was observed

in Low C3dg (p= .028), indicating that the disease activity differentiates the two subgroups, but not as strong as the disease activity subgroups. This result is contrary to the study conducted by Troldborg et al., who did not find C3dg to be correlated to the disease activity [23]. According to the present study, High activity and Low C3dg seem to be associated with the previously discussed parameters. However, C3dg may be able to distinguish SLE patient subgroups more precisely in other parameters, such as the liquid biomarkers, described in table X.

In the STRING analysis of the statistically significant proteins, two proteins within the activity disease group (C4A and Low-affinity IgG Fc region receptor III-A (FCGR3A)) were found to have a direct correlation to SLE according to KEGG pathways. Moreover, four proteins in the C3dg disease group (C4B, ACTN1, FCGR3A, and Low-affinity IgG Fc region receptor III-B (FCGR3B)) were found.

ACTN1, found in the C3dg disease group, is a protein necessary for the attachment of actin filaments to a variety of intracellular structures and is thought to have a relation to Lupus Nephritis (Becker-Merok et al.). A study by Becker-Merok et al. found significantly higher amounts of anti-α-actinin antibodies to be correlated to anti-ds-DNA in patients with Lupus Nephritis (Becker-Merok et al.). The present study found ACTN1 to be downregulated in High C3dg (p = .028), which brings up the question of whether the Low C3dg subgroup has more anti-α-actinin antibodies, as this subgroup was found to have a higher concentration of anti-ds-DNA. Thereby, it could indicate that patients with Low C3dg levels may be more prone to develop lupus nephritis. C4 is an effector protein in the immune system and can be degraded to two isoforms, C4A and C4B, which contribute to the susceptibility to autoimmune diseases. C4A binds more covalently to immune complexes than C4B which contributes to the formation of covalent ester bonds to antigens. A deficiency in the effector proteins (such as C1q, C1s, and C4) in the classical activation pathway is known to be one of the biggest genetic risk factors in SLE. Piereira et al. investigated the isoform C4A and found it to be negatively correlated to SLE disease activity via damage index (Piereira et al.). This is in contradiction with the results from the present study, which found C4A to be significantly higher expressed in the subgroup High activity (p = .022). Furthermore, C4B was found to be upregulated in High C3dg (p = .041) in the present study. This might indicate that a complete consumption of C4 has not happened within these subgroups. The two proteins FCGR3A and FCGR3B are both IgG Fcγ receptors, which are crucial in a variety of different immune responses, such as in the innate immune system [25]. Activation of both FCGR3A and FCGR3B initiates activation of immune cells which leads to inflammation. FCGR3A has the ability to mediate phagocytosis and antibody-dependent cell-mediated cytotoxicity, whereas FCGR3B can bind the immune complexes in the peripheral circulation and thereby act as an immune trap. In the present study, FCGR3B was upregulated in Low C3dg (p = .044) indicating a higher activation of immune cells and thereby more inflammation compared to patients with High C3dg. As one of the main functions of the IgG Fcγ receptors is clearance of immune complexes, these are highly associated with SLE, where especially FCGR3A is thought to have a crucial role. In the present study, FCGR3A was observed to be upregulated in High activity (p = .022) as well as upregulated in Low C3dg (p = .039). This might suggest a correlation between the two subgroups High activity and Low C3dg regarding complement activation and thereby disease activity.

5. Conclusion:

In the investigation of finding possible clinical and biological characteristics to differentiate the four SLE subgroups (High activity, Low activity, High C3dg, and Low C3dg), several parameters were analyzed using the provided clinical data and MS-based proteomics. Based on the clinical dataset, the activity subgroups differentiated on two parameters (ACR 1 and SLEDAI score), whereas the C3dg subgroups differentiated on five parameters (ACR 9, C3, C4, anti-ds-DNA, and SLEDAI score). Furthermore, the MS-based proteomic data from the activity patient group revealed 4 out of 31 significant proteins to be related to the pathogenesis of SLE. Among these, IGKV6–21 and FCGR3A were found to be associated with high disease activity. In the C3dg patient group, 10 out of 63 significant proteins were identified according to their relevance for SLE, of which the following proteins

were found to be associated with high disease activity: FCGR3A, FCGR3B, ACTN1, PRDX6, ITGA2B, and PKM. These proteins were all upregulated in the Low C3dg subgroup, indicating that a low concentration of C3dg might be associated with high disease activity. Moreover, the C3dg subgroups clustered more in the PCA plot compared to the activity subgroups.

This suggests that the concentration of C3dg in plasma could be used to allocate SLE patients based on their clinical, immunological, and biological characteristics. Furthermore, C3dg could act as a biomarker within Companion diagnostics type strategies serve as a more specific treatment endpoint for SLE patients. However, further validation of this observation is needed to translate these results for clinical use in the improvement of diagnostics and personalized treatment.

Conflicts of Interest: The authors have disclosed that they have no conflict of interest.

References:

1. Fanouriakis, A.; Kostopoulou, M.; Alunno, A.; Aringer, M.; Bajema, I.; Boletis, J.N.; Cervera, R.; Doria, A.; Gordon, C.; Govoni, M.; et al. 2019 Update of the EULAR recommendations for the management of systemic lupus erythematosus. *Ann. Rheum. Dis.* **2019**, *78*, 736–745, doi:10.1136/annrheumdis-2019-215089.

2. Troldborg, A.; Steffensen, R.; Trendelenburg, M.; Hauser, T.; Winther, K.G.; Hansen, A.G.; Stengaard-Pedersen, K.; Voss, A.; Thiel, S. Ficolin-3 Deficiency Is Associated with Disease and an Increased Risk of Systemic Lupus Erythematosus. *J. Clin. Immunol.* **2019**, *39*, 421–429, doi:10.1007/s10875-019-00627-2.

3. Kaul, A.; Gordon, C.; Crow, M.K.; Touma, Z.; Urowitz, M.B.; Van Vollenhoven, R.; Ruiz-Irastorza, G.; Hughes, G. Systemic lupus erythematosus. *Nat. Rev. Dis. Prim.* **2016**, *2*, 1–22, doi:10.1038/nrdp.2016.39.

4. Mak, A.; Cheung, M.W.L.; Chiew, H.J.; Liu, Y.; Ho, R.C. man Global Trend of Survival and Damage of Systemic Lupus Erythematosus: Meta-Analysis and Meta-Regression of Observational Studies from the 1950s to 2000s. *Semin. Arthritis Rheum.* **2012**, *41*, 830–839, doi:10.1016/j.semarthrit.2011.11.002.

5. *Characteristics, Benlysta Summar of Product*; 2011;

6. Dall'Era, M.; Cisternas, M.G.; Smilek, D.E.; Straub, L.; Houssiau, F.A.; Cervera, R.; Rovin, B.H.; MacKay, M. Predictors of long-term renal outcome in lupus nephritis trials: Lessons learned from the euro-lupus nephritis cohort. *Arthritis Rheumatol.* **2015**, *67*, 1305–1313, doi:10.1002/art.39026.

7. Muñoz, L.E.; Lauber, K.; Schiller, M.; Manfredi, A.A.; Herrmann, M. The role of defective clearance of apoptotic cells in systemic autoimmunity. *Nat. Rev. Rheumatol.* **2010**, *6*, 280–289, doi:10.1038/nrrheum.2010.46.

8. Mahajan, A.; Herrmann, M.; Muñoz, L.E. Clearance deficiency and cell death pathways: A model for the pathogenesis of SLE. *Front. Immunol.* **2016**, *7*, 1–12, doi:10.3389/fimmu.2016.00035.

9. Sturfelt, G.; Truedsson, L. Complement and its breakdown products in SLE. *Rheumatology* **2005**, *44*, 1227–1232, doi:10.1093/rheumatology/keh719.

10. Troldborg, A.; Thiel, S.; Laska, M.J.; Deleuran, B.; Jensenius, J.C.; Stengaard-Pedersen, K. Levels in Plasma of the Serine Proteases and Associated Proteins of the Lectin Pathway Are Altered in Patients with Systemic Lupus Erythematosus. *J. Rheumatol.* **2015**, *42*, 948–951, doi:10.3899/jrheum.141163.

11. Troldborg, A.; Thiel, S.; Trendelenburg, M.; Friebus-Kardash, J.; Nehring, J.; Steffensen, R.; Hansen, S.W.K.; Laska, M.J.; Deleuran, B.; Jensenius, J.C.; et al. The Lectin Pathway of Complement Activation in Patients with Systemic Lupus Erythematosus. *J. Rheumatol.* **2018**, *45*, 1136–1144, doi:10.3899/jrheum.171033.

12. M. K. Meyer, M. Andersen, T. V. Stausbo, K. J. Elbæk, G. N. Andersen, A.S. Polymyalgia Rheumatica Treatment Serum Biomarkers versus Rheumatoid Arthritis. In Proceedings of the Ann Rheum Dis Suppl 2017.

13. Poulsen, T.B.G.; Damgaard, D.; Jørgensen, M.M.; Senolt, L.; Blackburn, J.M.; Nielsen, C.H.; Stensballe, A. Identification of Novel Native Autoantigens in Rheumatoid Arthritis. *Biomedicines* **2020**, *8*, doi:10.3390/biomedicines8060141.

14. Kruse Meyer, M.; Andersen, M.; Bennike, T.B.; Birkelund, S.; Andersen, G.N.; Stensballe, A. Effect of IL-6R Inhibition with Tocilizumab on the Proteome of Peripheral Blood Mononuclear Cells from a Rheumatoid Arthritis Patient. *J. Proteomics Bioinform.* **2015**, *8*, doi:10.4172/jpb.1000380.

15. Bennike, T.B.; Ellingsen, T.; Glerup, H.; Bonderup, O.K.; Carlsen, T.G.; Meyer, M.K.; Bøgsted, M.; Christiansen, G.; Birkelund, S.; Andersen, V.; et al. Proteome Analysis of Rheumatoid Arthritis Gut Mucosa. *J. Proteome Res.* **2017**, *16*, 346–354, doi:10.1021/acs.jproteome.6b00598.

16. Meyer, M.; Andersen, M.; Ring, T.; Andersen, G.; Ehlers, L.; Rasmussen, C.; Stensballe, A. Personalized rheumatic medicine through dose reduction reduces the cost of biological treatment – a retrospective intervention analysis. *Scand. J. Rheumatol.* **2019**, *48*, 398–407, doi:10.1080/03009742.2019.1585940.

17. Kringelbach, T.M.; Glintborg, B.; Hogdall, E. V; Johansen, J.S.; Hetland, M.L.; Biomarker Protocol Study Group Identification of new biomarkers to promote personalised treatment of patients with inflammatory rheumatic disease: protocol for an open cohort study. *BMJ Open* **2018**, *8*, e019325, doi:10.1136/bmjopen-2017-019325.

18. Liu, C.C.; Kao, A.H.; Manzi, S.; Ahearn, J.M. Biomarkers in systemic lupus erythematosus: Challenges and prospects for the future. *Ther. Adv. Musculoskelet. Dis.* **2013**, *5*, 210–233, doi:10.1177/1759720X13485503.

19. Liu, C.-C.; Kao, A.H.; Manzi, S.; Ahearn, J.M. Biomarkers in systemic lupus erythematosus: challenges and prospects for the future. *Ther. Adv. Musculoskelet. Dis.* **2013**, *5*, 210–33, doi:10.1177/1759720X13485503.

20. Qi, S.; Chen, Q.; Xu, D.; Xie, N.; Dai, Y. Clinical application of protein biomarkers in lupus erythematosus and lupus nephritis. *Lupus* 2018, *27*, 1582–1590.

21. Mak, A.; Kow, N.Y.; Ismail, N.H.; Anuar, N.D.; Rutt, N.H.; Cho, J.; Rosli, N.S.B.M.; Dharmahidari, B. Detection of putative autoantibodies in systemic lupus erythematous using a novel native-conformation protein microarray platform. *Lupus* **2020**, *29*, 1948–1954, doi:10.1177/0961203320959696.

22. Poulsen, T.B.G.; Karamehmedovic, A.; Christhoper, A.; Jørgensen, M.M.; Yu, X.; Fang, X.; Blackburn, J.M.; Nielsen, C.H.; Kragstrup, T.W.; Stensballe, A. Protein array-based companion diagnostics in precision medicine. *Expert Rev. Mol. Diagn.* **2020**, *In revisio*.

23. Troldborg, A.; Jensen, L.; Deleuran, B.; Stengaard-Pedersen, K.; Thiel, S.; Jensenius, J.C. The C3dg Fragment of Complement Is Superior to Conventional C3 as a Diagnostic Biomarker in Systemic Lupus Erythematosus. *Front. Immunol.* **2018**, *9*, 581, doi:10.3389/fimmu.2018.00581.

24. Arora, S.; Isenberg, D.A.; Castrejon, I. Measures of Adult Systemic Lupus Erythematosus: Disease Activity and Damage. *Arthritis Care Res. (Hoboken).* **2020**, *72*, 27–46, doi:10.1002/acr.24221.

25. Chen, J.Y.; Wang, C.M.; Chang, S.W.; Cheng, C.H.; Wu, Y.J.J.; Lin, J.C.; Yang, B.; Ho, H.H.; Wu, J. Association of FCGR3A and FCGR3B copy number variations with systemic lupus erythematosus and rheumatoid arthritis in Taiwanese patients. *Arthritis Rheumatol.* **2014**, *66*, 3113–3121, doi:10.1002/art.38813.