Microarray testing in patients with systemic lupus erythematosus identifies a high prevalence of CpG DNA-binding antibodies

Tammo Brunekreef [ID] ,[1] Maarten Limper,[1] Rowena Melchers,[2] Linda Mathsson-Alm,[3] Jorge Dias,[3] Imo Hoefer,[4] Saskia Haitjema,[4] Jacob M van Laar,[1] Henny Otten[2]

[1]Rheumatology & Clinical Immunology, UMC Utrecht, Utrecht, The Netherlands
[2]Center of Translational Immunology, UMC Utrecht, Utrecht, The Netherlands
[3]ImmunoDiagnostics Division, Thermo Fisher Scientific, Uppsala, Sweden
[4]Clinical Diagnostic Laboratory, UMC Utrecht, Utrecht, The Netherlands

Correspondence to
Dr Henny Otten; h.g.otten@umcutrecht.nl

ABSTRACT

Objective Many autoantibodies are known to be associated with SLE, although their role in clinical practice is limited because of low sensitivity and weak associations with clinical manifestations. There has been great interest in the discovery of new autoantibodies to use in clinical practice. In this study, we investigated 57 new and known antibodies and their potential for diagnostics or risk stratification.

Methods Between 2014 and 2017, residual sera of all anti-dsDNA tests in the UMC Utrecht were stored in a biobank. This included sera of patients with SLE, patients with a diagnosis of another immune-mediated inflammatory disease (IMID), patients with low (non-IMID) or medium levels of clinical suspicion of SLE but no IMID diagnosis (Rest), and self-reported healthy blood bank donors. Diagnosis and (presence of) symptoms at each blood draw were retrospectively assessed in the patient records with the Utrecht Patient-Oriented Database using a newly developed text mining algorithm. Sera of patients were analysed for the presence of 57 autoantibodies with a custom-made immunofluorescent microarray. Signal intensity cut-offs for all antigens on the microarray were set to the 95th percentile of the non-IMID control group. Differences in prevalence of autoantibodies between patients with SLE and control groups were assessed.

Results Autoantibody profiles of 483 patients with SLE were compared with autoantibody profiles of 1397 patients from 4 different control groups. Anti-dsDNA was the most distinguishing feature between patients with SLE and other patients, followed by antibodies against Cytosine-phosphate-Guanine (anti-CpG) DNA motifs (p<0.0001). Antibodies against CMV (cytomegalovirus) and ASCA (anti-*Saccharomyces cerevisiae* antibodies) were more prevalent in patients with SLE with (a history of) lupus nephritis than patients with SLE without nephritis.

Conclusion Antibodies against CpG DNA motifs are prevalent in patients with SLE. Anti-CMV antibodies are associated with lupus nephritis.

INTRODUCTION

SLE is a systemic immune-mediated inflammatory disease (IMID) with a wide variety of associated symptoms and immunological

KEY MESSAGES

WHAT IS ALREADY KNOWN ABOUT THIS SUBJECT?
⇒ Many different autoantibodies have been investigated in SLE, although only a limited number of autoantibodies are being used in clinical practice.

WHAT DOES THIS STUDY ADD?
⇒ This study describes a high prevalence of anti-CpG antibodies in patients with SLE and a correlation of anti-CMV antibodies with lupus nephritis.

HOW MIGHT THIS IMPACT ON CLINICAL PRACTICE OR FUTURE DEVELOPMENTS?
⇒ Further research into anti-CpG autoantibodies might open up new diagnostic possibilities and provide new insights into the complex pathophysiology of SLE.

manifestations.[1 2] This heterogeneity in symptoms and lack of diagnostic criteria make SLE difficult to diagnose as there is no single laboratory test that can reliably confirm or exclude its presence.

Autoantibodies are a common immunological feature in patients with SLE and are known to be involved in its pathogenesis. Autoantibodies can bind antigen, so to form immune complexes, which can deposit in the skin or small vessels of the kidney and together with complement factors can cause inflammation.[3] Historically, autoantibodies have always been a component of classification criteria for SLE, although not necessarily required for classification.[4 5] However, in the new EULAR/ACR classification criteria, ANAs are now required to be present in order to classify a patient as having SLE.[1] Out of all ANAs, autoantibodies against double-stranded DNA (anti-dsDNA) have the strongest association with SLE and are present in up to 75% of all patients with SLE.[6] In clinical practice, suspicion of SLE is

the only reason for testing for anti-dsDNA antibodies, as tests for anti-dsDNA antibodies are known to have a high diagnostic specificity for SLE, although sensitivity varies.[7] Other common SLE-associated autoantibodies include antibodies against phospholipids (anti-cardiolipin, anti-beta2GP1), anti-Smith (anti-Sm), anti-C1q and anti-Ro/La antibodies. Some of these autoantibodies have also been reported to be associated with organ-specific manifestations, for example anti-C1q with renal involvement.[8]

Although many different autoantibodies have been investigated in SLE, only a limited number of autoantibodies are being used in clinical practice. Due to these limitations, there has been great interest in the discovery of new autoantibodies that can be used in clinical practice. In this study, we used a new microarray containing 57 autoantigens, some of which have never been explored in large cohorts, to analyse over 1500 serum samples of patients who were suspected of having SLE to explore their potential role in the diagnostic process and their association with symptoms of SLE.

PATIENTS AND METHODS
Patients
This study was performed in the University Medical Center Utrecht (UMC Utrecht), a tertiary care university hospital in the Netherlands. Between 2014 and 2017, residual material of all blood samples that were tested for the presence of anti-dsDNA in the UMC Utrecht were collected in a biobank and stored at −80°C. Additional serum samples from self-reported healthy blood bank donors (BBD) were provided by Thermo Fisher Scientific.

Clinical data
Clinical data of all patients from whom these samples were collected, were retrospectively gathered from the electronic health records (EHR) to assess the diagnosis and symptoms at the time each sample was taken. Since blood draw does not always happen on the same day as the visit to the clinician, all available data from the EHR up to 14 days after blood draw date were evaluated. All collected data were pseudonymised for privacy reasons.

Clinical data were extracted from the EHR through the Utrecht Patient Oriented Database (UPOD). UPOD is an infrastructure of relational databases comprising logistic and medical data for all patients treated at the UMC Utrecht since 2004.[9] A text mining algorithm was designed to extract the diagnosis and symptoms of patients automatically from the data of the EHR. To this end, the algorithm evaluated all physician's notes and letters of physicians to other physicians. Text mining has several advantages over manual extraction, it has excellent traceability and reproducibility, and is very well suited for large populations and more flexible than manual extraction as additional research questions can be formulated.

Details of the text mining algorithm used in this study have been described in detail.[10] In short, the text mining algorithm searches for pre-defined key words indicating the presence of a certain diagnosis or symptom in the written documents. Multiple diagnoses could be assigned to a single patient at the same time. Patients who did not get assigned any of these diagnoses were recorded as a 'non-diagnosis' group. The diagnoses assigned by the text mining algorithm do not necessarily comply with classification criteria for SLE and other IMIDs since these are mainly used to define a homogeneous population for research. Although most patients with a clinical diagnosis fulfil these criteria, in some cases classification criteria can be too restrictive for clinical practice. In this study, we assessed the clinical diagnoses in the EHR, rather than defining all patients according to classification criteria. It is noteworthy that, due to the working mechanism of the text mining algorithm, the recorded symptoms should be interpreted as having ever been present, rather than present at the moment of sample collection.

Patient groups
To assess the diagnostic value of the autoantibodies, we compared the presence of the autoantibodies in patients with SLE with four different control groups: In the first group, we included patients with a clinical diagnosis of an IMID other than SLE, as recorded by the treating physician in the EHR. This group was named the IMID control group. Patients without an IMID diagnosis recorded in the EHR were split into two different groups: in the second group, we selected patients most closely resembling healthy patients in our cohort as a primary control group, called the 'non-IMID' control group. These patients were seen at the outpatient clinic for manifestations associated with SLE, but were consequently not diagnosed with SLE or another IMID. Patients with either arthritis, nephritis or pleuritis, or two or more SLE-related symptoms were excluded from this group and recorded as the third group, called 'Rest'. These patients had more symptoms associated with and were clinically more suspect of having an IMID, although not enough for the physician to diagnose them as such. It should be noted that all patients included in the biobank had visited the outpatient clinic, which indicates that these patients are not true healthy controls. However, these patients resemble the non-disease population in clinical practice the most. In the fourth group, we included serum samples from self-reported healthy BBD, supplied by Thermo Fisher Scientific.

The outcome of the text mining algorithm was evaluated to select samples for further serological evaluation. Due to the large number of patients in the non-diagnosis group, a proportion of these samples was randomly selected for further serological evaluation. All available samples from patients with a diagnosis recorded by the algorithm were included for serological evaluation. If patients had multiple serum samples stored in the biobank, the first and last sample were analysed for the presence of autoantibodies.

Serological data

A custom-made chip-based microarray (Thermo Fisher Scientific, Uppsala Sweden) was used to analyse the serum samples for the presence of 57 IgG autoantibodies. The microarray contained nine control spots for calibration. Each antigen on the microarray was deposited in triplicate. An overview of the antigens included in this study is presented in online supplemental table 1. These antigens were selected by literature research prior to this study and by expert opinion. The antigens used in the microarray are all commercially available or in-house antigens supplied by Thermo Fisher Scientific. The correlation between the antigens on the microarray and their commercially available counterparts was assessed previous to this study and was found to be generally high (Spearman's correlation coefficients 0.58–0.88). Reproducibility of the microarray was studied extensively during development. An analysis of 72 slides, including all printed markers except guide spots, resulted in an overall coefficient of variance of approximately 5%. A full description of the microarray technology has been published previously.[11]

All selected samples were thawed prior to analysis. These samples were diluted 1:40 in dilution buffer, after which 30 µL was added to the reaction wells of the microarray. Samples were incubated for 2 hours at room temperature in a humidity chamber, after which the slides were washed and 30 µL of polyclonal Fluorescence DyLight550 labelled rabbit $F(ab')_2$-anti-human IgG was added to the reaction sites. After 30 min of incubating, the slides were washed again and the fluorescence intensity was measured by a Tecan LS Reloaded laser scanner. Each assay contained 96 samples and 12 control samples. Fluorescence intensities of all individual spots, as well as lock background intensity, were recorded and processed with the GenePix Pro Microarray Analysis software. This software evaluates the homogeneity within individual spots and between spots coated with the same antigen following pre-established quality criteria. All samples were processed by the same scanner and handled by the same laboratory technician.

Statistical analysis

The first available sample of each patient was used to compare the prevalence of autoantibodies between groups. Cut-offs of levels of immunofluorescent intensity to determine positive and negative results for all antigens on the microarray chips were set to the 95th percentile of the non-IMID controls. This cut-off was chosen, rather than a cut-off using the 95th percentile of the BBD patients, because the methods of sample collection and storage for the non-IMID group were identical to the rest of the cohort, but slightly different for the BBD, and because the non-IMID patients better represent the non-diseased patients seen in the hospital than the BBD. Categorical analysis was performed with the χ^2 test. Correction for multiple testing was done by Bonferroni correction. Statistical analysis was performed using SPSS (V.25.0.0.2).

RESULTS

Demographics

Blood samples of 2038 patients were collected in the biobank, with a total of 4607 collected samples. An overview of the baseline characteristics at the time of first sample collection per diagnosis is presented in table 1A. Patients with multiple diagnoses are represented in multiple rows. At the time of their first sample collection, a total of 1123 diagnoses were assigned to a total of 836 patients; 484 of these patients were assigned the diagnosis SLE. Also, 1202 patients were not assigned any diagnosis. The average age of the patients at the time the first sample was collected was 44.9 years (SD 16,9); 63.4% were female. Multiple samples were available in the biobank for 550 patients, and follow-up samples were not included in this analysis. A mean of 5.3 samples were collected for patients with SLE (range 1–26); for all other patients, this was a mean of 1.2 samples (range 1–14).

In the present study, a total of 1880 serum samples were analysed, including 483 samples of patients with SLE, 346 patients with another IMID, 690 patients with no diagnosis and 361 self-reported healthy BBD (figure 1). Seven samples of patients with SLE or another IMID were not analysed on the microarray: one sample was missing from the biobank; six samples were mistakenly not selected, because they were not assigned an IMID diagnosis during the selection of samples to analyse on the microarray, which was done using a pre-final iteration of the text-mining algorithm. Demographics of all groups are shown in table 1B.

Overlapping diagnoses

It is known that different IMIDs often overlap with regard to diagnosis and symptoms. In our cohort, 248 out of 836 (29.7%) patients were assigned more than one diagnosis, indicating that overlap of multiple diseases was common in our cohort. An overview of all overlapping diagnoses is presented in online supplemental table 2. Moreover, 40.5% of all patients with SLE had overlap with at least one other diagnosis, 13.2% of all patients with SLE had secondary APS, whereas 4.1% had secondary Sjogren's syndrome. Overlap of SLE with cutaneous lupus erythematosus (CLE), lupus-like disease (LLD, also referred to as incomplete lupus or iSLE), rheumatoid arthritis (RA) and mixed connective tissue disease (MCTD) was also recorded frequently.

Symptoms of patients with SLE

Fatigue was the most common symptom with a prevalence of 60.3%. Lupus nephritis, pericarditis and pleuritis were assigned to 33.3%, 12.8% and 14.0% of all patients with SLE, respectively. Psychosis was rare as a manifestation of SLE, with a prevalence of 4.2%. A small number of patients (2.3%) had no symptoms at the time of their first visit according to the text mining algorithm. An overview of all symptoms is presented in table 2.

Table 1 Demographics

(A) Demographics per diagnosis (patients with multiple diagnoses are included in multiple rows)

Diagnosis	Patients with diagnosis at sample 1	Patients with diagnosis in total follow-up	Female (%)	Mean age ±SD	Number of samples in biobank
Total of all diagnosed patients	1123	1218	63	44.9±16.9	4607
SLE	484	510	86	43.3±14.7	2726
CLE	143	160	85	48.5±16.1	410
APS	124	143	88	43.5±12.0	479
pSS	80	91	89	51.9±15.4	168
RA	68	72	65	44.6±14.0	103
LLD	57	70	93	41.4±15.2	169
MCTD	41	43	83	33.2±16.9	91
IIM	31	31	87	50.2±21.2	42
SSc	30	31	61	50.9±13.6	34
JIA	22	22	100	18.6±5.3	33
PMR	19	19	52	56.8±18.7	21
UCTD	11	13	91	39.4±14.5	20
GPA	7	7	57	46.9±12.5	7
GCA	6	6	33	61.6±13.0	6
No diagnosis	1202	1217	50	45.1±17.5	1324

(B) Demographics per patient group (age is the age at the time the first sample of the patient was collected)

Group	Samples (n)	Female (%)	Mean age±SD
SLE	483	86	43.3±14.7
Non-IMID	218	51	43.9±17.4
Rest	472	46	46.0±17.3
BBD	361*	50	46.7±14.1
IMID	346	78	46.2±17.6

*Sex and age of 2 blood bank donors is unknown.

CLE, cutaneous lupus erythematosus; GCA, giant cell arteritis; GPA, granulomatosis with polyangiitis; IIM, Idiopathic inflammatory myositis; JIA, juvenile idiopathic arthritis; LLD, lupus-like disease; MCTD, mixed connective tissue disease; PMR, polymyalgia rheumatica; pSS, (primary) Sjogren's syndrome; SSc, systemic sclerosis; UCTD, undifferentiated connective tissue disease.
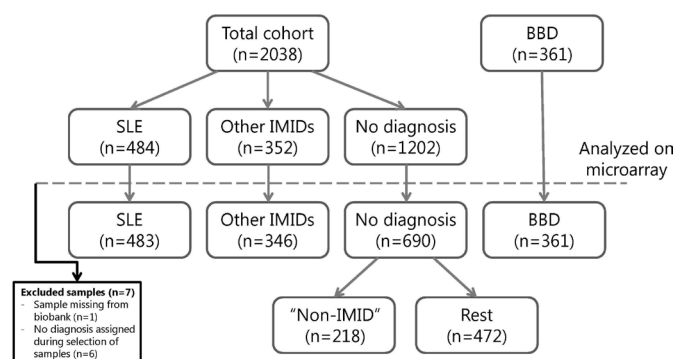
Figure 1 Flowchart of samples analysed on microarray per patient group. Overview of number of samples analysed on microarray for all patient groups. Samples of 1 patient with SLE and 6 patients with a diagnosis of another IMID were not analysed on the microarray system:1 sample was missing from the biobank; 6 samples were mistakenly not selected.

Analysis of autoantibodies

Measurements of anti-dsDNA antibodies from the regular laboratory measurement and the microarray correlated moderately (Spearman's ρ 0.578). Using the cut-offs as defined in the methods, the prevalence of multiple autoantibodies was significantly higher in patients with SLE, compared with non-IMID controls (table 3, see online supplemental file 3 for ROC curves and AUC values). Antibodies against dsDNA were the most common antibody found in patients with SLE, with a prevalence of 41.2%, followed by antibodies against single-stranded Cytosine-phosphate-Guanine (CpG) oligodinucleotides (CpG ODN 2216), with a prevalence of 40.2%. Antibodies against dsDNA and CpG were also the most distinguishing feature between patients with SLE and the rest group, the BBD and all other patients with another IMID (figure 2). Antibodies against dsDNA and CpG correlated moderately (Spearman's ρ 0.749, figure 3). Of all patients,

Table 2 Prevalence of features of SLE at the time of the first sample and at the time of the latest follow-up sample

Feature	At sample 1 (N, (%))	At latest follow-up sample (N, (%))
Total no of patients	484	369
Symptoms		
Alopecia	121 (25.0)	120 (32.5)
Arthritis	114 (23.6)	109 (29.5)
Arthralgia	157 (32.4)	160 (43.4)
Cranial nerve disorder	23 (4.8)	35 (9.5)
Digital ulceration	3 (0.6)	3 (0.8)
Fatigue	287 (59.3)	291 (78.9)
Mucosal ulcers	153 (31.6)	172 (46.6)
Myositis	15 (3.1)	29 (7.9)
Nephritis	161 (33.3)	190 (51.4)
New rash	144 (29.8)	170 (46.1)
Pericarditis	62 (12.8)	64 (17.3)
Pleurisy	68 (14.0)	75 (20.3)
Psychosis	20 (4.1)	28 (7.6)
Raynaud	161 (33.3)	157 (42.5)
Seizure	38 (7.9)	45 (12.2)
Vasculitis	66 (13.6)	82 (22.2)
Visual disturbance	3 (0.6)	8 (2.2)
Laboratory features		
ANA	86 (59.2)*	18 (58.1)†
dsDNA	170 (35.1)	141 (38.2)
Haematuria	77 (15.9)	65 (17.6)
Haemolytic anaemia	4 (0.8)	7 (1.9)
Leucopenia	17 (3.5)	12 (3.3)
Low complement	145 (30.0)	128 (34.7)
Proteinuria	14 (2.9)	9 (2.4)
Pyuria	179 (37.0)	150 (40.7)
Thrombocytopenia	19 (3.9)	7 (1.9)
Urinary casts	7 (1.4)	1 (0.3)
No features	3 (0.6)	0 (0)

Symptoms were assessed by text mining, laboratory assessments are directly taken from the EHR.
*For the first sample, 147 ANA tests were performed, 87 were positive or strong positive, 41 weak positive and 19 negative.
†For the follow-up samples, 31 ANA tests were performed, 18 were positive or strong positive, 10 weak positive and 3 negative.
dsDNA, anti-double-stranded antibodies.

7.4% was positive for anti-CpG antibodies and negative for anti-dsDNA antibodies. In patients with SLE, 8.9% of patients was positive for anti-CpG antibodies and negative for anti-dsDNA antibodies. Also, 15.1% of patients with SLE negative for anti-dsDNA were positive for anti-CpG antibodies, as well as 12.7% of patients with SLE with no anti-dsDNA and/or lowered complement factors (online supplemental file 3).

Prevalence of most autoantibodies was similar in the non-IMID and BBD control groups, although the auto-antibodies anti-Ro52/60 and anti-La were more prevalent in the non-IMID group and (nearly) absent in the BBD group (table 3). These differences suggest that patients who come to the outpatient clinic have a different serological profile than healthy blood donors, even without the patient being labelled as having a specific diagnosis. Prevalence of the antibodies most prevalent in patients with SLE, such as antibodies against RipP1, SmBB and histones, was slightly higher in the rest group compared with the non-IMID group. This indicates that, although these patients are not diagnosed with an IMID, they do not only display more SLE-related symptoms than the non-IMID group, but their serological profile is also a little bit closer to that of patients with SLE. The overall prevalence of autoantibodies in patients with other IMIDs than SLE is markedly higher than the prevalence in the non-IMID group, although many autoantibodies known to be associated with SLE are still significantly more prevalent in patients with SLE (table 3).

Associations with nephritis
Antibodies against CpG, SmBB, Histones, Nucleosomes, Mi2, CMV (cytomegalovirus) and ASCA (anti-*Saccharomyces cerevisiae* antibody) were more prevalent in patients with SLE with nephritis compared with patients without nephritis, although only differences in antibodies against CMV (31.7% vs 17.8%) and ASCA (8.1% vs 1.6%) remained significant after Bonferroni correction for multiple testing.

DISCUSSION
Our study focused on exploring the potential role of a number of autoantibodies, both well-known and lesser-known, in the diagnostic process and their associations with symptoms of SLE. We evaluated serum samples from more than 1500 patients for the presence of 57 autoantigens using a new microarray. We show a moderate association of anti-CMV and anti-ASCA antibodies with lupus nephritis and identified antibodies against CpG DNA motifs to be strongly associated with SLE.

In this study, we show that autoantibodies against CpG were highly discriminatory between patients with SLE and non-IMID controls and patients with other IMIDs. As an individual marker, the diagnostic potential of anti-CpG antibodies for SLE appears to be better than most other autoantibodies currently used in clinical practice, only secondary to anti-dsDNA. Furthermore, although this difference is not significant after Bonferroni correction, we show an increased prevalence of anti-CpG antibodies in patients with SLE with nephritis, compared with patients without nephritis, potentially further increasing their diagnostic applicability (table 3). CpG motifs are unmethylated cytosine-phosphate-guanine

Table 3 Prevalence of autoantibodies in patients with SLE, non-IMID controls, rest group, BBD controls and patients with other IMIDs, and prevalence of autoantibodies in patients with SLE with nephritis

Antibody target	Prevalence in SLE n (%)	SLE vs non-IMID controls (p value)	Prevalence rest n (%)	SLE vs rest (p value)	Prevalence BBD controls n (%)	SLE vs BBD controls (p value)	Prevalence IMIDs n (%)	SLE vs IMIDs (p value)	Prevalence in pts with nephritis n (%)	SLE nephritis vs non-nephritis (p value)
dsDNA	199 (41.2)	3.9E-22	26 (5.5)	1.3E-38	4 (1.1)	2.0E-41	42 (12.1)	1.0E-19	76 (47.2)	NS
CpG	194 (40.2)	3.0E-21	42 (8.9)	4.1E-29	26 (7.2)	3.7E-27	55 (15.9)	5.6E-14	77 (47.8)	1.5E-02*
RipP1	161 (33.3)	7.8E-16	55 (11.7)	1.2E-15	18 (5.0)	2.1E-23	49 (14.2)	3.9E-10	61 (37.9)	NS
SmBB	147 (30.4)	9.6E-14	41 (8.7)	2.9E-17	12 (3.3)	2.2E-23	38 (11)	3.3E-11	59 (36.6)	3.6E-02*
Histones	134 (27.7)	6.5E-12	40 (8.5)	1.2E-14	35 (9.7)	9.0E-11	60 (17.3)	4.9E-04	55 (34.2)	2.6E-02*
RipP0	133 (27.5)	8.9E-12	36 (7.6)	7.6E-16	19 (5.3)	8.0E-17	49 (14.2)	4.5E-06	44 (27.3)	NS
Nucleosome	129 (26.7)	3.1E-11	32 (6.8)	2.0E-16	16 (4.4)	2.1E-17	47 (13.6)	5.2E-06	55 (34.2)	8.8E-03*
RipP2	127 (26.3)	5.8E-11	41 (8.7)	9.0E-13	4 (1.1)	1.6E-23	46 (13.3)	5.6E-06	44 (27.3)	NS
Ro60	119 (24.6)	6.5E-10	30 (6.4)	7.0E-7	0 (0)	2.6E-24	53 (15.3)	1.1E-03*	46 (28.6)	NS
CMV	108 (22.4)	1.6E-08	45 (9.5)	6.6E-8	9 (2.5)	1.4E-16	37 (10.7)	1.3E-05	51 (31.7)	5.1E-04
RNPA	94 (19.5)	7.4E-07	43 (9.1)	5.0E-6	14 (3.9)	2.0E-11	56 (16.2)	NS	27 (16.8)	NS
Ro52	89 (18.4)	2.7E-06	27 (5.7)	1.9E-9	1 (0.3)	2.9E-17	50 (14.5)	NS	32 (19.9)	NS
RNP70	77 (15.9)	5.6E-05	24 (5.1)	4.9E-8	7 (1.9)	1.8E-11	37 (10.7)	3.0E-02*	24 (14.9)	NS
SMP	74 (15.3)	1.1E-04	23 (4.9)	9.1E-8	7 (1.9)	6.6E-11	17 (4.9)	2.3E-06	31 (19.3)	NS
LaSSB	71 (14.7)	2.3E-04	19 (4.0)	1.7E-8	0 (0)	2.7E-14	30 (8.7)	8.9E-03*	27 (16.8)	NS
C1q	69 (14.3)	3.7E-04	20 (4.2)	9.3E-8	16 (4.4)	2.5E-06	16 (4.6)	6.1E-06	29 (18.0)	NS
RNPC	69 (14.3)	3.7E-04	22 (4.7)	4.1E-7	18 (5.0)	1.1E-05	30 (8.7)	1.4E-02*	20 (12.4)	NS
H2Bp	60 (12.4)	2.7E-03*	18 (3.8)	1.2E-6	15 (4.2)	3.0E-05	8 (2.3)	1.7E-07	26 (16.1)	NS
C3b	57 (11.8)	5.2E-03*	29 (6.1)	2.2E-3*	19 (5.3)	1.0E-03*	29 (8.4)	NS	22 (13.7)	NS
Ku	51 (10.6)	1.7E-02*	27 (5.7)	6.3E-3*	6 (1.7)	3.5E-07	22 (6.4)	3.5E-02*	17 (10.6)	NS
RA33	50 (10.4)	2.1E-02*	17 (3.6)	4.4E-5	9 (2.5)	9.4E-06	20 (5.8)	2.0E-02*	19 (11.8)	NS
Actinin	50 (10.4)	2.1E-02*	44 (9.3)	NS	22 (6.1)	2.8E-02*	43 (12.4)	NS	15 (9.3)	NS
Cardiolipin	49 (10.1)	2.5E-02*	36 (7.6)	NS	16 (4.4)	2.1E-03*	21 (6.1)	3.7E-02*	15 (9.3)	NS
H2Bpac	44 (9.1)	NS	18 (3.8)	9.0E-4*	16 (4.4)	8.9E-03*	16 (4.6)	1.4E-02*	26 (16.1)	NS
H4p	42 (8.7)	NS	23 (4.8)	1.9E-2*	11 (3.0)	8.2E-04	13 (3.8)	4.8E-03*	16 (9.9)	NS
Pentraxin3	41 (8.5)	NS	35 (7.4)	NS	14 (3.9)	7.3E-03*	14 (4)	1.1E-02*	15 (9.3)	NS
CRP	35 (7.2)	NS	16 (3.4)	8.0E-3*	4 (1.1)	2.6E-05	10 (2.9)	6.3E-03*	15 (9.3)	NS
ASCA	18 (3.7)	NS	19 (4.0)	NS	10 (2.8)	NS	15 (4.3)	NS	13 (8.1)	3.6E-04
Mi2	31 (6.4)	NS	27 (5.7)	NS	14 (3.9)	NS	17 (4.9)	NS	16 (9.9)	2.6E-02*

Only autoantibodies with significant differences between groups are shown, antibodies not shown in this table did not significantly differ between groups. Due to the fixed specificity of 95%, used to determine the cut-offs for all autoantibodies, the prevalence of all autoantibodies in the non-IMID controls is 5% and is therefore not shown in this table. Statistical analysis by χ^2 test. P values are given (E-05=10^{-5}),

*No significant difference after Bonferroni correction.

BBD, blood bank donor; IMID, immune-mediated inflammatory disease; NS, not significant.
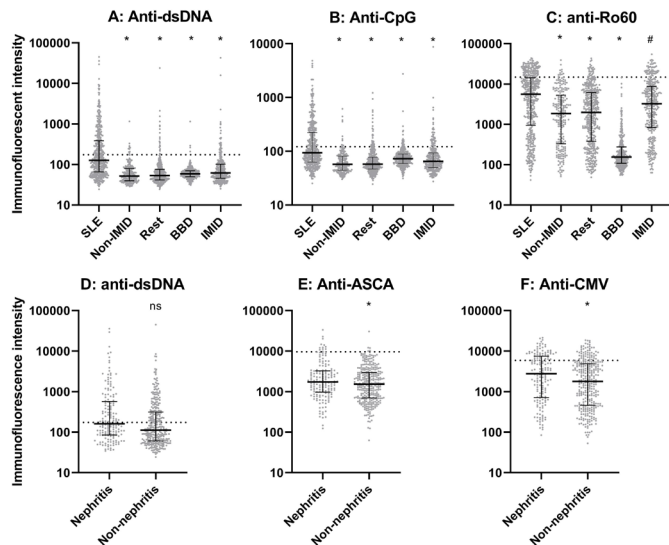
Figure 2 Levels of immunofluorescent intensity. The 5th–95th percentile of immunofluorescent intensity is presented for (A) anti-dsDNA, (B) anti-CpG, (C) anti-Ro60 antibodies for patients with SLE (n=483), non-IMID controls (n=218), the 'rest' group (n=472), blood bank donors (n=361) and patients with other AID (n=346). The 5th–95th percentile of immunofluorescent intensity is presented for (D) anti-dsDNA, (E) anti-ASCA, (F) anti-CMV antibodies for patients with SLE with and without nephritis. The dotted lines represent the cut-off between negative and positive results. *p<0001, #p=0001, ns=not significant. Prevalence of autoantibodies in the respective groups vs SLE, based on cut-offs (**A**–**C**). For comparisons between other groups, see online supplemental file 3.

oligodinucleotides, DNA patterns that are 20 times more prevalent in bacterial DNA than in mammalian DNA. CpG is recognised by the innate immune system as a pathogen-associated molecular pattern and can induce an interferon response by stimulating intracellular toll-like
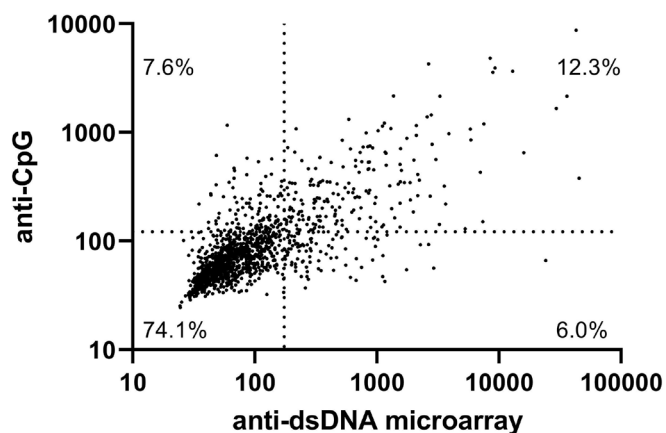
Figure 3 Correlation of anti-dsDNA and anti-CpG antibodies. Scatterplot of anti-dsDNA and anti-CpG antibodies in all patients, immunofluorescent intensity measured using the microarray. Spearman's ρ=0.749. The dotted lines represent the cut-off between negative and positive results.

receptor 9 (TLR9).[12 13] In patients with SLE, an interferon signature is seen in 60%–80% of patients, as well as an increase in the number of neutrophil extracellular traps (NETs), which both can be induced by stimulation of TLR9.[14–16] Mitochondrial DNA (mtDNA) is known to contain relatively high amounts of CpG motifs, compared with nuclear DNA, likely due to the bacterial evolutionary origin of the organelle.[17 18] mtDNA is also expelled from the cell in the process of NETosis, possibly providing a source of antigen for anti-CpG antibodies.[19] Recently, Becker et al suggested that the Farr assay commonly used for detection of dsDNA also detects antibodies against CpG-rich mtDNA. Furthermore, they speculate about the existence of antibodies preferentially binding to mtDNA rather than genomic DNA.[20] Our results suggest that antibodies against CpG motifs might fill this role.

Antibodies specific against CpG have only been described in one human study in which patients were injected with CpG as standalone treatment or as adjuvant for NY-ESO-1-specific cancer vaccines. Also, 21 out of 37 patients developed antibodies against CpG; no correlations with anti-dsDNA, antinuclear antibodies or autoimmune manifestations were seen.[21] Previous studies have described that some anti-dsDNA antibodies preferentially bind to CpG-containing dsDNA and that immune complexes of DNA and anti-DNA antibodies are relatively rich of immunostimulatory CpG motifs, but did not describe specific anti-CpG antibodies.[22 23] As our study was not designed to determine the exact specificities of antibodies, we cannot rule out that the results for anti-CpG antibodies in our study can at least partially be explained by cross-reactivity of anti-dsDNA antibodies with the CpG motifs on the microarray. In our study, we do find a correlation between anti-CpG and anti-dsDNA antibodies; however, the CpG antigen used in this study was single-stranded DNA, making cross-reactivity of anti-dsDNA antibodies less likely. Furthermore, we show that anti-CpG antibodies are present in 15.1% of patients with SLE who do not have anti-dsDNA antibodies, indicating potential additional value in the diagnostic process of SLE.

Our study also shows that IgG anti-CMV antibodies are more prevalent in patients with SLE (22.4%), particularly in patients who have experienced lupus nephritis (31.7%). CMV is a known cause of opportunistic infections in patients with SLE who are being treated with immunosuppressive medication, and an aetiological role of CMV in the onset of SLE has been suggested.[24] Zhang et al showed that 5.3% of patients with lupus nephritis had a CMV infection or reactivation, measured by CMV pp65 antigenemia, during hospitalisation.[25] Furthermore, Su et al showed that almost all patients with SLE have IgG anti-CMV antibodies, but that patients who were also positive for IgM anti-CMV antibodies had the highest titres of anti-CMV IgG and had higher SLEDAI scores.[26] Unfortunately, in our study we only have data of IgG anti-CMV antibodies, making it impossible to distinguish active from past infection, but this finding contributes to the

expanding evidence of the relation between CMV and SLE.

We also found significant differences in the prevalence of ASCA antibodies in patients with and without lupus nephritis, although the prevalence of these autoantibodies is probably too low to be of clinical relevance. Although ASCA antibodies are most commonly associated with Crohn's disease, they have also been found in serum of patients with SLE. Interestingly, we did not find associations between anti-dsDNA or anti-C1q and nephritis, which have previously been described in literature.[27 28] We speculate that this might be because the patients with lupus nephritis in our cohort often do not necessarily have active lupus nephritis, but rather a history of lupus nephritis, and these antibodies are most strongly associated with active nephritis.

Our study is unique in the number of patients with SLE included and in the use of multiple relevant control groups. Whereas most studies investigate differences in biomarkers between diseased and healthy controls, our study contains healthy controls as well as all three groups of patients visiting the outpatient clinic: patients with SLE, patients with another IMID and patients with signs and symptoms, but who cannot be diagnosed with a circumscriptive diagnosis. Although these patients do not receive a diagnosis, we show that there are clear differences in their serological profile compared with healthy blood donors, most clearly exemplified by the differences in presence of anti-Ro and anti-La antibodies. We argue that in studies investigating the diagnostic characteristics of autoantibodies, it would be most helpful to compare patients with SLE with patients who visit the outpatient clinic but do not receive a diagnosis, rather than from healthy controls, because this will provide more insight into the clinical relevancy.

Our study has several strengths: We used a dedicated custom microarray system to study the presence of autoantibodies in a large cohort of patients with SLE, which we compared with multiple relevant control groups. All autoantibodies included on this microarray system have been reported to be present in serum of patients with SLE in one or more studies (see online supplemental table 1). Furthermore, intra-observer and inter-observer variability are known to occur in review of medical records in patients with SLE.[29] In our study, we removed this variability by assessing all relevant information in an automatised and therefore uniform manner.

Our study also has some limitations. Although we previously showed that the text mining algorithm performs well in recognising patients with SLE, the text mining algorithm is not infallible, resulting in misclassification of a small number of patients. However, our text mining algorithm performs very well with a very high sensitivity and specificity in detecting patients with SLE (>90%), which should be adequate in a large cohort like this.[10] Ideally, the results of this study should be validated in patients meeting the classification criteria for SLE, rather than patients with a clinical diagnosis of SLE. Our study

did not include the use of immunosuppressive medication in the analysis, although some types of medication, such as belimumab and rituximab, are known to be able to influence levels of autoantibodies. Since patients with SLE, particularly with severe symptoms, are more likely to receive these types of medication, this could introduce a bias, resulting in a lower prevalence of autoantibodies in this group.

In conclusion, autoantibodies against CpG were shown to be strongly associated with SLE and can possibly help physicians in diagnosing SLE whenever a patient does not have anti-dsDNA antibodies. Our study presents insight into correlations of autoantibodies and symptoms of SLE, which in the future might be used in patient stratification. Our finding of a high prevalence of anti-CpG antibodies raises questions about their potential role in pathophysiology. Cross-reactivity of anti-CpG and anti-dsDNA antibodies, the preferred source of anti-CpG DNA, as well as the relation of anti-CpG antibodies with NETosis and CpG-rich mitochondrial DNA should be investigated. Further research into anti-CpG autoantibodies can possibly provide new insights into the complex pathophysiology of this disease, leading to advances towards a reliable risk stratification and new treatment strategies.

ORCID iD
Tammo Brunekreef http://orcid.org/0000-0003-1173-5937

REFERENCES

1 Aringer M, Costenbader K, Daikh D, *et al*. 2019 European League against Rheumatism/American College of Rheumatology classification criteria for systemic lupus erythematosus. *Ann Rheum Dis* 2019;78:1151–9.

2 Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. *J Rheumatol* 2002;29:288–91.

3 Yung S, Chan TM. Mechanisms of kidney injury in lupus nephritis – the role of anti-dsDNA antibodies. *Front Immunol* 2015;6:1–11.

4 Tan EM, Cohen AS, Fries JF, *et al*. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis & Rheumatism* 1982;25:1271–7.

5 Petri M, Orbai A-M, Alarcón GS, *et al*. Derivation and validation of the systemic lupus international collaborating clinics classification criteria for systemic lupus erythematosus. *Arthritis & Rheumatism* 2012;64:2677–86.

6 Isenberg DA, Manson JJ, Ehrenstein MR, *et al*. Fifty years of anti-ds DNA antibodies: are we approaching journey's end? *Rheumatology* 2007;46:1052–6.

7 Didier K, Bolko L, Giusti D, *et al*. Autoantibodies associated with connective tissue diseases: what meaning for clinicians? *Front Immunol* 2018;9:1–20.

8 Beurskens FJ, van Schaarenburg RA, Trouw LA. C1Q, antibodies and anti-C1q autoantibodies. *Mol Immunol* 2015;68:6–13.

9 ten Berg MJ, Huisman A, van den Bemt PMLA, *et al*. Linking laboratory and medication data: new opportunities for pharmacoepidemiological research. *Clin Chem Lab Med* 2007;45:13–19.

10 Brunekreef TE, Otten HG, van den Bosch SC, *et al*. Text mining of electronic health records can accurately identify and characterize patients with systemic lupus erythematosus. *ACR Open Rheumatol* 2021;3:65–71.

11 Hansson M, Mathsson L, Schlederer T, *et al*. Validation of a multiplex chip-based assay for the detection of autoantibodies against citrullinated peptides. *Arthritis Res Ther* 2012;14:R201.

12 Klinman DM, Yi AK, Beaucage SL, *et al*. Cpg motifs present in bacteria DNA rapidly induce lymphocytes to secrete interleukin 6, interleukin 12, and interferon gamma. *Proc Natl Acad Sci U S A* 1996;93:2879–83.

13 Avalos AM, Latz E, Mousseau B, *et al*. Differential cytokine production and bystander activation of autoreactive B cells in response to CpG-A and CpG-B oligonucleotides. *J Immunol* 2009;183:6262–8.

14 Crow MK. Type I interferon in the pathogenesis of lupus. *J.i.* 2014;192:5459–68.

15 Papayannopoulos V. Neutrophil extracellular traps in immunity and disease. *Nat Rev Immunol* 2018;18:134–47.

16 Liu L, Mao Y, Xu B, *et al*. Induction of neutrophil extracellular traps during tissue injury: involvement of sting and Toll-like receptor 9 pathways. *Cell Prolif* 2019;52:e12579–12.

17 Zhang Q, Raoof M, Chen Y, *et al*. Circulating mitochondrial DAMPs cause inflammatory responses to injury. *Nature* 2010;464:104–7.

18 Roger AJ, Muñoz-Gómez SA, Kamikawa R. The origin and diversification of mitochondria. *Current Biology* 2017;27:R1177–92.

19 Yousefi S, Stojkov D, Germic N, *et al*. Untangling "NETosis" from NETs. *Eur J Immunol* 2019;49:221–7.

20 Becker Y, Loignon R-C, Julien A-S, *et al*. Anti-mitochondrial autoantibodies in systemic lupus erythematosus and their association with disease manifestations. *Sci Rep* 2019;9:4530.

21 Karbach J, Neumann A, Wahle C, *et al*. Therapeutic administration of a synthetic CpG oligodeoxynucleotide triggers formation of anti-CpG antibodies. *Cancer Res* 2012;72:4304–10.

22 Sano H, Morimoto C. DNA isolated from DNA/anti-DNA antibody immune complexes in systemic lupus erythematosus is rich in guanine-cytosine content. *J Immunol* 1982;128:1341–5.

23 Uccellini MB, Busto P, Debatis M, *et al*. Selective binding of anti-DNA antibodies to native dsDNA fragments of differing sequence. *Immunol Lett* 2012;143:85–91.

24 Sekigawa I, Nawata M, Seta N, *et al*. Cytomegalovirus infection in patients with systemic lupus erythematosus. *Clin Exp Rheumatol* 2002;20:559–64.

25 Zhang L, Tao J, Wen Y, *et al*. Cytomegalovirus infection in patients with lupus nephritis: clinical and laboratory features and therapeutic considerations. *Clin Exp Med* 2017;17:467–75.

26 Su BY-J, Su C-Y, Yu S-F, *et al*. Incidental discovery of high systemic lupus erythematosus disease activity associated with cytomegalovirus viral activity. *Med Microbiol Immunol* 2007;196:165–70.

27 Trendelenburg M, Lopez-Trascasa M, Potlukova E, *et al*. High prevalence of anti-C1q antibodies in biopsy-proven active lupus nephritis. *Nephrology Dialysis Transplantation* 2006;21:3115–21.

28 Alba P *et al*. Anti-dsDNA, anti-Sm antibodies, and the lupus anticoagulant: significant factors associated with lupus nephritis. *Ann Rheum Dis* 2003;62:556–60.

29 FitzGerald JD, Grossman JM. Validity and reliability of retrospective assessment of disease activity and flare in observational cohorts of lupus patients. *Lupus* 1999;8:638–44.