Human Immunology 83 (2022) 509–514

Contents lists available at ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

Research article

# Microarray analysis of autoantibodies can identify future Systemic Lupus Erythematosus patients

Tammo E. Brunekreef [a], Leon C. Reteig [b,c], Maarten Limper [a], Saskia Haitjema [b], Jorge Dias [d], Linda Mathsson-Alm [d], Jacob M. van Laar [a], Henny G. Otten [c,*]

[a] Department of Rheumatology & Clinical Immunology, UMC Utrecht, Netherlands
[b] Central Diagnostic Laboratory, UMC Utrecht, Netherlands
[c] Center of Translational Immunology, UMC Utrecht, Netherlands
[d] Thermo Fisher Scientific, Uppsala, Sweden

ARTICLE INFO

*Article history:*
Received 25 January 2022
Revised 25 March 2022
Accepted 31 March 2022
Available online 11 April 2022

*Keywords:*
Systemic Lupus Erythematosus
Early diagnosis
Autoantibodies
Prediction model

ABSTRACT

*Objective:* Reliable early ascertainment in patients with SLE is important to prevent the accumulation of irreversible organ damage. Autoantibodies are often present in the serum of patients before the first symptoms arise, therefore they are of potential use as early diagnostic tools.
*Methods:* We used a custom-made antibody microarray containing 57 autoantigens to analyze serum samples of 1519 patients previously tested for anti-dsDNA and 361 samples of self-reported healthy blood bank donors (BBD). The 1519 patients included 483 patients with SLE, 346 patients with other immune mediated inflammatory diseases (IMID), 218 patient controls without relevant clinical symptoms (Non-IMID), and 472 patients that did not fit in any of the previous groups (Rest). The Non-IMID and BBD groups were used individually to create multivariable prediction models to distinguish samples of patients with SLE from these control groups. We subsequently used these models to predict the outcome for samples of patients who developed SLE while in follow-up (pre-SLE).
*Results:* Out of 1036 patients with no diagnosis of SLE at the moment of sample collection, 17 patients developed SLE while in follow-up (mean time to diagnosis 7.2 months). The best performing model (AUC 0.83) identified 9 out of 17 (53%) pre-SLE samples as SLE, with a specificity of 94%.
*Conclusion:* Serum samples of patients who will develop SLE in the future already show a shift of the autoantibody profile prior to diagnosis. In this study, we show that these autoantibody profiles can be used to identify these future SLE patients.


## 1. Introduction

Systemic lupus erythematosus (SLE) is a heterogeneous immune mediated inflammatory disease (IMID), which can often be difficult to diagnose. Prolonged activity of the disease can lead to accumulation of organ damage, sometimes irreversible [1]. This underlines the importance of identification of patients in early stages of the disease, in order to start adequate treatment before significant organ damage occurs [2].

In practice, however, the time from the first symptom to the diagnosis of SLE usually is multiple months to several years [3,4]. This delay can largely be attributed to the difficulty in the diagnostic process and lack of early diagnostic tools. The clinical diagnosis of SLE relies on the recognition of a pattern of symptoms and immunological manifestations by the physician, as diagnostic criteria do not exist. Classification criteria have been developed to identify relatively homogeneous groups of patients for inclusion in research studies, however a patient does not have to fulfill these criteria to be clinically diagnosed with SLE [5]. Patients who are eventually diagnosed with SLE initially often present with non-specific symptoms, such as arthralgia or fatigue, frequently resulting in the delay of the diagnostic process until additional symptoms arise [6,7].

Diagnostic tools are mainly aimed at determining the presence of autoantibodies. Numerous autoantibodies have been described to be related to SLE, although only a limited number of these, such as antibodies against anti-double-stranded DNA (anti-dsDNA), have a place in the regular diagnostic process [5,8]. Previous stud-

* Corresponding author at: PO Box 85500, 3508 GA Utrecht, The Netherlands.
*E-mail addresses:* h.g.otten@umcutrecht.nl, t.e.brunekreef-2@umcutrecht.nl, h.g.otten@umcutrecht.nl (H.G. Otten).

https://doi.org/10.1016/j.humimm.2022.03.010
0198-8859/

ies indicate autoantibodies might have great diagnostic potential in the early stages of SLE, as autoantibodies can precede the first clinical manifestations of SLE by several years, similar to anti-citrullinated protein antibodies (ACPA) in rheumatoid arthritis [9–11]. In this study we measured the autoantibody profile of 1880 serum samples on a custom-made antibody microarray containing 57 autoantigens. The acquired data was used to develop multivariable prediction models to identify future SLE patients.

## 2. Methods

### 2.1. Patients

This study was performed in the University Medical Center Utrecht (UMC Utrecht), a tertiary care university hospital in the Netherlands. Residual material of all blood samples that were tested for the presence of anti-dsDNA in the UMC Utrecht between 2014 and 2017 was collected in a biobank and stored at −80 °C. Additionally, serum samples from self-reported healthy blood bank donors (BBD) were provided by Thermo Fisher Scientific. This study was in accordance with the guidelines approved by the Medical Ethical Committee and approved by the biobank committee of the UMC Utrecht.

#### 2.1.1. Clinical data

Clinical data were retrospectively gathered from the electronic health records (EHR) via the Utrecht Patient Oriented Database [12]. We used a text-mining algorithm to assess the diagnosis and symptoms of all included patients [13]. All collected data were pseudonymized before analysis. The text-mining algorithm was used on the data recorded in the EHR up to July 2018. All data up to two weeks after the date of sample collection were used by the algorithm to determine the diagnosis at the time of each individual sample, to also incorporate diagnoses made with the results of this specific test.

#### 2.1.2. Patient groups

The first available serum samples of 1519 patients were included in this study. This included patients with a diagnosis of SLE, patients with another IMID and patients without an IMID at the moment the sample was collected. The group of patients with an IMID other than SLE included patients with a diagnosis of (at least) one of the following diseases: antiphospholipid syndrome (APS), cutaneous lupus erythematosus (CLE), giant cell arteritis (GCA), juvenile idiopathic arthritis (JIA), lupus like disease (LLD, also known as "incomplete lupus or iSLE"), mixed connective tissue disease (MCTD), polymyositis (PM), dermatomyositis (DM), polymyalgia rheumatica (PMR), Sjögren's syndrome (SS), rheumatoid arthritis (RA), systemic sclerosis (SSc) and undifferentiated connective tissue disease (UCTD). Patients with SLE as well as another overlapping diagnosis were included in the SLE group. It is noteworthy that the text mining algorithm assessed the clinical diagnosis recorded by the physician in the EHR, therefore patients with a diagnosis assigned by the text mining algorithm do not necessarily meet the corresponding classification criteria.

In order to create a control group with low clinical suspicion of SLE to function as a negative control group for the development of classification models, we split the group of patients without an IMID diagnosis into two groups: Patients with no more than two symptoms and no arthritis, nephritis or pleuritis were classified as the Non-IMID group, all other patients were classified as the Rest group. Additionally, the BBD were regarded as a separate group.

#### 2.1.3. Pre-SLE patients

To identify samples of SLE patients obtained before the diagnosis of SLE was made (pre-SLE), three different strategies were used: First, if patients had multiple samples included in the biobank and only received the diagnosis of SLE in later samples, the first available sample of this patient was recorded as pre-SLE. Second, the text mining algorithm was adapted to also include the clinical data from the electronic health records for up to 2 years after the last sampling date, to assess whether a diagnosis had been made in this period. Third, treating physicians were asked if they could identify patients who had been in their care for a longer period of time but only later on received the diagnosis of SLE. The EHR of all patients identified by these methods were manually reviewed to verify that these samples were obtained before the clinical diagnosis of SLE was made.

### 2.2. Serological data

We used the first available sample of all included patients to evaluate the serological profile of the patients cross-sectionally. We tested serum samples for the presence of 57 IgG autoantibodies, using a custom-made chip-based antigen microarray (Thermo Fisher Scientific, Uppsala Sweden). A list of all included autoantibodies is included in supplementary file 1. All antigens were coated in triplicate on the microarray. The presence of autoantibodies was measured with immunofluorescence after 2 h of incubation with the blood sample. A detailed description of the methods used for this analysis has been described previously [14].

### 2.3. Statistical analysis

#### 2.3.1. Visualization of data

We used Uniform Manifold Approximation and Projection (UMAP) as a method for dimension reduction to better visualize the serological data [15]. Patients were labeled according to the patient groups as described above.

#### 2.3.2. Multivariable prediction models

Using the data from the antibody microarray, we developed prediction models to distinguish samples of patients with SLE from samples of patients of two different control groups; the Non-IMID group and the BBD group. All models were implemented in python (version 3.7.5.) using the statsmodels and scikit-learn modules [16,17].

We analyzed the predictive value of the autoantibodies measured on the microarray with a penalized (Lasso) logistic regression model. We used Receiver Operating Characteristic (ROC) curves to calculate the area under the curve (AUC) to evaluate performance of these models. The Lasso model is a logistic regression model with an added penalization term ($\lambda$) which constrains the size of the estimated coefficients and performs variable selection [18]. As the strength of penalization increases, more and more coefficients will be pushed to zero, resulting in a more parsimonious model. The models were compared to a univariable logistic regression model using only anti-dsDNA.

#### 2.3.3. Lasso model

We used a Box-Cox transformation to normalize the data. Although the image processing software of the laser scanner used to assess the levels of immunofluorescent activity of all autoantibodies coated on the microarray compensates for background signal intensity, mean and median levels of some antigen spots for all patients were markedly higher than spots of other antigens, even in samples that had relatively low levels of immunofluorescence for specific antigens, compared to other samples. To limit the influence this has on the logistic regression model, all values after Box-

Cox transformation were transformed to the Z-score for each antigen. We estimated the optimal penalization strength λ through cross-validation (see: "Model evaluation" below). A grid search was performed for 100 different values of λ, ranging from $\lambda_{max}$, i.e. the smallest value of lambda where all regression coefficients are set to 0, to 1 thousandth of $\lambda_{max}$ [19]. The λ selected for the final model corresponded to the value of λ for the best performing model (in terms of cross-validated AUC), plus one standard error, such that the final model would retain the fewest variables without compromising performance [20].

#### 2.3.4. Model evaluation

To estimate the performance of the models, we internally validated them using 5-fold cross-validation, repeated 5 times. Folds were stratified such that the relative frequency of SLE patients was approximately equal in each fold. All preprocessing steps (Box-cox transformation and Z-scoring) were repeated in each fold to prevent information leakage between the training and testing sets.

The resulting models were re-fitted on all the available data that was used during cross-validation and subsequently used to classify the pre-SLE samples as either SLE or non-SLE. Patients from the Rest and IMID group were used as negative controls to evaluate specificity with ROC-curves, since these patients were not used in the creation of the models. We experimented with different cut-offs for the models, as the preferential cut-off for the model is dependent on the intended use.

## 3. Results

### 3.1. Patients

We included serum samples of 1519 patients from the biobank, as well as 361 serum samples from heatlhy blood bank donors. At the time of sample collection, 483 patients were diagnosed with SLE, and 346 patients were diagnosed with another IMID. Serum samples from 690 patients without an IMID diagnosis were included, 218 of these patients fulfilled the criteria of the non-IMID control group, the remaining 472 patients were categorized as "Rest". Demographics for all patient groups are shown in Table 1.

### 3.2. Dimension reduction

The antibody profile of the different patient groups was visualized by reducing the full 57-dimensional representation to two dimensions using UMAP (Fig. 1). Patients with SLE appear to cluster at the top of the Figure, however, this cluster cannot be separated from the IMID and non-IMID control groups. The BBD patients appear to cluster on the bottom of the Figure, suggesting that the serological profile of this group differs from patients with SLE or other IMIDs, as well as from undiagnosed patients that are suspected of having such diseases. The pre-SLE patients are spread out across the Figure and do not appear to cluster together.

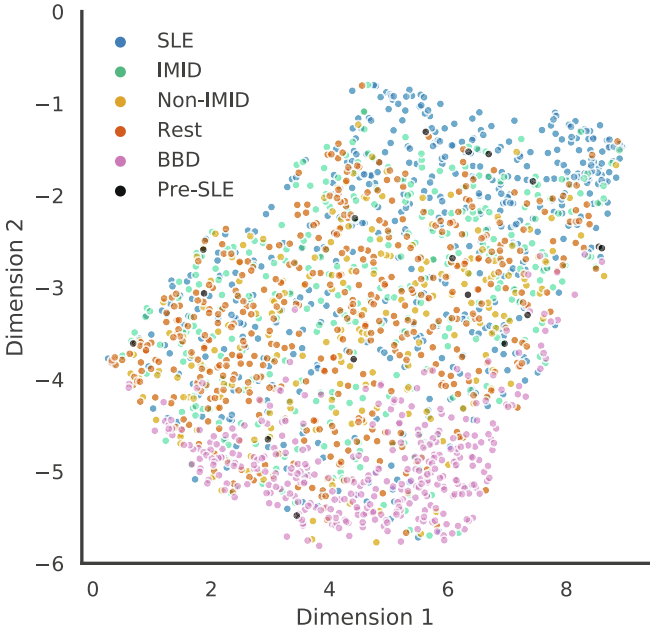


**Fig. 1.** Uniform Manifold Approximation and Projection (UMAP) dimension reduction UMAP dimension reduction plot based on all 57 included autoantigens. All patient groups as well as the pre-SLE samples are visualized by different colors. The SLE patients appear to cluster slightly around the top part of the figure and the BBD samples cluster around the bottom part of the figure. All other patient groups do not appear to form clear clusters.

### 3.3. Development of prediction models

#### 3.3.1. Non-IMID control group

A logistic regression model using only anti-dsDNA to predict patients as either SLE patients or non-IMID controls resulted in an AUC of 0.80. An unpenalized multivariable logistic regression model including all 57 parameters resulted in an AUC of 0.83. The Lasso model resulted in a maximum AUC of 0.84, with 42 antibodies included. The AUC of the final model using a higher value for λ (within 1 standard error of the λ for the maximum AUC) changed only maginally (AUC = 0.83, Fig. 2a). This model retained 29 non-zero coefficients. Antibodies against dsDNA, SmBB', Histones, CpG, and Ro60 were most strongly associated with increased odds of having SLE, the strongest negative association was with anti-GBM (anti-glomerular basement membrane) (suppl. Fig. 1a).

#### 3.3.2. BBD control group

Using the BBDs as the control group, the logistic model using only anti-dsDNA to predict patients as either SLE patients or BBD controls, resulted in an AUC of 0.80. Interestingly, the best AUC for a single antibody was achieved using anti-Ro60 (AUC: 0.90). The maximum AUC of the Lasso model was 0.98. Similar to the model using the non-IMID control group, the AUC was only reduced marginally with stronger penalization (remaining 0.98, Fig. 2b). Fifteen antibodies had non-zero coefficients in this final

**Table 1**
Baseline characteristics.

| Group | Samples analyzed | Female | Mean age ± SD | Anti-dsDNA positive |
|---|---|---|---|---|
| | | | | 170 (35%) |
| | | | | 26 (8%) |
| Non-IMID | 218 | 111 (51%) | 43,9 ± 17,4 | 7 (3%) |
| Rest | 472 | 248 (53%) | 46,0 ± 17,3 | 15 (3%) |
| BBD | 361 | 180 (50%)* | 46,7 ± 14,1 | X |

* Sex and age of two blood bank donors were not recorded.

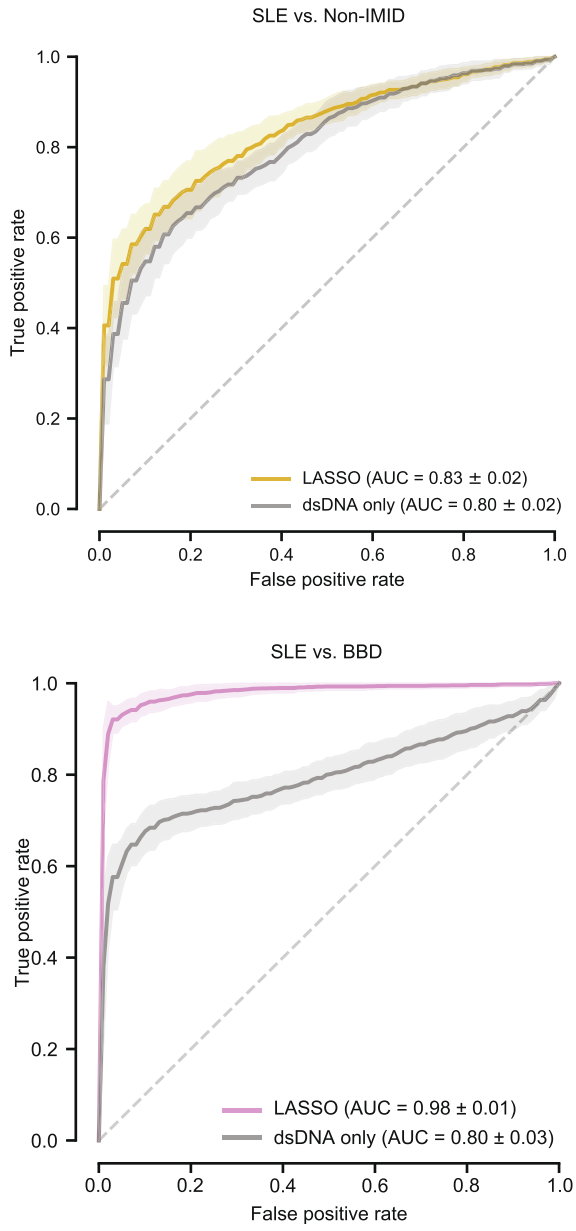


**Fig. 2.** Receiver Operating Characteristic (ROC) curves a: ROC-curve of the Lasso model for identification of SLE samples using the non-IMID patients as a negative control group. This model retained 29 non-zero coefficients. The area under the curve (AUC) is 0.83 (SD 0.02). A model using only anti-dsDNA as a coefficient resulted in an AUC of 0.80 (SD 0.02). b: ROC-curve of the Lasso model for identification of SLE samples using the BBD as a negative control group. This model retained 15 non-zero coefficients. The area under the curve (AUC) is 0.98 (SD 0.01). A model using only anti-dsDNA as a coefficient resulted in an AUC of 0.80 (SD 0.03).

model. Anti-dsDNA and Ro60 were the strongest contributing variables for categorization of SLE in this model. Interestingly, here anti-histone antibodies were negatively associated with SLE (suppl. Fig. 1b).

### *3.3.3. Prediction of the development of SLE*

Fifty potential pre-SLE patients were identified with a follow-up of up to 6 years. Manual review of the EHR showed that 23 patients were not diagnosed with SLE, ten patients were diagnosed with SLE at the time of sample collection. A total of 17 pre-SLE samples were identified and confirmed by manual review to be collected before the diagnosis of SLE was made. Ten patients had no previous IMID diagnosis, seven patients had another IMID diagnosis before the diagnosis SLE was made. The time between sample date and time of diagnosis varied between 3 weeks and 33 months (mean 7.2 months). Six patients were diagnosed with SLE within 2 months of sample collection. All of the pre-SLE patients tested for ANA (antinuclear antibodies) were positive or weak positive for ANA but only 35% were positive for anti-dsDNA in regular laboratory testing at the moment of sample collection (Table 2). In patients with SLE, 87% of patients were positive or weak positive for ANA and 39% for anti-dsDNA at the moment of sample collection. It is noteworthy that this is a cross-sectional measurement and does not necessarily mean that these patients have never tested positive for ANA or anti-dsDNA.

The SLE vs non-IMID Lasso model performed best for the prediction of pre-SLE (AUC 0.83). The model using the BBD group as controls had an AUC of 0.57. To evaluate the specificity of the non-IMID model, we used the samples from the "Rest" group as negative controls, as they were not used in the development of the model. In this analysis, ten patients were excluded from the Rest group, as they were identified as pre-SLE patients. When a relatively low cut-off was chosen (threshold score of 0.50), this model identified 88% of future SLE patients (Table 3, Fig. 3). However, although this model had a negative predictive value (NPV) of 99% this model had a positive predictive value (PPV) of only 6%, limiting its potential clinical applicability. However, the percentage patients other than pre-SLE patients predicted as pre-SLE, dropped drastically when the threshold was increased from 0.50 to 0.84. This stricter cut-off with a threshold score of 0.84resulted in a sensitivity of 53%, with a PPV of 25% and a NPV of 98%.

## 4. Discussion

Early identification of patients with SLE is important since prolonged disease activity due to inadequate treatment leads to organ damage [1]. Because of their early manifestation in the course of the disease, tests for autoantibodies are very well suited to aid the diagnostic process. In our study, we created prediction models based on the antibody profile of either patients with SLE, or blood bank donors or patients with a limited number of SLE related symptoms (non-IMID group) respectively. We subsequently showed that these models, when applied to a different group of (non-SLE) patients, can be used to predict which patients will develop SLE in the future. This finding could be used to expand the currently limited opportunities for early diagnostics in SLE and to identify patients at high risk of development of SLE in the future.

In our study, we investigated different cut-offs for the prediction of pre-SLE patients, ranging from a model with high sensitivity but low specificity to a model with high specificity but low sensitivity. The preferential cut-off for the model is dependent on the intended use, as the model with high sensitivity can best be used to identify patients who are at low risk of development of SLE (with an NPV of 99%). Although the model with higher specificity has a

**Table 2**
Baseline characteristics of the pre-SLE patients.

| Pre-SLE characteristics | |
|---|---|
| Female | 12/17 (70.5%) |
| Age (mean) | 39.6 (range 14–67) |
| Time between sampling and diagnosis | 7.2 months (range 1–33) |
| Diagnosis before SLE | 2 LLD |
| | 2 APS |
| | 1 CLE |
| | 1 RA |
| | 1 pSS |
| | 10 no diagnosis |
| Anti-dsDNA positive | 6/17 (35%) |

**Table 3**
Identification of future SLE patients.

| Group (n) | Positive with threshold score 0.50N (%) | Positive with threshold score 0.84N (%) |
|---|---|---|
| Pre-SLE (17) | 15 (88%) | 9 (53%) |
| IMID (346) | 216 (62%) | 45 (13%) |
| LLD (28) | 20 (71%) | 7 (25%) |
| Rest (462) | 227 (49%) | 27 (6%) |
| BBD (361) | 152 (42%) | 2 (1%) |

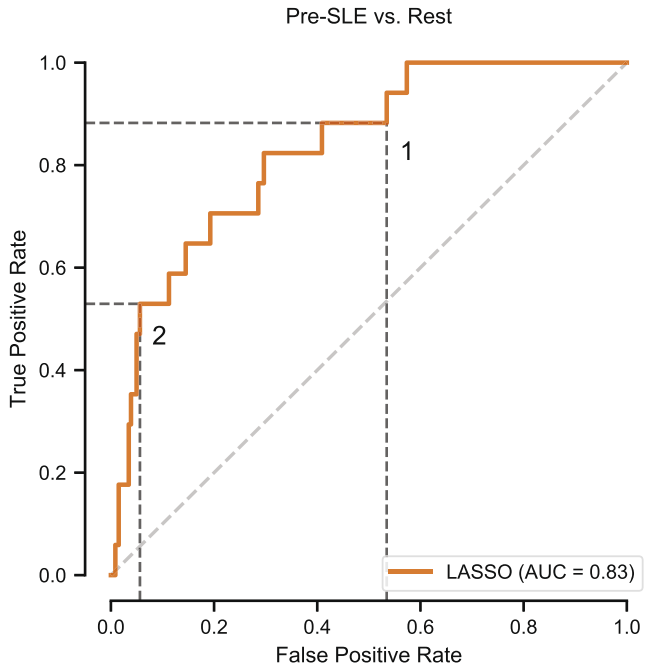


**Fig. 3.** ROC curve for identification of future SLE patients ROC curve of the non-IMID prediction model scores for the identification of pre-SLE patients, with patients from the rest-group as negative controls. This model includes 17 pre-SLE patients and 462 rest group patients, as 10 patients from the rest group were pre-SLE patients. The dotted lines resemble two different cut-offs, with a threshold score of 0.50 (1) and 0.84 (2). The results of these two different cut-offs are presented in Table 3.

limited sensitivity of 53%, this could potentially be used to determine patients at high risk of development of SLE who would remain in follow-up, whereas patients at low risk could be released from care. Implementation of this strategy would limit the number of patients in follow-up to approximately 10% of the previous total, out of which approximately one in four patients without an established IMID diagnosis will eventually develop SLE.

Previous studies investigating the prediction of the development of SLE are scarce. A recent review by Lambers et al. showed that 5–57% of patients with lupus like disease will eventually be diagnosed with SLE, although many patients will go into spontaneous remission [21]. They found that the presence of anti-dsDNA, anticardiolipin antibodies, and hypocomplementemia were associated with progression to SLE, although not strongly enough to accurately predict the development of SLE. In our study, only 2 out of 28 (7.1%) LLD patients progressed to SLE. However, in our study 15 out of the 17 (88.2%) pre-SLE patients did not previously receive the diagnosis LLD, with most patients having no previous diagnosis at all. This suggests that the detection of a unique profile of autoantibodies in pre-SLE patients in many cases precedes the clinical manifestations needed for the establishment of the diagnosis.

The wide variety of autoantibodies reported to be associated with SLE, in combination with the fact that autoantibodies are an early immunological manifestation of SLE, make autoantibodies well suited for a multivariable approach for diagnostic purposes. Although multivariable diagnostic models based on microarray data have been successful in different diseases, the use in the field of SLE is limited. Putterman e.a. used an antibody microarray with approximately 200 antigens to create multivariable models to differentiate between serum samples of SLE patients and healthy controls, resulting in a model with a sensitivity of 94% and specificity of 75% [22]. This results in a higher NPV than PPV, the authors, therefore, emphasize the excluding capabilities of the model, naming it the SLE-key® rule-out serological test. This study showed cross-sectional differences in the antibody profile of patients with SLE compared to healthy controls, however it did not assess the predictive value for patients who might develop SLE in the future. In a similar study, Fattal e.a. found that an SLE antibody profile is independent of disease activity or renal involvement [23]. Similar to this study, in our study, we did not find differences in antibody profiles between SLE patients with or without renal involvement (data not shown).

Interestingly, we saw some clear differences in results upon using the prediction models using the Non-IMID control group vs. the BBD control group, as well as differences in the relative contribution of specific autoantibodies to each respective model. Whereas autoantibodies known to be strongly associated with SLE (i.e., anti-dsDNA, anti-SmBB') were among the strongest contributing factors for the non-IMID model, in the BBD-model anti-Ro60, and in a lesser extent anti-La, were among the top contributing variables. These differences can potentially be explained by some level of autoimmunity in the non-IMID group, although not enough to amount to an IMID diagnosis, as these are patients who visited the Rheumatology clinic, most likely to be evaluated for the presence of a potential autoimmune disease. There also is a possibility that differences in storage technique may also have influenced the results from the BBD group, as these were from another source than the samples of all other patients. There also is the potential for a selection bias as a request to test for anti-dsDNA was the entry criterion for the patient biobank used in this study.

Our study also shows that the performance of diagnostic prediction models relies heavily on including blood bank donors versus non-IMID patients as control groups in order to create and test the models. This is illustrated by the difference in AUC of the non-IMID and BBD prediction models (Fig. 2) and in the UMAP results, in which the non-IMID patients are spread out between patients with IMIDs, whereas the BBDs appear to form a separate cluster (Fig. 1). These differences between the different control groups are probably caused by some level of unclassified autoimmunity in the non-IMID group, resulting in higher levels of autoantibodies. Diagnostic models based on the comparison between patients with SLE and healthy controls may prove to be suboptimal in clinical practice because of these differences in autoantibody profiles. We argue that studies using true healthy controls are very well suited to evaluate serological abnormalities in patients with SLE, but that patients without a diagnosis from the Rheumatology clinic might be a better control group to select for the development of diagnostic prediction models, as this is the "negative" group in the clinical setting. This may also explain the slightly lower performance from the SLE-key® rule out test in a prospective real-world setting, compared to the original study setting [24].

The strengths of our study include the large number of included patients, including multiple relevant control groups, and the use of a custom-made antibody microarray measuring 57 autoantibodies.

The model can potentially be further optimized by inclusion of other autoantibodies, additional analysis of IgM autoantibodies and complement levels. Our study is limited by the relatively small number of pre-SLE patients. It is possible that, with longer follow-up, more patients in this study will develop SLE in the future and are not yet identified as pre-SLE patients.

Our study shows that patients who will develop SLE in the future already exhibit an SLE-like antibody profile and that models based on autoantibody profiles can help identify patients at high risk of development of SLE in a real-life setting. Patients at high risk could be kept in follow-up, whereas patients at low risk might be released from care. Further prospective research is needed to confirm the additional diagnostic value of this antibody profile.

## Contributors

HO, JMvL, ML, LM-A and JD contributed to the conceptualization and design of the study. SH extracted relevant information from the Electronic Health Records. TB and LR performed data analysis and data visualisation. All authors contributed to data interpretation. TB wrote the manuscript. All authors critically reviewed the manuscript.

## Funding

This work was supported by Thermo Fisher Scientific (Uppsala, Sweden).

## Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Henny Otten reports financial support was provided by Thermo Fisher Scientific ImmunoDiagnostics Division. Jorge Dias reports a relationship with Thermo Fisher Scientific ImmunoDiagnostics Division that includes: employment. Linda Mathsson-Alm reports a relationship with Thermo Fisher Scientific ImmunoDiagnostics Division that includes: employment. Jacob van Laar reports a relationship with Thermo Fisher Scientific Inc that includes: funding grants. Jacob van Laar reports a relationship with AstraZeneca that includes: funding grants. Jacob van Laar reports a relationship with MSD that includes: funding grants.

## Acknowledgments

We thank Ms R Melchers for technical support with the analysis of the antibody microarrays.

## Data availability statement

Data are available on reasonable request. The code for the analysis used in this study will be published in a public, open access repository.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.humimm.2022.03.010.

## References

[1] J. Nossent, E. Kiss, B. Rozman, G. Pokorny, P. Vlachoyiannopoulos, M. Olesinska, et al., Disease activity and damage accrual during the early disease course in a multinational inception cohort of patients with systemic lupus erythematosus, Lupus 19 (8) (2010) 949–956.
[2] M.W.P. Tsang-A-Sjoe, I.E.M. Bultink, M. Heslinga, A.E. Voskuyl, Both prolonged remission and Lupus Low Disease Activity State are associated with reduced damage accrual in systemic lupus erythematosus, Rheumatology (Oxford) 56 (1) (2017) 121–128.
[3] A. Doria, M. Zen, M. Canova, S. Bettio, N. Bassi, L. Nalotto, et al. SLE diagnosis and treatment: When early is early. Autoimmun. Rev. 10(1) (2010) 55–60. Available from: http://dx.doi.org/10.1016/j.autrev.2010.08.014.
[4] C. Morgan, A.R. Bland, C. Maker, J. Dunnage, I.N. Bruce, Individuals living with lupus: findings from the LUPUS UK Members Survey 2014, Lupus 27 (4) (2018) 681–687.
[5] M. Aringer, K. Costenbader, D. Daikh, R. Brinks, M. Mosca, R. Ramsey-Goldman, et al., 2019 European League Against Rheumatism/American College of Rheumatology classification criteria for systemic lupus erythematosus, Ann. Rheum. Dis. 78 (9) (2019) 1151–1159.
[6] K.H. Costenbader, P.H. Schur, We need better classification and terminology for "people at high risk of or in the process of developing lupus", Arthritis Care Res. 67 (5) (2015) 593–596.
[7] M. Mosca, K.H. Costenbader, S.R. Johnson, V. Lorenzoni, G.D. Sebastiani, B.F. Hoyer, et al., Brief Report: How do patients with newly diagnosed systemic lupus erythematosus present? A multicenter cohort of early systemic lupus erythematosus to inform the development of new classification criteria, Arthritis Rheumatol. 71 (1) (2019) 91–98.
[8] K. Didier, L. Bolko, D. Giusti, S. Toquet, A. Robbins, F. Antonicelli, et al., Autoantibodies associated with connective tissue diseases: What meaning for clinicians?, Front Immunol. 9 (2018) 1–20.
[9] C. Eriksson, H. Kokkonen, M. Johansson, G. Hallmans, G. Wadell, S. Rantapää-Dahlqvist, Autoantibodies predate the onset of systemic lupus erythematosus in northern Sweden, Arthritis Res. Ther. 13 (1) (2011) 1–8.
[10] M.R. Arbuckle, M.T. McClain, M.V. Rubertone, R.H. Scofield, G.J. Dennis, J.A. James, et al., Development of autoantibodies before the clinical onset of Systemic Lupus Erythematosus, N. Engl. J. Med. 349 (16) (2003) 1526–1533.
[11] M.M.J. Nielen, D. Van Schaardenburg, H.W. Reesink, R.J. Van De Stadt, I.E. Van Der Horst-Bruinsma, M.H.M.T. De Koning, et al., Specific autoantibodies precede the symptoms of rheumatoid arthritis: A study of serial measurements in blood donors, Arthritis Rheum. 50 (2) (2004) 380–386.
[12] M.J. Ten Berg, A. Huisman, P.M.L.A. Van Den Bemt, A.F.A.M. Schobben, A.C.G. Egberts, W.W. Van Solinge, Linking laboratory and medication data: New opportunities for pharmacoepidemiological research, Clin. Chem. Lab Med. 45 (1) (2007) 13–19.
[13] T.E. Brunekreef, H.G. Otten, S.C. Bosch, I.E. Hoefer, J.M. Laar, M. Limper, et al., Text mining of electronic health records can accurately identify and characterize patients with systemic lupus erythematosus, ACR Open Rheumatol. (2021) 1–7.
[14] T. Brunekreef, M. Limper, R. Melchers, L. Mathsson-Alm, J. Dias, I. Hoefer, et al., Microarray testing in patients with systemic lupus erythematosus identifies a high prevalence of CpG DNA-binding antibodies, Lupus Sci. Med. [Internet] 8 (1) (2021). http://www.ncbi.nlm.nih.gov/pubmed/34725184.
[15] 15. McInnes L, Healy J, Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv [Internet]. (2018 Feb 9); Available from: http://arxiv.org/abs/1802.03426.
[16] S. Seabold, J. Perktold, Statsmodels: econometric and statistical modeling with python, in: Proceedings of the 9th Python in Science Conference [Internet]. (2010) p. 92–6. Available from: https://conference.scipy.org/proceedings/scipy2010/seabold.html.
[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., Preface, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[18] R. Tibshirani, Regression shrinkage and selection via the Lasso, J. R. Stat. Soc. Ser. B 58 (1) (1996) 267–288.
[19] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent Available from J. Stat. Softw. [Internet] 33 (1) (2010) 1–22. http://www.jstatsoft.org/v33/i01/.
[20] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning [Internet]. New York, NY: Springer New York; (2009). (Springer Series in Statistics). Available from: http://link.springer.com/10.1007/978-1-4419-9863-7_941.
[21] W.M. Lambers, J. Westra, M.F. Jonkman, H. Bootsma, K. de Leeuw, Incomplete systemic lupus erythematosus: what remains after application of American College of Rheumatology and Systemic Lupus International Collaborating Clinics criteria?, Arthritis Care Res 72 (5) (2020) 607–614.
[22] C. Putterman, A. Wu, A. Reiner-Benaim, D.S. Batty, I. Sanz, J. Oates, et al., SLE-key® rule-out serologic test for excluding the diagnosis of systemic lupus erythematosus: Developing the ImmunArray iCHIP® Available from J. Immunol. Methods [Internet] 429 (2016) 1–6. https://linkinghub.elsevier.com/retrieve/pii/S0022175915300697.
[23] I. Fattal, N. Shental, D. Mevorach, J.-M. Anaya, A. Livneh, P. Langevitz, et al., An antibody profile of systemic lupus erythematosus detected by antigen microarray, Immunology 130 (3) (2010) 337–343.
[24] M. Radin, S. Sciascia, I. Cecchi, G. Mengozzi, D. Roccatello, The potential role of SLE-key test in identifying patients with Systemic Lupus Erythematosus: Results from a prospective, real-world experience, Eur. J. Intern. Med. 68 (July) (2019) e12–e14.