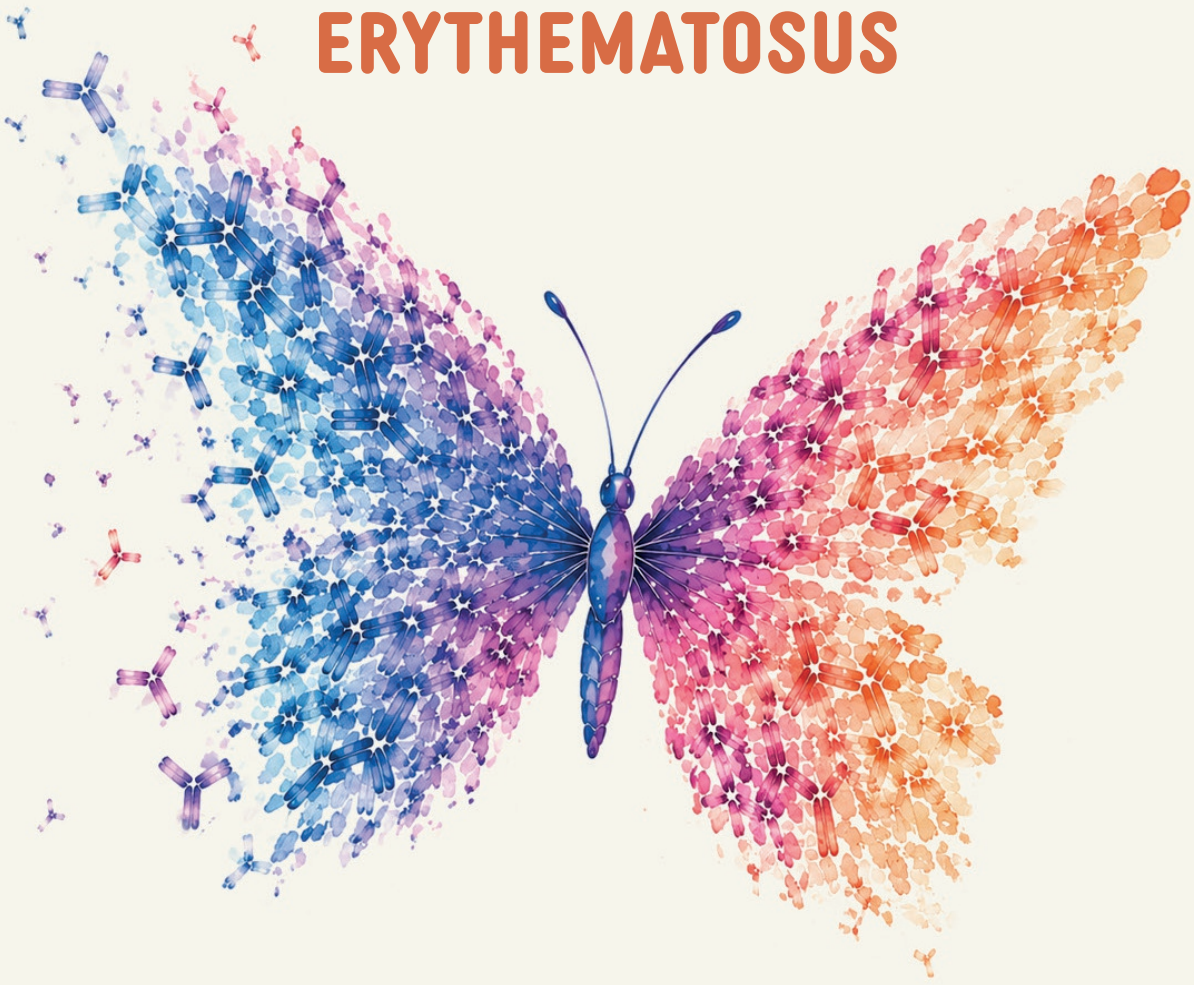
EARLY DIAGNOSTICS AND CLASSIFICATION IN PATIENTS WITH SYSTEMIC LUPUS ERYTHEMATOSUS



TAMMO BRUNEKREEF

STELLINGEN

behorende bij het proefschrift

Early Diagnostics and Classification in Patients with Systemic Lupus Erythematosus

1. The wide variety of clinical manifestations can make SLE very challenging to diagnose, often resulting in a significant delay between the constitution of the first SLE-related symptoms and the eventual diagnosis – *This thesis*
2. Researchers and healthcare professionals should work together to find a way for recording accurate data that are reusable in research, while limiting the associated burden of administration on healthcare professionals – *This thesis*
3. Patients who will develop SLE in the future already exhibit an SLElike antibody profile and models based on autoantibody profiles can help identify patients at high risk of development of SLE in a real-life setting – *This thesis*
4. Patients with symptoms associated with CTDs but who remain undiagnosed after the initial diagnostic process and do not have specific ANAs associated with their suspected CTD can safely be released from care – *This thesis*
5. Classification of subgroups of patients with SLE based on their clinical manifestations should be abandoned in favour of classification based on their immunological manifestations for the purpose of treatment selection – *This thesis*
6. DON'T PANIC – *Douglas Adams*
7. We live in a society exquisitely dependent on science and technology, in which hardly anyone knows anything about science and technology – *Carl Sagan*
8. Outside of a dog, a book is man's best friend. Inside of a dog, it's too dark to read – *Groucho Marx*
9. The path of the good ancestor lies before us. It is our choice whether or not to take it – *Roman Krznaric*
10. It's never lupus – *Hugh Laurie, House*

Tammo Brunekreef
3 november 2026

Early Diagnostics and Classification in Patients with Systemic Lupus Erythematosus

Tammo Brunekreef

ISBN: 978-94-6534-817-9
Cover design: Stefanie van den Herik | HerikMedia.nl
Lay-out & Print: ProefschriftMaken | proefschriftmaken.nl

© 2026 Tammo Brunekreef
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronically, mechanically, by photocopying, recording or otherwise, without the prior written permission of the author.

Early Diagnostics and Classification in Patients with Systemic Lupus Erythematosus

Vroege diagnose en classificatie bij patiënten met Systemische Lupus Erythematosus
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. ir. W. Hazeleger,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op

dinsdag 3 november 2026 des middags te 2.15 uur

door

Tammo Ernst Brunekreef

geboren op 25 juli 1993
te Utrecht

Promotor:

Prof. dr. J.M. van Laar

Copromotoren:

Dr. H.G. Otten
Dr. M. Limper

Beoordelingscommissie:

Prof. dr. D. Hamann (voorzitter)
Prof. dr. H.L. Leavis
Dr. S.C. Mastbergen
Prof. dr. B. Oldenburg
Prof. dr. A.E. Voskuyl

TABLE OF CONTENTS

Chapter 1	General introduction and thesis outline	7
Part 1	**Use of data from regular care for research**	21
Chapter 2	Text Mining of Electronic Health Records Can Accurately Identify and Characterize Patients with Systemic Lupus Erythematosus	23
Part 2	**Autoantibodies as targets for early diagnostics**	43
Chapter 3	Microarray testing in patients with Systemic Lupus Erythematosus identifies a high prevalence of CpG DNA-binding antibodies	45
Chapter 4	Microarray analysis of autoantibodies can identify future Systemic Lupus Erythematosus patients	71
Chapter 5	The Predictive Value of Antinuclear Antibodies in Patients with Suspected Connective Tissue Disease for the Development of Connective Tissue Diseases	91
Chapter 6	Association of autoantibodies with the IFN signature and NETosis in patients with Systemic Lupus Erythematosus	105
Chapter 7	Recent Developments in the use of Biomarkers as Diagnostic Tools in Patients with Systemic Lupus Erythematosus; A Narrative Review	125
Chapter 8	General Discussion	151
Nederlandse samenvatting		168
English Summary		171
Acknowledgements (Dankwoord)		174
About the author		178

1

General introduction and thesis outline

GENERAL INTRODUCTION

Systemic lupus erythematosus

Systemic lupus erythematosus (SLE) is a rare systemic autoimmune disease. The estimated prevalence of SLE worldwide is approximately 43.7 per 100.000 persons.[1] Approximately 90% of all patients with SLE are female, and the disease frequently debuts at a relatively young age between 20 and 50 years old.[2,3]

SLE is associated with a wide range of clinical symptoms, ranging from relatively mild symptoms and abnormalities in laboratory results, to potentially life-threatening symptoms.[4] For example, Patient A might exhibit a relatively mild course of the disease, with a characteristic UV-sensitive butterfly rash, fatigue and arthralgia as main symptoms of the disease, whereas in Patient B, SLE might cause inflammation of the internal organs, resulting in potentially life-threatening complications of the disease such as nephritis or pericarditis.[5,6] And in patient C, the disease might be characterized by neuropsychiatric symptoms, such as psychosis and severe headaches, as well as thrombosis and complications during pregnancy due to the closely related antiphospholipid syndrome (APS).[7] Despite experiencing vastly different symptoms, the symptoms of all three of these patients can be clinical manifestations of SLE.

The diagnosis of SLE does not rely on the presence of one single clinical or immunological manifestation, but rather on the simultaneous or consecutive presence of a number of clinical and immunological manifestations out of a set of associated manifestations. SLE could therefore be considered a syndromal diagnosis, without a single communal pathophysiological mechanism that can be used to diagnose SLE. In contrast, the diagnosis of SLE is made when patients exhibit a combination of symptoms and immunological manifestations.[4]

Pathophysiology of SLE

Rather than having a single communal pathophysiological pathway, SLE is considered to have a complex multifactorial pathophysiology, which includes immunological, hormonal, and genetic components.[8]

The process of apoptosis has been suggested to be one of the core drivers of the immunological pathophysiology of SLE.[9] Apoptosis is the process of regulated programmed cell death, in which damaged or redundant somatic cells are processed. In normal situations, this happens in an organized manner, in which components of the dying cell are broken down and distributed in small apoptotic blebs. These blebs can then be phagocytosed and cleared by macrophages.[10,11] This is in contrast to the unorganized process of necrosis, the process in which cell death occurs due to acute damage to cells, resulting in

rupturing of the cells, and spilling of unmodified intracellular components, causing local inflammation.[11]

In SLE, both the process of apoptosis itself and the clearing of apoptotic blebs can be impaired.[10] Disorganized apoptosis, as well as ineffective clearing of apoptotic material, can result in prolonged or increased exposure of intracellular content to the immune system. This exposure to autoantigens, which are usually confined intracellularly, can result in several immunological manifestations, including the production of autoantibodies, activation of the innate immune system and the production of pro-inflammatory cytokines (Figure 1). Furthermore, incomplete or delayed clearing of apoptotic material can cause secondary necrosis, additionally resulting in local inflammation.[12]

Autoantibodies are nearly omnipresent in patients with SLE and are also part of the classification criteria for SLE.[4] These autoantibodies often target antigens from the nucleus of somatic cells, such as double stranded DNA (anti-dsDNA) and ribonucleoproteins (anti-RNP and anti-Sm).[13,14] These antigens are typically located intracellularly, and therefore out of reach of the immune system. However, in the process of disorganized apoptosis or necrosis, these autoantigens are expelled in the extracellular space where they can be phagocytosed by dendritic cells, which in turn can stimulate B-cells to produce autoantibodies against autoantigens. These autoantibodies can react with their antigen, resulting in the formation of immune complexes. These immune complexes can cause direct organ damage by accumulating in small vessels, causing local inflammation and can also stimulate the production of additional pro-inflammatory cytokines.[15,16]

One group of pro-inflammatory cytokines influenced by immune complexes are Type I interferons (IFN).[17] Type I interferons are primarily produced by plasmacytoid dendritic cells (pDCs) and are part of the normal innate immune response against viral infection, in which they cause activation of both the innate and adaptive immune system.[16,18,19] Overproduction of type I interferons causes a measurable upregulation of IFN related genes, commonly called an IFN signature.[20] This IFN signature can often be seen in patients with SLE, as well as in patients with other systemic autoimmune diseases, such as Sjögren's syndrome.[21] Previous research projects, in which patients were treated for malignancies with IFN-α, a type I IFN, showed that increased exposure to type I IFN can cause patients to develop a lupus-like disease with formation of autoantibodies to nuclear antigens, suggesting an important role in the pathophysiology of SLE.[22]

Another core immunological feature of SLE is complement activation.[23,24] The complement system consists of a group of proteins present in plasma and on cell surfaces. The complement system plays an important feature of the normal innate immune system by augmenting the effectiveness of the innate and adaptive immune system. Furthermore, it is also involved in the disposal of immune complexes and apoptotic debris by opsoni-

zation.[23] The formation of immune complexes and ineffective process of apoptosis in SLE can cause activation of the complement system, resulting in increased inflammation. As complement factors are present in the plasma of patients and therefore relatively easy to measure, complement factors are often used as a marker to evaluate disease activity.[23,25,26]

Similar to apoptosis, NETosis is a process of organised cell death.[27,28] During NETosis, neutrophil extracellular traps (NETs) are formed, by neutrophilic granulocytes extruding their nuclear material to form a web-like structure. Unlike during apoptosis, in which the intracellular components are expelled in organised blebs, during NETosis the intracellular components are expelled in a far less organised process, resulting in exposure of the immune system to antigens that are usually confined intracellularly.[29] NETs are considered a physiological manifestation, as these NETs trap and kill bacteria with the use of complement activation. However, patients with SLE have been shown to have increased NET formation and an impaired degradation of NETs, resulting in prolonged exposure of the nuclear material in these NETs.[29] These NETs both contain many of the antigens of SLE-related autoantibodies, but have also been shown to stimulate production of type I interferons.[30]

Genetic and hormonal factors are also considered to play a role in the pathophysiology of SLE. Although little single genetic variations are known to directly cause SLE, variants of a number of genes have been associated with SLE. Many of these associated genes code are involved in immunological functions, such as Fcγ receptors (i.e. FcGR2A–FcGR3A), interferon response (i.e. STAT4), and clearance of complement and immune complexes (i.e. ITGAM).[31] However, these genetic factors cannot fully explain the prevalence of SLE. For example, a comprehensive review by Block found that only approximately 40% of monozygotic twins with at least one case of SLE were clinically concordant, despite being genetically identical.[32] Although the prevalence in monozygotic twins was notably higher than in dizygotic twins (approximately 4%), this suggests that genetic factors influence the immune system resulting in an increased predisposition to develop SLE, although not directly causing it.[33]

Similarly, sex hormones have been shown to influence the function of the immune response.[34] Oestrogen stimulates both the innate and the adaptive immune system in a multitude of ways. For example, oestrogen increases TLR7 and TLR9 signalling in dendritic cells, resulting in type 1 IFN production, and promotes antibody secretion in B-cells.[35,36] Although not directly causing the symptoms of SLE, they are thought to lead to an increased predisposition of factors contributing to the development of SLE, potentially explaining the increased prevalence of SLE in females compared to males.

Although many of these immunological manifestations are associated with SLE, they do not fully explain the variability in symptoms and immunological manifestations in patients with SLE. Patients with a similar course of the disease can have a different immunological phenotype, whereas patients with similar immunological phenotypes can have different symptoms.[37,38] The lack of a single communal pathophysiology raises the question whether SLE should be considered one single disease with different clinical phenotypes and a shared multifactorial pathophysiology, or rather a collection of different sub-diseases with a different immunological pathophysiology, resulting in similar phenotypes, which currently are all collected under the banner of SLE.[39]

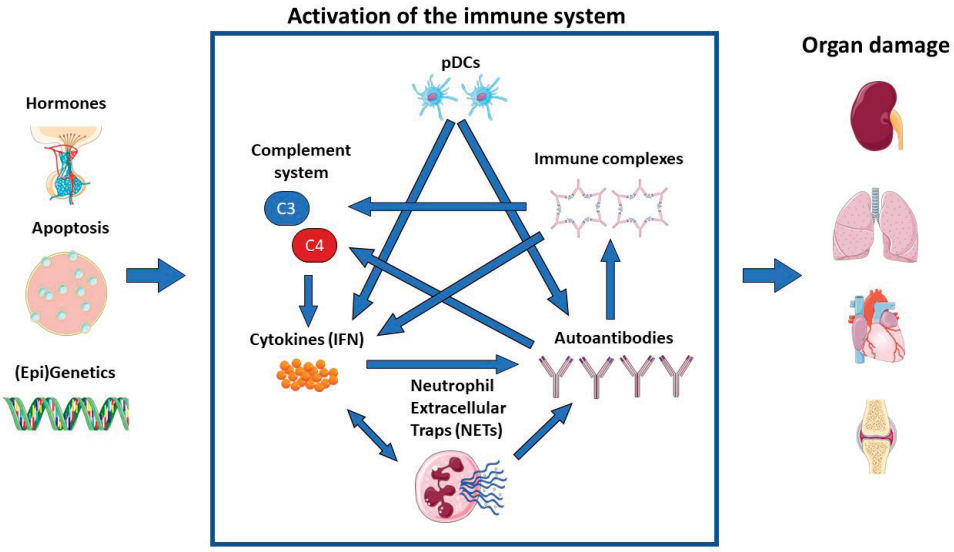


Figure 1: A simplified overview of the pathophysiology of SLE

Before the diagnosis of SLE

The wide variety of clinical manifestations can make SLE very challenging to diagnose, often resulting in a significant delay between the constitution of the first SLE-related symptoms and the eventual diagnosis. This delay commonly ranges from several months to multiple years.[5] The delay can partially be attributed to the initial presentation of the disease, as non-specific symptoms are usually the symptoms which are constituted first.[40,41] These symptoms can also be caused by a wide range of diseases, therefore clinical suspicion for a rare disease such as SLE is often low, as there are far more prevalent causes for these symptoms, such as viral infections. SLE is often only diagnosed after additional, more SLE-specific symptoms constitute. This is unfortunate, as this delay not only results in a higher burden of disease and increased uncertainty for patients, but a longer time to diagnosis is also associated with the accrual of potentially irreversible damage.[42]

After the diagnosis of SLE

For some patients with SLE, the disease is characterised by infrequent flares with relatively mild symptoms, and these patients usually achieve quiescent disease activity with relatively minimal medication. However for other patients, their disease is characterised by frequent flares with possibly life-threatening symptoms, such as nephritis or pericarditis.[4,43] Clinical tools to predict the course of disease for a newly diagnosed SLE patient are severely limited, resulting in great uncertainty about the future for these patients.[44,45] Furthermore, the inability to identify patients with severe disease in an early stage might result in undertreatment of patients because of late recognition of these symptoms, ultimately resulting in organ damage.[42]

Similar to the inability to recognize the course of the disease, there also is a need for means to predict which patients react well to different types of treatment.[46,47] Because of the complex pathophysiology of SLE, there are also different types of treatment, targeting different aspects of the immune system.[48] For example, corticosteroids are used as general immunosuppressants, hydroxychloroquine and anifrolumab target the interferon-related pathways, and rituximab and belimumab are mainly targeting B-lymphocytes.[49–52] Identification of different immunological pathophysiologic pathways in patients with SLE would help to identify patients which would respond well to certain types of targeted therapy, such as anti-CD20 or anti-interferon recombinant antibodies.

Targets for (early) diagnostics and prediction models

Insight in the complex pathophysiology of SLE can help to meet this need for better diagnostic tools, prediction of the course of the disease and response to therapy. Out of all immunological manifestations, autoantibodies are considered the most strongly associated immunological manifestations in SLE. Anti-dsDNA antibodies are commonly detected in the serum of patients with SLE and many more autoantibodies have been associated with SLE in previous studies.[14,53] Some antibodies are known to often be present years before diagnosis, suggesting potential as a prime target for early diagnostic tools.[54] Some antibodies also have a clear association with specific symptoms, suggesting possible use for the prediction of the symptoms and course of the disease. For example, anti-C1q antibodies are strongly associated with renal flares and anti-ribosomal P protein antibodies are associated with neuropsychiatric symptoms.[45,55] However, in practice individual antibodies lack the sensitivity and specificity to accurately establish the diagnosis of SLE or to predict the course of the disease for individual patients. The diagnostic potential of autoantibodies might be utilized more effectively by evaluating patterns of autoantibodies, rather than individual antibodies, and expanding prediction models to include other immunological manifestations as well, such as the IFN signature and NETs.

Challenges with scientific research

SLE is a rare disease, therefore performing research can be challenging, as large groups of study participants are often needed to answer research questions. Due to the rarity of the disease, studies are commonly performed with small groups, limiting the strength of the evidence of their findings.[56] Additionally, clinical trials often employ stringent methodological criteria for inclusion and exclusion. One of the main benefits of this is limiting the number of possible confounders, making the resulting data easier to interpret. However, the resulting curated study population does not always accurately represent the entire population of patients with SLE, potentially resulting in limited generalizability of the findings to the real world population.[57]

Besides data from clinical trials, data recorded in regular care can also serve as an important source of data for research, primarily in retrospective studies using patient chart review as the means for data collection. Although this method of data collection is much cheaper than clinical trials, and better represent the real-world population, it also has some drawbacks. Studies using data collected in regular care are more prone to biases, confounders and missing information, hampering the interpretability of the results.[58]

As the data collected in regular care is primarily recorded by physicians with the aim of providing the best care for patients, they are commonly not structured for research purposes, making it difficult to extract the relevant data for research.[59] This data extraction by patient chart review can be very time-consuming, especially when considering large groups of patients. However, in the last several years, technological developments have made it possible to automate part of the process of data collection from the health care records.[60,61] This has the potential to greatly reduce the time needed for the data collection, as well as eliminate inter- and intra-observer variability, further facilitating the use of regular care data for research.

Outline and goal of this thesis

SLE is a heterogeneous disease, both in heterogeneity of symptoms and of immunological manifestations. Due to the large number of associated symptoms, there often is a significant delay between the constitution of symptoms and the eventual diagnosis. The course of the disease can vary from relatively infrequent flares with mild symptoms to frequent flares with possibly life-threatening complications, requiring intensive treatment. There is a need for reliable early diagnostic tools, in order to limit the diagnostic delay and potential for accrual of organ damage during the period of undertreatment. Similarly, there is a need to identify predictors for the course of the disease, as well as predictors for response to targeted therapy in order to identify and adequately treat patient with a severe course of the disease.

Previous research has shown that several immunological manifestations, such as autoantibodies, have great potential for early diagnostic purposes. The goal of this thesis is to expand the field of early diagnostics in SLE. It focuses on expanding the arsenal of tools to collect and use data from regular care to facilitate research projects for rare diseases, and explores the use of autoantibodies and other immunological manifestations as part of prediction models for the development of SLE.

Part 1 – Use of data from regular care for research

In a research field where large-scale studies are scarce due to the rarity of the studied disease, it is important to utilize the real-life data from regular care to their full potential. The current method of manual patient chart review, which is seen as the golden standard for this end, has several drawbacks. In the first part of this thesis, we explored a novel method of extracting relevant information from the data recorded in the Electronic Health Records (EHR) during regular care. In **Chapter 2**, we present a text mining algorithm, designed to extract relevant information from the EHR, that can be used as an alternative for manual patient chart review.

Part 2 – Autoantibodies as targets for early diagnostics

Previous studies have shown a wide variety of autoantibodies to be associated with SLE. As some of these autoantibodies can be present in the blood of patients, even before the first symptoms constitute, these are prime targets for the development of early diagnostic tools. In **Chapter 3**, we show that, amongst well-known autoantibodies such as anti-dsDNA and anti-Sm, anti-CpG antibodies are prevalent in a large number of patients with SLE, potentially revealing a new target for diagnostic tools. In **Chapter 4**, we present a model based on profiles of multiple autoantibodies that can be used to identify patients with SLE, even before diagnosis.

In **Chapter 5**, we describe our effort to evaluate the predictive value of antinuclear antibodies (ANAs) in patients for the development of multiple different connective tissue diseases, including SLE. As the pathophysiology of SLE is multifactorial and complex, we investigated associations of several of these different immunological manifestations. In **Chapter 6**, we examined the relationships between autoantibodies and the type 1 interferon (IFN) signature, as well as the release of NETs. In **Chapter 7**, we provide an overview of recent developments in the field of biomarkers for the use of diagnostics in SLE.

Lastly, in **Chapter 8** the main findings of this thesis and potential improvements to the early diagnostics of SLE are discussed.

REFERENCES

1. Tian J, Zhang D, Yao X, Huang Y, Lu Q. Global epidemiology of systemic lupus erythematosus: a comprehensive systematic analysis and modelling study. *Ann Rheum Dis*. 2023;82(3):351-356. doi:10.1136/ard-2022-223035
2. Pons-Estel GJ, Alarcón GS, Scofield L, Reinlib L, Cooper GS. Understanding the Epidemiology and Progression of Systemic Lupus Erythematosus. *Semin Arthritis Rheum*. 2010;39(4):257. doi:10.1016/j.semarthrit.2008.10.007
3. Petri M. Epidemiology of systemic lupus erythematosus. *Best Pract Res Clin Rheumatol*. 2002;16(5):847-858. doi:10.1053/berh.2002.0259
4. Aringer M, Costenbader KH, Daikh DI, et al. 2019 EULAR/ACR Classification Criteria for Systemic Lupus Erythematosus. *Arthritis Rheumatol* 2019;71(9):1400-1412. doi:10.1002/art.40930
5. Cornet A, Andersen J, Myllys K, Edwards A, Arnaud L. Living with systemic lupus erythematosus in 2020: a European patient survey. *Lupus Sci Med*. 2021;8(1):e000469. doi:10.1136/lupus-2020-000469
6. Teh CL, Phui VE, Ling GR, Ngu LS, Wan SA, Tan CHH. Causes and predictors of mortality in biopsy-proven lupus nephritis: the Sarawak experience. *Clin Kidney J*. 2018;11(1):56-61. doi:10.1093/ckj/sfx063
7. Pons-Estel GJ, Andreoli L, Scanzi F, Cervera R, Tincani A. The antiphospholipid syndrome in patients with systemic lupus erythematosus. *J Autoimmun*. 2017;76:10-20. doi:10.1016/j.jaut.2016.10.004
8. Tsokos GC, Lo MS, Reis PC, Sullivan KE. New insights into the immunopathogenesis of systemic lupus erythematosus. *Nat Rev Rheumatol*. 2016;12(12):716-730. doi:10.1038/nrrheum.2016.186
9. Mahajan A, Herrmann M, Muñoz LE. Clearance Deficiency and Cell Death Pathways: A Model for the Pathogenesis of SLE. *Front Immunol*. 2016 Feb 8:7:35. doi: 10.3389/fimmu.2016.0003510. Munoz L, Van Bavel C, Franz S, Berden J, Herrmann M, Van Der Vlag J. Apoptosis in the pathogenesis of systemic lupus erythematosus. *Lupus*. 2008;17(5):371-375. doi:10.1177/0961203308089990
11. Yang Y, Jiang G, Zhang P, Fan J. Programmed cell death and its role in inflammation. *Mil Med Res*. 2015;2:12. doi:10.1186/s40779-015-0039-0
12. Rogers C, Fernandes-Alnemri T, Mayes L, Alnemri D, Cingolani G, Alnemri ES. Cleavage of DFNA5 by caspase-3 during apoptosis mediates progression to secondary necrotic/pyroptotic cell death. *Nat Commun*. 2017;8(1):14128. doi:10.1038/ncomms14128
13. Yaniv G, Twig G, Shor DBA, et al. A volcanic explosion of autoantibodies in systemic lupus erythematosus: A diversity of 180 different antibodies found in SLE patients. *Autoimmun Rev*. 2015;14(1):75-79. doi:10.1016/j.autrev.2014.10.003
14. Dema B, Charles N. Autoantibodies in SLE: Specificities, Isotypes and Receptors. *Antibodies (Basel)*. 2016;5(1):2. doi:10.3390/antib5010002
15. Sterner RM, Hartono SP, Grande JP. The Pathogenesis of Lupus Nephritis. *J Clin Cell Immunol*. 2014;5(2):205. doi:10.4172/2155-9899.1000205
16. Ohl K, Tenbrock K. Inflammatory Cytokines in Systemic Lupus Erythematosus. *J Biomed Biotechnol*. 2011;2011:432595. doi:10.1155/2011/432595
17. Rönnblom L, Leonard D. Interferon pathway in SLE: one key to unlocking the mystery of the disease. *Lupus Sci Med*. 2019;6(1):e000270. doi:10.1136/lupus-2018-000270
18. Fitzgerald-Bocarsly P, Dai J, Singh S. Plasmacytoid dendritic cells and type I IFN: 50 years of convergent history. *Cytokine Growth Factor Rev*. 2008;19(1):3-19. doi:10.1016/j.cytogfr.2007.10.006

19. Katze MG, He Y, Gale M. Viruses and interferon: a fight for supremacy. *Nat Rev Immunol*. 2002;2(9):675-687. doi:10.1038/nri888
20. Obermoser G, Pascual V. The interferon-α signature of systemic lupus erythematosus. *Lupus*. 2010;19(9):1012-1019. doi:10.1177/0961203310371161
21. Rönnblom L, Eloranta ML. The interferon signature in autoimmune diseases. *Curr Opin Rheumatol*. 2013;25(2):248-253. doi:10.1097/BOR.0b013e32835c7e32
22. Ioannou Y, Isenberg DA. Current evidence for the induction of autoimmune rheumatic manifestations by cytokine therapy. *Arthritis Rheum*. 2000;43(7):1431-1442. doi:10.1002/1529-0131(200007)43:7<1431::AID-ANR3>3.0.CO;2-E
23. Leffler J, Bengtsson AA, Blom AM. The complement system in systemic lupus erythematosus: an update. *Ann Rheum Dis*. 2014;73(9):1601-1606. doi:10.1136/annrheumdis-2014-205287
24. Weinstein A, Alexander RV, Zack DJ. A Review of Complement Activation in SLE. *Curr Rheumatol Rep*. 2021;23(3):16. doi:10.1007/s11926-021-00984-1
25. Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. *J Rheumatol*. 2002;29(2):288-291.
26. Yee CS, Farewell VT, Isenberg DA, et al. The use of Systemic Lupus Erythematosus Disease Activity Index-2000 to define active disease and minimal clinically meaningful change based on data from a large cohort of systemic lupus erythematosus patients. *Rheumatology (Oxford)*. 2011;50(5):982-988. doi:10.1093/rheumatology/keq376
27. Remijsen Q, Kuijpers TW, Wirawan E, Lippens S, Vandenabeele P, Vanden Berghe T. Dying for a cause: NETosis, mechanisms behind an antimicrobial cell death modality. *Cell Death Differ*. 2011;18(4):581-588. doi:10.1038/cdd.2011.1
28. Brinkmann V, Reichard U, Goosmann C, et al. Neutrophil Extracellular Traps Kill Bacteria. *Science*. 2004;303(5663):1532-1535. doi:10.1126/science.1092385
29. Jorch SK, Kubes P. An emerging role for neutrophil extracellular traps in noninfectious disease. *Nat Med*. 2017;23(3):279-287. doi:10.1038/nm.4294
30. Villanueva E, Yalavarthi S, Berthier CC, et al. Netting neutrophils induce endothelial damage, infiltrate tissues, and expose immunostimulatory molecules in systemic lupus erythematosus. *J Immunol.* 2011;187(1):538-552. doi:10.4049/jimmunol.1100450
31. Sestak AL, Furnrohr BG, Harley JB, Merrill JT, Namjou B. The genetics of systemic lupus erythematosus and implications for targeted therapy. *Ann Rheum Dis*. 2011;70(Suppl 1):i37-i43. doi:10.1136/ard.2010.138057
32. Block SR. A brief history of twins. *Lupus*. 2006;15(2):61-64. doi:10.1191/0961203306lu2263ed
33. Bogdanos DP, Smyk DS, Rigopoulou EI, et al. Twin studies in autoimmune disease: Genetics, gender and environment. *J Autoimmun*. 2012;38(2-3):J156-J169. doi:10.1016/j.jaut.2011.11.003
34. Moulton VR. Sex Hormones in Acquired Immunity and Autoimmune Disease. *Front Immunol*. 2018;9:2279. doi:10.3389/fimmu.2018.02279
35. Kim JW, Kim HA, Suh CH, Jung JY. Sex hormones affect the pathogenesis and clinical characteristics of systemic lupus erythematosus. *Front Med*. 2022;9:906475. doi:10.3389/fmed.2022.906475
36. Klein SL, Flanagan KL. Sex differences in immune responses. *Nat Rev Immunol*. 2016;16(10):626-638. doi:10.1038/nri.2016.90
37. Fairhurst A, Wandstrat AE, Wakeland EK. Systemic Lupus Erythematosus: Multiple Immunological Phenotypes in a Complex Genetic Disease. In: *Adv Immunol*. Vol 92. Academic Press; 2006:1-69. doi:10.1016/S0065-2776(06)92001-X
38. Guthridge JM, Lu R, Tran LTH, et al. Adults with systemic lupus exhibit distinct molecular phenotypes in a cross-sectional study. *eClinicalMedicine*. 2020;20. doi:10.1016/j.eclinm.2020.100291

39. Agmon-Levin N, Mosca M, Petri M, Shoenfeld Y. Systemic lupus erythematosus one disease or many? *Autoimmun Rev*. 2012;11(8):593-595. doi:10.1016/j.autrev.2011.10.020
40. Leuchten N, Milke B, Winkler-Rohlfing B, et al. Early symptoms of systemic lupus erythematosus (SLE) recalled by 339 SLE patients. *Lupus*. 2018;27(9):1431-1436. doi:10.1177/0961203318776093
41. Mosca M, Costenbader KH, Johnson SR, et al. 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*. 2019;71(1):91-98. doi:10.1002/art.40674
42. Kernder A, Richter JG, Fischer-Betz R, et al. Delayed diagnosis adversely affects outcome in systemic lupus erythematosus: Cross sectional analysis of the LuLa cohort. *Lupus*. 2021;30(3):431-438. doi:10.1177/0961203320983445
43. Tselios K, Gladman DD, Touma Z, Su J, Anderson N, Urowitz MB. Disease course patterns in systemic lupus erythematosus. *Lupus*. 2019;28(1):114-122. doi:10.1177/0961203318817132
44. Petri MA, van Vollenhoven RF, Buyon J, et al. Baseline predictors of systemic lupus erythematosus flares: data from the combined placebo groups in the phase III belimumab trials. *Arthritis Rheum*. 2013;65(8):2143-2153. doi:10.1002/art.37995
45. Stojan G, Petri M. Anti-C1q in systemic lupus erythematosus. *Lupus*. 2016;25(8):873-877. doi:10.1177/0961203316645205
46. Carter LM, Alase A, Wigston Z, et al. Gene Expression and Autoantibody Analysis Revealing Distinct Ancestry-Specific Profiles Associated With Response to Rituximab in Refractory Systemic Lupus Erythematosus. *Arthritis Rheumatol*. 2023;75(5):697-710. doi:10.1002/art.42404
47. Ayoub I, Wolf BJ, Geng L, et al. Prediction models of treatment response in lupus nephritis. *Kidney Int*. 2022;101(2):379-389. doi:10.1016/j.kint.2021.11.014
48. Fanouriakis A, Kostopoulou M, Alunno A, et al. 2019 update of the EULAR recommendations for the management of systemic lupus erythematosus. *Ann Rheum Dis*. 2019;78(6):736-745. doi:10.1136/annrheumdis-2019-215089
49. Dubey AK, Handu SS, Dubey S, Sharma P, Sharma KK, Ahmed QM. Belimumab: First targeted biological treatment for systemic lupus erythematosus. *J Pharmacol Pharmacother*. 2011;2(4):317-319. doi:10.4103/0976-500X.85930
50. Merrill JT, Neuwelt CM, Wallace DJ, et al. Efficacy and safety of rituximab in moderately-to-severely active systemic lupus erythematosus: The randomized, double-blind, phase ii/iii systemic lupus erythematosus evaluation of rituximab trial. *Arthritis Rheum*. 2010;62(1):222-233. doi:10.1002/art.27233
51. Dima A, Jurcut C, Chasset F, Felten R, Arnaud L. Hydroxychloroquine in systemic lupus erythematosus: overview of current knowledge. *Ther Adv Musculoskelet Dis*. 2022;14:1759720X211073001. doi:10.1177/1759720X211073001
52. Morand EF, Furie R, Tanaka Y, et al. Trial of Anifrolumab in Active Systemic Lupus Erythematosus. *N Engl J Med*. 2020;382(3):211-221. doi:10.1056/NEJMoa1912196
53. Isenberg DA, Manson JJ, Ehrenstein MR, Rahman A. Fifty years of anti-ds DNA antibodies: are we approaching journey's end? *Rheumatology (Oxford)*. 2007;46(7):1052-1056. doi:10.1093/rheumatology/kem112
54. Arbuckle MR, McClain MT, Rubertone MV, et al. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N Engl J Med*. 2003;349(16):1526-1533. doi:10.1056/NEJMoa021933
55. Choi MY, FitzPatrick RD, Buhler K, Mahler M, Fritzler MJ. A review and meta-analysis of anti-ribosomal P autoantibodies in systemic lupus erythematosus. *Autoimmun Rev*. 2020;19(3):102463. doi:10.1016/j.autrev.2020.102463

56. Merrill JT, Manzi S, Aranow C, et al. Lupus community panel proposals for optimising clinical trials: 2018. *Lupus Sci Med*. 2018;5(1):e000258. doi:10.1136/lupus-2018-000258
57. Monti S, Grosso V, Todoerti M, Caporali R. Randomized controlled trials and real-world data: differences and similarities to untangle literature data. *Rheumatology (Oxford)*. 2018;57(Supplement_7):vii54-vii58. doi:10.1093/rheumatology/key109
58. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Routinely collected data and comparative effectiveness evidence: promises and limitations. *Can Med Assoc J*. 2016;188(8):E158-E164. doi:10.1503/cmaj.150653
59. Tayefi M, Ngo P, Chomutare T, et al. Challenges and opportunities beyond structured data in analysis of electronic health records. *Wiley Interdiscip Rev Comput Stat*. 2021;13(6):e1549. doi:10.1002/wics.1549
60. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc*. 2016;23(5):1007-1015. doi:10.1093/jamia/ocv180
61. Knevel R, Liao KP. From real-world electronic health record data to real-world results using artificial intelligence. *Ann Rheum Dis*. 2023;82(3):306-311. doi:10.1136/ard-2022-222626

Part 1

Use of data from regular care for research

2

Text Mining of Electronic Health Records Can Accurately Identify and Characterize Patients with Systemic Lupus Erythematosus

Authors:
T.E. Brunekreef[a], H.G. Otten[b], S.C. van den Bosch[b], I.E. Hoefer[c], J.M. van Laar[a], M. Limper[a], S. Haitjema[c]

Affiliations:
[a] Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht, The Netherlands
[b] Center for Translational Immunology, University Medical Center Utrecht, Utrecht, The Netherlands
[c] Central Diagnostic Laboratory, University Medical Center Utrecht, Utrecht, The Netherlands

Published
ACR Open Rheumatol. 2021 Feb;3(2):65-71. doi: 10.1002/acr2.11211.

ABSTRACT

Background

Electronic Health Records (EHR) are increasingly being recognized as a major source of data reusable for medical research and quality monitoring, although patient identification and assessment of symptoms (characterization) remain challenging, especially in complex diseases such as systemic lupus erythematosus (SLE). Current coding systems are unable to assess information recorded in the physician's free text notes. This study shows that text mining can be used as a reliable alternative.

Methods

In a multidisciplinary research team of data scientists and medical experts, a text mining algorithm on 4607 patient records was developed to assess the diagnosis of 14 different immune-mediated inflammatory diseases and presence of 18 different symptoms from the EHR. The text mining algorithm includes key words in the EHR, while mining the context for exclusion phrases. The accuracy of the text mining algorithm was assessed by manually checking the EHR of 100 random patients suspected of having SLE for diagnoses and symptoms, and comparing the outcome with the outcome of the text mining algorithm.

Results

After evaluation of 100 patient records, the text mining algorithm had a sensitivity of 96.4% and a specificity of 93.3% assessing presence of SLE. The algorithm detected potentially life-threatening symptoms (nephritis, pleuritis) with good sensitivity (80%-82%) and very high specificity (97%-97%).

Conclusion

We present a text mining algorithm that can accurately identify and characterize patients with SLE using routinely collected data from the EHR. Our study shows that using text mining, data from the EHR can be reused in research and quality control.

INTRODUCTION

Electronic health records (EHR) are increasingly being recognized as a major source of data reusable for medical research and quality monitoring. In the EHR, clinical information for the patient's diagnostic and therapeutic trajectory is collected. This includes logistic information, such as time and date of appointments, as well as results from laboratory tests, lists of medication and arguably most important, the physicians' notes of the patients' visits. Some of the data are recorded using the structured International Classification of Diseases (ICD)-10 (or ICD-9) codes, yet these are too rigid to reflect clinically relevant subtleties, often unreliable and have a large inter-observer variability.[1–3] However, crucial information about diagnosis and symptoms is often recorded in the physician's unstructured free text notes.

For physicians, the main advantage of using free text is that it offers the opportunity for nuance and expression.[4] This nuance and expression is very important in diseases with different phenotypes and courses of disease, such as auto-immune diseases. For research and quality control however, accumulation of relevant information in free text in the EHR presents a challenge. Manual extraction of information from free text is time consuming and is therefore not practical in studies with large amounts of patients. In addition, manual extraction is prone to errors and limits reproducibility.[5,6] Moreover, a new or altered research question will result in having to review all data again. Because of this, other methods of retrieving this data are currently being developed.

Text mining methods are increasingly being used to collect information from free text. With text mining, it is possible to scrutinize free text for key words or phrases regarding variables of interest, which can be captured in a structured way. However, recent studies show that identification of patients with a certain diagnosis improves when also using free text, compared to coding systems alone.[7,8] Text mining can also be used to characterize patients by evaluating the presence of risk factors or multiple manifestations of disease, whereas coding systems are mainly used for diagnoses and a small selection of symptoms.[9–11] This can be especially helpful in studying the heterogeneous clinical course of patients with immune mediated inflammatory diseases (IMIDs).

One IMID in which this characterization is very important is systemic lupus erythematosus (SLE). SLE is a systemic IMID with a large but varied set of associated symptoms, ranging from relatively mild dermatological manifestations to possibly life-threatening nephritis. SLE can be difficult to diagnose, as it shares multiple symptoms and often overlaps with other IMIDs, such as Antiphospholipid Syndrome (APS). Serological examination can be utilized to aid in the diagnostic process. Auto-antibodies against double-stranded DNA (anti-dsDNA) are highly specific for SLE and are only examined in practice whenever there

is clinical suspicion of SLE, since they are not associated with any other disease, although it is known that they are also present in a small percentage of the healthy population.[12]

Reliable identification of the patients is a necessity for clinical research and quality control reusing data from the EHR. Furthermore, for diseases in which there is a large heterogeneity in the clinical course, patient characterization can be of great added value. In this study we present a rule-based text mining algorithm capable of reliably assessing the diagnosis and clinical manifestations of a cohort of patients suspected of having SLE.

PATIENTS AND METHODS

Study population

This study was performed in the University Medical Center Utrecht (UMC Utrecht), a tertiary care center in the Netherlands. To ensure a balanced cohort of patients with and without the diagnosis of SLE, all patients who were tested for anti-dsDNA antibodies between February 2014 and July 2017 were included in a cohort, irrespective of the result of the test. Anti-dsDNA testing is an important component of the diagnostic process for all patients who are suspected of having SLE and is also part of the annual laboratory follow-up for all patients with SLE, according to the hospital protocol. The date of each anti-dsDNA test in this cohort was regarded as an individual time point for which the diagnosis and symptoms were assessed in this study. Of some patients, samples were collected at multiple time points. Informed consent was not collected in this study, the requirement for obtaining informed consent was waived by the Biobank Research Ethics Committee due to the large number of patients, many of which no longer are under control in the UMC Utrecht.

Electronic Health Records

The UMC Utrecht uses ChipSoft HiX as its electronic health record system. Within this software system all relevant medical data can be accessed, reviewed and recorded by the physician. This includes among other things laboratory data, medication data, physician's notes by all of the physicians treating the patient in the hospital, including from different departments, and letters to the patient's general practitioner (GP), which are used in the current study. Medication data and laboratory data are recorded in a structured format and are therefore easily usable in research. Although the physician's notes and letters to the GP usually follow a certain format, they consist mostly of free text and are therefore considered unstructured.

Because drawing of blood did not always happen on the same day as the visit to the clinician, clinical data for up to 14 days after sample collection were used. Data were extracted from the EHR through the Utrecht Patient Oriented Database (UPOD). The structure and

function of UPOD is described elsewhere.[13] In brief, UPOD is an infrastructure of relational databases comprising data on patient characteristics, hospital discharge diagnoses, medical procedures, medication orders and laboratory tests for all patients treated at the UMC Utrecht since 2004. This study was in accordance with the guidelines approved by the Medical Ethical Committee and approved by the biobank committee of the UMC Utrecht. All collected data were pseudonymized for privacy reasons.

Text mining

To search the EHR for relevant data, a rule-based text mining algorithm for the Dutch language was developed in R. The algorithm looks for specific key words that indicate that something (e.g. a diagnosis) is present in the text. Key words were constructed for SLE, as well as several other IMIDs, which can cause similar symptoms as SLE and could therefore explain the requested anti-dsDNA test requested by the physician. These key words were used to identify patients with a clinical diagnosis of an IMID, as recorded in the EHR, rather than identify patients fulfilling classification criteria for a specific IMID.

Key words for diagnoses and symptoms

Key words were formulated for SLE as well as several other auto-immune diseases which can overlap with or can mimic symptoms of SLE. The list of symptoms for which key words were made, was based on the individual clinical components of the SLE disease activity index (SLEDAI).[14] Hematological and serological components of the SLEDAI can be extracted from laboratory results directly, therefore there is no need to use text mining for these manifestations. Ambiguous terms or abbreviations that can indicate multiple things, for example "DM" can indicate the diagnosis of either dermatomyositis or diabetes mellitus, were excluded from the list of key words. When seen in the context of the total report in the EHR, the true meaning of such an ambiguous term is usually relatively easy to interpret for a person, but not for a text mining algorithm. The complete list of diagnoses and symptoms for which key words were formulated is available in supplementary Tables 1a and 1b.

Key words were formulated based on terms used in literature, clinical practice and known synonyms using regular expression. Results of the most recent version of the algorithm were compared to manual review in an iterative process. Discordances were noted and used to improve the algorithm. Over 100 patient records were manually reviewed in five iterations to improve performance. To facilitate incorporation of expert knowledge within the algorithm, the process of constructing key words was supervised by a specialized clinical immunologist (ML).

Context mining of negative and neutral records

In the EHR, key words can be present in either a positive, negative or neutral context. In a positive context, the key words indicate that the disease or symptom is present and

should be recorded as such. However, when mentioned in a negative context (e.g. 'diagnosis X is unlikely'), this instance should function as an indication that a diagnosis is less likely. 70 variations of negative context were constructed. Using both the positive and negative instances of the key words, a calculation was made to assess the likelihood of actual presence (see below). This calculation does not include neutral records, e.g. records mentioned in the differential diagnosis, because a diagnosis mentioned in the differential diagnosis is neither definitively present nor absent and therefore does not influence the likelihood of the definitive diagnosis.

Confidence

To assess the likeliness of presence of each individual diagnosis or symptom, we calculated a "confidence" score:

Established diagnoses are often repeated in the EHR, therefore the number of times a disease that is present is mentioned in the EHR should be relatively high, resulting in a high level of confidence. New patients will have fewer records of their diagnoses, but since the total number of documents included is small as well, the resulting confidence level will still be high. To limit the number of false positives, negative records were assigned more weight than positive records, since positive records might occasionally be found in the EHR of non-patients, whereas negative records in the EHR of patients are scarce.

Based on iterative manual checking for diagnoses given by the text mining program during development of the algorithm, we identified a cut-off of a confidence level of 14% to be appropriate to limit the number of false positives and false negatives. If there was no single diagnosis with a confidence level of 14% or higher for a sample, this was defined as a "no diagnosis" category. Patients with more than one diagnosis with a confidence level of 14% or higher were recorded as having multiple diagnoses, for example if there is a confidence level of 60% for both SLE and APS, this is recorded as the patient having both diagnoses.

For symptoms, a similar approach was used. However, unlike the diagnosis, the presence of symptoms can be different at each visit to the physician and is therefore mentioned less often in the total number of documents, resulting in lower confidence levels. This lower level of confidence mainly affects non-life-threatening symptoms, since critical symptoms such as nephritis or pericarditis are commonly repeated in the patient history if ever present. Similar to the diagnosis, negative records of symptoms lowered the confidence level and neutral records were ignored. With the goal of creating an overview of the symptoms in the patient's history, we chose to set the level for inclusion of symptoms to a confidence level of 1%, meaning there had to be more positive records than negative records. This approach means that assigned symptoms should be interpreted as having been present at least once and not necessarily as present at the time of record.

Assessment of accuracy of the text mining algorithm

To assess the accuracy of the final version of the text mining algorithm, 100 patients from the cohort were randomly selected, independent of the records used to determine the cut-off for the level of confidence, the diagnoses were manually gathered from the EHR by a clinical immunologist with expertise in the field of SLE (ML). The clinical immunologist was blinded for the outcome of the text mining algorithm during this process. The diagnoses reported by the clinical immunologist were used as the gold standard to compare with the diagnoses reported by the text mining algorithm and used to calculate sensitivity and specificity. A similar approach was used to determine the accuracy of the text mining algorithm for symptoms. Due to low prevalence of several symptoms, we only analyzed the accuracy for symptoms with a prevalence of at least 10% in the 100 randomly selected patients.

RESULTS

Patients and demographics

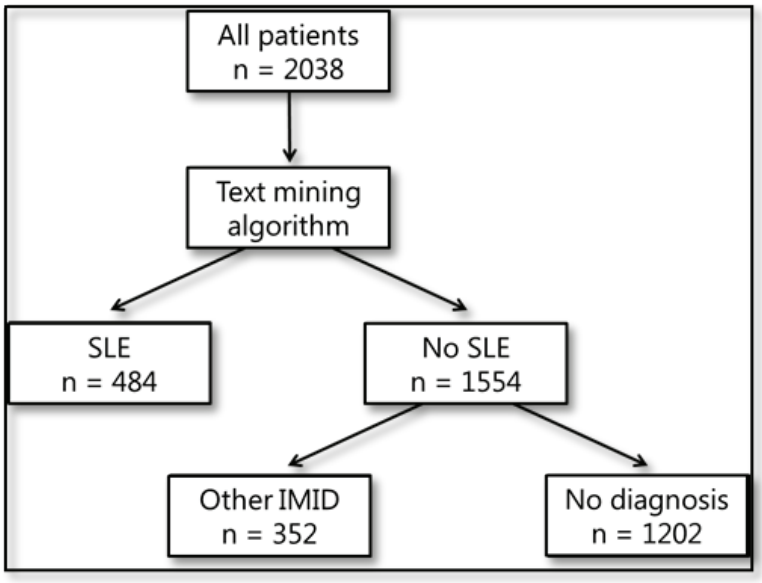
Between 2014 and 2017, 2038 patients were tested for the presence of anti-dsDNA antibodies for a total of 4607 times. The mean age of all patients at the time of collection of their first visit was 44.9 years, 63.4% of all patients were female.

Diagnoses

The text mining algorithm assigned the diagnosis SLE to 484 patients (23.7%) at the first record, making it the most common diagnosis (Figure 1, supplementary Table 2). One thousand two hundred and two (1202) patients (59.0%) were assigned no diagnosis at the first record. Twenty-six patients were assigned SLE after the first record, bringing the total number of patients with SLE during the entire follow-up to 510 (25%). The diagnosis SLE was assigned in 2726 records (59%).

Overlap of multiple diagnoses was common in our cohort. Overlap of APS and SLE was the most common, in 64 patients at the time of the first sample. Our algorithm reported both a diagnosis of Cutaneous Lupus Erythematosus (CLE) and SLE in 59 patients. Lupus-Like Disease (LLD) is also called incomplete SLE, and some patients with LLD will eventually develop SLE. Combination of both LLD and SLE was recorded in 28 patients.

Figure 1 Flowchart of identification of patients with SLE at first record.



The first records of all 2038 patients in the cohort were evaluated by the text mining algorithm, which assigned the diagnosis SLE to 484 patients. 352 patients were assigned another IMID: Antiphospholipid syndrome (APS), Cutaneous lupus erythematosus (CLE), Giant cell arteritis (GCA), Granulomatosis with polyangiitis (GPA), Idiopathic inflammatory myopathy (IIM), Juvenile idiopathic arthritis (JIA), Lupus like disease (LLD, also known as incomplete SLE), Mixed connective tissue disease (MCTD), Polymyalgia rheumatica (PMR), (primary) Sjogren's syndrome (pSS), Rheumatoid arthritis (RA), Systemic sclerosis (SSc) or Undifferentiated connective tissue disease (UCTD).

Symptoms

The prevalence of all symptoms is presented in Table 1. Fatigue was the most common symptom in both patients with and without SLE. Nephritis was reported in 33.3% of all patients with SLE.

Table 1: Prevalence of symptoms.

Feature	Prevalence (N, (%))
Total number of patients	2038
Symptoms	
Alopecia	292 (14.3)
Arthritis	328 (16.1)
Arthralgia	322 (15.8)
Cranial Nerve Disorder	141 (6.9)
CVA	317 (15.6)
Digital ulceration	14 (0.7)
Fatigue	1030 (50.5)

Table 1: Prevalence of symptoms. (*continued*)

Feature	Prevalence (N, (%))
Mucosal ulcers	323 (15.6)
Myositis	55 (2.7)
Nephritis	203 (10.0)
Rash	310 (15.2)
Pericarditis	134 (6.6)
Pleurisy	117 (5.7)
Psychosis	59 (2.9)
Raynaud	432 (21.2)
Seizure	139 (6.8)
Vasculitis	270 (13.2)
Visual Disturbance	20 (1.0)

Prevalence of clinical symptoms in all patients at the time of the first record. Since laboratory parameters can be assessed by using structured data and no text mining is required to assess these data, these parameters are not shown. "Rash" includes photosensitivity and malar rash.

Performance of the algorithm

The results of the comparison of the outcome of the text mining algorithm and judgement of the clinical immunologist is shown in Table 2. In 71 out of the 100 randomly selected patients, the diagnoses of the text mining algorithm and of the clinical immunologist matched completely. In 9 cases, the text mining algorithm recorded overlap of SLE with CLE and/or LLD, whereas the clinical immunologist recorded only SLE to be present. Seven patients were assigned another diagnosis overlapping with SLE by the text mining algorithm, for example overlap of SLE with APS, whereas the clinical immunologist only reported the diagnosis SLE. The algorithm misdiagnosed 5 patients, which were determined to have no diagnosis according to the clinical immunologist. Conversely, the algorithm missed a diagnosis in 8 cases.

The clinical immunologist assigned SLE as a diagnosis to fifty-five of the 100 patients. Fifty-three out of these 55 were correctly recorded by the text mining algorithm and 3 false positives were recorded, resulting in a sensitivity and specificity of 96.4% and 93.3% respectively, with a positive predictive value (PPV) of 94.6%. The algorithm correctly reported 22 out of the 26 patients with no assigned diagnosis, and mistakenly reported 8 patients not to have an IMID, resulting in a sensitivity and specificity of 84.6% and 89.2% respectively for identifying patients without an IMID diagnosis. (Table 3).

Ten out of the 18 studied symptoms were present in at least 10% of the 100 patients. The sensitivity of the text mining algorithm was variable for different symptoms, ranging from poor (0.38 for mucosal ulcerations) to good (0.84 for fatigue). Specificity generally was very good, with a specificity of over 0.90 for most symptoms (Table 4). The algorithm

detected possibly life-threatening symptoms pleuritis and nephritis with good sensitivity and very high specificity. 2x2 Tables for all symptoms are available in supplementary Table 3.

Table 2: Performance of the text mining algorithm for diagnosis.

Outcome	N
Complete match	71
Algorithm reported overlapping diagnosis of LLD/CLE in context of SLE, whereas immunologist did not report overlap	9
Algorithm missed an overlapping diagnosis	4*
Algorithm assigned an extra overlapping diagnosis	4*
Diagnosis wrongly assigned by algorithm	5
Diagnosis missed by algorithm	8

Comparison of 100 randomly selected patients, diagnosis assessed by text mining algorithm compared to assessment of a clinical immunologist.

* In one case, the text mining algorithm missed one overlapping diagnosis and also misdiagnosed another overlapping diagnosis. This case is represented in both rows.

Table 3: Performance of the text mining algorithm for diagnosis of SLE.

Diagnosis	N (according to manual review)	Sensitivity	Specificity	PPV	NPV
SLE	53	0.96	0.93	0.95	0.95
APS	11	0.72	1	1	0.97
No diagnosis	26	0.85	0.89	0.73	0.94

Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for all diagnoses with a prevalence of over 10% in the 100 randomly selected patients (n=90). Ten patients had another diagnosis than SLE or APS.

Table 4: Performance of the text mining algorithm for symptoms.

Symptom	N (according to manual review)	Sensitivity	Specificity	PPV	NPV
Alopecia	26	0.65	0.95	0.81	0.89
Arthritis	41	0.41	0.92	0.69	0.77
Arthralgia	45	0.56	0.93	0.86	0.72
Fatigue	67	0.84	0.64	0.82	0.66
Mucosal ulcers	21	0.38	0.81	0.35	0.83
Nephritis	39	0.82	0.97	0.94	0.89
Rash	49	0.51	0.90	0.66	0.83
Pleuritis	10	0.80	0.97	0.73	0.98
Raynaud	26	0.69	0.88	0.67	0.89
Vasculitis	17	0.53	0.94	0.64	0.91

Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for all symptoms with a prevalence of over 10% in the 100 randomly selected patients.

DISCUSSION

Patient identification is a vital part of research and quality control reusing data from the EHR. Because the reliability of traditional coding systems is limited for these purposes, we investigated an alternative method of patient identification as well as patient characterization. In this study, we developed a rule-based text mining algorithm to assess the presence of a diagnosis of autoimmune diseases in the EHR, as well as to assess SLE-related disease manifestations.

Our text mining algorithm has very good accuracy for the diagnosis of SLE when compared with the assessment of a medical specialist, with a high sensitivity and specificity. Moreover, it outperforms other algorithms for identifying patients with SLE presented in previous studies. For example, Jorge et al. and Barnado et al. used algorithms processing a combination of structured parameters, such as number of corresponding ICD-codes, positive lab results and use of medication (PPV 92%, sensitivity 47% and PPV 91%, sensitivity 40% respectively).[15,16] Jorge et al. also developed an algorithm which included both structured parameters and natural language processing, but this algorithm performed slightly worse than the algorithm using only structured parameters (PPV 90%, sensitivity 41%).[15] Murray et al. used a combination of these structured parameters and also included the use of a single search term for "lupus" in the EHR notes, but reached similar results to the studies who did not include this single free text search term (PPV 85/98%, sensitivity 97/84%).[17] A limitation of these machine learning models is that they have limited portability, restricting their widespread use, although recently improvements in this regard have been made.[15] A possible explanation for why our algorithm outperforms these models is that our algorithm does not include ICD-codes and does include context mining in regards to text processing. Furthermore, our text mining algorithm does not consider our primary inclusion criteria, anti-dsDNA, whereas these studies selected patients based on SLE-associated ICD-codes, which were subsequently assessed by the corresponding algorithms, possibly introducing a bias.

It is noteworthy that our algorithm assesses the clinical diagnosis recorded in the EHR, rather than the diagnosis according to classification criteria. Although classification criteria provide a point of reference to clinicians as to which patients should be given a clinical diagnosis, the main use of classification criteria is to define a homogeneous group of patients for research.[18] In our study, we chose to focus on the clinical diagnosis.

Ideally, we would have also compared the outcome of the algorithm with the recorded ICD-coding. Unfortunately, in the Netherlands the ICD coding is not recorded directly into the EHR, but rather a derivative of the financial coding system, called "DBC", which uses different coding than the ICD-coding. For registration purposes, an ICD-code is later linked to all different DBC codes. This indirect coding introduces two translational steps

with loss of information and chances for mistakes. Furthermore, these coding systems are not always suited for specific diagnoses, i.e. cutaneous lupus erythematosus is often coded as "inflammatory dermatosis", which covers several other diseases as well. We compared the outcome of the 100 manually reviewed patients with the recorded DBC code. For SLE, this resulted in similar percentages of sensitivity and specificity, however for most other diseases, the DBC coding was often inaccurate or incomplete in cases of overlapping diagnoses. For example, a DBC that could be related to the diagnosis APS was only recorded in three out of the eleven patients in the 100 reviewed patients.

To our knowledge, only one other study investigated assessment of symptoms by text mining in SLE. This study by Gianfrancesco et al. study focused solely on lupus nephritis.[19] Our study shows that patient characterization and assessment of a wide range of symptoms is also possible with the use of text mining in a heterogeneous disease like SLE. The sensitivity of our algorithm varied for different symptoms, which implies that the key words used for some symptoms could be improved. However, sensitivity for possible life-threatening symptoms was good, possibly due to more frequent repetition in the medical history in the EHR. Our algorithm performed slightly better than a text mining algorithm used by Gianfrancesco et al. (sensitivity 79%, specificity 86%) to detect lupus nephritis in the EHR.[20] It should be noted that the gold standard used to compare with a text mining algorithm, manual review of EHR records, is prone to human errors. For example, in several cases where there were discrepancies between the outcome produced by the algorithm and manual review, further investigation revealed that a symptom was overlooked by manual review and the algorithm was correct. Yet, for this article only the initial assessment by the clinical immunologist was used for analysis.

Due to the low prevalence of certain symptoms, we were not able to adequately test the performance of the algorithm for these symptoms, which therefore were excluded from analysis. We were also not able to reliably assess some clinical features that usually are recorded only once in the physician's notes. Smoking status, family history and use of alcohol and drugs is usually only extensively discussed and recorded during the first visit of the patient to the physician. Because this is only mentioned in one visit, this results in low levels of confidence and also requires that the first visit of the patient is included in the data supplied to the text mining program. Improvements in identifying smokers with the use of text mining recently have been made in our hospital, although this method is not yet included in this study.[21]

With increasingly smart and flexible text mining programs being developed, we predict that the use of clinical data from the EHR in clinical research and quality control will increase in the future. In computer science, the phrase "garbage in, garbage out" is often used to indicate the necessity of quality input data to produce a quality outcome. Although administration is often seen as a burden by healthcare professionals, their role

in this process is vital. Researchers and healthcare professionals should work together to find a way for recording accurate data that are reusable in research, while limiting the associated burden of administration on healthcare professionals.[22]

Text mining algorithms have great flexibility and adaptability, allowing the use in different research projects. Although the text mining algorithm presented in this study is a relatively simple and transparent rule-based algorithm and can still be improved upon, it can relatively easily be adapted to better suit research questions from other cohort studies, in contrast to machine learning algorithms. Further research is needed to evaluate the portability of our algorithm to other hospitals, as this would, for example, pave the way for a national registry for rare diseases, using routinely collected clinical data. Moreover, we plan to investigate whether the performance of the algorithm can be improved when our method is combined with machine learning methods using more than only free text data, similar to those used in previous studies.[15–17]

There are several hurdles to be overcome before an algorithm like this can be implemented on a large scale. It is likely that the algorithm would have to be amended for use in hospitals using another EHR-program, since the structure of the necessary data might be different, which would likely also influence the level of confidence used for cut-off. Furthermore, our algorithm was developed in a Dutch-speaking environment, language-specific adaptations are required for implementation in other countries.

Our study presents a method to assess the diagnosis and clinical manifestations of patients with SLE recorded in the EHR for large groups of patients with great accuracy. Our algorithm can both be used for patient identification and characterization of patients with SLE. Our study demonstrates that text mining can be used for performing large scale research and quality control with clinical data from the EHR in heterogeneous diseases such as SLE.

ACKNOWLEDGEMENTS

This work was supported by Thermo Fisher Scientific.

AUTHOR CONTRIBUTION

H.O., J.L, and M.L. contributed to the conceptualization and design of the study. S.B and S.H. developed the text mining algorithm. I.H., and S.H. were extracted the relevant information from the Electronic Health Records. T.B. and M.L. performed verification of the outcome of the text mining algorithm by patient chart review. T.B. performed data analysis and data visualization. All authors contributed to data interpretation. T.B. wrote the manuscript. All authors critically reviewed the manuscript. No potential conflicts of interest relevant to this article were reported.

SUPPLEMENTARY TABLES

Supplementary Table S1a

Diagnoses
Antiphospholipid syndrome (APS)
Cutaneous lupus erythematosus (CLE)
Giant cell arteritis (GCA)
Granulomatosis with polyangiitis (GPA)
Idiopathic inflammatory myopathy (IIM)
Juvenile idiopathic arthritis (JIA)
Lupus like disease (LLD, also known as incomplete SLE)
Mixed connective tissue disease (MCTD)
Polymyalgia rheumatica (PMR)
(primary) Sjogren's syndrome (pSS)
Rheumatoid arthritis (RA)
Systemic sclerosis (SSc)
Undifferentiated connective tissue disease (UCTD)
Systemic lupus erythematosus (SLE)

Supplementary Table S1a

Symptoms
Alopecia
Arthritis
Arthralgia
Cranial Nerve Disorder
CVA
Digital ulceration
Fatigue
Mucosal ulcers
Myositis
Nephritis
Rash
Pericarditis
Pleuritis
Psychosis
Raynaud
Seizure
Vasculitis
Visual Disturbance

Supplementary Tables S1a & S1b: Complete list for diagnoses (a) and symptoms (b) of interest.

List of 14 diagnoses and 18 symptoms for which key words were formulated. CVA: Cerebrovascular accident.

Supplementary Table S2:

Number of assigned diagnoses at the first record of all patients.

Disease	Number
APS	124
CLE	143
GCA	6
GPA	7
IIM	31
JIA	22
LLD	57
MCTD	41
PMR	19
pSS	80
RA	68
SLE	484
SSc	30
UCTD	11
No diagnosis	1202

REFERENCES

1. Stausberg J, Lehmann N, Kaczmarek D, Stein M. Reliability of diagnoses coding with ICD-10. *Int J Med Inform*. 2008;77(1):50-57. doi:10.1016/j.ijmedinf.2006.11.005
2. Moores KG, Sathe NA. A systematic review of validated methods for identifying systemic lupus erythematosus (SLE) using administrative or claims data. *Vaccine*. 2013;31(S10):K62-K73. doi:10.1016/j.vaccine.2013.06.104
3. Otsa K, Talli S, Harding P, et al. Administrative database as a source for assessment of systemic lupus erythematosus prevalence: Estonian experience. *BMC Rheumatol*. 2019;3(1):1-6. doi:10.1186/s41927-019-0074-7
4. Lovis C, Baud RH, Planche P. Power of expression in the electronic patient record: structured data or narrative text? *Int J Med Inform*. 2000;58-59:101-110.
5. Klein DO, Rennenberg R, Gans R, Enting R, Koopmans R, Prins MH. Limited external reproducibility restricts the use of medical record review for benchmarking. *BMJ Open Qual*. 2019;8(2):e000564. doi:10.1136/bmjoq-2018-000564
6. Hanskamp-Sebregts M, Zegers M, Vincent C, Van Gurp PJ, Vet HCWD, Wollersheim H. Measurement of patient safety: A systematic review of the reliability and validity of adverse event detection with record review. *BMJ Open*. 2016;6(8):1-9. doi:10.1136/bmjopen-2016-011078
7. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc*. 2016;23(5):1007-1015. doi:10.1093/jamia/ocv180
8. Ford E, Nicholson A, Koeling R, et al. Optimising the use of electronic health records to estimate the incidence of rheumatoid arthritis in primary care: what information is hidden in free text? *BMC Med Res Methodol*. 2013;13(1):105. doi:10.1186/1471-2288-13-105
9. Jonnalagadda SR, Adupa AK, Garg RP, Corona-Cox J, Shah SJ. Text Mining of the Electronic Health Record: An Information Extraction Approach for Automated Identification and Subphenotyping of HFpEF Patients for Clinical Trials. *J Cardiovasc Transl Res*. 2017;10(3):313-321. doi:10.1007/s12265-017-9752-2
10. Jonnagaddala J, Liaw ST, Ray P, Kumar M, Chang NW, Dai HJ. Coronary artery disease risk assessment from unstructured electronic health records using text mining. *J Biomed Inform*. 2015;58:S203-S210. doi:10.1016/j.jbi.2015.08.003
11. Karystianis G, Nevado AJ, Kim CH, Dehghan A, Keane JA, Nenadic G. Automatic mining of symptom severity from psychiatric evaluation notes. *Int J Methods Psychiatr Res*. 2018;27(1):e1602. doi:10.1002/mpr.1602
12. Kavanaugh AF, Solomon DH. Guidelines for immunologic laboratory testing in the rheumatic diseases: Anti-DNA antibody tests. *Arthritis Rheum*. 2002;47(5):546-555. doi:10.1002/art.10558
13. Ten Berg MJ, Huisman A, Van Den Bemt PMLA, Schobben AFAM, Egberts ACG, Van Solinge WW. Linking laboratory and medication data: New opportunities for pharmacoepidemiological research. *Clin Chem Lab Med*. 2007;45(1):13-19. doi:10.1515/CCLM.2007.009
14. Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. *J Rheumatol*. 2002;29(2):288-291.
15. Jorge A, Castro VM, Barnado A, et al. Identifying lupus patients in electronic health records: Development and validation of machine learning algorithms and application of rule-based algorithms. *Semin Arthritis Rheum*. 2019;49(1):84-90. doi:10.1016/j.semarthrit.2019.01.002
16. Barnado A, Casey C, Carroll RJ, Wheless L, Denny JC, Crofford LJ. Developing Electronic Health Record Algorithms That Accurately Identify Patients With Systemic Lupus Erythematosus. *Arthritis Care Res (Hoboken)*. 2017;69(5):687-693. doi:10.1002/acr.22989

17. Murray SG, Avati A, Schmajuk G, Yazdany J. Automated and flexible identification of complex disease: Building a model for systemic lupus erythematosus using noisy labeling. *J Am Med Inform Assoc*. 2019;26(1):61-65. doi:10.1093/jamia/ocy154
18. Bertsias GK, Pamfil C, Fanouriakis A, Boumpas DT. Diagnostic criteria for systemic lupus erythematosus: Has the time come? *Nat Rev Rheumatol*. 2013;9(11):687-694. doi:10.1038/nrrheum.2013.103
19. Norgeot B, Glicksberg BS, Trupin L, et al. Assessment of a Deep Learning Model Based on Electronic Health Record Data to Forecast Clinical Outcomes in Patients With Rheumatoid Arthritis. *JAMA Netw Open*. 2019;2(3):e190606. doi:10.1001/jamanetworkopen.2019.0606
20. Gianfrancesco MA, Tamang S, Schmajuk G, Yazdany J. Application of text mining methods to identify lupus nephritis from electronic health records. In: *Lupus Science and Medicine*. Vol 60. Lupus Foundation of America; 2019:A142.1-A142. doi:10.1136/lupus-2019-lsm.187
21. Groenhof TKJ, Koers LR, Blasse E, et al. Data mining information from EHRs produced high yield and accuracy for current smoking status. *J Clin Epidemiol*. Published online November 2019. doi:10.1016/j.jclinepi.2019.11.006
22. Bezemer T, de Groot MC, Blasse E, et al. A Human(e) Factor in Clinical Decision Support Systems. *J Med Internet Res*. 2019;21(3):e11732. doi:10.2196/11732

Part 2

Autoantibodies as targets for early diagnostics

3

Microarray testing in patients with Systemic Lupus Erythematosus identifies a high prevalence of CpG DNA-binding antibodies

Authors:
T.E. Brunekreef[a], M. Limper[a], R. Melchers[b], L. Mathsson-Alm[c], J. Dias[c], I. Hoefer[d], S. Haitjema[d],
J. M. van Laar[a], H.G. Otten[b]
Affiliations:
[a] Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht, The Netherlands
[b] Center for Translational Immunology, University Medical Center Utrecht, Utrecht, The Netherlands
[c] Thermo Fisher Scientific, Uppsala, Sweden;
[d] Central Diagnostic Laboratory, University Medical Center Utrecht, Utrecht, The Netherlands

Published
Lupus Sci Med. 2021 Oct;8(1):e000531. doi: 10.1136/lupus-2021-000531.

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 (BBD). Diagnosis and (presence of) symptoms at each blood draw were retrospectively assessed in the patient records with the Utrecht Patient-Oriented Database (UPOD) using a newly developed text mining algorithm. Sera of patients were analyzed 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 SLE patients with (a history of) lupus nephritis than SLE patients without nephritis.

Conclusion

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

INTRODUCTION

Systemic lupus erythematosus (SLE) is a systemic immune mediated inflammatory disease (IMID) with a wide variety of associated symptoms and immunological 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, antinuclear antibodies (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 analyze 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. The requirement for informed consent was waived by the biobank committee of the UMC Utrecht. This study was conducted in accordance with the Helsinki Declaration of 1975 as revised in 2014 and approved by the biobank committee of the UMC Utrecht.

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 pseudonymized 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 homogenous 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 2 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 blood bank donors (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 9 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 supplementary 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 labeled rabbit

$F(ab')_2$-anti-human IgG was added to the reaction sites. After 30 minutes 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 Chi-square test. Correction for multiple testing was done by Bonferroni correction. Statistical analysis was performed using SPSS (version 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. 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, 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 analyzed, including 483 samples of patients with SLE, 346 patients with another IMID, 690 patients with no diagnosis and 361 self-reported healthy blood bank donors (Figure 1). 7 samples of patients with SLE or another IMID were not analyzed on the microarray: 1 sample was missing from the biobank; 6 samples were mistakenly not selected, because they were not assigned an

IMID-diagnosis during the selection of samples to analyze 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 1 diagnosis, indicating that overlap of multiple diseases was common in our cohort. An overview of all overlapping diagnoses is presented in supplementary Table 2. 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 SLE patients

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 SLE patients 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 1a:

Demographics per diagnosis. Patients with multiple diagnoses are included in multiple rows. Age is the age at the time the first sample of the patient was collected. *: Sex and age of 2 blood bank donors is unknown.

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

Table 1a:

Demographics per diagnosis. Patients with multiple diagnoses are included in multiple rows. Age is the age at the time the first sample of the patient was collected. *: Sex and age of 2 blood bank donors is unknown. (*continued*)

Diagnosis	Patients with diagnosis at sample 1	Patients with diagnosis in total follow-up	Female	Mean age ± SD	Number of samples in biobank
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

SLE: Systemic lupus erythematosus, CLE: Cutaneous lupus erythematosus, pSS: (primary) Sjogren's Syndrome, LLD: Lupus-like disease, MCTD: Mixed connective tissue disease, IIM: Idiopathic inflammatory myositis, SSc: Systemic sclerosis, JIA: Juvenile idiopathic arthritis, PMR: Polymyalgia rheumatica, UCTD: Undifferentiated connective tissue disease, GPA: Granulomatosis with polyangiitis, GCA: Giant cell arteritis.

Table 1b:

Demographics per patient group. Age is the age at the time the first sample of the patient was collected. *: Sex and age of 2 blood bank donors is unknown.

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

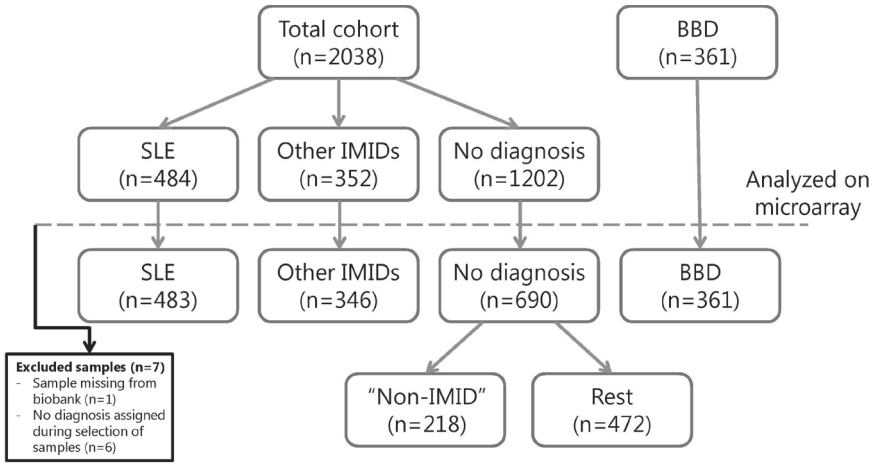


Figure 1: Flowchart of samples analyzed on microarray per patient group.

Overview of number of samples analyzed on microarray for all patient groups. Samples of 1 patient with SLE and 6 patients with a diagnosis of another IMID were not analyzed on the microarray system: 1 sample was missing from the biobank; 6 samples were mistakenly not selected.

Table 2: Prevalence of features of SLE

Prevalence of features of SLE at the time of the first sample and at the time of the latest follow-up sample. Symptoms were assessed by text mining, laboratory assessments are directly taken from the EHR.

Feature	At sample 1 (N, (%))	At latest follow-up sample (N, (%))
Total number 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)
Hemolytic anemia	4 (0,8)	7 (1,9)
Leukopenia	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)

ANA: Anti-nuclear antibodies; dsDNA: anti-double-stranded antibodies. *: 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.

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 to non-IMID controls (Table 3, see suppl. file 3 for ROC curves and AUC values). Antibodies against dsDNA were the most common antibody found in patient 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, 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. 15.1% of SLE patients negative for anti-dsDNA were positive for anti-CpG antibodies, as well as 12.7% of SLE patient with no anti-dsDNA and/or lowered complement factors (suppl. file 3).

Prevalence of most autoantibodies was similar in the non-IMID and BBD control groups, although the autoantibodies 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 to 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 SLE patients. 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 SLE patients with nephritis compared to 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.

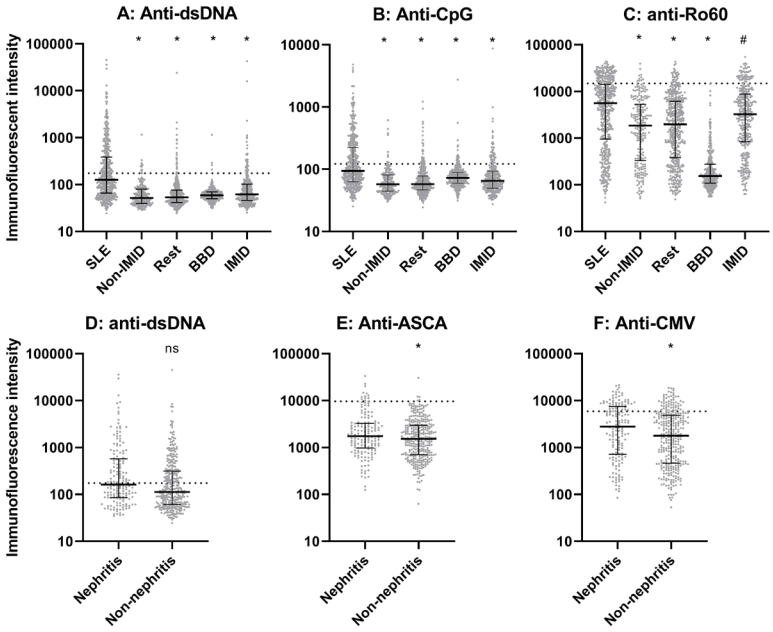


Figure 2: Levels of immunofluorescent intensity.

The 5-95th 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 5-95th percentile of immunofluorescent intensity is presented for (d) anti-dsDNA, (e) anti-ASCA, (f) anti-CMV antibodies for SLE-patients with and without nephritis. The dotted lines represent the cut-off between negative and positive results.

*: p<0,001, #: p=0,001, ns = not significant. Prevalence of autoantibodies in the respective groups vs SLE, based on cut-offs (2A-2C). For comparisons between other groups, see suppl. file 3.

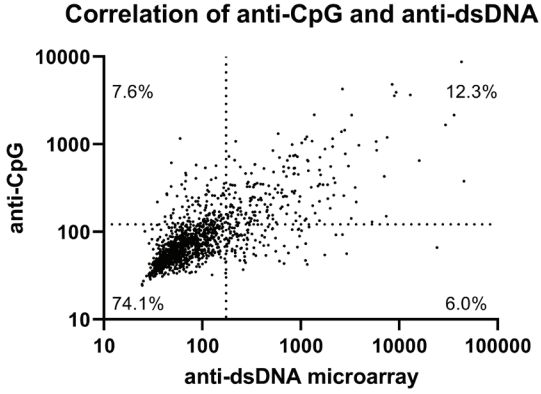


Figure 3: Correlation of anti-dsDNA and anti-CpG antibodies.

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.

Table 3: Prevalence of autoantibodies.

Prevalence of autoantibodies in patients with SLE, non-IMID controls, rest group, BBD controls and patients with other IMID's, and prevalence of autoantibodies in patients with SLE with nephritis. 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 Chi-square test. P-values are given (E-05 = 10^-5). * = no significant difference after Bonferroni correction. NS = Not significant. *SLE: Systemic lupus erythematosus, BBD: Blood bank donors, IMID: Immune mediated inflammatory disease.*

Antibody target	Prevalence in SLE No. (%)	SLE vs non-IMID controls (p-value)	Prevalence rest No. (%)	SLE vs rest (p-value)	Prevalence BBD controls No. (%)	SLE vs BBD controls (p-value)	Prevalence IMIDs No. (%)	SLE vs IMIDs (p-value)	Prevalence in pts with nephritis No. (%)	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

Table 3: Prevalence of autoantibodies.

Prevalence of autoantibodies in patients with SLE, non-IMID controls, rest group, BBD controls and patients with other IMID's, and prevalence of autoantibodies in patients with SLE with nephritis. 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 Chi-square test. P-values are given (E-05 = 10^-5). * = no significant difference after Bonferroni correction. NS = Not significant. *SLE: Systemic lupus erythematosus, BBD: Blood bank donors, IMID: Immune mediated inflammatory disease.* (*continued*)

Antibody target	Prevalence in SLE No. (%)	SLE vs non-IMID controls (p-value)	Prevalence rest No. (%)	SLE vs rest (p-value)	Prevalence BBD controls No. (%)	SLE vs BBD controls (p-value)	Prevalence IMIDs No. (%)	SLE vs IMIDs (p-value)	Prevalence in pts with nephritis No. (%)	SLE nephritis vs non-nephritis (p-value)
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*

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 auto-antibodies 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 SLE patients with nephritis, compared to patients without nephritis, potentially further increasing their diagnostic applicability (Table 3). CpG motifs are unmethylated cytosine-phosphate-guanine oligodinucleotides, DNA patterns that are 20 times more prevalent in bacterial DNA than in mammalian DNA. CpG is recognized by the innate immune system as a pathogen associated molecular pattern (PAMP) and can induce an interferon response by stimulating intracellular toll-like 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 to 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. 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 SLE patients 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 etiological role of CMV in the onset of SLE has been suggested.[24] Zhan et al. showed that 5.3% of patients with lupus nephritis had a CMV infection or reactivation, measured by CMV pp65 antigenemia, during hospitalization.[25] Furthermore, Su et al. showed that almost all SLE patients have IgG anti-CMV antibodies, but that patients who were also positive for IgM anti-CMV antibodies had the highest titers 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 with 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 SLE patients. 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 lupus nephritis patients 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 3 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 to 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 outpa-

tient 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 supplementary Table 1). Furthermore, intra- 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 automatized and therefore uniform manner.

Our study also has some limitations: Although we previously showed that the text mining algorithm performs well in recognizing 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.

FUNDING

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

COMPETING INTERESTS

LM and JD are employed by Thermo Fisher Scientific. JL reports grants from Astra Zeneca, MSD, Thermo Fisher Scientific. TB, ML, IH, SH, RM and HO have declared no conflicts of interest.

CONTRIBUTORSHIP STATEMENT

H.O., J.L,, M.L., L.M-A and J.D. contributed to the conceptualization and design of the study. I.H and S.H. extracted relevant information from the Electronic Health Records. R.M. prepared and performed the serological analysis of the microarrays. T.B. performed data analysis and data visualization. All authors contributed to data interpretation. T.B. wrote the manuscript. All authors critically reviewed the manuscript.

ACKNOWLEDGEMENTS

None

DATA SHARING STATEMENT

The data underlying this article cannot be shared publicly for privacy reasons. The data will be shared on reasonable request to the corresponding author.

RESEARCH ETHICS APPROVAL

This study was approved by the Toetsingscommissie Biobanken UMC Utrecht, TCBio protocolnr. 16-824.

SUPPLEMENTARY TABLES

Supplementary Table 1

Overview of antigens used on the microarray.

Abbreviation	Antigen	Source	Native/ Recombinant	Reference for inclusion in the study
Actinin	Actinin	MyBiosource	Recombinant	Renaudineau et al.; ARTHRITIS & RHEUMATISM Vol. 54, No. 8, August 2006, pp 2523–2532; DOI 10.1002/art.22015; Seret et al.; Journal of Autoimmunity 61 (2015) 54e61
ASCA	Anti-*Saccharomyces cerevisiae* antibody	Thermo Fisher Scientific	Native	Mankaï et al.; Rheumatol Int (2013) 33:665–669 DOI 10.1007/s00296-012-2431-3
Beta2GP1	Beta 2 glycoprotein 1	Thermo Fisher Scientific	Native	Szodoray et al.; Autoimmunity, 42:6, 497-506, DOI: 10.1080/08916930902882731
C1q	Complement protein C1q	Complement Technology	Native	Trendelenburg et al.; Nephrol Dial Transplant (2006) 21: 3115–3121 doi:10.1093/ndt/gfl436; Trendelenburg; Springer Semin Immun (2005) 27:276–285
C3b	Complement protein C3b	Complement Technology	Native	Kenyon et al.; Immunol 2011; 187:2101-2111; doi: 10.4049/jimmunol.1003468
Cardiolipin	Cardiolipin	Undisclosed due to contractual agreement	Native	Szodoray et al.; Autoimmunity, 42:6, 497-506, DOI: 10.1080/08916930902882731
CCP1arg	Cyclic citrullinated peptide (arginized)	Innovagen	Peptide	Budhram et al.; Lupus (2014) 23, 1156–1163; Skare et al.; Clin Rheumatol (2013) 32:1065–1070 DOI 10.1007/s10067-013-2213-7
CCP1cit	Cyclic citrullinated peptide	Innovagen	Peptide	Budhram et al.; Lupus (2014) 23, 1156–1163; Skare et al.; Clin Rheumatol (2013) 32:1065–1070 DOI 10.1007/s10067-013-2213-7
CENP	Centromere protein	Thermo Fisher Scientific	Recombinant	Respaldiza et al.; Scandinavian Journal of Rheumatology, 35:4, 290-294, DOI: 10.1080/03009740600588376
CMV	Cytomegalovirus	MyBiosource	Recombinant	Sekigawa et al.; Clin Exp Rheumatol 2002: 20: 559-564.
CollagenII	Collagen type 2	Undisclosed due to contractual agreement	Native	He et al.; Lupus (2015) 24, 1370–1376

Supplementary Table 1

Overview of antigens used on the microarray. (*continued*)

Abbreviation	Antigen	Source	Native/ Recombinant	Reference for inclusion in the study
CpG	Cytosine-phosphate-guanidine containint ssDNA (ODN 2216, Invivogen)	InvivoGen	DNA	Yasuda et al.; J Immunol. 2009 September 1; 183(5): 3109–3117. doi:10.4049/jimmunol.0900399
CRP	C-reactive protein	MyBiosource	Native	Meyer; Joint Bone Spine 77 (2010) 384–389
DFS70	Dense Fine Speckled, 70kDa molecular weight	Thermo Fisher Scientific	Recombinant	Conrad et al.; Clinic Rev Allerg Immunol (2017) 52:202–216 DOI 10.1007/s12016-016-8564-5
dsDNA	double-stranded desoxyribonucleic acid (DNA)	Thermo Fisher Scientific	DNA	Rekvig; Clinical and Experimental Immunology, 179: 5–10
Enolasearg	Enolase (arginized)	Innovagen	Recombinant	Mosca et al.; J Rheumatol 2006;33;695-697
Enolasecit	Enolase	Innovagen	Recombinant	Mosca et al.; J Rheumatol 2006;33;695-697
EphB2	Ephrin type-B recptor 2	Undisclosed due to contractual agreement	Native	Azzouz et al.; PLoS ONE 11(9): e0160283. doi:10.1371/journal.pone.0160283
FcER	Fc epsilon receptor	Thermo Fisher Scientific	Recombinant	Fiebiger et al.; J. Clin. Invest. Volume 101, Number 1, January 1998, 243–251
Fibrillarin	Fibrillarin	Thermo Fisher Scientific	Recombinant	Plawecki et al.; PLoS ONE 11 (9): e0160879. doi:10.1371/journal.pone.0160879; Tormey et al.; International Immunology, Vol. 16, No. 9, pp. 1295–1304
Ficolin	Ficolin	MyBiosource	Recombinant	Colliard et al.; Arthritis Care & Research Vol. 70, No. 8, August 2018, pp 1263–1268 DOI 10.1002/acr.23449
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase	Undisclosed due to contractual agreement	Native	Delunardo et al.; ARTHRITIS & RHEUMATOLOGY; Vol. 68, No. 11, November 2016, pp 2708–2716.; Takasaki et al.; International Immunology, Vol. 16, No. 9, pp. 1295–1304
GBM	Glomerular basement membrane	Thermo Fisher Scientific	Recombinant	Koutsonikoli et al.; Clinical and Experimental Immunology, 188: 79–85
H2Bp	Histone H2B protein	Thermo Fisher Scientific	Peptide	Dieker et al.; PLoS ONE 11(10): e0165373. doi:10.1371/journal.pone.0165373

Supplementary Table 1

Overview of antigens used on the microarray. (*continued*)

Abbreviation	Antigen	Source	Native/ Recombinant	Reference for inclusion in the study
H2Bpac	Histone H2B protein (acetylized)	Thermo Fisher Scientific	Peptide	Dieker et al.; PLoS ONE 11(10): e0165373. doi:10.1371/journal.pone.0165373
H4p	Histone H4 protein	Thermo Fisher Scientific	Peptide	Dieker et al.; PLoS ONE 11(10): e0165373. doi:10.1371/journal.pone.0165373
H4pac	Histone H4 protein (acetylized)	Thermo Fisher Scientific	Peptide	Dieker et al.; PLoS ONE 11(10): e0165373. doi:10.1371/journal.pone.0165373
Histones	Histones	Thermo Fisher Scientific	Native	Dieker et al.; PLoS ONE 11(10): e0165373. doi:10.1371/journal.pone.0165373
IFNLambda	Interferon-Lambda	MyBiosource	Recombinant	Ching et al.; PLoS ONE 7(2): e32001. doi:10.1371/journal.pone.0032001
IFNOmega	Interferon-Omega	MyBiosource	Recombinant	Ching et al.; PLoS ONE 7(2): e32001. doi:10.1371/journal.pone.0032001
Jo1	histidyl transfer ribonucleic acid (tRNA) ligase	Thermo Fisher Scientific	Recombinant	Zandman-Goddard et al.; Clinical & Developmental Immunology, June 2005; 12(2): 107–111
Ku	Complex of 70 and 80kDa proteins involved in DNA break repair and recombination	Diarect	Native	Cavazzana et al.; Lupus (2008) 17, 727–732
La/SSB	La Protein	Thermo Fisher Scientific	Recombinant	Hochberg et al.; Medicine, Vol. 64 No. 5 (1985)
MBL	Mannose binding lectin	MyBiosource	Recombinant	Pradhan et al.; Rheumatol Int (2013) 33:1193–1199; DOI 10.1007/s00296-012-2519-9; Seelen et al.; Clin Exp Immunol 2003; 134:335–343
Mi2	Mi-2 protein	Thermo Fisher Scientific	Recombinant	Seelig et al.; ARTHRITIS & RHEUMATISM Vol. 38, No. 10, October 1995, pp 1389-1399
Nucleosome	Nucleosome	Arotec Diagnostics	Native	Gutiérrez-Adrianzén et al.; J Rheumatol 2006;33;1538-1544
PCNA	Proliferating cell nuclear antigen	Thermo Fisher Scientific	Recombinant	Mahler et al.; Autoimmunity Reviews 11 (2012) 771–775

Supplementary Table 1

Overview of antigens used on the microarray. (*continued*)

Abbreviation	Antigen	Source	Native/ Recombinant	Reference for inclusion in the study
Pentraxin3	Pentraxin 3	MyBiosource	Recombinant	Augusto et al., Rheumatology 2009;48:442–444; doi:10.1093/rheumatology/ken507
PmScl100	PM/Scl complex	Thermo Fisher Scientific	Recombinant	Mahler et al.; Autoimmunity Reviews 6 (2007) 432–437
RA33	Heterogeneous ribonucleoprotein A2/B1 (hnRNPA2B1)	Thermo Fisher Scientific	Native	Steiner et al.; Molecular Biology Reports 23:167-171, 1996.
RipP0	Ribosomal protein P0	Thermo Fisher Scientific	Recombinant	Heinlen et al.; Mol Med (Berl). 2010 July ; 88(7): 719–727. doi:10.1007/s00109-010-0618-1.; Shi et al.; Lupus (2015) 24, 588–596
RipP0pept	Ribosomal protein P0 peptide	Thermo Fisher Scientific	Peptide	Heinlen et al.; Mol Med (Berl). 2010 July ; 88(7): 719–727. doi:10.1007/s00109-010-0618-1.; Shi et al.; Lupus (2015) 24, 588–596
RipP1	Ribosomal protein P1	Thermo Fisher Scientific	Recombinant	Heinlen et al.; Mol Med (Berl). 2010 July ; 88(7): 719–727. doi:10.1007/s00109-010-0618-1.; Shi et al.; Lupus (2015) 24, 588–596
RipP2	Ribosomal protein P2	Thermo Fisher Scientific	Recombinant	Heinlen et al.; Mol Med (Berl). 2010 July ; 88(7): 719–727. doi:10.1007/s00109-010-0618-1.; Shi et al.; Lupus (2015) 24, 588–596
RNAPolIII	RNA polymerase III	Thermo Fisher Scientific	Recombinant	Cavazzana et al.; Autoimmunity Reviews 8 (2009) 580–584
RNP70	Ribonucleoprotein 70	Thermo Fisher Scientific	Recombinant	Migliorini et al.; Autoimmunity, 38:1, 47-54, DOI: 10.1080/08916930400022715
RNPA	Ribonucleoprotein A	Thermo Fisher Scientific	Recombinant	Migliorini et al.; Autoimmunity, 38:1, 47-54, DOI: 10.1080/08916930400022715
RNPC	Ribonucleoprotein C	Thermo Fisher Scientific	Recombinant	Migliorini et al.; Autoimmunity, 38:1, 47-54, DOI: 10.1080/08916930400022715
Ro52	52kDa Ro Protein	Thermo Fisher Scientific	Recombinant	Hochberg et al.; Medicine, Vol. 64 No. 5 (1985)

Supplementary Table 1

Overview of antigens used on the microarray. (*continued*)

Abbreviation	Antigen	Source	Native/ Recombinant	Reference for inclusion in the study
Ro60	60kDa Ro Protein	Thermo Fisher Scientific	Recombinant	Hochberg et al.; Medicine, Vol. 64 No. 5 (1985)
RPP25ThTo	Ribonuclease P and MRP subunit P25	Thermo Fisher Scientific	Recombinant	Mahler et al.; Autoimmunity Reviews 14 (2015) 254–257
Scl70	Topoisomerase I	Thermo Fisher Scientific	Recombinant	Mahler et al.; Autoimmunity Reviews 9 (2010) 756–760
SmBB	Small nuclear ribonucleoprotein polypeptides B and B'	Thermo Fisher Scientific	Native	Salmhofer et al.; JEADV 2007, 21, 1047–1053
SMP	Small nuclear ribonucleoprotein	Undisclosed due to contractual agreement	Recombinant	Migliorini et al.; Autoimmunity, 38:1, 47-54, DOI: 10.1080/08916930400022715
TIF1gamma	Transcriptional intermediary factor 1-gamma	Diarect	Recombinant	Hoshino et al.; Rheumatology 2010;49:1726–1733 doi:10.1093/rheumatology/keq153
TPO	Thyroid peroxidase	Thermo Fisher Scientific	Recombinant	Posselt et al.; Sao Paulo Med J. 2017; 135(6):535-40
tTG	Tissue transglutaminase	Thermo Fisher Scientific	Native	Marai et al.; L upus (2004) 13, 241-244

Supplementary Table 2

Overlap of IMIDs in the cohort

Disease	Number	No overlap	APS	CLE	GCA	GPA	IIM	JIA	LLD	MCTD	PMR	pSS	RA	SLE	SSc	UCTD
APS	124	52	X	5	0	0	1	0	12	1	0	1	2	64	0	0
CLE	143	67	5	X	0	0	4	0	9	2	0	8	5	59	0	1
GCA	6	5	0	0	X	0	0	0	0	0	1	0	0	0	0	0
GPA	7	6	0	0	0	X	0	0	0	0	0	1	0	1	0	0
IIM	31	18	1	4	0	0	X	0	0	4	1	2	0	7	2	1
JIA	22	16	0	0	0	0	0	X	0	3	0	0	1	5	0	0
LLD	57	14	12	9	0	0	0	0	X	2	1	1	2	28	1	1
MCTD	41	15	1	2	0	0	4	3	2	X	0	4	1	20	1	0
PMR	19	13	0	0	1	0	1	0	1	0	X	1	2	0	0	0
pSS	80	37	1	8	0	1	2	0	1	4	1	X	9	20	4	0
RA	68	32	2	5	0	0	0	1	2	1	2	9	X	21	0	0
SLE	484	288	64	59	0	1	7	5	28	20	0	20	21	X	4	5
SSc	30	21	0	0	0	0	2	0	1	1	0	4	0	4	X	0
UCTD	11	4	0	1	0	0	1	0	1	0	0	0	0	5	0	X

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(9):1151-1159. doi:10.1136/annrheumdis-2018-214819
2. Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. *J Rheumatol*. 2002;29(2):288-291.
3. Yung S, Chan TM. Mechanisms of kidney injury in lupus nephritis - the role of anti-dsDNA antibodies. *Front Immunol*. 2015;6(SEP):1-11. doi:10.3389/fimmu.2015.00475
4. Tan EM, Cohen AS, Fries JF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum*. 1982;25(11):1271-1277. doi:10.1002/art.1780251101
5. Petri M, Orbai AM, Alarcõn GS, et al. Derivation and validation of the systemic lupus international collaborating clinics classification criteria for systemic lupus erythematosus. *Arthritis Rheum*. 2012;64(8):2677-2686. doi:10.1002/art.34473
6. Isenberg DA, Manson JJ, Ehrenstein MR, Rahman A. Fifty years of anti-ds DNA antibodies: Are we approaching journey's end? *Rheumatology (Oxford)*. 2007;46(7):1052-1056. doi:10.1093/rheumatology/kem112
7. Didier K, Bolko L, Giusti D, et al. Autoantibodies associated with connective tissue diseases: What meaning for clinicians? *Front Immunol*. 2018;9(MAR):1-20. doi:10.3389/fimmu.2018.00541
8. Beurskens FJ, van Schaarenburg RA, Trouw LA. C1q, antibodies and anti-C1q autoantibodies. *Mol Immunol*. 2015;68(1):6-13. doi:10.1016/j.molimm.2015.05.010
9. Ten Berg MJ, Huisman A, Van Den Bemt PMLA, Schobben AFAM, Egberts ACG, Van Solinge WW. Linking laboratory and medication data: New opportunities for pharmacoepidemiological research. *Clin Chem Lab Med*. 2007;45(1):13-19. doi:10.1515/CCLM.2007.009
10. Brunekreef TE, Otten HG, Bosch SC, et al. Text Mining of Electronic Health Records Can Accurately Identify and Characterize Patients With Systemic Lupus Erythematosus. *ACR Open Rheumatol*. 2021;0(0):1-7. doi:10.1002/acr2.11211
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(5):R201. doi:10.1186/ar4039
12. Klinman DM, Yi AK, Beaucage SL, Conover J, Krieg AM. CpG motifs present in bacterial DNA rapidly induce lymphocytes to secrete interleukin 6, interleukin 12, and interferon γ. *Proc Natl Acad Sci U S A*. 1996;93(7):2879-2883. doi:10.1073/pnas.93.7.2879
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(10):6262-6268. doi:10.4049/jimmunol.0901941
14. Crow MK. Type I Interferon in the Pathogenesis of Lupus. *J Immunol*. 2014;192(12):5459-5468. doi:10.4049/jimmunol.1002795
15. Papayannopoulos V. Neutrophil extracellular traps in immunity and disease. *Nat Rev Immunol*. 2018;18(2):134-147. doi:10.1038/nri.2017.105
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(3):1-12. doi:10.1111/cpr.12579
17. Zhang Q, Raoof M, Chen Y, et al. Circulating mitochondrial DAMPs cause inflammatory responses to injury. *Nature*. 2010;464(7285):104-107. doi:10.1038/nature08780
18. Roger AJ, Muñoz-Gómez SA, Kamikawa R. The Origin and Diversification of Mitochondria. *Curr Biol*. 2017;27(21):R1177-R1192. doi:10.1016/j.cub.2017.09.015

19. Yousefi S, Stojkov D, Germic N, et al. Untangling "NETosis" from NETs. *Eur J Immunol*. 2019;49(2):221-227. doi:10.1002/eji.201747053
20. Becker Y, Loignon R claude, Julien A sophie, et al. Anti-mitochondrial autoantibodies in systemic lupus erythematosus and their association with disease manifestations. *Sci Rep*. 2019;9(1):4530. doi:10.1038/s41598-019-40900-3
21. Karbach J, Neumann A, Wahle C, Brand K, Gnjatic S, Jäger E. Therapeutic administration of a synthetic CpG oligodeoxynucleotide triggers formation of anti-CpG antibodies. *Cancer Res*. 2012;72(17):4304-4310. doi:10.1158/0008-5472.CAN-12-0257
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(3):1341 LP - 1345.
23. Uccellini MB, Busto P, Debatis M, Marshak-Rothstein A, Viglianti GA. Selective binding of anti-DNA antibodies to native dsDNA fragments of differing sequence. *Immunol Lett*. 2012;143(1):85-91. doi:10.1016/j.imlet.2012.01.003
24. Sekigawa I, Nawata M, Seta N, Yamada M, Iida N, Hashimoto H. Cytomegalovirus infection in patients with systemic lupus erythematosus. *Clin Exp Rheumatol*. 2002;20(4):559-564.
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(4):467-475. doi:10.1007/s10238-017-0456-3
26. Su BYJ, Su CY, Yu SF, Chen CJ. Incidental discovery of high systemic lupus erythematosus disease activity associated with cytomegalovirus viral activity. *Med Microbiol Immunol*. 2007;196(3):165-170. doi:10.1007/s00430-007-0040-7
27. Trendelenburg M, Lopez-Trascasa M, Potlukova E, et al. High prevalence of anti-C1q antibodies in biopsy-proven active lupus nephritis. *Nephrol Dial Transplant*. 2006;21(11):3115-3121. doi:10.1093/ndt/gfl436
28. Alba P, Bento L, Cuadrado MJ, et al. Anti-dsDNA, anti-Sm antibodies, and the lupus anticoagulant: significant factors associated with lupus nephritis. *Ann Rheum Dis*. 2003;62(6):556-560. doi:10.1136/ard.62.6.556
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(8):638-644. doi:10.1191/096120399680411443

4

Microarray analysis of autoantibodies can identify future Systemic Lupus Erythematosus patients

Authors:
T.E. Brunekreef[a], L.C. Reteig[b,c], M. Limper[a], S. Haitjema[b], J. Dias[d], L. Mathsson-Alm[d], J.M. van Laar[a], H.G. Otten[c]
Affiliations:
[a] Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht, The Netherlands
[b] Central Diagnostic Laboratory, University Medical Center Utrecht, Utrecht, The Netherlands
[c] Center for Translational Immunology, University Medical Center Utrecht, Utrecht, The Netherlands
[d] Thermo Fisher Scientific, Uppsala, Sweden;

Published
Hum Immunol. 2022 Jun;83(6):509-514. doi: 10.1016/j.humimm.2022.03.010

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 analyse 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.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a heterogeneous immune mediated inflammatory disease (IMID), which can sometimes 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 sometimes 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, 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 double-stranded DNA (anti-dsDNA), have a place in the regular diagnostic process.[5,8] Previous studies 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.

METHODS

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.

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.

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.

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.

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 hours of incubation with the blood sample. The correlations between the antigens on the microarray and their commercially available counterparts were assessed previous to this study and were found to be generally high (Spearman's correlation coefficients 0.58-0.88). Reproducibility of the microarray was also studied 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 detailed description of the methods used for this analysis has been described previously.[14]

STATISTICAL ANALYSIS

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.

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 (λ) 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 the anti-dsDNA on the microarray.

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 max, i.e. the smallest value of lambda where all regression coefficients are set to 0, to 1 thousandth of 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]

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.

RESULTS

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.

Table 1: Baseline characteristics

Baseline characteristics of all included patient groups. The Pre-SLE samples are recorded in the IMID and Rest patient groups (see Table 2).

Group	Samples analyzed	Female	Mean age ± SD	Anti-dsDNA positive	Samples analyzed
SLE	483	413 (86%)	43,3 ± 14,7	170 (35%)	483
IMID	346	270 (78%)	46,2 ± 17,6	26 (8%)	346
Non-IMID	218	111 (51%)	43,9 ± 17,4	7 (3%)	218
Rest	472	248 (53%)	46,0 ± 17,3	15 (3%)	472
BBD	361	180 (50%)*	46,7 ± 14,1	X	361

Dimension reduction

The antibody profile of the different patient groups was visualized by reducing the full 57-dimensional representation to two dimensions using UMAP (Figure 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.

DEVELOPMENT OF PREDICTION MODELS

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, Figure 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. Figure 1a).

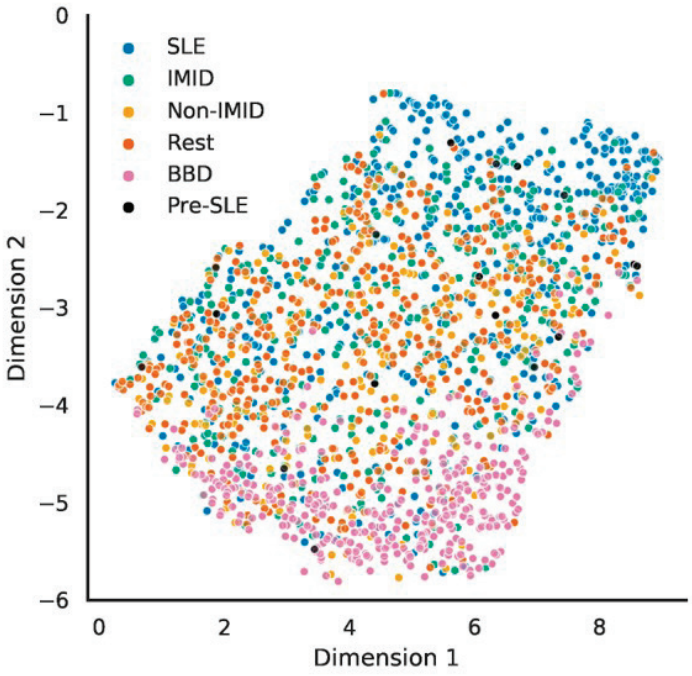


Figure 1: Uniform Manifold Approximation and Projection (UMAP) dimension reduction

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.

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, Figure 2b). Fifteen antibodies had non-zero coefficients in this final 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. Figure 1b).

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, and ten patients were already 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 (59%) had no previous IMID diagnosis, seven patients (41%) 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 (35%) were diagnosed with SLE within 2 months of sample collection. All of the fourteen pre-SLE patients tested for ANA (antinuclear antibodies) were positive or weak positive for ANA but only six patients (35%) were positive for anti-dsDNA in regular laboratory testing at the moment of sample collection (Table 2). In patients with SLE, 128 out of 147 (87%) patients tested for ANAs were positive or weak positive for ANA and 170 out of all 483 SLE patients (35%) were positive 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, Figure 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 0.84 resulted in a sensitivity of 53%, with a PPV of 25% and a NPV of 98%.

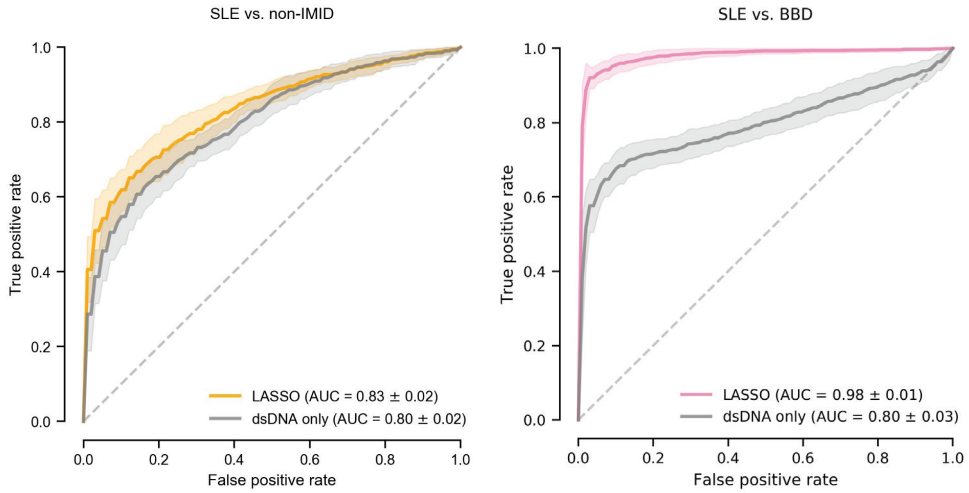


Figure 2: Receiver Operating Characteristic (ROC) curves

Figure 2a: 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).

Figure 2b: 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).

Table 2: Baseline characteristics of the pre-SLE patients

Baseline characteristics of pre-SLE patients. Seven patients had a previous diagnosis before being diagnosed with SLE, 10 others did not. LLD: Lupus-like disease, APS: anti-phospholipid syndrome, CLE: Cutaneous lupus erythematosus, RA: Rheumatoid arthritis, pSS: (primary) Sjögren's Syndrome.

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

Number of patients per patient group predicted as pre-SLE by the non-IMID prediction model, using two different cut-off scores of 0.50 and 0.84. When the rest group is used as a negative control group to calculate sensitivity and specificity for the pre-SLE patients (positive), this results in a sensitivity of 53% and a specificity of 94% for the 0.84 cut-off.

Group (n)	Positive with threshold score 0.50 N (%)	Positive with threshold score 0.84 N (%)
Pre-SLE (17)	15 (88%)	9 (53%)
IMID (346)	216 (62%)	45 (13%)
LLD (28)	20 (71%)	7 (25%)
Rest (462)	227 (47%)	27 (6%)
BBD (361)	152 (42%)	2 (1%)

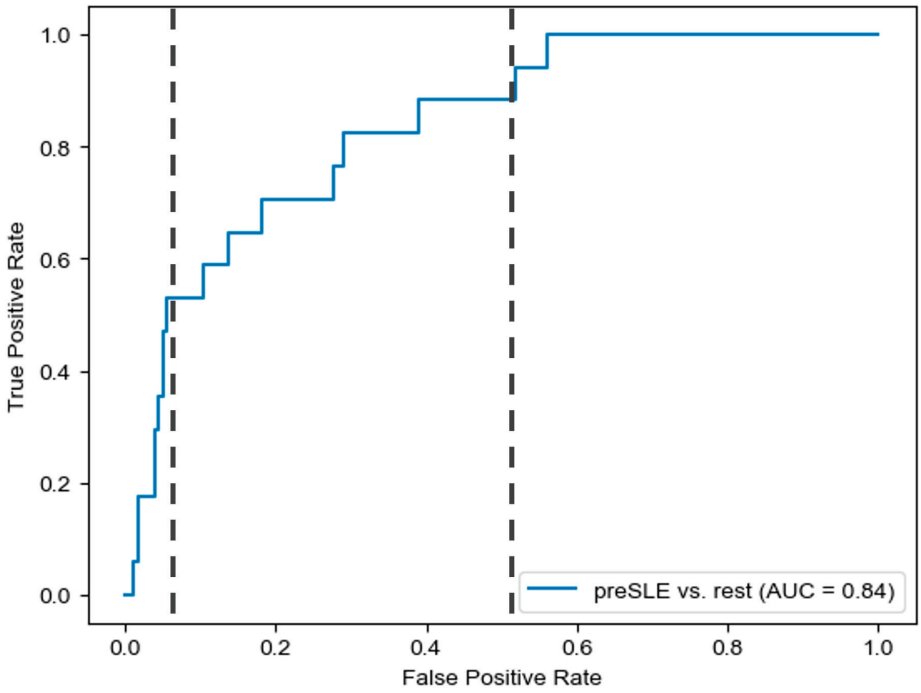


Figure 3: ROC curve for identification of future SLE patients

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.

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%). The major drawback of this strategy would be that many patients would be classified as "high-risk" patients, possibly remaining in follow-up despite never developing SLE. Although the model with higher specificity has a 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. From a cost-effectiveness point of view, implementation of such a strategy would include a rise in the initial diagnostic costs, but would potentially result in lower long-term costs by reducing the number of patients in follow-up. Furthermore, the costs of the microarray could potentially be reduced. In this study we used an investigative microarray containing multiple autoantibodies which were not used in the final model. These could be omitted from future iterations of the microarray to make it more cost-effective.

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 previ-

ous 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).

In our study, we used the anti-dsDNA on the microarray for comparison to the Lasso model, to show the added value of a multivariable approach over a single antibody approach. As the microarray anti-dsDNA is included in the Lasso model, and anti-dsDNA measured by ELISA is not, this eliminates the possibility of variability between different tests for this comparison. Although the non-IMID model using the anti-dsDNA ELISA had a slightly higher AUC (0.82) than the model using the anti-dsDNA of the microarray (0.75), at the recommended cut-off this would result in similar sensitivity and specificity for both models.

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 (Figure 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 (Figure 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. Unfortunately, we were not able to assess the effect of immunological treatment on the presence of autoantibodies because of missing or incomplete medication data in the EHR. Our study is also 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).

COMPETING INTERESTS

LM-A and JD are employed by Thermo Fisher Scientific. JL reports grants from Astra Zeneca, MSD, Thermo Fisher Scientific.

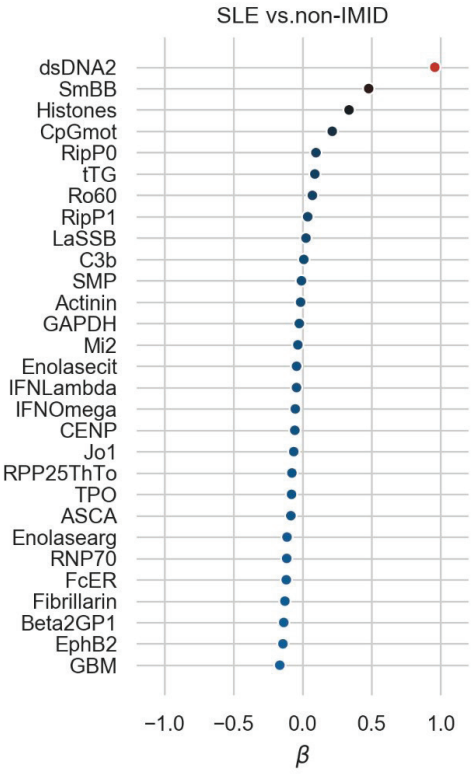
ACKNOWLEDGMENTS

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

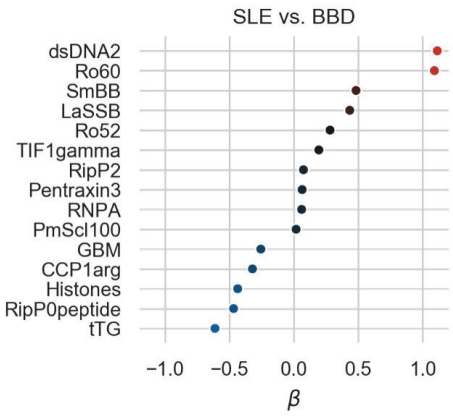
DATA AVAILABILITY STATEMENT

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

SUPPLEMENTARY MATERIAL



Supplementary Figure 1a



Supplementary Figure 1b

Supplementary Figure 1: Feature importance of the Lasso models.

Suppl. Figure 1a: Feature importance (the strength of the effect of the individual variable on the dependent variable, standardized β) of the Lasso model for identification of SLE samples using the non-IMID patients as a negative control group.

Suppl. Figure 1b: Feature importance of the Lasso model for identification of SLE samples using the BBD as a negative control group.

REFERENCES

1. Nossent J, Kiss E, Rozman B, et al. Disease activity and damage accrual during the early disease course in a multinational inception cohort of patients with systemic lupus erythematosus. *Lupus*. 2010;19(8):949-956. doi:10.1177/0961203310366572
2. Tsang-A-Sjoe MWP, Bultink IEM, Heslinga M, Voskuyl AE. Both prolonged remission and Lupus Low Disease Activity State are associated with reduced damage accrual in systemic lupus erythematosus. *Rheumatology (Oxford)*. 2017;56(1):121-128. doi:10.1093/rheumatology/kew377
3. Doria A, Zen M, Canova M, et al. SLE diagnosis and treatment: When early is early. *Autoimmun Rev*. 2010;10(1):55-60. doi:10.1016/j.autrev.2010.08.014
4. Morgan C, Bland AR, Maker C, Dunnage J, Bruce IN. Individuals living with lupus: findings from the LUPUS UK Members Survey 2014. *Lupus*. 2018;27(4):681-687. doi:10.1177/0961203317749746
5. 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(9):1151-1159. doi:10.1136/annrheumdis-2018-214819
6. Costenbader KH, Schur PH. We need better classification and terminology for "people at high risk of or in the process of developing lupus." *Arthritis Care Res (Hoboken)*. 2015;67(5):593-596. doi:10.1002/acr.22484
7. Mosca M, Costenbader KH, Johnson SR, 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*. 2019;71(1):91-98. doi:10.1002/art.40674
8. Didier K, Bolko L, Giusti D, et al. Autoantibodies associated with connective tissue diseases: What meaning for clinicians? *Front Immunol*. 2018;9(MAR):1-20. doi:10.3389/fimmu.2018.00541
9. Eriksson C, Kokkonen H, Johansson M, Hallmans G, Wadell G, Rantapää-Dahlqvist S. Autoantibodies predate the onset of systemic lupus erythematosus in northern Sweden. *Arthritis Res Ther*. 2011;13(1):1-8. doi:10.1186/ar3258
10. Arbuckle MR, McClain MT, Rubertone M V., et al. Development of Autoantibodies before the Clinical Onset of Systemic Lupus Erythematosus. *N Engl J Med*. 2003;349(16):1526-1533. doi:10.1056/nejmoa021933
11. Nielen MMJ, Van Schaardenburg D, Reesink HW, et al. Specific Autoantibodies Precede the Symptoms of Rheumatoid Arthritis: A Study of Serial Measurements in Blood Donors. *Arthritis Rheum*. 2004;50(2):380-386. doi:10.1002/art.20018
12. Ten Berg MJ, Huisman A, Van Den Bemt PMLA, Schobben AFAM, Egberts ACG, Van Solinge WW. Linking laboratory and medication data: New opportunities for pharmacoepidemiological research. *Clin Chem Lab Med*. 2007;45(1):13-19. doi:10.1515/CCLM.2007.009
13. Brunekreef TE, Otten HG, Bosch SC, et al. Text Mining of Electronic Health Records Can Accurately Identify and Characterize Patients With Systemic Lupus Erythematosus. *ACR Open Rheumatol*. 2021;0(0):1-7. doi:10.1002/acr2.11211
14. Brunekreef T, Limper M, Melchers R, et al. Microarray testing in patients with systemic lupus erythematosus identifies a high prevalence of CpG DNA-binding antibodies. *Lupus Sci Med*. 2021;8(1). doi:10.1136/lupus-2021-000531
15. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv*. Published online February 9, 2018.

16. Seabold S, Perktold J. Statsmodels: Econometric and Statistical Modeling with Python. In: *Proceedings of the 9th Python in Science Conference*. 2010:92-96. doi:10.25080/Majora-92bf1922-011
17. Pedregosa F, Varoquaux G, Gramfort A, et al. Journal of Machine Learning Research: Preface. *J Mach Learn Res*. 2011;12:2825-2830.
18. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *J Roy Sta Soc: Ser B (Methodol)*. 1996;58(1):267-288.
19. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010;33(1):1-22. doi:10.18637/jss.v033.i01
20. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. Springer New York; 2009. doi:10.1007/978-0-387-84858-7
21. Lambers WM, Westra J, Jonkman MF, Bootsma H, de Leeuw K. Incomplete Systemic Lupus Erythematosus: What Remains After Application of American College of Rheumatology and Systemic Lupus International Collaborating Clinics criteria? *Arthritis Care Res (Hoboken)*. 2020;72(5):607-614. doi:10.1002/acr.23894
22. Putterman C, Wu A, Reiner-Benaim A, et al. SLE-key® rule-out serologic test for excluding the diagnosis of systemic lupus erythematosus: Developing the ImmunArray iCHIP®. *J Immunol Methods*. 2016;429:1-6. doi:10.1016/j.jim.2015.12.003
23. Fattal I, Shental N, Mevorach D, et al. An antibody profile of systemic lupus erythematosus detected by antigen microarray. *Immunology*. 2010;130(3):337-343. doi:10.1111/j.1365-2567.2010.03245.x
24. Radin M, Sciascia S, Cecchi I, Mengozzi G, Roccatello D. 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*. 2019;68(July):e12-e14. doi:10.1016/j.ejim.2019.07.026

5

The Predictive Value of Antinuclear Antibodies in Patients with Suspected Connective Tissue Disease for the Development of Connective Tissue Diseases

Authors:
T.E. Brunekreef[a], H.B. van den Heuvel[a], M. Limper[a], H. Leavis[a], J.M. van Laar[a], H.G. Otten[b]

Affiliations:
[a] Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht, The Netherlands
[b] Center for Translational Immunology, University Medical Center Utrecht, Utrecht, The Netherlands

Published
Clin Exp Rheumatol. 2024 Jan;42(1):24-29. doi: 10.55563/clinexprheumatol/k2xvu8.

ABSTRACT

Introduction

Early confirmation of the diagnosis of connective tissue diseases (CTD) is important, as prolonged disease activity can result in irreversible organ damage. Although antinuclear antibodies (ANAs) have been shown to precede the diagnosis of SLE, this has not been investigated in large cohorts for other CTDs. In this study, we investigated whether the presence of antinuclear autoantibodies in undiagnosed patients suspected of having CTDs is predictive of development of a future CTD.

Methods

We screened the Electronic Health Records of a cohort of 1030 patients, who were tested for ANAs and their specificity in 2013/2014, to evaluate whether new CTD diagnoses had been recorded by a clinician between the original blood draw date and 2020. We compared the prevalence of ANAs in patients who developed a new CTD diagnosis during follow-up with patients with similar symptoms at baseline who did not receive a subsequent CTD diagnosis and with patients with an established CTD at baseline.

Results

Sixteen out of 1030 patients had developed a new CTD in the studied time period. The mean time period between baseline blood draw and subsequent CTD diagnosis of these patients was approximately 2.3 years. Eleven out of 16 (69%) newly diagnosed patients had positive ANA screening tests, compared to 54% of patients with a CTD diagnosis at baseline (p=ns) and 30% of symptomatic undiagnosed patients ($p<0.001$). This resulted in a positive predictive value (PPV) of 7% and a negative predictive value (NPV) of 98% for the development of a new CTD in undiagnosed symptomatic patients. For specific ANAs associated with the suspected CTD, the PPV was 12%, with a NPV of 98%.

Discussion

Progression to a CTD diagnosis is rare in undiagnosed patients. Undiagnosed patients with symptoms associated with a CTD who progress to a CTD more often have ANAs than patients with similar symptoms who do not progress to a CTD. ANA testing can be used to more stringently select patients who should remain in follow-up.

INTRODUCTION

Early confirmation of the diagnosis of connective tissue diseases (CTDs) is important to limit uncertainty for patients and prevent delay of the diagnosis and subsequent treatment, which can result in irreversible damage caused by prolonged disease activity.[1–3] In practice however, not all patients will exhibit sufficient symptoms to be diagnosed with a CTD at their initial visit to the Rheumatology outpatient clinic. Some CTD-associated symptoms, such as sicca symptoms or Raynaud's phenomenon, are commonly considered to be a prelude to a CTD diagnosis, however it is currently not possible to predict which symptomatic patients will progress to a CTD and which patients' symptoms will resolve over time.[4,5] This results in many patients remaining in follow-up, despite never developing a CTD. Reliable identification of future patients would result in earlier treatment of CTD patients and limit the number of patients without CTDs remaining in unnecessary follow-up.

Antinuclear antibodies (ANAs) are an important tool in the diagnostic process of many CTDs, as many CTDs are associated with either one individual or a combination of ANAs.[6] ANAs have been included in classification criteria of several CTDs and testing for ANAs is therefore often part of the routine diagnostic workup of patients suspected of having a CTD.[7–9] Furthermore, previous studies have shown that ANAs can already be present in the serum of CTD patients several years before diagnosis, similar to rheumatoid factor and anti-cyclic citrullinated peptides antibodies in patients with rheumatoid arthritis.[10–13] For example, anti-Ro and anti-La antibodies have been shown to be present in the serum of patients with Sjogren's syndrome over 10 years before diagnosis and anti-dsDNA antibodies have been shown in patients with SLE several years before the first clinical manifestations of the disease constitute.[14,15] This would suggest great potential of ANA testing for early diagnostics and prediction of development of CTDS as well. However, most of these studies retrospectively analyzed serum samples of confirmed CTD patients without clinically relevant control groups. As ANAs are also present in approximately 5-20% of healthy individuals, this raises the question whether these ANAs are useful in clinical practice as predictors of future development of CTDs.[16,17]

Currently, it is unclear whether positive ANA tests in patients suspected of having a CTD but with insufficient features to be diagnosed should be seen as pre-clinical immunological manifestations of a future CTD or as clinically irrelevant. In this study, we investigated whether the presence of ANAs in undiagnosed patients with symptoms associated with CTDs is predictive of a future CTD diagnosis.

METHODS

This study continues a previous study by our group, performed between April 2013 and May 2014 in the University Medical Center Utrecht (UMC Utrecht), in which information about the diagnosis and presence of ANAs of included patients, present at the time of inclusion, were recorded.[18] We reevaluated the electronic health record (EHR) of all patients in this cohort to investigate whether new clinical diagnoses had been made between the original serum sample collection date and January 1st 2020. This study was in accordance with the guidelines of the Medical Ethics Committee of the UMC Utrecht. The requirement for collection of informed consent was waived by the Medical Ethics Committee.

Patients

Between April 2013 and May 2014, 1030 individual patients in the UMC Utrecht were tested for ANAs, as requested by the treating physician. These patients included both patients with a confirmed CTD and undiagnosed patients suspected of having a CTD. For this cohort of patients, the reason for the ANA test request was recorded by a medical specialist after a review of the EHR.

Antinuclear antibody analysis

In the original study by our group, all collected serum samples were analyzed for the presence of ANAs using two different screening tests (ANA-IIF using Hep-20-10 cell lines (EuroImmun, Lubeck, DE) and CTD screen (Thermo Fisher Scientific, Freiburg, DE)). If either of the screening ANA tests was positive, the sample was subsequently analyzed to identify the individual antibody specificities using the Profile 3 line blot (EuroImmun) and CTD single analytes (Thermo Fisher Scientific).

For this study, ANA tests performed in regular care after the 2013-2014 study had ended, were evaluated to investigate whether the prevalence of ANAs changes over time. These tests were only performed with the ANA-IIF screening test and, if positive, followed by analysis of specific ANAs with the Profile 3 line blot.

Evaluation of clinical diagnosis

We used a text mining algorithm to evaluate the clinical diagnosis recorded in the EHR at baseline for all patients, as well as the clinical diagnosis recorded in the EHR on January 1st 2020. The CTD diagnoses included in this study were rheumatoid arthritis (RA), juvenile idiopathic arthritis (JIA), systemic lupus erythematosus (SLE), (primary) Sjogren's syndrome (pSS), systemic sclerosis (SSc), polymyositis or dermatomyositis (PM/DM), mixed connective tissue disease (MCTD) and undifferentiated connective tissue disease (UCTD). In all cases with discrepancies between the results from the baseline evaluation and follow-up evaluation, patient charts were manually reviewed to confirm presence of

a new CTD diagnosis. A full description of the text mining algorithm has been published previously.[19]

Patients were divided into four groups: Patients with an established CTD diagnosis or diagnosis made at the time of the baseline ANA test (Group A), patients who received a new CTD diagnosis during follow-up (Group B), patients with signs and symptoms suggestive of a CTD at baseline, but without a CTD diagnosis during follow-up (Group C) and patients for whom the ANA test had been requested for a reason unlikely to be caused by a CTD (Group D). Patients with a diagnosis made within one month after sample collection were included into Group A, rather than Group B, because they were already in the diagnostic process to confirm the diagnosis. Therefore, these patients did not develop a new CTD to our definition, but rather got a fitting diagnosis for symptoms they were already experiencing.

Statistics

Statistical analysis was performed with SPSS (IBM, version 26.0.0.1). Categorical analysis was performed with Chi-square test or Fisher's exact test, whenever appropriate.

RESULTS

A total of 1030 patients were included in this study. At baseline, the mean age of all patients was 41 years old (SD 20.9), 61.2% were female. In the original study, an established CTD was reported as the reason for ANA testing for 162 patients. Patient chart review performed for this study showed that 49 patients, originally classified in other groups, were diagnosed within one month after sample collection or had previously been diagnosed with a CTD, although this had not been the reason for ANA testing. For this study, this resulted in a total of 211 patients with a CTD diagnosis at baseline. 383 patients had at least one sign or symptom strongly associated with a CTD, such as arthralgia or myalgia, sicca or skin abnormalities compatible with lupus. 228 patients had symptoms that could be related to a CTD but are associated with a CTD to a lesser extent, such as psychiatric disease, uveitis or peripheral neuropathy, resulting in low clinical suspicion. 65 patients had abnormal liver function tests and were tested for the presence of autoimmune hepatitis, and 143 patients were tested for miscellaneous reasons, such as screening for a heart transplantation, humoral immunodeficiency or erythema nodosum. For this study, these last three groups were combined in Group D.

Out of all patients, sixteen had developed a new CTD diagnosis between baseline and 2020. Fourteen of these sixteen patients originally had symptoms strongly associated with a CTD at baseline but did not exhibit symptoms sufficient for a CTD diagnosis yet. One patient had been diagnosed with dermatomyositis and received an overlapping diagnosis of systemic sclerosis, and one patient was tested for ANAs in the original study because of abnormal liver function tests and was later diagnosed with rheumatoid arthritis. This resulted in 210 patients in Group A (established CTD diagnosis), 16 patients in Group B (future CTD patients), 369 patients in Group C (suspect symptoms at baseline, but no future CTD diagnosis), and 435 patients in Group D (patients with low clinical suspicion of a CTD at baseline).

All patients in Group B were female, for patients in Group A, Group C, and Group D the percentage of female patients was 75%, 65% and 50% respectively (Table 1).

The reason for the ANA test at baseline and the diagnosed CTD received during follow-up of all patients from Group A are shown in Table 2. The mean time between the ANA test and diagnosis was approximately 2.3 years, ranging from 52 days up to 5.6 years.

Antinuclear antibodies

Out of all 383 patients with signs and symptoms of a CTD at baseline, but no diagnosis (yet), 121 patients (32%) tested positive for ANAs at baseline. Nine of these patients developed a new CTD during follow-up (positive predictive value (PPV) = 7%). Five patients of this group tested negative for ANAs at baseline but developed a new CTD, resulting in a negative predictive value (NPV) of 98%).

Table 1: Demographics

Demographics of all patient groups.

Group	Nummer of patients	Mean age	Female (%)
Group A	210	40.8 (±22.3)	75%
Group B	16	41.7 (±16.9)	100%
Group C	369	40.1 (±20.0)	65.3%
Group D	435	41.7 (±21.2)	49.7%
Total	1030	41.0 (±20.9)	61.1%

56 (15%) out of the 383 patients with signs and symptoms of a CTD at baseline, but no CTD diagnosis, had specific ANAs compatible with their suspected CTD. Seven out of these 56 (12%) patients developed a future CTD, compared to seven (2%) of the patients without compatible specific ANAs (p=0.002), resulting in a PPV of 12% and a NPV of 98%. This shows that progression to a CTD is uncommon in patients with compatible specific ANAs and very rare in patients without compatible specific ANAs.

When looking at all 4 groups, we found that 114 (54%) patients from Group A tested positive for ANAs with one or both screening ANA tests, in Group B, 11 patients (69%) tested positive for ANAs. For patients in group C and D, this were 112 (30%) and 67 (15%) patients respectively, showing that patients with either an established or future diagnosis of a CTD more often had positive results on ANA screening test, compared to the other patient groups (Figure 1, Table 3).

When we excluded patients with an (eventual) diagnoses of RA and JIA from the analysis, as there are no specific ANAs associated with these diseases and therefore ANA testing is not part of the regular diagnostic process, this resulted in 80 out of 97 (83%) of patients in Group A and 10 out of 13 (77%) patients in Group B testing positive for ANAs with a screening test (p=ns). Seven (54%) patients from Group B also had specific ANAs compatible with their future diagnosis, compared to 49 (13%) of patients from Group C having specific ANAs and presenting symptoms associated with the same CTD (p=0.001). In Group A, 67 (69%) patients had specific ANAs associated with their CTD diagnosis (vs Group B p=ns). A full overview of what specific antibodies were found in each group can be found in supplementary Table 1.

Development of autoantibodies during follow-up

A total of 434 screening ANA tests were performed for 207 patients during follow-up. Thirteen out of 140 patients (9%) with a negative screening ANA test (ANA IIF) at baseline tested positive during follow-up, including one patient with a future diagnosis of SLE. Three of these patients previously tested positive with the CTD screen test. Conversely, 10 out of 67 patients (15%) with a positive screening ANA test at baseline had a negative screening ANA test during follow-up. 262 line blot tests for the detection of specific ANAs were performed for 140 patients during follow-up. Eight patients (6%) had new positive results for one specific autoantibody, only one of which (0.7%) had potential clinical relevance (anti-SSA in a future SLE patient).

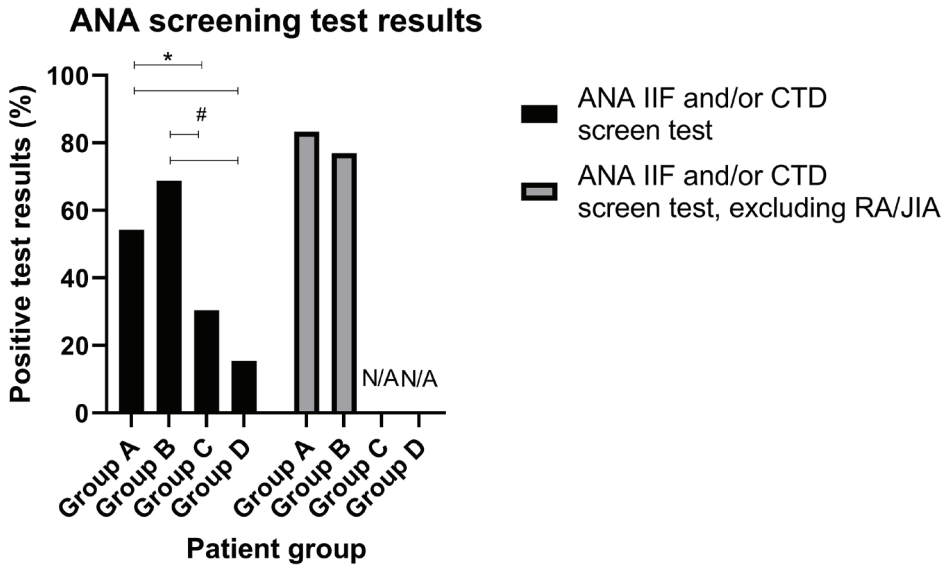


Figure 1: Prevalence of ANAs

Prevalence of positive ANA test results at baseline per patient group. The results of the ANA IIF and CTD screen tests are combined and recorded as positive when one or both tests had a positive result. The gray bar shows the prevalence when patients with RA or JIA were excluded from group A and B. Significant differences between groups are highlighted.

* = Significant difference in prevalence of ANAs between Group A and Group C ($p<0.001$) and between Group A and Group D respectively ($p<0.001$).

= Significant difference in prevalence of ANAs between Group B and Group C ($p<0.001$) and between Group B and Group D respectively ($p<0.001$).

N/A= not applicable

Table 2: Patients from group B

This Table shows the age, presenting symptoms, eventual diagnosis and the time between inclusion in the original study and diagnosis in days.

Age	Reason for baseline ANA test	Diagnosis in 2020	Time between ANA tests and diagnosis (days)
22	Serositis	MCTD	128
50	Raynaud	SSc	322
45	Dermato/polymyositis*	PM/DM and SSc	902
57	Liver function disorder	RA	239
68	Sicca	pSS and SSc	604
25	Cytopenia	SLE	1980
38	Arthralgia/myalgia	RA	610
53	Skin abnormalities compatible with SLE	PM and CLE	52
47	Sicca	SLE	362

Table 2: Patients from group B

This Table shows the age, presenting symptoms, eventual diagnosis and the time between inclusion in the original study and diagnosis in days. (*continued*)

Age	Reason for baseline ANA test	Diagnosis in 2020	Time between ANA tests and diagnosis (days)
29	APS	SLE and APS	1127
68	Sicca	pSS	1371
61	Interstitial lung disease	RA	1693
30	Cytopenia	SLE	2051
27	Arthralgia/myalgia	pSS and LLD	1257
33	Proteinuria	SLE and APS	478
14	Proteinuria	SLE	94

=* One patient with a diagnosis of dermatomyositis got the diagnosis systemic sclerosis during follow-up. This patient was included in the future patients group (group B), as patients were categorized based on clinical diagnoses made by the treating physician, regardless of ANA status.

RA: rheumatoid arthritis, SLE: systemic lupus erythematosus, pSS: (primary) Sjogren's syndrome, SSc: systemic sclerosis, PM/DM: polymyositis or dermatomyositis, MCTD: mixed connective tissue disease.

Table 3: Prevalence of antinuclear antibodies in all patient groups

Number and percentage of patients with positive ANA screening tests in all patient groups. Prevalence of ANAs in all different CTDs included in group B are shown in the bottom part of the Table.

Patient group	Patients (N)	ANA positive (ANA IIF and/or CTD screen)	ANA IIF positive	CTD screen positive	Double positive
Group A	210	114 (54,3%)	109 (51,9%)	81 (38,6%)	76 (36,2%)
Group B	16	11 (68,8%)	11 (68,8%)	7 (43,8%)	7 (43,8%)
Group C	369	112 (30,4%)	84 (22,8%)	63 (17,1%)	35 (9,5%)
Group D	435	67 (15,4%)	56 (12,9%)	22 (5,1%)	11 (2,5%)
Group A:	210	114 (54,3%)	109 (51,9%)	81 (38,6%)	76 (36,2%)
JIA	48	15 (31,3%)	15 (31,3%)	5 (10,4%)	5 (10,4%)
MCTD	6	6 (100%)	6 (100%)	6 (100%)	6 (100%)
PM/DM	11	7 (63,6%)	6 (54,5%)	3 (27,3%)	2 (18,2%)
pSS	31	27 (87,1%)	24 (77,4%)	27 (87,1%)	24 (77,4%)
RA	65	19 (29,2%)	19 (29,2%)	4 (6,2%)	4 (6,2%)
SLE	29	26 (89,7%)	25 (86,2%)	22 (75,9%)	21 (72,4%)
SSc	19	13 (68,4%)	13 (68,4%)	13 (68,4%)	13 (68,4%)
UCTD	1	1 (100%)	1 (100%)	1 (100%)	1 (100%)
Total	210	114 (54,3%)	109 (51,9%)	81 (38,6%)	76 (36,2%)
Total excluding RA and JIA	97	80 (82,5%)	75 (77,3%)	72 (74,2%)	67 (69,1%)

DISCUSSION

The main goal of this study was to investigate whether ANA test results in patients with suspected CTDs who remained undiagnosed at the time of that test could be used to predict a future CTD diagnosis. In our study, only a small percentage of suspected CTD patients progressed from undiagnosed patients with symptoms associated with CTDs to patients with a diagnosed CTD. Although our study shows an increased prevalence of ANAs in future CTD patients compared to suspected CTD patients who do not progress to a definitive CTD, we show that only a small percentage of positive ANA tests occur in future patients, indicating that most positive ANA tests performed in this patient group are clinically irrelevant. As ANA testing has a low PPV and a high NPV for development of a future CTD, this indicates that positive ANA test results do not predict development of a future CTD, but a negative ANA test results makes development of a future CTD highly unlikely.

This appears contradictory to previous studies, which have shown that ANAs can already be present in the serum of CTD patients several years before diagnosis, suggesting great positive predictive values. However, our study differs from previous studies in the way that it is a prospective cohort study, whereas most of these studies investigating the predictive value of autoantibodies for disease evolution in CTDs are retrospective.[10,11,14,15,20] This difference is also expressed in the composition of the cohort: whereas these studies look retrospectively at previous test results of patients with a confirmed CTD diagnosis, they often lack a good control group with a similar phenotype pre-diagnosis, whereas in our study these patients are also included, which makes our population more representative of that of a Rheumatology outpatient population. We conclude that although positive ANA tests are more prevalent in future patients, the positive predictive value is limited because of the moderate prevalence of ANAs in non-progressing patients suspected of having a CTD.

Despite the substantial follow-up time of approximately six years, the number of undiagnosed patients progressing to a CTD diagnosis was lower than we originally expected. Our results show that only a small fraction of undiagnosed patients out of all patients with a suspicion of a CTD will develop a CTD if not diagnosed immediately. Currently however, many patients remain in follow-up because of the uncertainty whether their symptoms should be interpreted as a prelude for a CTD or as clinically irrelevant symptoms that resolve over time. Presumably, this is especially the case for patients who test positive for ANAs. However, our study suggests that this caution is largely unwarranted, as very little patients develop an CTD in this follow-up and the PPV of a positive ANA screening test is only 7%. A possible explanation is that ANAs can be present in up to 20% of non-diseased individuals, resulting in clinically irrelevant positive test results. Our results also show that specific ANAs associated with the future diagnosis are almost as prevalent in future

patients as a general positive ANA test results and are less prevalent in patients who do not receive an eventual diagnosis. A new risk stratification strategy could involve only following undiagnosed patients in the outpatient clinic when they have positive results for ANAs associated with their suspected CTD based on clinical signs and symptoms. This is similar to the recommendation for patients with Raynaud's phenomenon, for whom a recent study showed that progression to systemic sclerosis mainly happens in patients with positive ANA tests.[21]

As mentioned previously, several studies have shown that autoantibodies can be present in the serum of CTD patients several years before diagnosis. In particular, the study by Arbuckle et al. shows that a variety of autoantibodies are often present in the serum of SLE patients many years before the diagnosis of the disease and that there is a gradual increase in seropositive patients for all autoantibodies over time.[15] The study by Arbuckle et al. also shows that a large percentage of patients are already seropositive for specific ANAs at the time of the first manifestations of SLE.[15] Consequently, this would suggest that seroconversion of ANAs precedes the first ANA test of the patient in clinical practice, as this testing only occurs after the already symptomatic patients visit the physician. In our study, we show that repeated ANA tests during follow-up are of little clinical value, as seroconversion is rare in follow-up and most of the newly found autoantibodies are clinically irrelevant. The low clinical value of repeated ANA tests was also shown in a recent study by Yeo et al., in which they showed that of the 19% of patients who seroconverted from a negative test result to a positive result, only 1.1% was clinically relevant.[22]

The main strengths of our study are the prospective design with clinically relevant control groups and the follow-up of approximately 6 years. The main limitations of our study are the low number of patients developing new CTDs and the potential of loss to follow-up. As the new diagnosis were assessed using the data from the EHR in our hospital, there is a possibility that patients who did not receive a CTD diagnosis at baseline were released from care but later diagnosed with an CTD in another hospital.

In conclusion, our study shows that progression from an undiagnosed patient seen at the outpatient clinic to an CTD is rare. Although very rarely progressing to a CTD, patients with specific ANAs associated with the suspected CTD are at higher risk of developing this disease in the future and patients with negative ANA tests are highly unlikely to develop a new CTD, which could be used to more stringently select patients to keep in follow-up. Repeated ANA tests however are not useful, as seroconversion is likely to precede the visit to the outpatient clinic. Patients with symptoms associated with CTDs but who remain undiagnosed after the initial diagnostic process and do not have specific ANAs associated with their suspected CTD can safely be released from care.

REFERENCES

1. Tsang-A-Sjoe MWP, Bultink IEM, Heslinga M, Voskuyl AE. Both prolonged remission and Lupus Low Disease Activity State are associated with reduced damage accrual in systemic lupus erythematosus. *Rheumatology (Oxford)*. 2017;56(1):121-128. doi:10.1093/rheumatology/kew377
2. Kernder A, Richter JG, Fischer-Betz R, et al. Delayed diagnosis adversely affects outcome in systemic lupus erythematosus: Cross sectional analysis of the LuLa cohort. *Lupus*. 2021;30(3):431-438. doi:10.1177/0961203320983445
3. Rosa JE, García MV, Luissi A, et al. Rheumatoid Arthritis Patient's Journey: Delay in Diagnosis and Treatment. *J Clin Rheumatol*. 2020;26(7S Suppl 2):S148-S152. doi:10.1097/RHU.0000000000001196
4. Hughes M, Allanore Y, Chung L, Pauling JD, Denton CP, Matucci-Cerinic M. Raynaud phenomenon and digital ulcers in systemic sclerosis. *Nat Rev Rheumatol*. 2020;16(4):208-221. doi:10.1038/s41584-020-0386-4
5. Kassan SS, Moutsopoulos HM. Clinical manifestations and early diagnosis of Sjögren syndrome. *Arch Intern Med*. 2004;164(12):1275-1284. doi:10.1001/archinte.164.12.1275
6. Didier K, Bolko L, Giusti D, et al. Autoantibodies associated with connective tissue diseases: What meaning for clinicians? *Front Immunol*. 2018;9(MAR):1-20. doi:10.3389/fimmu.2018.00541
7. Aringer M, Costenbader K, Daikh D, et al. 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus. *Arthritis Rheumatol*. 2019;71(9):1400-1412. doi:10.1002/art.40930
8. van den Hoogen F, Khanna D, Fransen J, et al. 2013 Classification Criteria for Systemic Sclerosis: An American College of Rheumatology/European League Against Rheumatism Collaborative Initiative. *Arthritis Rheum*. 2013;65(11):2737-2747. doi:10.1002/art.38098
9. Shiboski CH, Shiboski SC, Seror R, et al. 2016 American College of Rheumatology/European League Against Rheumatism Classification Criteria for Primary Sjögren's Syndrome: A Consensus and Data-Driven Methodology Involving Three International Patient Cohorts. *Arthritis Rheumatol*. 2017;69(1):35-45. doi:10.1002/art.39859
10. Burbelo PD, Gordon SM, Waldman M, et al. Autoantibodies are present before the clinical diagnosis of systemic sclerosis. *PLoS One*. 2019;14(3):1-18. doi:10.1371/journal.pone.0214202
11. Ma WT, Chang C, Gershwin ME, Lian ZX. Development of autoantibodies precedes clinical manifestations of autoimmune diseases: A comprehensive review. *J Autoimmun*. 2017;83:95-112. doi:10.1016/j.jaut.2017.07.003
12. Rantapää-Dahlqvist S, de Jong BAW, Berglin E, et al. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. *Arthritis Rheum*. 2003;48(10):2741-2749. doi:10.1002/art.11223
13. Nielen MMJ, Van Schaardenburg D, Reesink HW, et al. Specific Autoantibodies Precede the Symptoms of Rheumatoid Arthritis: A Study of Serial Measurements in Blood Donors. *Arthritis Rheum*. 2004;50(2):380-386. doi:10.1002/art.20018
14. Theander E, Jonsson R, Sjöström B, Brokstad K, Olsson P, Henriksson G. Prediction of Sjögren's syndrome years before diagnosis and identification of patients with early onset and severe disease course by autoantibody profiling. *Arthritis Rheumatol*. 2015;67(9):2427-2436. doi:10.1002/art.39214
15. Arbuckle MR, McClain MT, Rubertone M V., et al. Development of Autoantibodies before the Clinical Onset of Systemic Lupus Erythematosus. *N Engl J Med*. 2003;349(16):1526-1533. doi:10.1056/nejmoa021933

16. Krzemień P, Kasperczyk S, Banach M, et al. Analysis of the impact of sex and age on the variation in the prevalence of antinuclear autoantibodies in Polish population: a nationwide observational, cross-sectional study. *Rheumatol Int*. 2022;42(2):261-271. doi:10.1007/s00296-021-05033-9
17. Tan EM, Feltkamp TEW, Smolen JS, et al. Range of antinuclear antibodies in "healthy" individuals. *Arthritis Rheum*. 1997;40(9):1601-1611. doi:10.1002/art.1780400909
18. Otten HG, Brummelhuis WJ, Fritsch-Stork R, et al. Measurement of antinuclear antibodies and their fine specificities: Time for a change in strategy? *Clin Exp Rheumatol*. 2017;35(3):0462-0470.
19. Brunekreef TE, Otten HG, Bosch SC, et al. Text Mining of Electronic Health Records Can Accurately Identify and Characterize Patients With Systemic Lupus Erythematosus. *ACR Open Rheumatol*. 2021;0(0):1-7. doi:10.1002/acr2.11211
20. Radin M, Rubini E, Cecchi I, et al. Disease evolution in a long-term follow-up of 104 undifferentiated connective tissue disease patients. *Clin Exp Rheumatol*. 2022;40(3):575-580. doi:10.55563/clinexprheumatol/7vp1bo
21. Bellando-Randone S, Del Galdo F, Lepri G, et al. Progression of patients with Raynaud's phenomenon to systemic sclerosis: a five-year analysis of the European Scleroderma Trial and Research group multicentre, longitudinal registry study for Very Early Diagnosis of Systemic Sclerosis (VEDOSS). *Lancet Rheumatol*. 2021;3(12):e834-e843. doi:10.1016/S2665-9913(21)00244-7
22. Yeo AL, Le S, Ong J, et al. Utility of repeated antinuclear antibody tests: a retrospective database study. *Lancet Rheumatol*. 2020;2(7):e412-e417. doi:10.1016/S2665-9913(20)30084-9

6

Association of autoantibodies with the IFN signature and NETosis in patients with Systemic Lupus Erythematosus

Authors:
E.D. Kaan[a,b,c,1], T.E. Brunekreef[b,1], J. Drylewicz[a], L.L. van den Hoogen[a,b,2], M. van der Linden[a,3], H.L. Leavis[a,b], J.M. van Laar[b], M. van der Vlist[a,c], H.G. Otten[a,d], M. Limper[b]
Affiliations:
[a] Center for Translational Immunology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
[b] Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
[c] Oncode Institute, Utrecht, The Netherlands
[d] Central Diagnostic Laboratory, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
[1] These authors contributed equally to this article
[2] Present affiliation: department of Rheumatology, Radboud UMC and Sint Maartenskliniek, Nijmegen, The Netherlands
[3] Present affiliation: Citryll B.V., Oss, The Netherlands.

Published
J Transl Autoimmun. 2024 Jun 20:9:100246. doi: 10.1016/j.jtauto.2024.100246.

ABSTRACT

Objective

Systemic lupus erythematosus (SLE) is an autoimmune disease characterized by a variety of disease symptoms and an unpredictable clinical course. To improve treatment outcome, stratification based on immunological manifestations commonly seen in patients with SLE such as autoantibodies, type I interferon (IFN) signature and neutrophil extracellular trap (NET) release may help. It is assumed that there is an association between these immunological phenomena, since NET release induces IFN production and IFN induces autoantibody formation via B-cell activation. Here we studied the association between autoantibodies, the IFN signature, NET release, and clinical manifestations in patients with SLE.

Methods

We performed principal component analysis (PCA) and hierarchical clustering of 57 SLE-related autoantibodies in 25 patients with SLE. We correlated each autoantibody to the IFN signature and NET inducing capacity.

Results

We observed two distinct clusters: one cluster contained mostly patients with a high IFN signature. Patients in this cluster often present with cutaneous lupus, and have higher anti-dsDNA concentrations. Another cluster contained a mix of patients with a high and low IFN signature. Patients with high and low NET inducing capacity were equally distributed between the clusters. Variance between the clusters is mainly driven by antibodies against histones, RibP2, RibP0, EphB2, RibP1, PCNA, dsDNA, and nucleosome. In addition, we found a trend towards increased concentrations of autoantibodies against EphB2, RibP1, and RNP70 in patients with an IFN signature. We found a negative correlation of NET inducing capacity with anti-FcER (r= -0.530; p=0.007) and anti-PmScl100 (r= -0.445; p=0.03).

Conclusion

We identified a subgroup of patients with an IFN signature that express increased concentrations of antibodies against DNA and RNA-binding proteins, which can be useful for further patient stratification and a more targeted therapy. We did not find positive associations between autoantibodies and NET inducing capacity. Our study further strengthens the evidence of a correlation between RNA-binding autoantibodies and the IFN signature.

INTRODUCTION

Systemic lupus erythematosus (SLE) is an autoimmune disease characterized by a variety of disease symptoms and an unpredictable clinical course. While some patients experience a relatively mild disease, others develop a complicated disease course, with frequent flares that can precede organ damage. Ultimately, SLE can lead to premature death, as the result of disease activity or because of treatment side effects.[1] This underlines the urgency to identify patients that develop a complicated disease course, and the need for tailored therapy.[2,3]

Risk stratification is a tool to predict progression to severe disease. In SLE a clinically applicable stratification tool or biomarker still remains high on the agenda.[2] Several cytokine phenotypes and gene signatures have been proposed as a tool for risk stratification, but these have not found their way to clinical practice yet.[4–9] The main immunological mechanisms in SLE are formation of autoantibodies, the presence of an interferon (IFN) signature, and development of neutrophil extracellular traps (NETs).[10] Autoantibodies are a nearly omnipresent immunological manifestation in SLE patients. Antinuclear antibodies are present in most patients, although there is individual variation in the specific autoantigens that these autoantibodies recognize. Some autoantibodies are associated with specific disease manifestations. For example, anti-C1q antibodies are associated with lupus nephritis, and a rise in the serum concentration of anti-dsDNA antibodies is often seen in patients during flares.[11,12]

Besides autoantibody production, approximately half of all SLE patients show increased expression of genes regulated by proinflammatory type 1 IFNs, known as the IFN signature.[13,14] The IFN signature is often used as a proxy for the presence of type 1 IFN, considering the absence of a reliable method to measure type 1 IFNs in vivo. Type 1 IFNs play an important role in the pathophysiology of SLE, as randomized controlled trials show that treatment with type 1 IFN receptor blocking antibody anifrolumab leads to a decrease on the British Isles Lupus Assessment Group (BILAG)-based Composite Lupus Assessment (BICLA) in patients with SLE.[15,16] In these trials, anifrolumab suppresses interferon regulated genes in patients with a high IFN signature. Surprisingly, treatment effect was observed in patients with a high IFN signature as well as in patients with a low IFN signature. This indicates that SLE activity is not solely based on the presence of type I IFNs, and that the contribution of an IFN signature to disease activity differs between patients. Unfortunately, the small sample size of IFN low patients did not allow a direct comparison of the therapeutical effect of anifrolumab between patients with a low and high IFN signature. More information is therefore needed to determine the use of the IFN signature for personalized treatment.

NETs are important drivers of type I IFN production and a source of auto-antigens in patients with SLE.[17] NETs are formed as a first line defense mechanism of neutrophilic granulocytes by physically capturing microbes in structures containing DNA and antimicrobial peptides via a process called NETosis. NETs are relatively rich in antigens for autoantibodies commonly seen in SLE, such as double-stranded DNA. Indeed, NETosis is associated with anti-nuclear and anti-dsDNA antibodies, and the IFN signature in SLE.[18] The extent to which these immunological mechanisms and pathways are involved in the pathophysiology of SLE differs between patients.

Since NETs are a rich source of antigens for autoantibodies, and NETosis induces IFNα production, it is expected that these immunological mechanisms in SLE would to some extent correlate with each other. Indeed, NETosis is associated with the IFN signature, ANA and anti-dsDNA formation in SLE and type 1 IFNs are known to affect B-cell function and survival, which could lead to increased autoantibody production.[18–20] In the present study, we explored correlations between a wide range of established and lesser known SLE-associated autoantibodies and the presence of an IFN signature and NET release to further study their complex interaction in patients with SLE. Better understanding of these underlying immunological manifestations and their correlations could lead to better tools for patient stratification and more targeted treatment options.

MATERIALS AND METHODS

Patients

Blood samples and clinical data of patients with SLE as well as healthy controls were collected in the University Medical Center (UMC) Utrecht, The Netherlands in 2014 and 2015. The study protocol was approved by the medical ethical committee of the UMC Utrecht (reference number 12-296). All participating patients provided written informed consent. All patients fulfilled the American College of Rheumatology 1997 criteria for SLE. None of the included patients had signs of infection at the time of sample collection. Patients did not receive B lymphocyte suppressive therapy (belimumab or rituximab) in the 6 months prior to inclusion in this study. The IFN signature and NET inducing capacity were determined in patients from this cohort and were previously published.[18,21] Serum samples from these patients were stored in the UMC Utrecht biobank.

In 2019, we tested the presence of 57 SLE-related autoantibodies in biobanked serum of 483 patients with SLE, using a custom-made antibody microarray system.[22] 25 of these patients were also included in the previously reported cohort. We combined the datasets for the present study. Figure 1 shows a visual representation of the inclusion process.

Assessment of IFN signatures, NET inducing capacity and autoantibodies

A complete description of the methods for determining the IFN signature and NET inducing capacity has been previously published.[18,21] In short, mRNA expression of four IFN regulated genes (IFI44L, IFITM1, SERPING1 and LY6E) was determined relative to housekeeping gene GUSB in CD14+ monocytes from SLE patients and healthy controls. The relative gene expression was normalized against expression in a group of healthy controls (HC), and the individual z-scores were averaged to determine the IFN signature. The IFN signature was considered high if the averaged z-score was above 2SD of the mean IFN signature in the HD group. NET inducing capacity was studied by stimulating healthy donor neutrophils with 10% plasma from SLE patients for 4 hours. NET inducing capacity was measured with quantitative live imaging. Youden's J-statistic of the receiver operating characteristics (ROC) curve of NET-release in patients with SLE compared to HC's was used to determine whether a patient has a high or low NET inducing capacity.

We determined serum concentrations of 57 IgG autoantibodies with a custom-made antigen microarray system, using indirect immunofluorescence (Thermo Fisher Scientific, Uppsala Sweden). A full description of the methods has been described previously.[22]

Statistical analysis

We performed a principal component analysis to reduce the dimensionality of SLE patients with a high or low IFN signature, and a high or low amount of NET release, followed by a k-mean clustering approach (using principal components explaining 90% of the variance in the data) to define clusters.

We also performed a hierarchical clustering analysis to confirm the clusters, using the furthest neighbor clustering method, measuring the squared Euclidean distance of the autoantibodies measured on the antibody microarray. The scores of immunofluorescent intensity were log transformed and expressed as Z-scores to correct for differences in mean concentrations between the different autoantibodies. Data were visualized in a heatmap. Hierarchical clustering and principal component analysis were performed in R (www.r-project.org, version 4.2.2).

We used Mann-Whitney-U tests to compare concentrations of immunofluorescence of different autoantibodies between patients with high and low scores for IFN signature and NET inducing capacity. P-values were corrected for multiple testing according to the Bonferroni Correction ($\alpha = 0.05/57 = 0.0009$). We used Spearman's correlation to analyse the relationship between IFN signature, NET inducing capacity and concentrations of autoantibodies. Statistical analysis was performed using SPSS (version 25.0.0.2). Figures were made using GraphPad Prism (version 8.3.0).

RESULTS

Baseline characteristics

The median age of all included patients was 43 years, 92% of the patients were female. Most patients were in remission or had low disease activity at the time of sample collection, although three patients had active disease (SLEDAI 6, 9 and 16). Nearly all patients (96%) tested positive for anti-nuclear antibodies (ANA). At the time of inclusion, six patients (24%) tested positive for anti-double stranded DNA antibodies (anti-dsDNA), although 22 patients (88%) had ever tested positive for anti-dsDNA. Nine patients (36%) had decreased concentrations of C3 and/or C4. A full overview of the baseline characteristics is presented in Table 1.

Patients with a high IFN signature, or high NET inducing capacity had a higher median SLEDAI and anti-dsDNA concentrations, whereas median complement concentrations in these groups were lower compared to their counterparts with low IFN signature and low NET inducing capacity. All three patients who never tested positive for anti-dsDNA had a low IFN signature and low NET inducing capacity.

Patients with a high IFN signature and/or high NET inducing capacity were more likely to have active disease as well as receive medication other than hydroxychloroquine (HCQ).

Table 1: Patient characteristics

[a]Categorical variables are represented as medians with range. SELENA-SLEDAI: Safety of Estrogens in Lupus Erythematosus National Assessment – Systemic Lupus Erythematosus Disease Activity Index; ANA: antinuclear antibodies; ENA extractable nuclear antigen antibodies; anti-dsDNA: anti-double-stranded DNA; HCQ: hydroxychloroquine; AZA: azathioprine; MMF: mycophenolate mofetil; IFN: interferon; NET: neutrophil extracellular traps.

	All patients	**IFN signature**			**NET inducing capacity**	
	N=25	**Low n=7**	**High n=15**	**Missing n=3**	**Low n=10**	**High n=15**
Age in years[a]	43 (21-65)	43 (21-59)	44 (22-65)		43 (21-59)	43 (25-65)
Female sex (%)	23 (92%)	7 (100%)	14 (93%)		10 (100%)	13 (87%)
SELENA-SLEDAI score at inclusion[a]	2 (0-16)	0 (0-2)	2 (0-16)		1 (0-4)	2 (0-16)
Serology						
ANA positive (%)	24 (96%)	6 (86%)	15 (100%)		9 (90%)	15 (100%)
ENA positive (%)	17 (68%)	3 (43%)	13 (93%)		6 (60%)	11 (73%)
Anti-dsDNA (IU/mL)[a]	6.1 (0.9-497.0)	1.6 (0.9-21.0)	8.7 (1.5-497.0)		1.9 (0.9-27.0)	8.2 (1.5-497.0)
Anti-dsDNA ever positive	22 (88%)	4 (57%)	15 (100%)		7 (70%)	15 (100%)
C3(g/L) [a]	0.99 (0.54-1.51)	1.09 (0.96-1.51)	0.89 (0.64-1.41)		1.08 (0.68-1.51)	0.92 (0.54-1.41)
C4 (g/L) [a]	0.18 (0.06-0.34)	0.22 (0.18-0.34)	0.16 (0.06-0.25)		0.19 (0.11-0.34)	0.18 (0.06-0.25)
Current use of immunosuppressive medication (%)						
HCQ	18 (72%)	5 (71%)	11 (73%)		7 (70%)	11 (73%)
Prednisone	9 (36%)	1 (14%	6 (40%)		3 (30%)	6 (40%)
AZA	7 (28%)	1 (14%)	4 (27%)		3 (30%)	4 (27%)
MMF	1 (4%)	0	1 (7%)		0	1 (7%)

A distinct autoantibody pattern is present in a subgroup of patients with a high IFN signature, and is associated with development of a new rash

To investigate if patients with a IFN signature and/or NETosis have a specific autoantibody profile, we performed a principal component analysis (PCA) followed by a k-mean clustering (Fig. 2A). We found an undetectably low concentration of citrullinated α-enolase antibodies in one patient. This unusual value was defined as an outlier and consequently excluded from further clustering analyses. We tested whether excluding this patient influenced the clustering analyses, and found that the results represented in Figure 1 are not dependent of this patient. An overview of the studied autoantibodies is included as supplementary Table 1. We identified two distinct clusters: cluster 1 (n=9) contains mostly patients with a high IFN signature, and cluster 2 (n=15) contains a mix of patients with a high and low IFN signature. The two clusters are mainly separated by PC1 explaining 22.7% of the total variability. PC1 is mainly driven by antibodies against histones, RibP2, RibP0, EphB2, RibP1, PCNA, dsDNA, and nucleosome. The component loadings for PC1, PC2 and PC3 are shown in supplemental Figure S1. A hierarchical cluster analysis shows two distinct clusters, and one patient previously clustering with patients in cluster 1, crossed over to cluster 2. Hierarchical clustering confirmed the clustering of patients with a distinct autoantibody pattern and a high IFN signature (Fig. 2B).

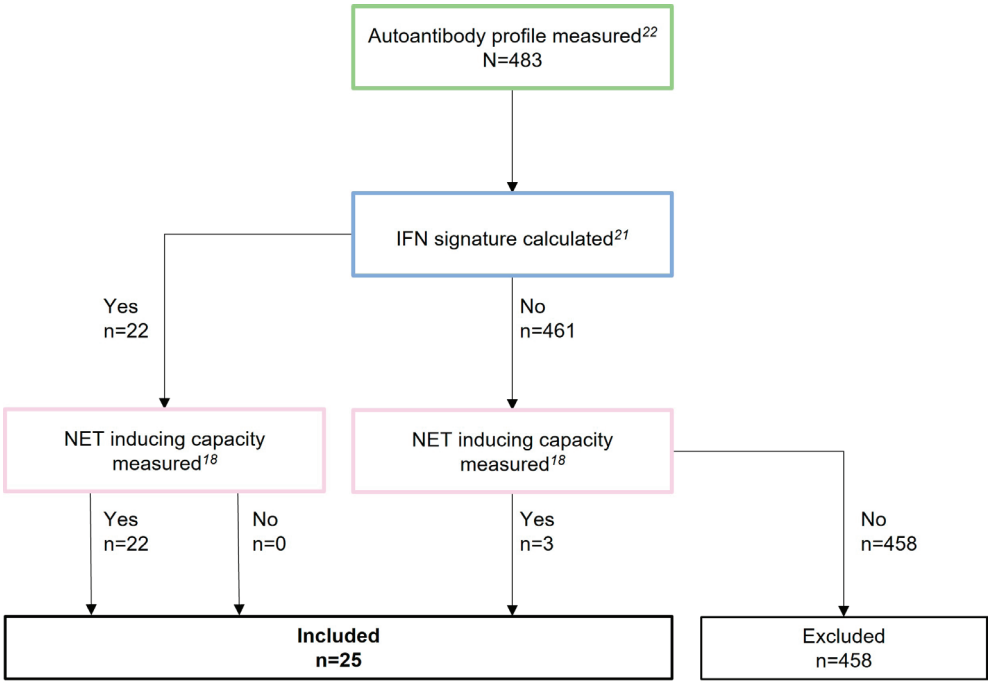


Figure 1: Flowchart representing the patient selection process. IFN: interferon; NET: neutrophil extracellular traps.

We evaluated whether patients presented with different clinical features and found an increased median anti-dsDNA and thrombocyte concentration in cluster 1 compared to cluster 2 (Fig. 2C). Although the median SELENA-SLEDAI score did not differ, cluster 1 contained all four patients that developed a new rash (Fig. 2D). Other clinical SELENA-SLEDAI features did not significantly differ between the clusters.

Autoantibodies against RibP1, RNP70 and EphB2 are associated with the IFN signature

We further analyzed the association between specific autoantibodies and the IFN signature. We compared the autoantibody concentration of patients with a high IFN signature to the concentration of patients with a low IFN signature and found a trend towards increased concentrations of autoantibodies against EphB2, RibP1, and RNP70 in patients with a high IFN signature (Figure 3A). Moreover, we found positive correlations between EphB2, RipP1, and RNP70 and the IFN signature (Figure 3B).

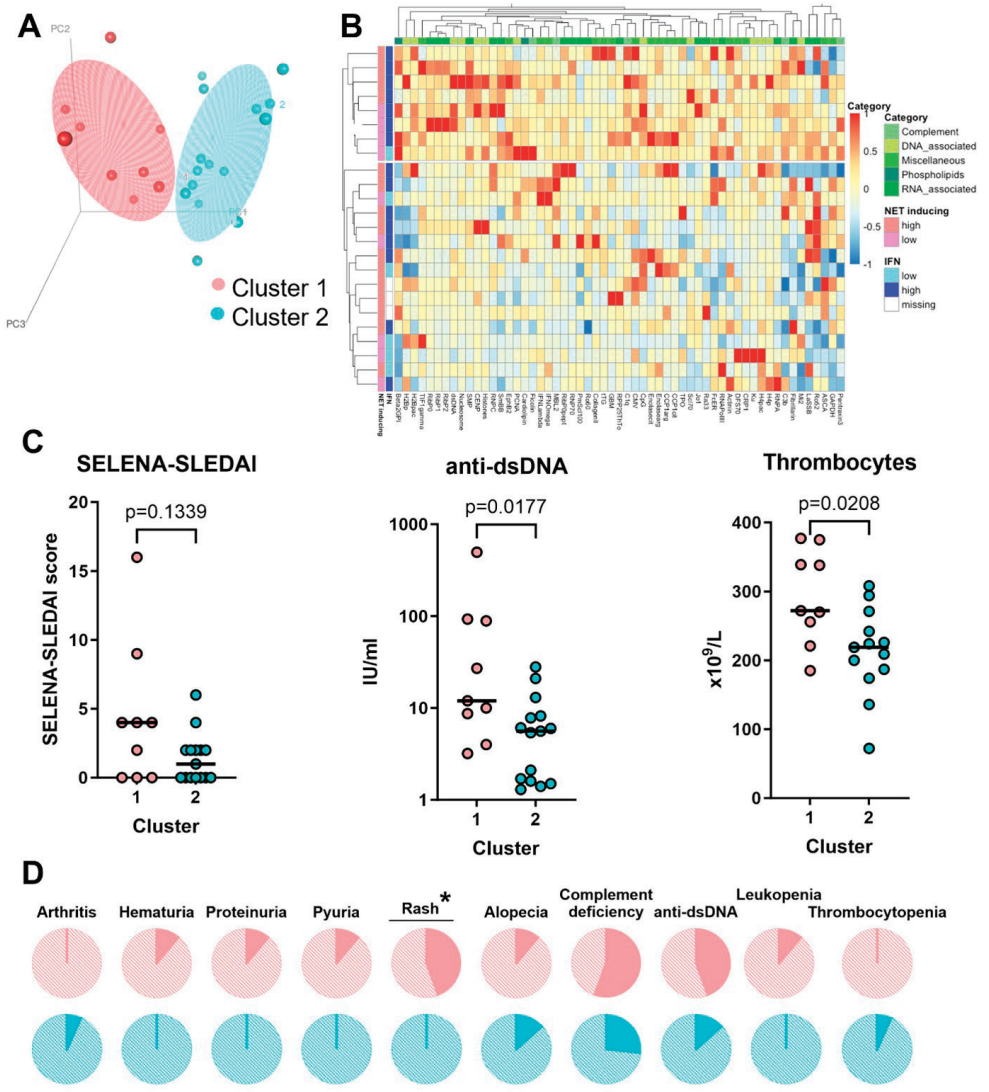


Figure 2: (A) Principal component analysis of 24 patients with SLE visualized in a 3D plot. **(B)** Heatmap of Z-scores for 57 autoantibodies in patients with high and low IFN signature, and high or low NET inducing capacity. Vertical axis shows clustering of patients based on IFN status (blue) and NET inducing capacity (orange/pink). Horizontal axis represents clustering based on autoantibody profiles. **(C)** SELENA-SLEDAI score, anti-dsDNA, and thrombocyte concentration in cluster 1 and 2 from Figure 2A. Lines represent median. Differences between clusters were tested with a Mann-Whitney-U test. **(D)** Representation of individual SELENA-SLEDAI items in patients in cluster 1 and cluster 2 from Figure 2A. Darker area represents % of patients positive for this item. Only items present in patients from our cohort are shown. *χ^2(=1, N=24) =8.00, p=0.005. IFN: interferon signature; NET: neutrophil extracellular trap.

Categorizing patients based on NET inducing capacity does not reveal distinct autoantibody clusters

When we categorized patients based on NET inducing capacity, we did not identify clustering of patients with certain autoantibody profiles (Fig. 2B). We additionally compared autoantibody concentrations in patients with a high NET inducing capacity to concentrations in patients with a low NET inducing capacity. Patients with a high NET inducing capacity tend to have decreased concentrations of antibodies against FcER and PmScl100 compared to patients with a low NET inducing capacity (Figure 4A). Antibodies against FcER and PmScl100 showed a negative correlation to NETosis (Figure 4B).

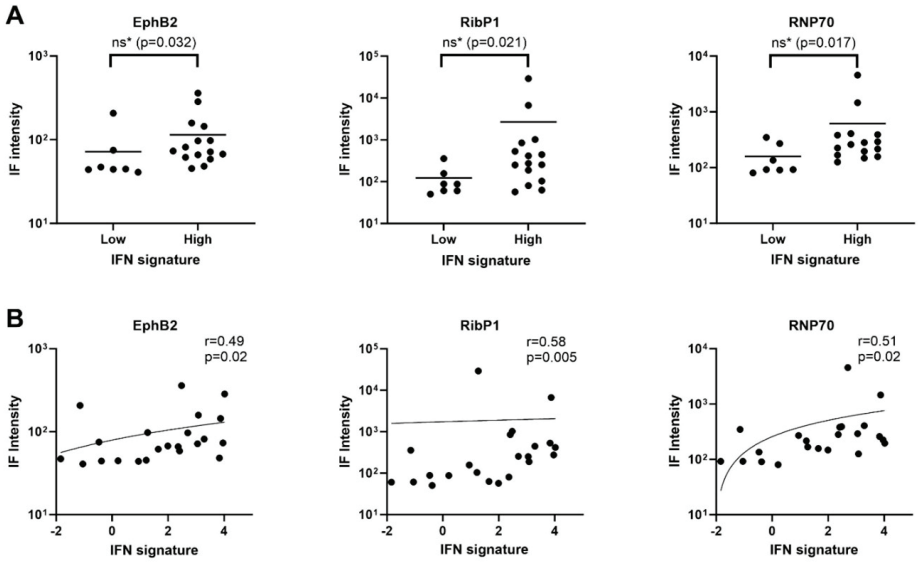


Figure 3: (**A)** Immunofluorescence (IF) intensity in IFN positive and IFN negative patiënts for EphB2, RipP1, and RNP70. Line represents mean IF intensity. **(B)** Correlation of EphB2, RipP1, and RNP70 IF intensity with IFN score.IF:immunofluorescence; IFN:interferon. *not significant after Bonferroni correction

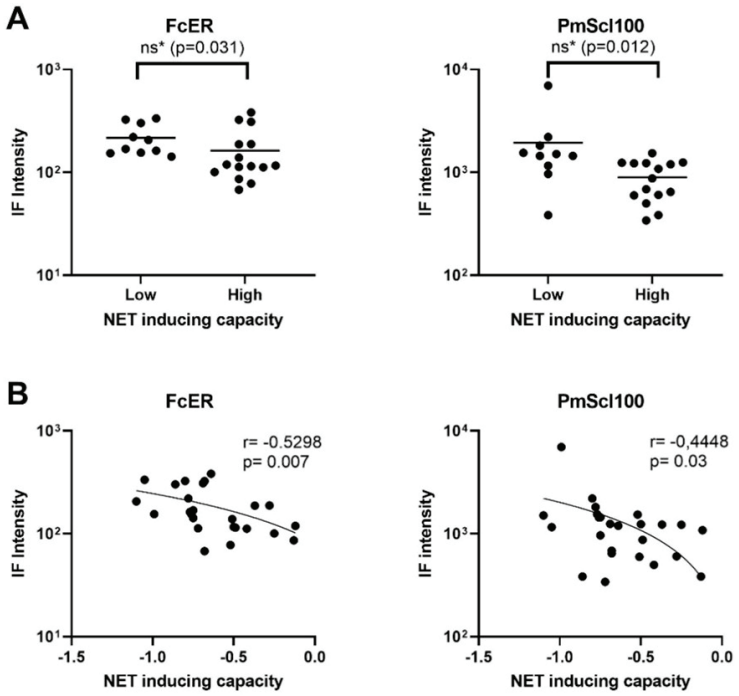


Figure 4: (A) Immunofluorescence (IF) intensity of autoantibodies against FcER and PmScl100 in patients with a low and high NET inducing capacity; Line represents mean IF intensity. **(B)** Correlation of FcER and PmScl100 IF intensity with NET inducing capacity. IF: immunofluorescence; NET: NETosis. *Not significant after Bonferroni correction for multiple testing.

DISCUSSION

In this study, we investigated the association between the IFN signature, NETosis and 57 autoantibodies in SLE patients, since further insight in the interplay of these immunological manifestations could lead to better options for patient stratification and personalized treatment.

Stratifying patients with SLE based on underlying immunological traits has gained attention in recent years, and has been attempted in several cohorts.[23–27] The use of different autoantibody panels and sample sizes however impairs comparability between studies. Recently, a clustering analysis based on 20 autoantibodies and 29 ANA patterns in a large cohort of patients with SLE identified four distinct autoantibody clusters.[28] One cluster contained patients positive for antibodies against (among others) dsDNA, RibP, histones, and PCNA. These antibodies were among the main drivers of variability in our PCA.

We found a subgroup of patients with an IFN signature who express increased concentrations of antibodies against DNA and RNA-binding proteins. In line with our findings, others have previously described an association between antibodies against RibP0, RNP, anti-dsDNA, and the IFN signature.[29–31] However, we could not confirm previously proposed associations between the IFN signature and anti-SSA, anti-SSB, and anti-Sm.[31]

The exact interplay of RNA-binding antibodies and the IFN signature is not fully understood. Interestingly, RNA-binding antibodies are more prone to forming circulating immune complexes than DNA-binding antibodies in SLE patients, which are more likely to bind directly to the target tissue.[32] As immune complexes are thought to be an important source of IFN via signaling through TLR7 and TLR9, this suggests that the RNA-related pathway involving TLR7 is a driver of IFN production in patients with SLE. Whether RNA-binding autoantibodies such as anti-U1 RNP, of which RNP70 is a subcomponent, play a causative role in the pathophysiology or are an epiphenomenon remains to be elucidated.

We found a trend towards increased concentrations of autoantibodies against EphB2, RibP1, and RNP70 in IFN-positive patients. Especially the presence of antibodies against the U1 RNP complex is associated with an interferon signature.[33,34] This association is not only seen in SLE, but across systemic autoimmune diseases, indicating that subgroups of patients with systemic autoimmunity share immunopathological phenotypes, despite different clinical disease manifestations.[35,36] The association between antibodies against ribosomal protein RibP1 and the IFN signature is less studied, although patients with anti-RibP antibodies have higher serum concentrations of IFN alpha than patients who were negative for anti-RibP antibodies.[37]

To our surprise, we found that the presence of antibodies directed against EphB2 is one of the main drivers of variability between patients with a high IFN signature and patients with a low IFN signature. We found a positive correlation between antibodies against EphB2 and the IFN signature, in addition to a previously described correlation between EphB2 protein and the IFN signature.[38] EphB2 protein is involved in atherosclerosis, and EphB2 expression in atherosclerotic plaques increases with plaque severity.[39,40] Antibodies against EphB2 have been observed in both SLE and systemic sclerosis patients. SLE patients with anti-EphB2 antibodies more often experience cardiovascular events compared to patients that do not express this antibody.[41] Whether EphB2, or antibodies against EphB2, play a causative role in atherosclerotic plaque formation in SLE is to be further elucidated, but its presence might partially explain the observed increased prevalence of cardiovascular events in patients with SLE compared to healthy patients.[42]

As NETs are rich sources of antigens for antibodies commonly seen in SLE, we expected a positive association between NETosis and several of these autoantibodies. Indeed, an association between NETosis and anti-dsDNA has been reported and was also found in the original cohort from which this sub analysis was performed.[18,43] Here, we did not find positive associations of autoantibodies with NETosis, but rather found a negative correlation of NETosis with anti-FcER and anti-PmScl100, indicating that we performed analyses for this present study in a subgroup from the original cohort. Although anti-FcER and anti-PmScl100 are mostly associated with chronic urticaria and systemic sclerosis respectively, both have also been described in patients with SLE.[44,45]

The main limitation in our study is the small sample size: we associated 57 autoantibodies with NETosis and the IFN signature in 25 and 22 patients respectively. For these analyses, we combined data from the same patients that were part of both studies, but we did not perform a power calculation beforehand. The low number of patients with active disease and the cross-sectional design of the study limits predicting flares and does not allow us to identify patients at risk for a complicated disease course. However, our findings are in line with current literature, indicating that our cohort might not be large enough to detect statistically significant differences after correction for multiple testing, but is still representative for SLE patients.

CONCLUSION

In this study, we show that combining data on different immunological phenomena in patients with SLE can help to further stratify patients. We identified a subgroup of patients with a high IFN signature that express DNA and RNA-binding antibodies, as well as antibodies against EphB2. Larger observational cohort studies are needed to further stratify patients based on a combination of genetic, proteomic and autoantibody profiles. This could lead to improved prediction of flares in SLE, as well as more targeted treatment options.

ETHICS APPROVAL

This study was conducted in accordance with the Declaration of Helsinki as revised in 2013. The medical ethical committee of the UMC Utrecht approved the study protocol (reference number 12-296). Use of serum samples was approved by the biobank committee of the UMC Utrecht (Toetsingscommissie Biobanken UMC Utrecht, TCBio protocol no. 16-824).

FUNDING

This work was supported by the Dutch Arthritis Society, grant number 18-1-403.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

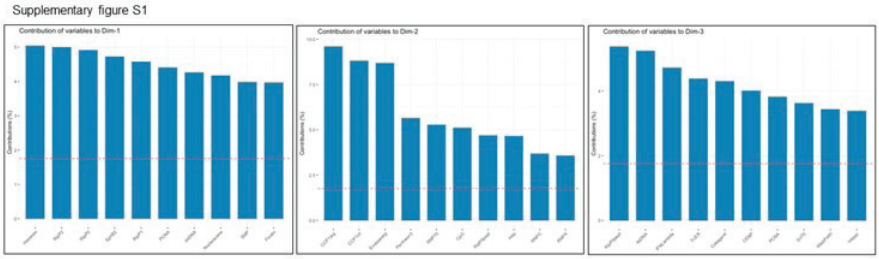
EK and TB performed data collection, analysis, interpretation and visualization. EK and TB wrote the manuscript. JD performed PCA and hierarchical clustering. LvdH determined the IFN signature. MvdL and LvdH measured plasma induced NET capacity. HL and JvL provided input on the data interpretation. MvdV, HO and ML conceptualized and designed the study, interpreted the data, wrote the manuscript, and provided overall supervision. All authors critically reviewed the manuscript. ML acts as guarantor for this study.

DECLARATION OF COMPETING INTEREST

JvL reports financial support was provided by Thermofischer. JvL reports a relationship with Abbvie, Eli Lilly, Galapagos, Magenta. that includes: consulting or advisory and funding grants. All other authors have no competing interests to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings presented in this study are available from the corresponding author upon request.



Supplementary Figure S1: contribution of variables to PC1 (left), PC2 (middle) and PC3 (right).

For Supplementary Table S1, see supplementary Table 1 of Chapter 3.

REFERENCES

1. Stojan G, Petri M. Epidemiology of systemic lupus erythematosus: an update. *Curr Opin Rheumatol*. 2018;30(2):144-150. doi:10.1097/BOR.0000000000000480
2. Gatto M, Zen M, Iaccarino L, Doria A. New therapeutic strategies in systemic lupus erythematosus management. *Nat Rev Rheumatol*. 2019;15(1):30-48. doi:10.1038/s41584-018-0133-2
3. Rodríguez-Almaraz E, Gutiérrez-Solís E, Rabadán E, et al. Something new about prognostic factors for lupus nephritis? A systematic review. *Lupus*. 2021;30(14):2256-2267. doi:10.1177/09612033211061475
4. Toro-Domínguez D, Martorell-Marugán J, Goldman D, Petri M, Carmona-Sáez P, Alarcón-Riquelme ME. Stratification of Systemic Lupus Erythematosus Patients Into Three Groups of Disease Activity Progression According to Longitudinal Gene Expression. *Arthritis Rheumatol*. 2018;70(12):2025-2035. doi:10.1002/art.40653
5. Kariuki SN, Franek BS, Kumar AA, et al. Trait-stratified genome-wide association study identifies novel and diverse genetic associations with serologic and cytokine phenotypes in systemic lupus erythematosus. *Arthritis Res Ther*. 2010;12(4):R151. doi:10.1186/ar3101
6. Haynes WA, Haddon DJ, Diep VK, et al. Integrated, multicohort analysis reveals unified signature of systemic lupus erythematosus. *JCI Insight*. 2020;5(4):e122312, 122312. doi:10.1172/jci.insight.122312
7. Guimarães J de AR, Furtado S da C, Lucas ACDS, Mori B, Barcellos JFM. Diagnostic test accuracy of novel biomarkers for lupus nephritis-An overview of systematic reviews. *PLoS One*. 2022;17(10):e0275016. doi:10.1371/journal.pone.0275016
8. Wahadat MJ, Schonenberg-Meinema D, van Helden-Meeuwsen CG, et al. Gene signature fingerprints stratify SLE patients in groups with similar biological disease profiles: a multicentre longitudinal study. *Rheumatology (Oxford)*. 2022;61(11):4344-4354. doi:10.1093/rheumatology/keac083
9. Barturen G, Babaei S, Català-Moll F, et al. Integrative Analysis Reveals a Molecular Stratification of Systemic Autoimmune Diseases. *Arthritis Rheumatol*. 2021;73(6):1073-1085. doi:10.1002/art.41610
10. Tsokos GC, Lo MS, Reis PC, Sullivan KE. New insights into the immunopathogenesis of systemic lupus erythematosus. *Nat Rev Rheumatol*. 2016;12(12):716-730. doi:10.1038/nrrheum.2016.186
11. Marto N, Bertolaccini ML, Calabuig E, Hughes GRV, Khamashta MA. Anti-C1q antibodies in nephritis: correlation between titres and renal disease activity and positive predictive value in systemic lupus erythematosus. *Ann Rheum Dis*. 2005;64(3):444-448. doi:10.1136/ard.2004.024943
12. ter Borg EJ, Horst G, Hummel EJ, Limburg PC, Kallenberg CG. Measurement of increases in anti-double-stranded DNA antibody levels as a predictor of disease exacerbation in systemic lupus erythematosus. A long-term, prospective study. *Arthritis Rheum*. 1990;33(5):634-643. doi:10.1002/art.1780330505
13. Bennett L, Palucka AK, Arce E, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med*. 2003;197(6):711-723. doi:10.1084/jem.20021553
14. Baechler EC, Batliwalla FM, Karypis G, et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci U S A*. 2003;100(5):2610-2615. doi:10.1073/pnas.0337679100
15. Furie RA, Morand EF, Bruce IN, et al. Type I interferon inhibitor anifrolumab in active systemic lupus erythematosus (TULIP-1): a randomised, controlled, phase 3 trial. *Lancet Rheumatol*. 2019;1(4):e208-e219. doi:10.1016/S2665-9913(19)30076-1

16. Morand EF, Furie R, Tanaka Y, et al. Trial of Anifrolumab in Active Systemic Lupus Erythematosus. *N Engl J Med*. 2020;382(3):211-221. doi:10.1056/NEJMoa1912196
17. Garcia-Romo GS, Caielli S, Vega B, et al. Netting neutrophils are major inducers of type I IFN production in pediatric systemic lupus erythematosus. *Sci Transl Med*. 2011;3(73):73ra20. doi:10.1126/scitranslmed.3001201
18. van der Linden M, van den Hoogen LL, Westerlaken GHA, et al. Neutrophil extracellular trap release is associated with antinuclear antibodies in systemic lupus erythematosus and anti-phospholipid syndrome. *Rheumatology (Oxford)*. 2018;57(7):1228-1234. doi:10.1093/rheumatology/key067
19. Akita K, Yasaka K, Shirai T, Ishii T, Harigae H, Fujii H. Interferon α Enhances B Cell Activation Associated With FOXM1 Induction: Potential Novel Therapeutic Strategy for Targeting the Plasmablasts of Systemic Lupus Erythematosus. *Front Immunol*. 2020;11:498703. doi:10.3389/fimmu.2020.498703
20. Kiefer K, Oropallo MA, Cancro MP, Marshak-Rothstein A. Role of type I interferons in the activation of autoreactive B cells. *Immunol Cell Biol*. 2012;90(5):498-504. doi:10.1038/icb.2012.10
21. van den Hoogen LL, Fritsch-Stork RDE, Versnel MA, Derksen RHW, van Roon JAG, Radstake TRD. Monocyte type I interferon signature in antiphospholipid syndrome is related to proinflammatory monocyte subsets, hydroxychloroquine and statin use. *Ann Rheum Dis*. 2016;75(12):e81. doi:10.1136/annrheumdis-2016-210485
22. Brunekreef T, Limper M, Melchers R, et al. Microarray testing in patients with systemic lupus erythematosus identifies a high prevalence of CpG DNA-binding antibodies. *Lupus Sci Med*. 2021;8(1):e000531. doi:10.1136/lupus-2021-000531
23. Artim-Esen B, Çene E, Şahinkaya Y, et al. Cluster analysis of autoantibodies in 852 patients with systemic lupus erythematosus from a single center. *J Rheumatol*. 2014;41(7):1304-1310. doi:10.3899/jrheum.130984
24. Budde P, Zucht HD, Vordenbäumen S, et al. Multiparametric detection of autoantibodies in systemic lupus erythematosus. *Lupus*. 2016;25(8):812-822. doi:10.1177/0961203316641770
25. Ching KH, Burbelo PD, Tipton C, et al. Two major autoantibody clusters in systemic lupus erythematosus. *PLoS One*. 2012;7(2):e32001. doi:10.1371/journal.pone.0032001
26. Diaz-Gallo LM, Oke V, Lundström E, et al. Four Systemic Lupus Erythematosus Subgroups, Defined by Autoantibodies Status, Differ Regarding HLA-DRB1 Genotype Associations and Immunological and Clinical Manifestations. *ACR Open Rheumatol*. 2022;4(1):27-39. doi:10.1002/acr2.11343
27. Pacheco JA, Rasmussen LV, Wiley K, et al. Evaluation of the portability of computable phenotypes with natural language processing in the eMERGE network. *Sci Rep*. 2023;13(1):1971. doi:10.1038/s41598-023-27481-y
28. Choi MY, Chen I, Clarke AE, et al. Machine learning identifies clusters of longitudinal autoantibody profiles predictive of systemic lupus erythematosus disease outcomes. *Ann Rheum Dis*. 2023;82(7):927-936. doi:10.1136/ard-2022-223808
29. Doedens JR, Jones WD, Hill K, et al. Blood-Borne RNA Correlates with Disease Activity and IFN-Stimulated Gene Expression in Systemic Lupus Erythematosus. *J Immunol Baltim Md 1950*. 2016;197(7):2854-2863. doi:10.4049/jimmunol.1601142
30. Hua J, Kirou K, Lee C, Crow MK. Functional assay of type I interferon in systemic lupus erythematosus plasma and association with anti-RNA binding protein autoantibodies. *Arthritis Rheum*. 2006;54(6):1906-1916. doi:10.1002/art.21890
31. Li QZ, Zhou J, Lian Y, et al. Interferon signature gene expression is correlated with autoantibody profiles in patients with incomplete lupus syndromes. *Clin Exp Immunol*. 2010;159(3):281-291. doi:10.1111/j.1365-2249.2009.04057.x

32. Lövgren T, Eloranta ML, Båve U, Alm GV, Rönnblom L. Induction of interferon-alpha production in plasmacytoid dendritic cells by immune complexes containing nucleic acid released by necrotic or late apoptotic cells and lupus IgG. *Arthritis Rheum*. 2004;50(6):1861-1872. doi:10.1002/art.20254
33. Chasset F, Ribi C, Trendelenburg M, et al. Identification of highly active systemic lupus erythematosus by combined type I interferon and neutrophil gene scores vs classical serologic markers. *Rheumatology (Oxford)*. 2020;59(11):3468-3478. doi:10.1093/rheumatology/keaa167
34. Weckerle CE, Franek BS, Kelly JA, et al. Network analysis of associations between serum interferon-α activity, autoantibodies, and clinical features in systemic lupus erythematosus. *Arthritis Rheum*. 2011;63(4):1044-1053. doi:10.1002/art.30187
35. Assassi S, Mayes MD, Arnett FC, et al. Systemic sclerosis and lupus: points in an interferon-mediated continuum. *Arthritis Rheum*. 2010;62(2):589-598. doi:10.1002/art.27224
36. Reynolds JA, Briggs TA, Rice GI, et al. Type I interferon in patients with systemic autoimmune rheumatic disease is associated with haematological abnormalities and specific autoantibody profiles. *Arthritis Res Ther*. 2019;21(1):147. doi:10.1186/s13075-019-1929-4
37. Mozo L, López P, Caminal-Montero L, Rodríguez-Carrio J, Suárez A. Anti-ribosomal P antibodies are associated with elevated circulating IFNα and IL-10 levels in systemic lupus erythematosus patients. *Lupus*. 2014;23(14):1477-1485. doi:10.1177/0961203314546020
38. Smith MA, Chiang CC, Zerrouki K, et al. Using the circulating proteome to assess type I interferon activity in systemic lupus erythematosus. *Sci Rep*. 2020;10(1):4462. doi:10.1038/s41598-020-60563-9
39. Sakamoto A, Ishibashi-Ueda H, Sugamoto Y, et al. Expression and function of ephrin-B1 and its cognate receptor EphB2 in human atherosclerosis: from an aspect of chemotaxis. *Clin Sci Lond Engl 1979*. 2008;114(10):643-650. doi:10.1042/CS20070339
40. Vreeken D, Bruikman CS, Cox SML, et al. EPH receptor B2 stimulates human monocyte adhesion and migration independently of its EphrinB ligands. *J Leukoc Biol*. 2020;108(3):999-1011. doi:10.1002/JLB.2A0320-283RR
41. Azzouz DF, Martin GV, Arnoux F, et al. Anti-Ephrin Type-B Receptor 2 (EphB2) and Anti-Three Prime Histone mRNA EXonuclease 1 (THEX1) Autoantibodies in Scleroderma and Lupus. *PLoS One*. 2016;11(9):e0160283. doi:10.1371/journal.pone.0160283
42. Tselios K, Gladman DD, Su J, Ace O, Urowitz MB. Evolution of Risk Factors for Atherosclerotic Cardiovascular Events in Systemic Lupus Erythematosus: A Longterm Prospective Study. *J Rheumatol*. 2017;44(12):1841-1849. doi:10.3899/jrheum.161121
43. Bruschi M, Bonanni A, Petretto A, et al. Neutrophil Extracellular Traps Profiles in Patients with Incident Systemic Lupus Erythematosus and Lupus Nephritis. *J Rheumatol*. 2020;47(3):377-386. doi:10.3899/jrheum.181232
44. Fiebiger E, Hammerschmid F, Stingl G, Maurer D. Anti-FcepsilonRIalpha autoantibodies in autoimmune-mediated disorders. Identification of a structure-function relationship. *J Clin Invest*. 1998;101(1):243-251. doi:10.1172/JCI511
45. Hanke K, Brückner CS, Dähnrich C, et al. Antibodies against PM/Scl-75 and PM/Scl-100 are independent markers for different subsets of systemic sclerosis patients. *Arthritis Res Ther*. 2009;11(1):R22. doi:10.1186/ar2614

7

Recent Developments in the use of Biomarkers as Diagnostic Tools in Patients with Systemic Lupus Erythematosus; A Narrative Review

Authors:
T.E. Brunekreef[a], H.G. Otten[b], J.M. van Laar[a], M. Limper[a]
Affiliations:
[a] Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht, The Netherlands
[b] Center for Translational Immunology, University Medical Center Utrecht, Utrecht, The Netherlands

Accepted for publication by EULAR Rheumatology Open

ABSTRACT

Systemic lupus erythematosus (SLE) is a heterogeneous systemic autoimmune disease with a complex a multifactorial pathophysiology. Due to difficulty recognizing patients with SLE in an early stage, diagnostic delay is very common, which can result in the accumulation of organ damage. Some immunological signs, such as the presence of autoantibodies, have previously been shown to be present in patients with SLE, even before the constitution of the first symptoms. This makes these immunological signs a prime target to improve the diagnostic process. In this article, we review recent studies investigating the use of biomarkers for diagnostics, including autoantibodies, other serum biomarkers, biomarkers in other bodily material, circulating RNAs and gene expression profiles. We discuss their place in the pathophysiology of SLE and their potential to improve and expedite the diagnostic process for patients with SLE.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a systemic autoimmune disease characterized by a wide range of symptoms and immunological manifestations.[1] Because of the heterogeneity of the disease, SLE can be difficult to diagnose, often resulting in a significant diagnostic and treatment delay.[2] This can result in the accumulation of organ damage, increasing the burden of the disease for patients.[3]

There is an unmet need for additional tools to improve and expedite the diagnostic process to diminish this delay. In the absence of diagnostic criteria for SLE, the EULAR/ACR classification criteria, developed primarily for research purposes, are commonly used in clinical practice as a substitute to aid clinical decision making, blurring the lines of what would be considered a diagnosis of SLE in clinical practice.[1] The pre-diagnostic phase of SLE ('pre-SLE') is characterized by the absence or limited presence of SLE-related clinical symptoms, which are insufficient to establish a clinical diagnosis of SLE. These symptoms are often non-specific and can be caused by a wide range of different causes. However, in pre-SLE patients some signs of autoimmunity, such as the presence of autoantibodies and interferons, have been shown to be present frequently during this pre-classification phase. We show an overview of these signs of autoimmunity in pre-SLE patients in Table 1.[4–19]

These signs of autoimmunity can potentially be used to distinguish patients who will eventually be diagnosed with SLE, the pre-SLE patients, from patients with similar signs and symptoms, but who will never develop SLE. Numerous studies have therefore focussed on different (immunological) biomarkers to aid in the diagnostic process for patients with SLE. Ideally, these studies would use cohorts of pre-SLE patients to investigate biomarkers that are both be accurate and present in pre-diagnostic phases of SLE. However, as pre-SLE cohorts are very limited, nearly all studies compare patients with an established diagnosis of SLE, fitting the 2019 EULAR/ACR or older classification criteria, with control groups. It is often assumed that the results from these studies are translatable to the pre-SLE phase as well, and therefore have potential relevance for the diagnostic process.

In this review we give an overview of recent studies investigating biomarkers in patients with an established clinical diagnosis of SLE and their potential for use as diagnostic tools. The use of biomarkers associated with specific organ involvement (i.e. skin or renal involvement) is beyond the scope of this review.

Table 1

Signs of autoimmunity in pre-SLE patients, as described in previous studies.

Immune factor	Marker	References
Autoantibodies	ANA	5,6,9,12
	Anti-Ro	5,6,7,8,9
	Anti-La	5,6,8,9
	APl	5,16,19
	Anti-dsDNA	5,6,8,9,12,14,17
	Anti-Sm	5,6,8,9
	Anti-RNP	5,6,8,9
	Anti-chromatin	8,9
	Rheumatoid factor	13
	Anti-PCNA	16
	Anti-C1q	16,17
	Increased IgG:IgM ratio	16,17
Interferon	IFN-α	8
	IFN-γ	9,18
	IFN-Score-B	11
	IFN-Score-A	11
Soluble markers	IL-6	9
	IL-12p70	9
	IP-10	9,10
	IL-5	9
	IL-10	19
	TGF-β	9,18
	Complement levels	14
	BLyS	18
	SCF	18

ANA: anti-nuclear antibodies. APl: anti-phospholipid antibodies. Anti-dsDNA: anti-double stranded DNA. Anti-RNP: anti-ribonucleoprotein.

METHODS

We performed a PubMed search with the following search terms: "lupus erythematosus, systemic"[MeSH Terms] AND ("biomarkers"[MeSH Terms] OR "antibodies"[MeSH Terms] OR "immunoglobulins"[MeSH Terms] OR "antibodies"[MeSH Terms] OR "chemokines"[MeSH Terms] OR "cytokines"[MeSH Terms] OR "polymorphism, single nucleotide"[MeSH Terms]) AND ("systemic lupus erythematosus"[Title/Abstract] AND ("biomarker"[Title/Abstract] OR "biomarkers"[Title/Abstract]) AND ("diagnosis"[Title/Abstract] OR "diagnostic"[Title/Abstract])).

Searches were restricted to include articles published from January 2021 up to July 2024.

Inclusion criteria

Studies that investigated the discriminating potential of one or multiple biomarkers in patients with SLE compared to healthy controls. Only studies that focused on adults, with full text availability and in written in English were included. Relevant studies cited within these articles were also included.

Exclusion criteria

Studies focussing primarily on the diagnosis of specific organ involvement, such as lupus nephritis or cutaneous lupus erythematosus, rather than diagnosis of SLE in general, were excluded, as this was beyond the scope of this review.

RESULTS

The search resulted in 201 articles. After application of the inclusion and exclusion criteria, 61 articles were included. Fifty-eight out of sixty-one included studies note that the patients included in their studies were patients with an established diagnosis of SLE, fitting classification criteria (most commonly the 1997 ACR criteria), or are using public datasets comprised of data from SLE patients fulfilling these classification criteria. Three studies do not explicitly mention which metrics they used to define SLE-patients in their study.

As the pathophysiology of SLE is multifactorial and very complex, we have divided the studies into five categories: Autoantibodies, other serum markers, markers in other bodily material, circulating RNAs and gene expression. A simplified overview of these markers and their interconnections is displayed in Figure 1.

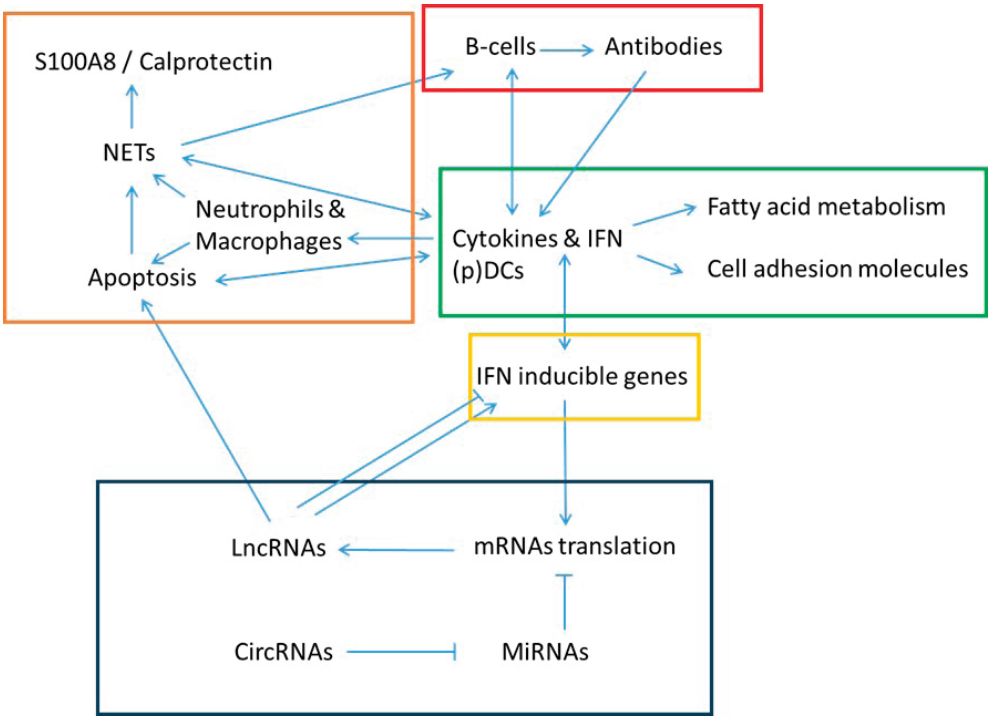


Figure 1

A simplified overview of the different areas of the pathophysiology of SLE and their interconnections, investigated by the articles mentioned in this study.

Serum markers

Autoantibodies

Over the past 80 years, over 100 different autoantibodies associated with SLE have been identified.[20] In the last couple of years however, research into the diagnostic potential of autoantibodies seems to have slowed down. We found 6 recent studies investigating the diagnostic use of (auto)antibodies in our search.[21–25]

With the use of phage immunoprecipitation sequencing, one study used high-throughput DNA sequencing to describe antibodies targeting a peptide from a random peptide library chip (SLE_p27, seq38567 GLYHSNASFRVP). These antibodies were highly distinctive (AUC 0.938) for SLE patients compared to healthy controls (HCs). It is noteworthy that SLE_p27 outperformed both anti-dsDNA and anti-Sm (AUC of 0.840 and 0.575 respectively).[21] Another study described the presence of anti-NET antibodies (ANETAs) in patients with SLE, which had moderate diagnostic potential.[22] Interestingly, these ANETAs did not inhibit the immunostimulatory effects of NETs, but rather protected them from degradation, as reported previously.[26] Although they did not find the exact targets

of these ANETAs, they notice that ANETA showed no more than a weak correlation with other tested autoantibodies, such as anti-dsDNA.

Other than focussing on the discovery of new autoantibodies in patients with SLE, some recent studies have focussed on identifying the exact epitopes that previously known autoantibodies react with. A small number of positive ANA-tests in SLE-patients, in particular those with an AC-25 or AC-26 pattern, can be explained by the presence of anti-NuMa antibodies. Approximately 15% of SLE patients tested positive for anti-NuMa antibodies and for half of these patients, these were the only autoantibodies that were detected, suggesting potential additional diagnostic value in patients that test negative for conventional autoantibodies.[23] Anti-mitochondrial antibodies (AMAs) targeting C1qBP or MFN-1 are more prevalent in patients with SLE compared to healthy controls, whereas levels of anti-C1qBP and anti-MFN-1 were not elevated in patients with other diseases associated with the presence of AMAs.[24] Compared to healthy controls, SLE-patients have increased levels of antibodies targeting the EBV envelope glycoprotein gp350, which might serve as a potential diagnostic biomarker (AUC 0.804). Additionally, the study shows that these antibodies do not appear to be cross-reacting with autoantigens, strengthening the hypothesis that gp350 is a prospective EBV vaccine candidate.[25]

One study investigated whether the ratio of immunoglobulin G (IgG) to immunoglobulin M ratio would be predictive of progression from incomplete SLE (iSLE) to SLE. They found that patients with iSLE had higher IgG/IgM ratios that healthy controls, although these ratios were not predictive of progression to SLE in patients with iSLE.[27]

Despite the numerous SLE-associated antibodies, the practical use of many of these antibodies is limited, as historically only a small number are deemed to have sufficient added diagnostic value to be incorporated into regular clinical practice.[1,28] Potentially also due to technical improvements leading to the increased accessibility of newer diagnostic methods such as the -omics, more recent research seems to mainly focus on the potential of other diagnostic methodologies.[29,30]

Other serum markers

Cytokines and serum proteins

We found 11 studies that have focussed on other serum markers, such as cytokines and other soluble serum proteins.[31–40] Among the cytokines that were found to be upregulated were cytokines that are involved in the induction of inflammation response (IL-6, TNF-alpha, CCL4 (or MIP-1β)), and B-cell activation (BAFF).[31,32] Interestingly, increased serum levels of IL1RA (interleukin-1 receptor antagonist), which is an anti-inflammatory cytokine, were found as well, with an AUC of 0.938, compared to 0.973 for anti-dsDNA. IL1RA functions by acting as an antagonist to the interleukin-1 receptor, inhibiting activa-

tion of the IL-1 receptor by the pro-inflammatory IL-1.[31,41] Serum IL1RA might be elevated in a response to the excessive inflammation seen in patients with SLE, as a compensatory modulating agent.

Other studied serum proteins included complement levels, calprotectin, a soluble protein abundantly present in the cytosol of neutrophils (and released in NETs as a major antifungal component), protein hormones (leptin and adiponectin) and several (trans)membrane proteins, in particular involved in cell adhesion (VCAM-1, E-selectin, L-selectin, SIGLEC1) or vesicular transport (Cav-3, Cav-1).[19,31,34,36,37,39,40] One study shows an association between persistent low levels of complement proteins C3 and C4 to be associated with progression from iSLE to SLE.[40] VCAM-1 and E-selectin are membrane proteins expressed on endothelial cells, whereas L-selectins are expressed on leukocytes and SIGLEC-1 is expressed on monocytes. These adhesion molecules have been shown to be involved in leukocyte rolling and their expression is likely increased as a result of inflammation causing endothelial damage.[42,43] Out of these biomarkers, SIGLEC1 had the highest AUC, with an AUC of 0.95, comparable to that of anti-dsDNA (AUC 0.90).[39]

Metabolomics

Metabolomics is a relatively new addition to the diagnostic field. Metabolomics are used to study the metabolite composition by measuring numerous metabolites simultaneously.[44] These metabolites can be influenced by diet or microbiota, but can also be changed as a result of inflammation.[45,46]

Six recent studies investigated changes in metabolomics in SLE patients compared to healthy controls.[47–52] In general, lipid and fatty acid metabolism seem most affected in SLE patients in these studies. Multiple studies suggest that changes in the lipid and fatty acid metabolism might result in the increased prevalence of cardiovascular disease in patients with SLE, which is a major cause of mortality in SLE patients. These six studies found multiple metabolites, either single metabolites or multiple metabolites combined into a multivariate model, with an AUC of 0.950 or larger, for SLE patients compared to healthy controls, which suggests great diagnostic potential.

Proteomics

Whereas metabolomics studies the composition on metabolites in patients, proteomics focusses on the proteome, the collection of proteins in these patients. Three recent studies investigated the proteome of patients with SLE, and found that a number of proteins were significantly changed in patients with SLE.[47,53,54] These proteins, such as IFIT3, TNF-RII and Amyloid A1, mainly reflected organ damage and inflammation, with multiple interferon-induced proteins being found differentially expressed between SLE patients and HCs.

Some of the studies mentioned above show correlations of the investigated biomarkers with conventional diagnostic markers, such as anti-dsDNA and complement factors C3 and C4. Three studies also compared the AUC of the investigated biomarker to that of anti-dsDNA and complement factors.[37,39,41] In all three cases the AUC of the investigated biomarkers were similar to that of anti-dsDNA and outperformed the AUC of complement factors.

Markers in other bodily materials

Although most diagnostics involve tests that are performed using a blood or serum sample, five studies have also investigated the use of other bodily materials as a diagnostic of SLE.[55–59] Three studies have investigated urine samples, whereas the diagnostic potential of the composition of pleural effusion and the microbiome was investigated by one study respectively.

Similar to blood samples, urine samples are easy to collect and are commonly used in current clinical practice. The main use of urine diagnostics in current clinical practice is to screen for lupus nephritis. For the purpose of this review, we only included studies investigating the use of diagnostics on urine samples for the diagnosis of SLE in general and excluded studies investigating the use of urine sample to diagnose lupus nephritis in patients with established SLE.

The first study found that levels of several pro-inflammatory proteins were elevated in the urine of SLE patients, compared to healthy controls.[55] The second study showed that patients with SLE also have increased urinary levels of beta-2 glycoprotein 1, which is also known as one of the targets of antiphospholipid antibodies. Interestingly, the authors found an AUC of 0.946 for SLE patients compared to healthy controls, with similar levels of urinary beta-2 glycoprotein 1 in patients with or without lupus nephritis.[56] The third study showed increased levels of S100A8 (S100 calcium-binding protein A8) in urine, as well as in serum and saliva of SLE patients.[57] Together with S100A9, S100A8 forms the heterodimer calprotectin, which itself also has been shown to be elevated in serum of SLE-patients. Similar to calprotectin, elevated serum levels of S100A8 are thought to be associated with NETosis, as it is abundantly present in the cytosol of neutrophils.[60]

A study investigating levels of complement C3 and C4 in pleural effusion found that C3 and C4 could be used to distinguish pleural effusion caused by SLE from pleural effusion of other causes, with lower levels of C3 and C4 being found in the pleural effusion of SLE patients.[58] Although these levels of C3 and C4 were significantly lower than in pleural effusion caused by malignancy or infection, they were not significantly different from levels of C3 and C4 seen in the pleural effusion caused by fluid overload.

One study investigated the composition of the microbiome of SLE patients using 16S rRNA analysis.[59] Using a machine learning model, they were able to create a predictive model with an AUC of 0.930. Presence of *escherichia fergusonii*, a pathogenic bacterium known to cause urinary tract infections, was found as the most important contributing factor to the model.[61]

Circulating RNAs

Several studies have investigated the diagnostic properties of circulating RNAs. These can broadly be divided into three different categories: micro-RNAs (miRNA), circular RNAs (circRNA) and long non-coding RNAs (LncRNA).

MiRNA

MiRNAs are short non-coding RNA fragments that regulate protein expression by blocking the translation of the corresponding messenger RNA (mRNA) by binding to it. Depending on the function of the protein coded by the corresponding mRNA, miRNAs can have both immunostimulating and immunosuppressive effects.[62]

One study found increased levels of miR-21 and miR-146a in patients with SLE, compared to healthy controls.[63] They note that miR-21 is associated with T-cell activation and miR-146a is associated with type 1 IFN response. A different study found increased levels of miR-181a and decreased levels of miR-223 in patients with SLE, compared to healthy controls.[64] MiR-181a plays a role in lymphocyte differentiation and miR-223 influences granulocyte differentiation.[65,66] Another study showed increased levels of miR-124-3p and miR-377-3p in patients with SLE.[67] The authors suggest that both these miRNAs might inhibit the transcription of EGR1, which functions as a transcription regulator and has been shown to be downregulated in patients with SLE, although they did not experimentally confirm this.[68] A. found miR-183-5p to be upregulated in SLE patients. MiR-183-5p targets Foxo1, which showed inverse correlation with miR-183-5p expression.[69] It has been suggested that Foxo1 downregulation is one of the important causes of Th17 dysregulation.[70]

Despite the significant alterations in levels of miRNAs found by these studies, in general they appear not to be very specific for SLE, as many of these have also been associated with different types of cancer and other diseases.[71] Therefore, these alterations might not represent changes specific to SLE, but rather a change from the normal homeostasis, as seen in a multitude of diseases. For example, as a review by Jenike and Halushka points out, miR-21 is one of the most abundant miRNAs and has been claimed as a biomarker in at least 29 different diseases, and therefore cannot be considered a useful biomarker.[71]

CircRNA

Whereas miRNAs are suggested to inhibit mRNA transcription, circular RNAs act as "sponges" for miRNAs, therefore limiting the effect modulating effect of miRNAs.[72] One

study found that circPTPN22 is downregulated in patients with SLE, as well as in patients with rheumatoid arthritis.[73] A follow-up study by the same research group found that circPTPN22 acts as a sponge for miR-4689, which targets S1PR1, an important regulator of immune cell migration, differentiation and proliferation.[74]

One study found increased expression levels of three circular RNAs (hsa_circ_0082688, hsa_circ_0082689 and hsa_circ_0008675). A model using a combination of two of these circular RNAs resulted in an AUC of 0.740 for distinguishing patients with SLE and HCs.[75] In another study by the same group, they did not reproduce these findings, but instead describe lowered levels of 4 different circular RNAs (hsa_circ_0000175, hsa_circ_0044235, hsa_circ_0068367 and hsa_circ_0001947).[76] Additionally, a different study found that hsa_circ_0006689 was overexpressed in patients with SLE compared to HCs with an AUC of 0.713, and another study found overexpression of circETS1 in patients with SLE, with an AUC of 0.873.[77,78] The lack of reproducibility and unclear pathophysiological mechanisms of these circRNAs limit their diagnostic potential.

LncRNA

Long non-coding RNAs (LncRNAs) are commonly defined as RNAs with a length of over 200 nucleotides that are not translated into functional proteins. Many lncRNAs are considered transcriptional noise, although increasing numbers of lncRNAs have been shown to have biological functions, mainly with a regulatory or architectural role in cell biology.[79]

We found 7 studies investigating the use of lncRNAs for the diagnosis of SLE.[80–86] The function of the lncRNAs mentioned in these studies is often unclear, but the lncRNAs reported in these studies are commonly suggested to have an inflammation-regulatory function. Two studies report lncRNAs associated with apoptosis pathways.[83,86] One study found increased expression levels of lncRNA SNHG1, which is found to negatively regulate tumor suppression genes such as p53, and found that lncRNA SNHG1 overexpression was associated with an increased apoptosis rate in PBMCs in SLE patients. Another study found an increased expression of lncRNA H19, which they presume to target micro-RNA 19-b (miR-19b). miR-19b has been shown to enhance proliferation of PBMCs and induce apoptosis resistance, whereas lncRNA H19 can inhibit the function of miR-19b.

Gene expression

Several recent studies have focussed on gene expression profiles in SLE.[87–98] Many of these studies used similar methodologies, using existing datasets from previous trials, that have been published in an online repository with the Gene Expression Omnibus (GEO) database. These new studies use the gene expression data series (GSE) from these trials to identify interesting differentially expressed genes (DEGs), validate the DEGs, often using other GSEs, and then validate the findings using RT-qPCR in a (small) local cohort. Following this, KEGG (Kyoto Encyclopedia of Genes and Genomes) and GO (Gene

Ontology) analyses are often used to further investigate the function of the genes found by these analyses.

Many of the DEGs reported by these studies are known interferon-inducible genes, such as IFI44, IFI44L, MX2, IFI27 and EPSTI1.[88–90,92–97] Some of these genes are also included in commonly used combinations to calculate the so-called "interferon signature".[99] One study used Mendelian Randomization to predict cytokine levels using publicly available GWAS data. Their main finding suggest an increase in levels of IL-4 in patients with SLE, identifying it as a potential biomarker, although they mention this has to be further investigated in experimental studies.[100]

It is noteworthy to highlight a study by Khunsriraksakul et al..[98] In this study, they performed a multi-ancestry and multi-trait meta-analysis of 12 SLE GWAS studies, which included over 20.000 SLE patients and almost 700.000 HCs. They found 16 new loci associated with SLE, and integrated the differentially expressed genes into a polygenic risk score (PRS) to evaluate the diagnostic potential. Nevertheless, the PRS on its own had an AUC of 0.65-0.69, whereas the more conventional combination of ANA and anti-dsDNA had an AUC of 0.72-0.73. Combining both only increased the AUC slightly to 0.75.

Methylation levels

Methylation of gene promotors is an important epigenetic factor in gene expression regulation, that can both increase or inhibit the expression of the associated gene, based on the binding site on the DNA.[101] Two studies investigated the methylation levels of specific genes in PBMCs of SLE patients, compared to HCs.[102,103] Both of these studies found increased methylation of tumor suppressor genes in PBMCs, potentially indicating decreased function of these genes. One of these studies showed increased methylation of the RUNX3 gene, a known tumor-suppressor gene that is frequently deleted or silenced in different types of cancers.[102,104,105] The other study found increased methylation levels of CDKN2A promotors, which encodes two tumor suppressor genes, p14 and p16, and has been shown to modulate production of inflammatory cytokines.[103]

DISCUSSION

There is a need for additional tools to expedite the diagnostic process for patients with SLE, as the current diagnostic process often leads to diagnostic and treatment delays.[2,3] In this review, we evaluated recent developments in the use of biomarkers for the purpose of diagnostics in patients with SLE.

Antibodies are one of the few immunological manifestations that have been shown to actually be present before the diagnosis of SLE is eventually made, as Arbuckle et al. showed

previously in a longitudinal cohort study.[5] Combined with the ease of investigating the presence of autoantibodies, they have therefore historically been a prime target for the discovery of new diagnostic biomarkers. However, over half of the studies investigating autoantibodies included in this review mainly focussed on elucidating the exact antigens of known autoantibodies, rather than the presence of undiscovered autoantibodies. Nevertheless, the number of SLE-associated antibodies has grown significantly since the study by Arbuckle et al..[28] It would be interesting to repeat a study similar to that of Arbuckle et al., that includes multiple SLE-related autoantibodies that were not included in the original study to investigate the sensitivity of other autoantibodies or combinations of autoantibodies in the pre-clinical phases of SLE.

An aberrant process of apoptosis and increased prevalence of NETs are known phenomena in patients with SLE. Previous studies have shown an increased prevalence of NETosis-prone low-density granulocytes (LDGs) in patients with SLE.[106] Interestingly, increased levels of calprotectin, a soluble protein abundantly present in the cytosol of neutrophils and important component of NETs, are found in the serum of SLE patients.[34] Increased levels of S100A8, one of the two major components of calprotectin, are also found in the serum, as well as the urine and saliva of SLE patients.[57,60] S100A8 might serve as an indirect marker of NETosis activity, which might be useful in the diagnostic process or can potentially be used to identify different subsets of SLE patients for treatment selection.

Cytokines play an important part in the regulation of the autoimmune pathophysiology. Production of cytokines are influenced by many factors and can also influence different parts of the immune system. The studies reviewed in this article show increased prevalence of multiple pro-inflammatory cytokines as well as altered levels of cell adhesion molecules such as selectins, potentially facilitating extravasation of leukocytes into tissues.[31–38] Additionally, changes in fatty acid metabolism were found in patients with SLE using metabolomics analysis. There are some caveats that should be taken into consideration while interpreting the metabolomics data. First, all six of these included studies were conducted in China. Because of globally different dietary habits and the influence that diet can have on the metabolic profile, combined with differences in genetic background, this might limit the global applicability of these models.[107] Second, metabolism can be altered by disease modifying drugs, including glucocorticoids such as prednisone, which are commonly given to patients with SLE.[108] Although these recent studies do report the use of medication, the use of different types of medication might influence their results.

It would be interesting to investigate whether these changes in metabolism, as well as the potentially related increased cardiovascular risk observed in SLE-patients can be observed during pre-SLE as well.[109] Overall, cytokines and metabolomic analyses result in promising AUCs when comparing patients with SLE to HCs. If future studies can confirm these findings in pre-SLE, this would confirm their potential as a diagnostic tool for SLE.

Levels of multiple different microRNAs, their presumed inhibitors circular RNAs, and lncRNAs were found to be altered in patients with SLE in comparison to healthy controls. Nevertheless, their overall specificity is lacking, as multiple of these circulating RNAs have been associated with a multitude of different diseases. For example, miR-21 has been associated with over 20 different diseases.[71] We hypothesize that changes in these circulating RNAs are mainly a reflection of non-disease-specific changes in the homeostasis, rather than a reflection of disease-specific changes. This would likely limit their use for diagnostics in SLE, as it might show signs of nonspecific disturbance of regular homeostasis, but would not aid the diagnostic process significantly.

A number of genes associated with an increased risk of development of SLE have been found in previous studies, although this only fully explains the presence of SLE in a small number of familial cases.[110] In particular, the increased expression of interferon (IFN)-inducible genes, commonly referred to as the IFN-signature, has been well-established in (a subgroup of) patients with SLE, as well as some other autoimmune diseases like Sjogren's syndrome.[111,112] Most of the studies investigating differential gene expression included in this review also found increased expression levels of IFN-related genes, such as IFI44. Nevertheless, the diagnostic accuracy of the IFN-signature has so far deemed to be insufficient to have been incorporated in the standard diagnostic process for SLE.

One study using a large multi-ethnic cohort managed to only marginally increase the diagnostic performance of conventional diagnostics using a polygenic risk score.[98] Potential reasons for this relative failure are differences between different ethnic backgrounds, resulting in too much heterogeneity in the cohort, or the prevalence of different pathophysiological profiles resulting in different gene expression profiles. Similar to the other areas studied, it would be interesting to investigate changes in genetic expression profiles in pre-SLE patients, to answer the question whether these changes are drivers of pathology, accumulating over time to eventually cause disease, or merely an epiphenomenon resulting from other pathophysiological manifestations, such as increased interferon production. Although the diagnostic performance of gene expression profiles might be limited, identification of these dysregulated pathways provides new potential drug targets.[113] Furthermore, identification of which pathways are dysregulated in individual patients has potential for use in precision medicine, tailored to the immunological changes in the individual.

It is noteworthy that many of the studies investigating gene expression profiles included in this review used datasets from previous cohorts, accessible via the GEO database.[87–97] Most of these studies do not mention how or why they selected the specific GSE they chose to use for their analysis. Future studies would benefit from describing this selection process more clearly to reduce the suggestion of a selection bias. Additionally, future studies should report more clearly what type of cells are used in the GSEs they used

for their study. GSEs in the GEO database are most often derived from peripheral blood mononuclear cells (PBMCs) or whole blood (WB), with some smaller datasets investigating DEGs in other cell lines, such as CD4+ T-cells or B-cells. Some studies included in this review are using data from pediatric SLE cohorts without stating so.[89,92,96] In one study, it seems that first-degree relatives of SLE patients, included in the cohort GSE24706, are classified as SLE patients.[97] Comparing results from different cell lines and different populations might influence results and should be clearly stated if done so.

In our manuscript, we tried to accurately display the findings that were relevant to our research question of all included studies. This meant that for all included studies we primarily highlighted the findings that were related to diagnostic potential for patients with SLE. We emphasize that for some of the included studies, showing diagnostic potential for patients with SLE was not the main goal of their study. This has resulted in the omission of important or main findings of some of the studies in our review, as they did not fit within our narrow research question. Additionally, we did not include articles discussing biomarkers primarily aimed at predicting specific organ involvement, such as lupus nephritis or CLE, as this was beyond the scope of our review. Although our review focusses on biomarkers predicting the development of SLE, it is noteworthy that approximately 15% of patients with CLE progress to SLE.[114] A review by Zhu et al. describes an association between the presence of ANAs, anti-dsDNA, anti-Sm, antiRNP autoantibodies and an elevated erythrocyte sedimentation rate with progression from CLE to SLE.[115] In our search, we found no additional studies describing biomarkers that predict the progression from CLE to SLE.

Overall, with our review we have not found any recent studies into biomarkers that are likely to change the diagnostic process in the short term. Nevertheless, some studies show promising results, meriting further investigation. For example, the peptide SLE_p27 outperformed anti-dsDNA and anti-Sm in diagnostic accuracy.[21] Similarly, the serum levels of SIGLEC1 showed a good sensitivity and specificity for SLE, although the authors note that SIGLEC1 is a surrogate marker for IFN-I pathway activation and elevated levels are also found in other inflammatory diseases or infections.[39] Additionally, multiple studies investigating metabolomics in patients with SLE found AUCs of 0.95 or higher for several biomarkers, although these different studies do not identify the same metabolomic biomarkers.[47–52]

Most of the studies included in this review show differences between healthy controls and patients with an established diagnosis of SLE, often already undergoing treatment which might also influence the results. It often is assumed that the findings of the studies comparing biomarkers in SLE-patients with HCs are transferrable to the early diagnostic phases, although this is rarely investigated properly. Studies investigating immunological changes in patients in the period leading up to their SLE-diagnosis are very limited.[5–19]

A small number of these studies have specifically investigated predictors of progression from incomplete SLE to definite SLE, fulfilling classification criteria.[14–19] A review by Lambers et al. provides a more extensive overview of these studies.[116] These studies have identified some potentially predictive biomarkers for progression in this group, such as increased levels of anti-dsDNA and lowered C3 levels. However, these studies represent a very selective population and are commonly using small cohorts with a limited number of pre-SLE patients.

One major improvement in diagnostic studies would be to perform studies using (serum) samples from pre-SLE patients, although this has significant methodological and logistic challenges, requiring large (population-based) cohorts with regular collection of bodily material. The 2023 EULAR SLE recommendations also named the establishment of pre-SLE cohorts as a key part of the future research agenda, to better be able to identify patients at risk to develop SLE and gain further insight into the disease development stage.[117] The creation of such population based cohorts would likely be a costly endeavour, but could contribute to research projects not only in the field of SLE, but for many different diseases and would therefore serve many purposes. Additionally, the increasing potential and accessibility of artificial intelligence models also have the potential to aid in the development of diagnostic multivariate models and the recognition of potential pre-SLE patients. To our knowledge, no data on newly established pre-SLE cohorts had been reported at the time of our search. It would be very interesting to further investigate promising biomarkers such as SLE_p27 and SIGLEC1 in pre-SLE cohorts. Although more elaborate to measure, changes in metabolomics in pre-SLE patients could also grant further insight in the pathophysiological process in the early stages of SLE.

REFERENCES

1. Aringer M, Costenbader KH, Daikh DI, et al. 2019 EULAR/ACR Classification Criteria for Systemic Lupus Erythematosus. *Arthritis Rheumatol*. 2019;71(9):1400-1412. doi:10.1002/art.40930
2. Kapsala NN, Nikolopoulos DS, Flouda SP, et al. From first symptoms to diagnosis of systemic lupus erythematosus: mapping the journey of patients in an observational study. *Clin Exp Rheumatol*. 2023;41(1):74-81. doi:10.55563/clinexprheumatol/x3s9td
3. Kernder A, Richter JG, Fischer-Betz R, et al. Delayed diagnosis adversely affects outcome in systemic lupus erythematosus: Cross sectional analysis of the LuLa cohort. *Lupus*. 2021;30(3):431-438. doi:10.1177/0961203320983445
4. Choi MY, Costenbader KH. Understanding the Concept of Pre-Clinical Autoimmunity: Prediction and Prevention of Systemic Lupus Erythematosus: Identifying Risk Factors and Developing Strategies Against Disease Development. *Front Immunol*. 2022;13:890522. doi:10.3389/fimmu.2022.890522
5. Arbuckle MR, McClain MT, Rubertone MV, et al. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N Engl J Med*. 2003;349(16):1526-1533. doi:10.1056/NEJMoa021933
6. Eriksson C, Kokkonen H, Johansson M, Hallmans G, Wadell G, Rantapää-Dahlqvist S. Autoantibodies predate the onset of systemic lupus erythematosus in northern Sweden. *Arthritis Res Ther*. 2011;13(1):R30. doi:10.1186/ar3258
7. Muñoz-Grajales C, Prokopec SD, Johnson SR, et al. Serological abnormalities that predict progression to systemic autoimmune rheumatic diseases in antinuclear antibody–positive individuals. *Rheumatology (Oxford)*. 2022;61(3):1092-1105. doi:10.1093/rheumatology/keab501
8. Munroe ME, Lu R, Zhao YD, et al. Altered type II interferon precedes autoantibody accrual and elevated type I interferon activity prior to systemic lupus erythematosus classification. *Ann Rheum Dis*. 2016;75(11):2014-2021. doi:10.1136/annrheumdis-2015-208140
9. Lu R, Munroe ME, Guthridge JM, et al. Dysregulation of innate and adaptive serum mediators precedes systemic lupus erythematosus classification and improves prognostic accuracy of autoantibodies. *J Autoimmun*. 2016;74:182-193. doi:10.1016/j.jaut.2016.06.001
10. Eriksson C, Rantapää-Dahlqvist S. Cytokines in relation to autoantibodies before onset of symptoms for systemic lupus erythematosus. *Lupus*. 2014;23(7):691-696. doi:10.1177/0961203314523869
11. Yusof MYM, Psarras A, El-Sherbiny YM, et al. Prediction of autoimmune connective tissue disease in an at-risk cohort: prognostic value of a novel two-score system for interferon status. *Ann Rheum Dis*. 2018;77(10):1432-1439. doi:10.1136/annrheumdis-2018-213386
12. Al Daabil M, Massarotti EM, Fine A, et al. Development of SLE among "potential SLE" patients seen in consultation: long-term follow-up. *Int J Clin Pract*. 2014;68(12):1508-1513. doi:10.1111/ijcp.12466
13. Heinlen LD, McClain MT, Merrill J, et al. Clinical criteria for systemic lupus erythematosus precede diagnosis, and associated autoantibodies are present before clinical symptoms. *Arthritis Rheum*. 2007;56(7):2344-2351. doi:10.1002/art.22665
14. Vilá LM, Mayor AM, Valentín AH, García-Soberal M, Vilá S. Clinical outcome and predictors of disease evolution in patients with incomplete lupus erythematosus. *Lupus*. 2000;9(2):110-115. doi:10.1191/096120300678828073
15. Li QZ, Zhou J, Wandstrat AE, et al. Protein array autoantibody profiles for insights into systemic lupus erythematosus and incomplete lupus syndromes. *Clin Exp Immunol*. 2007;147(1):60-70. doi:10.1111/j.1365-2249.2006.03251.x

16. Olsen NJ, Li QZ, Quan J, Wang L, Mutwally A, Karp DR. Autoantibody profiling to follow evolution of lupus syndromes. *Arthritis Res Ther*. 2012;14(4):R174. doi:10.1186/ar3927
17. Bhattacharya J, Pappas K, Toz B, et al. Serologic features of cohorts with variable genetic risk for systemic lupus erythematosus. *Mol Med*. 2018;24(1):24. doi:10.1186/s10020-018-0019-4
18. Munroe ME, Young KA, Kamen DL, et al. Discerning Risk of Disease Transition in Relatives of Systemic Lupus Erythematosus Patients Utilizing Soluble Mediators and Clinical Features. *Arthritis Rheumatol*. 2017;69(3):630-642. doi:10.1002/art.40004
19. Hallengren CS, Nived O, Sturfelt G. Outcome of incomplete systemic lupus erythematosus after 10 years. *Lupus*. 2004;13(2):85-88. doi:10.1191/0961203304lu477oa
20. Yaniv G, Twig G, Shor DBA, et al. A volcanic explosion of autoantibodies in systemic lupus erythematosus: A diversity of 180 different antibodies found in SLE patients. *Autoimmun Rev*. 2015;14(1):75-79. doi:10.1016/j.autrev.2014.10.003
21. Li X, Wang HX, Yin X, et al. Screening epitope peptides based on a phage-displayed random peptide and peptide microarrays to contribute to improving the diagnostic efficiency of systemic lupus erythematosus. *Immunol Lett*. 2023;259:30-36. doi:10.1016/j.imlet.2023.05.008
22. Zhan M, Wang Z, Bao H, et al. Antibodies against neutrophil extracellular traps (NETs) potentiate clinical performance of anti-double-stranded DNA antibodies in systemic lupus erythematosus. *Clin Immunol*. 2023;249:109297. doi:10.1016/j.clim.2023.109297
23. Arcani R, Bertin D, Bardin N, et al. Anti-NuMA antibodies: clinical associations and significance in patients with primary Sjögren's syndrome or systemic lupus erythematosus. *Rheumatology (Oxford)*. 2021;60(9):4074-4084. doi:10.1093/rheumatology/keaa881
24. Becker YLC, Gagné JP, Julien AS, et al. Identification of Mitofusin 1 and Complement Component 1q Subcomponent Binding Protein as Mitochondrial Targets in Systemic Lupus Erythematosus. *Arthritis Rheumatol*. 2022;74(7):1193-1203. doi:10.1002/art.42082
25. Li HL, Zhong LY, Kang YF, et al. Evaluation of serum Epstein-Barr virus envelope glycoproteins antibodies and their association with systemic autoimmune diseases. *J Med Virol*. 2024;96(4):e29595. doi:10.1002/jmv.29595
26. Hakkim A, Fürnrohr BG, Amann K, et al. Impairment of neutrophil extracellular trap degradation is associated with lupus nephritis. *Proc Natl Acad Sci U S A*. 2010;107(21):9813-9818. doi:10.1073/pnas.0909927107
27. Henning S, Westra J, Roozendaal C, et al. Immunoglobulin G/immunoglobulin M autoantibody ratios in incomplete systemic lupus erythematosus. *Scand J Rheumatol*. 2024;53(3):207-216. doi:10.1080/03009742.2024.2321700
28. Gómez-Bañuelos E, Fava A, Andrade F. An update on autoantibodies in systemic lupus erythematosus. *Curr Opin Rheumatol*. 2023;35(2):61-67. doi:10.1097/BOR.0000000000000922
29. Baysoy A, Bai Z, Satija R, Fan R. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol*. 2023;24(10):695-713. doi:10.1038/s41580-023-00615-w
30. Anchang CG, Xu C, Raimondo MG, et al. The Potential of OMICs Technologies for the Treatment of Immune-Mediated Inflammatory Diseases. *Int J Mol Sci*. 2021;22(14):7506. doi:10.3390/ijms22147506
31. Lindblom J, Beretta L, Borghi MO, Alarcón-Riquelme ME, Parodis I. Serum profiling identifies CCL8, CXCL13, and IL-1RA as markers of active disease in patients with systemic lupus erythematosus. *Front Immunol*. 2023;14:1257085. doi:10.3389/fimmu.2023.1257085
32. Zhao Y, Qi Y, Liu X, Cui Y, Zhao Z. Efficiency of Disease and Disease Activity Diagnosis Models of Systemic Lupus Erythematosus Based on Protein Array Analysis. *J Immunol Res*. 2022;2022:1830431. doi:10.1155/2022/1830431

33. He R, Tan X, Xiang J, et al. Semaphorin 4A as a Potential Biomarker for Diagnosis of Systemic Lupus Erythematosus. *Immunol Invest*. 2023;52(1):104-116. doi:10.1080/08820139.2022.2134024
34. Homa-Mlak I, Mazurek M, Majdan A, Mlak R, Majdan M, Mełecka-Massalska T. Serum Calprotectin - a NET Product - as a Biomarker of Disease Activity in Patients with Systemic Lupus Erythematosus: A Single-Center Case-Control Study from Poland. *Med Sci Monit*. 2022;28:e936534. doi:10.12659/MSM.936534
35. Kamel SM, Abdel Azeem Abd Elazeem MEMI, Mohamed RA, Kamel MM, Abdel Aleem Abdelaleem EA. High serum leptin and adiponectin levels as biomarkers of disease progression in Egyptian patients with active systemic lupus erythematosus. *Int J Immunopathol Pharmacol*. 2023;37:3946320231154988. doi:10.1177/03946320231154988
36. Li M, Zhang YJ, Liu DX, et al. Expression of caveolin family proteins in serum of patients with systemic lupus erythematosus. *Lupus*. 2021;30(11):1819-1828. doi:10.1177/09612033211035508
37. San Antonio E, Silván J, Sevilla-Montero J, et al. PSGL-1, ADAM8, and selectins as potential biomarkers in the diagnostic process of systemic lupus erythematosus and systemic sclerosis: an observational study. *Front Immunol*. 2024;15:1403104. doi:10.3389/fimmu.2024.1403104
38. Yang C, Li R, Xu WD, Huang AF. Increased levels of sirtuin-1 in systemic lupus erythematosus. *Int J Rheum Dis*. 2022;25(8):869-876. doi:10.1111/1756-185X.14360
39. Zorn-Pauly L, von Stuckrad ASL, Klotsche J, et al. Evaluation of SIGLEC1 in the diagnosis of suspected systemic lupus erythematosus. *Rheumatology (Oxford)*. 2022;61(8):3396-3400. doi:10.1093/rheumatology/keab875
40. Lambers WM, Westra J, Arends S, et al. Persistent low complement levels and increased interferon gene expression are predictive for disease progression in patients with incomplete systemic lupus erythematosus. *Joint Bone Spine*. 2022;89(5):105381. doi:10.1016/j.jbspin.2022.105381
41. Zhang X, Liu C, Yang J, et al. Potential biomarkers for diagnosis and assessment of disease activity in systemic lupus erythematosus. *Int Immunopharmacol*. 2022;111:109155. doi:10.1016/j.intimp.2022.109155
42. Ley K. The role of selectins in inflammation and disease. *Trends Mol Med*. 2003;9(6):263-268. doi:10.1016/S1471-4914(03)00071-6
43. Xiong YS, Cheng Y, Lin QS, et al. Increased expression of Siglec-1 on peripheral blood monocytes and its role in mononuclear cell reactivity to autoantigen in rheumatoid arthritis. *Rheumatology (Oxford)*. 2014;53(2):250-259. doi:10.1093/rheumatology/ket342
44. Patti GJ, Yanes O, Siuzdak G. Metabolomics: the apogee of the omic triology. *Nat Rev Mol Cell Biol*. 2012;13(4):263-269. doi:10.1038/nrm3314
45. Noerman S, Landberg R. Blood metabolite profiles linking dietary patterns with health—Toward precision nutrition. *J Intern Med*. 2023;293(4):408-432. doi:10.1111/joim.13596
46. Scrivo R, Casadei L, Valerio M, Priori R, Valesini G, Manetti C. Metabolomics Approach in Allergic and Rheumatic Diseases. *Curr Allergy Asthma Rep*. 2014;14(6):445. doi:10.1007/s11882-014-0445-5
47. He J, Tang D, Liu D, et al. Serum proteome and metabolome uncover novel biomarkers for the assessment of disease activity and diagnosing of systemic lupus erythematosus. *Clin Immunol*. 2023;251:109330. doi:10.1016/j.clim.2023.109330
48. Li S, Ding H, Qi Z, et al. Serum Metabolic Fingerprints Characterize Systemic Lupus Erythematosus. *Adv Sci (Weinh)*. 2024;11(2):e2304610. doi:10.1002/advs.202304610
49. Liang J, Han Z, Feng J, et al. Targeted metabolomics combined with machine learning to identify and validate new biomarkers for early SLE diagnosis and disease activity. *Clin Immunol*. 2024;264:110235. doi:10.1016/j.clim.2024.110235

50. Wang Y, Guo F, Guo Y, et al. Untargeted lipidomics reveals specific lipid abnormalities in systemic lupus erythematosus. *Clin Exp Rheumatol*. 2022;40(5):1011-1018. doi:10.55563/clinexprheumatol/ye2ua5
51. Zhang Q, Li X, Yin X, et al. Metabolomic profiling reveals serum L-pyroglutamic acid as a potential diagnostic biomarker for systemic lupus erythematosus. *Rheumatology (Oxford)*. 2021;60(2):598-606. doi:10.1093/rheumatology/keaa126
52. Zhang W, Zhao H, Du P, et al. Integration of metabolomics and lipidomics reveals serum biomarkers for systemic lupus erythematosus with different organs involvement. *Clin Immunol*. 2022;241:109057. doi:10.1016/j.clim.2022.109057
53. Li Y, Ma C, Liao S, et al. Combined proteomics and single cell RNA-sequencing analysis to identify biomarkers of disease diagnosis and disease exacerbation for systemic lupus erythematosus. *Front Immunol*. 2022;13:969509. doi:10.3389/fimmu.2022.969509
54. Zhou G, Wei P, Lan J, et al. TMT-based quantitative proteomics analysis and potential serum protein biomarkers for systemic lupus erythematosus. *Clin Chim Acta*. 2022;534:43-49. doi:10.1016/j.cca.2022.06.031
55. Kim JW, Baek WY, Jung JY, Kim HA, Lee SW, Suh CH. Longitudinal assessment of urinary ALCAM, HPX, and PRDX6 in Korean patients with systemic lupus erythematosus: implications for disease activity monitoring and treatment response. *Front Immunol*. 2024;15:1369385. doi:10.3389/fimmu.2024.1369385
56. Lee JS, Lee EJ, Yeom J, et al. Urine β-2-glycoprotein 1 as a biomarker for diagnosis of systemic lupus erythematosus. *Lupus*. 2021;30(8):1306-1313. doi:10.1177/09612033211014268
57. Kim JW, Jung JY, Lee SW, Baek WY, Kim HA, Suh CH. S100A8 in Serum, Urine, and Saliva as a Potential Biomarker for Systemic Lupus Erythematosus. *Front Immunol*. 2022;13:886209. doi:10.3389/fimmu.2022.886209
58. Chen DY, Huang YH, Chen YM, et al. ANA positivity and complement level in pleural fluid are potential diagnostic markers in discriminating lupus pleuritis from pleural effusion of other aetiologies. *Lupus Sci Med*. 2021;8(1):e000562. doi:10.1136/lupus-2021-000562
59. Song W, Wu F, Yan Y, et al. Gut microbiota landscape and potential biomarker identification in female patients with systemic lupus erythematosus using machine learning. *Front Cell Infect Microbiol*. 2023;13:1289124. doi:10.3389/fcimb.2023.1289124
60. Sprenkeler EGG, Zandstra J, van Kleef ND, et al. S100A8/A9 Is a Marker for the Release of Neutrophil Extracellular Traps and Induces Neutrophil Activation. *Cells*. 2022;11(2):236. doi:10.3390/cells11020236
61. Gaastra W, Kusters JG, van Duijkeren E, Lipman LJA. Escherichia fergusonii. *Vet Microbiol*. 2014;172(1):7-12. doi:10.1016/j.vetmic.2014.04.016
62. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front Endocrinol (Lausanne)*. 2018;9:402. doi:10.3389/fendo.2018.00402
63. Ibrahim MRK, Waly NG, Moness H, Ahmed SS, Ibrahem R. Serum miRNA-21, miRNA-146a and plasma cell free DNA as novel biomarkers for assessing systemic lupus erythematosus activity. *Mol Biol Rep*. 2023;50(12):10025-10036. doi:10.1007/s11033-023-08845-z
64. Abdul-Maksoud RS, Rashad NM, Elsayed WSH, Ali MA, Kamal NM, Zidan HE. Circulating miR-181a and miR-223 expression with the potential value of biomarkers for the diagnosis of systemic lupus erythematosus and predicting lupus nephritis. *J Gene Med*. 2021;23(5):e3326. doi:10.1002/jgm.3326
65. Li QJ, Chau J, Ebert PJR, et al. miR-181a Is an Intrinsic Modulator of T Cell Sensitivity and Selection. *Cell*. 2007;129(1):147-161. doi:10.1016/j.cell.2007.03.008

66. Johnnidis JB, Harris MH, Wheeler RT, et al. Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. *Nature*. 2008;451(7182):1125-1129. doi:10.1038/nature06607
67. Yan L, Jiang L, Wang B, et al. Novel microRNA biomarkers of systemic lupus erythematosus in plasma: miR-124-3p and miR-377-3p. *Clin Biochem*. 2022;107:55-61. doi:10.1016/j.clinbiochem.2022.05.004
68. Udhaya Kumar S, Thirumal Kumar D, Siva R, et al. Dysregulation of Signaling Pathways Due to Differentially Expressed Genes From the B-Cell Transcriptomes of Systemic Lupus Erythematosus Patients - A Bioinformatics Approach. *Front Bioeng Biotechnol*. 2020;8:276. doi:10.3389/fbioe.2020.00276
69. Zhou S, Zhang J, Luan P, et al. miR-183-5p Is a Potential Molecular Marker of Systemic Lupus Erythematosus. *J Immunol Res*. 2021;2021:5547635. doi:10.1155/2021/5547635
70. Ichiyama K, Gonzalez-Martin A, Kim BS, et al. The MicroRNA-183-96-182 Cluster Promotes T Helper 17 Cell Pathogenicity by Negatively Regulating Transcription Factor Foxo1 Expression. *Immunity*. 2016;44(6):1284-1298. doi:10.1016/j.immuni.2016.05.015
71. Jenike AE, Halushka MK. miR-21: a non-specific biomarker of all maladies. *Biomark Res*. 2021;9(1):18. doi:10.1186/s40364-021-00272-1
72. Yu CY, Kuo HC. The emerging roles and functions of circular RNAs and their generation. *J Biomed Sci*. 2019;26(1):29. doi:10.1186/s12929-019-0523-z
73. Jiang Z, Zhong Z, Miao Q, et al. circPTPN22 as a novel biomarker and ceRNA in peripheral blood mononuclear cells of rheumatoid arthritis. *Mol Med Rep*. 2021;24(2):617. doi:10.3892/mmr.2021.12256
74. Jiang Z, Li S, Jia Y, et al. CircPTPN22 modulates T-cell activation by sponging miR-4689 to regulate S1PR1 expression in patients with systemic lupus erythematosus. *Arthritis Res Ther*. 2023;25(1):206. doi:10.1186/s13075-023-03150-3
75. Luo Q, Li X, Fu B, et al. Expression profile and diagnostic value of circRNAs in peripheral blood from patients with systemic lupus erythematosus. *Mol Med Rep*. 2021;23(1):1. doi:10.3892/mmr.2020.11639
76. Luo Q, Ye Y, Zhang L, et al. Hsa_circ_0044235 and hsa_circ_0001947 as novel biomarkers in plasma of patients with new-onset systemic lupus erythematosus. *J Immunotoxicol*. 2023;20(1):2196453. doi:10.1080/1547691X.2023.2196453
77. Li W, Fan R, Zhou C, et al. Circular RNA expression profile of systemic lupus erythematosus and its clinical significance as a potential novel biomarker. *Genes Genomics*. 2022;44(11):1405-1414. doi:10.1007/s13258-022-01315-z
78. Ma S, Feng G, Li L, et al. Downregulation of circETS1 disrupts Th17/Treg homeostasis by inhibiting FOXP3 transcription: A new potential biomarker in systemic lupus erythematosus. *Lupus*. 2023;32(12):1430-1439. doi:10.1177/09612033231207545
79. Mattick JS, Amaral PP, Carninci P, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat Rev Mol Cell Biol*. 2023;24(6):430-447. doi:10.1038/s41580-022-00566-8
80. Peng XC, Ma LL, Miao JY, Xu SQ, Shuai ZW. Differential lncRNA profiles of blood plasma-derived exosomes from systemic lupus erythematosus. *Gene*. 2024;927:148713. doi:10.1016/j.gene.2024.148713
81. You Y, Zhao X, Wu Y, et al. Integrated Transcriptome Profiling Revealed That Elevated Long Non-Coding RNA-AC007278.2 Expression Repressed CCR7 Transcription in Systemic Lupus Erythematosus. *Front Immunol*. 2021;12:615859. doi:10.3389/fimmu.2021.615859

82. Zhu X, Chen Y, Yin Z, et al. Novel potential lncRNA biomarker in B cells indicates essential pathogenic pathway activation in patients with SLE. *Lupus Sci Med*. 2024;11(1):e001065. doi:10.1136/lupus-2023-001065
83. Jiang L, Qi A, Yang H, et al. LncRNA SNHG1 serves as a biomarker for systemic lupus erythematosus and participates in the disease progression. *APMIS*. 2024;132(7):507-514. doi:10.1111/apm.13410
84. Mahmoud RH, Fouad NA, Hefzy EM, et al. The potential role of serum expression profile of long non coding RNAs, Cox2 and HOTAIR as novel diagnostic biomarkers in systemic lupus erythematosus. *PLoS One*. 2022;17(8):e0268176. doi:10.1371/journal.pone.0268176
85. Wang Y, He J, Ma H, et al. NR_103776.1 as a novel diagnostic biomarker for systemic lupus erythematosus. *Ir J Med Sci*. 2024;193(1):211-221. doi:10.1007/s11845-023-03420-8
86. Yang H, Wang S, Wang F, Bai X, Ren J. Serum long non-coding Ribonucleic Acid H19 serves as a biomarker for systemic lupus erythematosus and participates in the disease progression. *Lupus*. 2024;33(7):675-684. doi:10.1177/09612033241243175
87. Chen W, Li ZY, Huang L, et al. Integrative Bioinformatics Analysis Identifies DDX60 as a Potential Biomarker for Systemic Lupus Erythematosus. *Dis Markers*. 2023;2023:8564650. doi:10.1155/2023/8564650
88. Fu Q, He Q, Dong Q, et al. The role of cyclic GMP-AMP synthase and Interferon-I-inducible protein 16 as candidatebiomarkers of systemic lupus erythematosus. *Clin Chim Acta*. 2022;524:69-77. doi:10.1016/j.cca.2021.11.003
89. Meng XW, Cheng ZL, Lu ZY, Tan YN, Jia XY, Zhang M. MX2: Identification and systematic mechanistic analysis of a novel immune-related biomarker for systemic lupus erythematosus. *Front Immunol*. 2022;13:978851. doi:10.3389/fimmu.2022.978851
90. Shen M, Duan C, Xie C, et al. Identification of key interferon-stimulated genes for indicating the condition of patients with systemic lupus erythematosus. *Front Immunol*. 2022;13:962393. doi:10.3389/fimmu.2022.962393
91. Sun HW, Zhang X, Shen CC. The shared circulating diagnostic biomarkers and molecular mechanisms of systemic lupus erythematosus and inflammatory bowel disease. *Front Immunol*. 2024;15:1354348. doi:10.3389/fimmu.2024.1354348
92. Wei M, Dong Q, Chen S, Wang J, Yang H, Xiong Q. Identification of Key Biomarkers in Systemic Lupus Erythematosus by a Multi-Cohort Analysis. *Front Immunol*. 2022;13:928623. doi:10.3389/fimmu.2022.928623
93. Yang Y, Zhang H, Xiao X, Guo M. Identification of EPSTI1 as a new potential biomarker for SLE based on GEO database. *Clin Rheumatol*. 2024;43(5):1531-1540. doi:10.1007/s10067-024-06881-z
94. Zhang W, Liang G, Zhou H, et al. Identification of potential biomarkers for systemic lupus erythematosus by integrated analysis of gene expression and methylation data. *Clin Rheumatol*. 2023;42(5):1423-1433. doi:10.1007/s10067-022-06495-3
95. Zhao X, Zhang L, Wang J, et al. Identification of key biomarkers and immune infiltration in systemic lupus erythematosus by integrated bioinformatics analysis. *J Transl Med*. 2021;19(1):35. doi:10.1186/s12967-020-02698-x
96. Zhao X, Duan L, Cui D, Xie J. Exploration of biomarkers for systemic lupus erythematosus by machine-learning analysis. *BMC Immunol*. 2023;24(1):44. doi:10.1186/s12865-023-00581-0
97. Zhong Y, Zhang W, Hong X, et al. Screening Biomarkers for Systemic Lupus Erythematosus Based on Machine Learning and Exploring Their Expression Correlations With the Ratios of Various Immune Cells. *Front Immunol*. 2022;13:873787. doi:10.3389/fimmu.2022.873787

98. Khunsriraksakul C, Li Q, Markus H, et al. Multi-ancestry and multi-trait genome-wide association meta-analyses inform clinical risk prediction for systemic lupus erythematosus. *Nat Commun*. 2023;14(1):668. doi:10.1038/s41467-023-36306-5
99. Nocturne G, Mariette X. Interferon signature in systemic autoimmune diseases: what does it mean? *RMD Open*. 2022;8(2):e002687. doi:10.1136/rmdopen-2022-002687
100. Jie J, Gong Y, Luo S, Yang X, Guo K. Genetically predicted associations between circulating cytokines and autoimmune diseases: a bidirectional two-sample Mendelian randomization. *Front Immunol*. 2024;15:1404260. doi:10.3389/fimmu.2024.1404260
101. Moore LD, Le T, Fan G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology*. 2013;38(1):23-38. doi:10.1038/npp.2012.112
102. Ehtesham N, Alesaeidi S, Mohammad Zadeh D, et al. Significant heightened methylation levels of RUNX3 gene promoter in patients with systemic lupus erythematosus. *Lupus*. 2024;33(6):547-554. doi:10.1177/09612033241241850
103. Gravand A, Alesaeidi S, Khoshbakht S, et al. Demethylation of CDKN2A in systemic lupus erythematosus and rheumatoid arthritis: a blood biomarker for diagnosis and assessment of disease activity. *Clin Rheumatol*. 2023;42(12):3387-3395. doi:10.1007/s10067-023-06736-z
104. Li QL, Ito K, Sakakura C, et al. Causal Relationship between the Loss of RUNX3 Expression and Gastric Cancer. *Cell*. 2002;109(1):113-124. doi:10.1016/S0092-8674(02)00690-6
105. Lau QC, Raja E, Salto-Tellez M, et al. RUNX3 Is Frequently Inactivated by Dual Mechanisms of Protein Mislocalization and Promoter Hypermethylation in Breast Cancer. *Cancer Res*. 2006;66(13):6512-6520. doi:10.1158/0008-5472.CAN-06-0369
106. Mistry P, Kaplan MJ. Cell death in the pathogenesis of systemic lupus erythematosus and lupus nephritis. *Clin Immunol*. 2017;185:59-73. doi:10.1016/j.clim.2016.08.010
107. Chen L, Zhernakova DV, Kurilshikov A, et al. Influence of the microbiome, diet and genetics on inter-individual variation in the human plasma metabolome. *Nat Med*. 2022;28(11):2333-2343. doi:10.1038/s41591-022-02014-8
108. Bordag N, Klie S, Jürchott K, et al. Glucocorticoid (dexamethasone)-induced metabolome changes in healthy males suggest prediction of response and side effects. *Sci Rep*. 2015;5(1):15954. doi:10.1038/srep15954
109. Wong CY, Ma BMY, Zhang D, Cheung W, Chan TM, Yap DYH. Cardiovascular risk factors and complications in patients with systemic lupus erythematosus with and without nephritis: a systematic review and meta-analysis. *Lupus Sci Med*. 2024;11(1):e001152. doi:10.1136/lupus-2024-001152
110. Vinuesa CG, Shen N, Ware T. Genetics of SLE: mechanistic insights from monogenic disease and disease-associated variants. *Nat Rev Nephrol*. 2023;19(9):558-572. doi:10.1038/s41581-023-00732-x
111. Baechler EC, Batliwalla FM, Karypis G, et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci U S A*. 2003;100(5):2610-2615. doi:10.1073/pnas.0337679100
112. Nguyen CQ, Peck AB. The Interferon-Signature of Sjögren's Syndrome: How Unique Biomarkers Can Identify Underlying Inflammatory and Immunopathological Mechanisms of Specific Diseases. *Front Immunol*. 2013;4:142. doi:10.3389/fimmu.2013.00142
113. Lindblom J, Toro-Domínguez D, Carnero-Montoro E, et al. Distinct gene dysregulation patterns herald precision medicine potentiality in systemic lupus erythematosus. *J Autoimmun*. 2023;136:103025. doi:10.1016/j.jaut.2023.103025
114. Curtiss P, Walker AM, Chong BF. A Systematic Review of the Progression of Cutaneous Lupus to Systemic Lupus Erythematosus. *Front Immunol*. 2022;13:866319. doi:10.3389/fimmu.2022.866319

115. Zhu JL, Black SM, Chong BF. Role of biomarkers in the diagnosis and prognosis of patients with cutaneous lupus erythematosus. *Ann Transl Med*. 2021;9(5):429-429. doi:10.21037/atm-20-5232
116. Lambers WM, Westra J, Bootsma H, de Leeuw K. From incomplete to complete systemic lupus erythematosus; A review of the predictive serological immune markers. *Semin Arthritis Rheum*. 2021;51(1):43-48. doi:10.1016/j.semarthrit.2020.11.006
117. Fanouriakis A, Kostopoulou M, Andersen J, et al. EULAR recommendations for the management of systemic lupus erythematosus: 2023 update. *Ann Rheum Dis*. 2024;83(1):15-29. doi:10.1136/ard-2023-224762

8

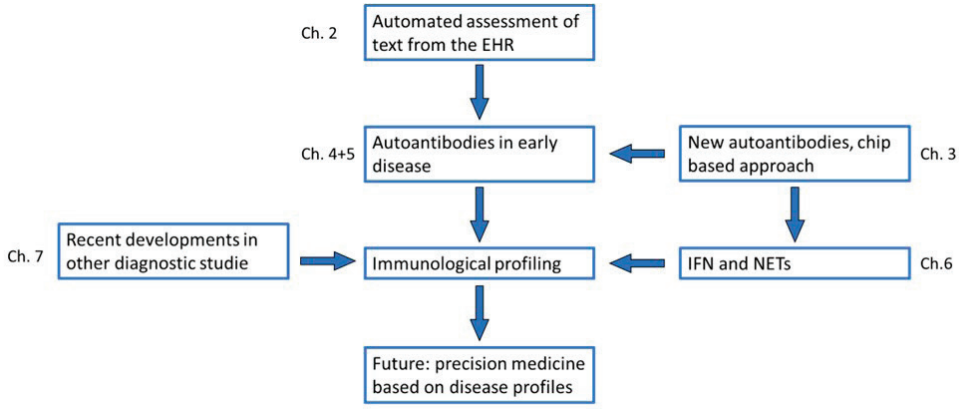
General Discussion

GENERAL DISCUSSION

General introduction to the discussion

The diagnostic process for patients with SLE is often difficult, frequently resulting in a delay between the constitution of the first symptoms to the eventual diagnosis of several years.[1,2] Better methods to diagnose patients with SLE in an early stage would reduce the burden of early disease and potentially prevent accumulation of organ damage.[3] After the diagnosis, there is often also uncertainty about the future. There is a need for biomarkers that can be used to improve the diagnostic process, and biomarkers that can predict the course of the disease and response to treatment, as both the course of the disease and whether specific medication will work well for individual patients are currently impossible to predict.

In this thesis, I explored a novel method to extrapolate relevant clinical information from the electronic health records (EHR) in an automated fashion for research purposes, and explored the use of autoantibodies to immunologically profile patients with SLE (Figure 1). The goal of the creation of such profiles is to identify profiles that can be used to improve the diagnostic process, aid in predicting the course of the disease, and hopefully pave the way to the implementation of precision medicine for patients with SLE.



Methodology / epidemiology

SLE has an estimated prevalence of approximately 43.7 per 100.000 persons worldwide, with some notable regional differences.[4–6] Due to the relative rarity of the disease, research into SLE can be challenging, as the number of patients able to participate in studies is limited. Therefore, it is of importance to effectively utilize the data that is available.

Instead of gathering new data for each research question, research can also be performed using data that is already available. Retrospective cohort studies in general are

performed by manual patient chart review of electronic health records (EHR). This has been the golden standard for this type of research for years, because of the reliability and ability to assess nuances and interpret the data in the EHR. The main downsides of the method are the large time investment needed to examine the records and inter-observer variability.

In **Chapter 2**, we explored the use of a novel method to extrapolate relevant clinical information from the electronic health records (EHR) in an automated fashion for research purposes. We show that a text mining algorithm can be used to assess the diagnosis recorded in the EHR of patients with autoimmune diseases. The text mining algorithm largely ameliorates the downsides associated with manual patient chart review, as it is much faster in scanning the data from the EHR and has no inter- or intra-observer variability. In theory, a well-functioning text mining algorithm also has the additional benefit of being relatively easy to adapt in order to facilitate data evaluation for amended or expanded research questions. With the use of manual patient chart review, this would require a complete revision of all patient charts. Additionally, a portable algorithm would facilitate replication or expansion of single centre retrospective cohort studies in different centres by greatly reducing the effort needed for such studies.

Text mining algorithms can be used to extract many different valuable resources from the EHR, ranging from relatively simple single outcomes such as the smoking status of patients, to the more elaborate task of screening cohorts of patients for potential inclusion in research trials.[7,8] We show that a text mining algorithm can even be used accurately to identify patients in a complex disease such as SLE. Nevertheless, the reliability of our algorithm, as well as the reliability of similar methods of automated data extraction of free text from the EHR, could use improvement. For example, our algorithm had a sensitivity and specificity of approximately 95% to discern the presence of SLE, meaning that 5% of patients were misclassified by the algorithm. Although this is only a small percentage, it shows that it is still inferior to the golden standard of manual patient chart review. In our study, we used a relatively simple rule-based algorithm to assess the diagnoses of the included patients. Recently, artificial intelligence (AI) tools that use Natural Language Processing (NLP), are increasingly being used for the same purpose as our algorithm, to extract relevant clinical information from the EHR.[9] Although this field is very much still in development, it is likely that these NLP models will further improve on and expand the useability of models made for extracting data from the EHR.[10]

A good functioning algorithm to extract relevant information from the EHR has a lot of potential. In our studies, we use it to identify and classify patients for research purposes. It could also potentially be used to create models to identify patients with early signs of a disease such as SLE or patients that exhibit early signs of adverse events or flares, or identify patients that might benefit from tapering treatment. Ideally, this would evaluate the

data from the EHR on a regular basis, as a real-life monitoring tool. Although these tools should not be used to make diagnoses on their own, they might assist the healthcare professional in identifying patients that exhibit symptoms fitting a specific pattern associated with an autoimmune disease. Implementation of algorithm-assisted decision making has been investigated in particular in image analysis. For example, studies have shown that such models can increase the diagnostic accuracy of stress echocardiographs and in diabetic retinopathy, compared to the clinicians alone.[11,12] It is reasonable to assume that similar methods could be used to evaluate the text from the EHR to assist clinicians in their decision making. The principle of using readily available clinical data for a real-life monitoring tool to assist in clinical decision making is currently being investigated by researchers from our department, who have developed a dynamic flare prediction model to predict flares during tapering of disease-modifying antirheumatic drugs (DMARDs) in rheumatoid arthritis. This model currently does not use free text recorded in the EHR, and could potentially be improved further by implementing automated evaluation of key features in the free text of the EHR.[13] Our study and similar studies show that relevant research data is readily available and extractable from the EHR. However, a major hurdle for the large-scale implementation of these models is portability. Automated data extraction models have problems with portability due to the differences in digital infrastructure in different hospitals, meaning that any digital solution needs to be amended in new situations to fit this digital infrastructure.

The large-scale implementation of such models would require systemic effort of the healthcare industry. Future EHRs should be designed in a uniform manner. Not only would this facilitate easy portability of data extraction methods for research, but it would also facilitate easy data exchange between healthcare providers. Currently, it is very difficult to exchange patient records between health care providers, for a large part due to the differences in EHR systems. This results in an inability for hospital personnel to easily access potentially clinically relevant data from the General Practitioner, as well as data from other hospitals the patient might have previously been in care, for example after moving or in emergency situations. Currently, this problem is ameliorated by the transfer of medical letters containing summaries of the patient's history. Although these are very useful, these summaries can be incomplete and miss important details, and because of the inability to import them easily, they might be harder to find in the EHR (for example under "imported documents", rather than under "medical letters"). A uniformly designed EHR would enable easy access to data from other healthcare providers, both for regular care and research purposes.

Implementation of these models would also require significant effort of healthcare personnel, as they are the main providers of input in the EHR. Both manual patient chart review and automated data extraction are heavily reliant on the quality of the data recorded in the EHR, as inaccurate or incomplete data recorded in the EHR will result in

inaccurate or incomplete outcomes during data extraction. Medical organisations, such as the EULAR, already have recommendations for what clinical information should be recorded during each patient visit. I would suggest that these recommendations should be revisited to discuss what additional data should be recorded for research purposes. It is important to also take the patients perspective into consideration in this effort, as the patients' needs are not always reflected properly on the research agenda of medical professionals. For example, research often focusses more on concretely measurable outcomes such as biomarkers, whereas patients with SLE structurally report the difficultly measurable fatigue as one of the main factors impacting their quality of life.[14,15]

If a uniform EHR is adopted, the resulting recommendations could be translated into a template that can be used at each patient visit to record these data. Although it is naïve to assume that such a template would result in no missing data, due to various reasons such as time constraints, it is likely that it would reduce missing data and provide semi-structured data for research. Uniformity of the clinical data should enable (inter) national cooperation between hospitals. The corresponding medical organisations could play a coordinating role for researchers wanting to do research on large datasets from multiple hospitals to further facilitate these efforts.

Autoantibodies and other biomarkers

Autoantibodies, and in particular anti-dsDNA antibodies, have long been a hallmark immunological manifestation associated with SLE. As early as 1957, it was reported that serum of patients with SLE reacted with DNA, leading to the suggestion that this reaction was the result of the presence of autoantibodies against DNA or nucleoproteins.[16] In the following years, over 100 different autoantibodies that are associated with SLE have been discovered.[17] Many of these antibodies target nuclear proteins, and can therefore be characterized as antinuclear antibodies (ANAs). The presence of ANAs is associated with a number of different autoimmune diseases, with different specific antibodies associated with different (although sometimes multiple) diseases. Despite these associations, presence of these specific ANAs alone is generally insufficient to make a diagnosis, as this often requires the presence of a number of other symptoms or immunological manifestations to be present as well.

Autoantibodies are believed to play a central role in the pathophysiology in autoimmune diseases. Previous studies have shown that ANAs can be present very early in the disease, frequently even preceding the first symptoms of the disease by several years. For example, Arbuckle et al. showed that ANAs were present in the serum of approximately 80% of SLE patients, several years before their eventual diagnosis, which suggests a great predictive value of ANAs for the development of SLE.[18]

In their article, Arbuckle et al. hypothesize that there are (at least) three phases in the development of autoantibodies, going from normal immunity to benign (asymptomatic) autoimmunity, to pathological autoimmunity and eventually reaching clinical illness. They suggest that some autoantibodies that are not as stringently linked to autoimmune diseases might present in the phase of benign autoimmunity, whereas more ominous usually autoantibodies present in the phase of pathological autoimmunity.

However, the presence of ANAs is not exclusive to (future) patients with autoimmune diseases, as ANAs can be found in up to 20% of healthy individuals in the general population, severely limiting their potential use as a screening tool in the general population.[19,20] According to the theory of Arbuckle et al., these healthy individuals are likely exhibiting benign autoimmunity. It is likely that many patients who visit the Rheumatology outpatient clinic experience some level of benign or even pathological autoimmunity, without being outright diagnosed with an autoimmune disease.

In **Chapter 5**, we tried to answer the question how many undiagnosed individuals with benign or pathological autoimmunity progress to developing an actual autoimmune disease. We found that only a small number of patients progress from having symptoms associated with an autoimmune disease to actually being diagnosed with that disease. However, we found that most patients that did develop an autoimmune disease had positive ANAs, and in particular more frequently had specific ANAs that were associated with the symptoms they were experiencing.

We also show that new antibodies are unlikely to form after the initial ANA test has been used, suggesting that the establishment of the autoantibodies precedes the symptoms of the patients, as that is usually the moment patients will visit a doctor. Furthermore, very little patients that developed new ANAs later in their disease course. Similar to a study by Yeo et al., which showed that only 1.1% of patients had a clinically relevant result of a repeated ANA test.[21] This indicates that repeated ANA testing is of limited use, as formation of relevant autoantibodies usually precedes the moment of initial testing for autoantibodies. A new diagnostic strategy could involve testing patients only once for ANAs when they are experiencing symptoms associated with an ANA-related disease. If the symptoms of the patient are insufficient to establish a diagnosis at that time and no specific ANAs are found that are associated with the suspected disease, the patient could likely be released from care safely and should not have to remain in follow-up. Although in our study only 12.5% of the patient with specific ANAs associated with their suspected disease eventually were diagnosed with this disease, this strategy would drastically lower the number of patients that are kept in follow-up.

Although the predictive value of individual ANAs for the development of autoimmune diseases might be limited, their potential might be amplified by using multiple ANAs

simultaneously to develop autoantibody profiles. In **Chapter 4**, we showed that many future SLE patients exhibit an autoantibody profile that can be used to identify them, even before the time of their eventual diagnosis. In our study, we used a lasso regression model to develop autoantibody profiles to distinguish patients with SLE and two different control groups. The Non-IMID group consisted of patients visiting the outpatient clinic with symptoms vaguely related to SLE, and the BBD group consisted of healthy blood bank donors.

The reason we used these two different groups is the presumed difference in (subclinical) autoimmunity between these two groups. The BBD group were healthy blood bank donors with little to no expected autoimmunity, whereas the Non-IMID group were actual patients visiting the hospital because of complaints that could, although not likely, be related to an autoimmune disease. We presumed the second group would display a higher level of (subclinical) autoimmunity, which would therefore influence the results of the final model. Furthermore, a model that would be able to identify (future) patients with SLE out of a group of patients suspected of having SLE, would have more clinical relevance than being able to identify these (future) patients out of the general population. Our assumption that the BBD group would display a different level of autoimmunity is corroborated in Figure 1 (page 67), where the UMAP graph shows the BBD group clustering distinctly from the Non-IMID group. This is an important distinction and something that sets our study apart from similar studies), as most studies compare both ends of the autoimmunity spectrum by comparing patients solely with healthy controls, rather than with non-diagnosed patients. [22–24]

In **Chapter 3**, we also investigated the contribution of individual autoantibodies to the regression model. Besides a number of autoantibodies well known to be associated with SLE, we also found a high prevalence of antibodies against CpG oligodeoxynucleotides (ODNs), with a strong contribution to the regression model. CpG ODNs are known to stimulate TLR9, which is expressed in both B-cells and plasmacytoid dendritic cells (pDCs). pDCs are known to be able to produce large amounts of type I interferons, and play an important role in the constitution of the IFN-signature in patients with SLE.[25] Interestingly however, other studies have shown that the TLR9-mediated type I IFN production of pDCs of SLE patients is attenuated compared to healthy controls, whereas TLR7-mediated stimulation is upregulated in patients with SLE and appears to have a positive feedback loop.[26–28] In **Chapter 6**, we further investigated the associations between different autoantibodies and the IFN signature. We found that antibodies against EphB2, RipP1, and RNP70 were positively associated with a high IFN score. RipP1 and RNP70 are both (part of) RNA-binding proteins. RNA is a ligand for TLR7, further suggesting that the IFN-signature in patients with SLE is largely driven by TLR7, rather than by TLR9.[29]

Studies have also shown that in B-cells, TLR9 activation appears to have a tolerogenic effect.[30–33] This tolerogenic effect of TLR9 in B-cells might be caused by antagonism of TLR7, as TLR9 stimulation restrains TLR7 signalling.[34,35] However, in patients with SLE, this immunomodulatory function of TLR9 in B-cells might be impaired, as Gies et al. showed that B-cells from SLE patients displayed defective TLR9 responses, resulting in more pro-inflammatory and less anti-inflammatory responses.[32] TLR9 appears to have a dual function in the pathogenesis of SLE, where it functions pro-inflammatory by stimulating IFN production in pDCs, although limited compared to healthy patients, but immunomodulating in B-cells by limiting TLR7 signalling. TLR7 and TLR9 are known target for pharmacological treatment in SLE, as hydroxychloroquine functions as a TLR7 and TLR9 antagonist.[36]

Despite, or more likely because of, their apparent role in the pathophysiology of SLE, autoantibodies have proven to be an important tool in the combatting of the disease. In **Chapter 7**, we reviewed recent developments in the field of autoantibodies, as well as other diagnostic biomarkers such as microRNAs and genetic markers. These biomarkers can be used to construct immunological profiles to improve diagnostics and result in earlier diagnoses and start of treatment. These immunological profiles also have the potential to provide insight into the pathophysiological mechanisms that are at play in individual patients, which could also facilitate further improvements in the treatment of individual patients: precision medicine.

Towards precision medicine in SLE

In 1999, reporters Robert Langreth and Michael Waldholz introduced the term "personalized medicine" in an article in The Wall Street Journal.[37] Their article, called "New Era of Personalized Medicine: Targeting Drugs for Each Unique Genetic Profile" (later republished in *The Oncologist*), describes the creation of a consortium of drug companies and several academic institutions, which aimed to distinguish the genetic variability among individuals to improve the safety and potency of drugs. The idea was that increased knowledge of the genetic makeup of the individual patients would help identify patients likely to respond well to specific drugs, as well as help to create drugs designed to target specific genetic profiles. Nowadays, personalized medicine is often called "precision medicine", and also encompasses the use of targeted therapies based on other patient characteristics besides genetics.

In order to utilize precision medicine, biomarkers that can predict the response to specific drugs or course of the disease are necessary, as well as a reliable test to assess these biomarkers. Such tests are commonly referred to as "companion diagnostics". Historically, these companion diagnostics have mainly been based on genetics and immunohistochemistry, and predominantly focussed on the field of (hemato-)oncology. Diagnostics

tests for HER2 overexpression in breast cancer and the BCR-ABL fusion gene in chronic myeloid leukemia are two examples of such companion diagnostics.

One of the earliest examples of targeted therapy is trastuzumab.[38–40] Trastuzumab is a monoclonal antibody targeting HER2, which is overexpressed on malignant cells in approximately 20% of patients with breast cancer. Trastuzumab is specifically given to patients with HER2 overexpression on tumor cells as targeted therapy, and therefore the use of trastuzumab is dependent on the result of the companion diagnostic. Treatment with trastuzumab has significantly improved the disease-free survival of these patients.[41] Because of this additional treatment option for a subgroup of patients with breast cancer, testing for HER2 expression has become standard of care.[42] The success of a drug like trastuzumab highlights the importance of the identification of subgroups of patients with characteristics, such as HER2 overexpression, that might respond better to specific targeted therapies.

However, the development of precision medicine tools gets more complicated when it involves a disease with multiple underlying pathophysiological pathways, like many autoimmune diseases. Efforts to identify predictors of the course of the disease or predictors of flares of disease activity have yielded some promising targets, however variability in measurement methods and definitions of outcome measurements have severely limited the applicability and generalizability of these biomarkers.[43–46] In an effort to combat this variability, definitions for clinical remission and lupus low disease activity state (LLDAS) have been developed.[47,48]

According to the 2023 EULAR recommendations for the treatment of SLE, the goal of treatment should be to reach remission (as defined by the DORIS criteria, requiring a SLEDAI of 0 and a physician global assessment scale of 0.5 or lower), or low disease activity if remission is not attainable.[15,47] The EULAR recommendations note that current pharmacological treatment strategies for SLE are directed by patient characteristics (such as age and ethnicity), type and severity of organ involvement, treatment-related harms, comorbidities, risk for progressive organ damage and patient preferences.[15] Although these guidelines do mention that black patients with lupus nephritis might be more responsive to treatment with mycophenolate than cyclophosphamide, it does not mention any other characteristics that might influence the response to treatment of individual patients. Current treatment strategies for SLE can be summarized as a one-size-fits-all basis of immunosuppressive treatment with hydroxychloroquine with a step-up regiment in case of persistent disease activity or organ-specific related damage. Precision medicine, in the sense of adjusting the therapy to patient-specific immunological or genetic characteristics, has not yet found a place in the current treatment selection for patients with SLE. However, it is likely that this will change in the future, as more specific targeted therapies are being developed. These targeted drugs, some of which already have been approved,

and some of which are currently in trial stages, are targeting different elements of the immune system such as interferons (IFN), BLyS, JAK and BTK inhitors, and regulatory T cells.[49–52] It is imaginable that these drugs will be utilized most efficiently in patients in which the targeted pathways are most central to the constitution of symptoms of SLE. The 2023 EULAR recommendation also note the need for research into biomarkers for the use of prediction of response to specific therapies, further indicating a shift towards precision medicine.

One of the drugs that has the potential to be used for precision medicine in patients with SLE is anifrolumab. Anifrolumab targets the type 1 IFN pathway by targeting the type 1 interferon receptor, inhibiting downstream signalling. An analysis of two phase III trials of anifrolumab shows a better response rate of patients with an elevated type 1 IFN gene signature (IFN-high), compared to patients with a low IFN gene signature (IFN-low), as measured by British Isles Lupus Assessment Group-based Composite Lupus Assessment (BICLA).[49] This would suggest a potential for the IFN signature to function as a companion diagnostic to select patients suited for treatment with anifrolumab, although this has not been translated to clinical practice as of yet. One possible reason could be the limited difference in disease activity reduction between IFN-high and IFN-low patients in these studies, although this could be influenced by an association between an elevated IFN gene signature and greater disease activity, resulting in higher disease activity in the IFN-high placebo group.[53] Another possible reason could be that one of the phase III trials (TULIP-1) failed to reach its primary endpoint of an improved treatment response in patients treated with anifrolumab, as measured by the SLE responder index-4 (SRI-4). This study did however show a significant improvement in the BICLA response in patients treated with anifrolumab, compared to placebo, a result that was replicated in the following TULIP-2 trial.[54,55] It is possible that this initial failure to reach the primary study end point has made health care professionals cautious to embrace anifrolumab and additional testing of the IFN signature, which currently is not part of the regular diagnostic workup for SLE patients.

With multiple drugs targeting different pathways, and potentially them all requiring testing of the involvement of these pathways to determine the preferred treatment for individual patients, this testing can quickly become expensive and elaborate. Although reducing the time to diagnosis and better treatment selection for individual patients will likely result in lower costs of health care utilization overall, the initial costs associated with this testing can be a hurdle barring implementation in clinical practice. However, there is a possibility that multiple pathways can be assessed simultaneously using autoantibody profiles as a proxy for testing of these different pathways individually. This would likely reduce both costs and time investment needed to facilitate the testing.

Although there are some associations between some antibodies and specific manifestations of SLE, such as an increased prevalence of Raynaud's phenomenon in patients with anti-U1RNP antibodies, and nephritis in patients with anti-c1q antibodies, this currently does not have consequences for the treatment of these patients.[56,57] Some suggest that the profile of autoantibodies are likely to be correlated with the underlying disease aetiology and mechanism.[58] In **Chapter 6**, we indeed show an increased prevalence of autoantibodies targeting ribosomal proteins in patients with a high IFN signature. This hypothesis is further corroborated by recent studies focussing on the associations between specific autoantibodies and other different immunological pathways.[59–61]

Recent improvements in the development of high throughput protein assays have given the opportunity to investigate key proteins in multiple pathways at once, facilitating the simultaneous analysis of multiple autoantibodies. These protein assays can be used as an investigatory tool to discover new biomarkers in hypothesis-generating studies, similarly to how we showed the presence of anti-CpG antibodies in patients with SLE in **Chapter 3**. Additionally, these assays allow the creation of immunological profiles that could represent the affected immunological pathways and therefore have potential to function as a companion diagnostic. Furthermore, profiles of autoantibodies can aid in the diagnostic process as well, as we showed in **Chapter 4**. Currently, we are investigating whether these immunological profiles can predict the general severity and course of the disease, as well as predict flares of disease activity.

So what has got to happen to make the step to personalized medicine in the treatment of patients with SLE? The first step is that immunological subtypes of patients with SLE need to be defined. These immunological subtypes should be based on profiles of autoantibodies, potentially expanded by the presence of chemokines and cytokines, and gene expression data.[62] Similar to the use of trastuzumab only in patients with HER2 positive breast cancer, certain types of medication could likely be used most effectively in patients with certain patient-specific characteristics. For example, patients with an immunological subtype that is largely IFN-driven might benefit more from treatment targeting the IFN-pathway than patients with an immunological subtype more related to B-cell dysfunction, even if they display similar symptoms. After these immunological subtypes have been defined, trials should be undertaken to investigate whether treatment selection based on these immunological subtypes results in better outcomes than the current step-up regime. If this proves to be successful, classification of subgroups of patients with SLE based on their clinical manifestations should be abandoned in favour of classification based on their immunological manifestations for the purpose of treatment selection. Two separate diseases might result in similar symptoms, but have a distinct pathophysiology and therefore require different treatment. It makes sense to target the immunological pathways that are demonstrably affected in specific patients, rather than focus solely on the symptoms these patients are exhibiting.

This would result in a future course for an SLE patient somewhat similar to this: Patients with suspected SLE visit the physician to undergo a physical examination and do a single blood test, which is used to identify the immunological phenotype of the patient. If they get an SLE diagnosis, they get a subclassification based on the immunological phenotype and following treatment is based on this phenotype, rather than the symptoms the patient exhibits. During follow-up, clinical data is collected in a (semi-)structured format by the physician, so that the data recorded during patient visits is re-usable for research purposes. A NLP algorithm regularly screens the patient records to identify patients that might be showing clinical signs of an upcoming flare, or conversely might suggest tapering medication to the physician for patients with a low probability of an upcoming flare. Additionally, by screening data from the EHR, the NLP algorithm might also identify new patients with suspected SLE in an early stage, who are already exhibiting symptoms and have visited a physician, but have not yet received a SLE diagnosis.

REFERENCES

1. Morgan C, Bland AR, Maker C, Dunnage J, Bruce IN. Individuals living with lupus: findings from the LUPUS UK Members Survey 2014. *Lupus*. 2018;27(4):681-687. doi:10.1177/0961203317749746
2. Kernder A, Richter JG, Fischer-Betz R, et al. Delayed diagnosis adversely affects outcome in systemic lupus erythematosus: Cross sectional analysis of the LuLa cohort. *Lupus*. 2021;30(3):431-438. doi:10.1177/0961203320983445
3. Tsang-A-Sjoe MWP, Bultink IEM, Heslinga M, Voskuyl AE. Both prolonged remission and Lupus Low Disease Activity State are associated with reduced damage accrual in systemic lupus erythematosus. *Rheumatology (Oxford)*. 2017;56(1):121-128. doi:10.1093/rheumatology/kew377
4. Tian J, Zhang D, Yao X, Huang Y, Lu Q. Global epidemiology of systemic lupus erythematosus: a comprehensive systematic analysis and modelling study. *Ann Rheum Dis*. 2023;82(3):351-356. doi:10.1136/ard-2022-223035
5. Fatoye F, Gebrye T, Mbada C. Global and regional prevalence and incidence of systemic lupus erythematosus in low-and-middle income countries: a systematic review and meta-analysis. *Rheumatol Int*. 2022;42(12):2097-2107. doi:10.1007/s00296-022-05183-4
6. Barber MRW, Drenkard C, Falasinnu T, et al. Global epidemiology of systemic lupus erythematosus. *Nat Rev Rheumatol*. 2021;17(9):515-532. doi:10.1038/s41584-021-00668-1
7. de Boer AR, de Groot MCH, Groenhof TKJ, et al. Data mining to retrieve smoking status from electronic health records in general practice. *Eur Heart J Digit Health*. 2022;3(3):437-444. doi:10.1093/ehjdh/ztac031
8. van Dijk WB, Fiolet ATL, Schuit E, et al. Text-mining in electronic healthcare records can be used as efficient tool for screening and data collection in cardiovascular trials: a multicenter validation study. *Journal of Clinical Epidemiology*. 2021;132:97-105. doi:10.1016/j.jclinepi.2020.11.014
9. Hossain E, Rana R, Higgins N, et al. Natural Language Processing in Electronic Health Records in relation to healthcare decision-making: A systematic review. *Comput Biol Med*. 2023;155:106649. doi:10.1016/j.compbiomed.2023.106649
10. Yang X, Chen A, PourNejatian N, et al. A large language model for electronic health records. *NPJ Digit Med*. 2022;5(1):1-9. doi:10.1038/s41746-022-00742-2
11. Upton R, Mumith A, Beqiri A, et al. Automated Echocardiographic Detection of Severe Coronary Artery Disease Using Artificial Intelligence. *JACC Cardiovasc Imaging*. 2022;15(5):715-727. doi:10.1016/j.jcmg.2021.10.013
12. Sayres R, Taly A, Rahimy E, et al. Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy. *Ophthalmology*. 2019;126(4):552-564. doi:10.1016/j.ophtha.2018.11.016
13. van der Leeuw MS, Messelink MA, Tekstra J, et al. Using real-world data to dynamically predict flares during tapering of biological DMARDs in rheumatoid arthritis: development, validation, and potential impact of prediction-aided decisions. *Arthritis Res Ther*. 2022;24(1):1-11. doi:10.1186/s13075-022-02751-8
14. Cornet A, Andersen J, Myllys K, Edwards A, Arnaud L. Living with systemic lupus erythematosus in 2020: a European patient survey. *Lupus Sci Med*. 2021;8(1):e000469. doi:10.1136/lupus-2020-000469
15. Fanouriakis A, Kostopoulou M, Andersen J, et al. EULAR recommendations for the management of systemic lupus erythematosus: 2023 update. *Ann Rheum Dis*. Published online October 12, 2023. doi:10.1136/ard-2023-224762

16. Holman HR, Kunkel HG. Affinity between the Lupus Erythematosus Serum Factor and Cell Nuclei and Nucleoprotein. *Science*. 1957;126(3265):162-163. doi:10.1126/science.126.3265.162
17. Yaniv G, Twig G, Shor DBA, et al. A volcanic explosion of autoantibodies in systemic lupus erythematosus: A diversity of 180 different antibodies found in SLE patients. *Autoimmun Rev*. 2015;14(1):75-79. doi:10.1016/j.autrev.2014.10.003
18. Arbuckle MR, McClain MT, Rubertone MV, et al. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N Engl J Med*. 2003;349(16):1526-1533. doi:10.1056/NEJMoa021933
19. Tan EM, Feltkamp TEW, Smolen JS, et al. Range of antinuclear antibodies in "healthy" individuals. *Arthritis Rheum*. 1997;40(9):1601-1611. doi:10.1002/art.1780400909
20. Krzemień P, Kasperczyk S, Banach M, et al. Analysis of the impact of sex and age on the variation in the prevalence of antinuclear autoantibodies in Polish population: a nationwide observational, cross-sectional study. *Rheumatol Int*. 2022;42(2):261-271. doi:10.1007/s00296-021-05033-9
21. Yeo AL, Le S, Ong J, et al. Utility of repeated antinuclear antibody tests: a retrospective database study. *Lancet Rheumatol*. 2020;2(7):e412-e417. doi:10.1016/S2665-9913(20)30084-9
22. Fattal I, Shental N, Mevorach D, et al. An antibody profile of systemic lupus erythematosus detected by antigen microarray. *Immunology*. 2010;130(3):337. doi:10.1111/j.1365-2567.2010.03245.x
23. Wandstrat A, Carrjohnson F, Branch V, et al. Autoantibody profiling to identify individuals at risk for systemic lupus erythematosus. *J Autoimmun*. 2006;27(3):153-160. doi:10.1016/j.jaut.2006.09.001
24. Klecka M, Thybo C, Macaubas C, et al. Autoantibody Profiling in Lupus Patients using Synthetic Nucleic Acids. *Sci Rep*. 2018;8(1):5554. doi:10.1038/s41598-018-23910-5
25. Kim JM, Park SH, Kim HY, Kwok SK. A Plasmacytoid Dendritic Cells-Type I Interferon Axis Is Critically Implicated in the Pathogenesis of Systemic Lupus Erythematosus. *Int J Mol Sci*. 2015;16(6):14158-14170. doi:10.3390/ijms160614158
26. Sakata K, Nakayamada S, Miyazaki Y, et al. Up-Regulation of TLR7-Mediated IFN-α Production by Plasmacytoid Dendritic Cells in Patients With Systemic Lupus Erythematosus. *Front Immunol*. 2018;9. Accessed January 7, 2024. https://www.frontiersin.org/articles/10.3389/fimmu.2018.01957
27. Kwok SK, Lee JY, Park SH, et al. Dysfunctional interferon-α production by peripheral plasmacytoid dendritic cells upon Toll-like receptor-9 stimulation in patients with systemic lupus erythematosus. *Arthritis Res Ther*. 2008;10(2):R29. doi:10.1186/ar2382
28. Mortezagholi S, Babaloo Z, Rahimzadeh P, et al. Evaluation of TLR9 expression on PBMCs and CpG ODN-TLR9 ligation on IFN-α production in SLE patients. *Immunopharmacol Immunotoxicol*. 2017;39(1):11-18. doi:10.1080/08923973.2016.1263859
29. Green NM, Moody KS, Debatis M, Marshak-Rothstein A. Activation of Autoreactive B Cells by Endogenous TLR7 and TLR3 RNA Ligands. *J Biol Chem*. 2012;287(47):39789-39799. doi:10.1074/jbc.M112.383000
30. Fillatreau S, Manfroi B, Dörner T. Toll-like receptor signalling in B cells during systemic lupus erythematosus. *Nat Rev Rheumatol*. 2021;17(2):98-108. doi:10.1038/s41584-020-00544-4
31. Tilstra JS, John S, Gordon RA, et al. B cell-intrinsic TLR9 expression is protective in murine lupus. *J Clin Invest*. 2020;130(6):3172-3187. doi:10.1172/JCI132328
32. Gies V, Schickel JN, Jung S, et al. Impaired TLR9 responses in B cells from patients with systemic lupus erythematosus. *JCI Insight*. 2018;3(5). doi:10.1172/jci.insight.96795

33. Çakan E, Ah Kioon MD, Garcia-Carmona Y, et al. TLR9 ligand sequestration by chemokine CXCL4 negatively affects central B cell tolerance. *J Exp Med*. 2023;220(12). doi:10.1084/jem.20230944
34. Desnues B, Macedo AB, Roussel-Queval A, et al. TLR8 on dendritic cells and TLR9 on B cells restrain TLR7-mediated spontaneous autoimmunity in C57BL/6 mice. *Proc Natl Acad Sci U S A*. 2014;111(4):1497-1502. doi:10.1073/pnas.1314121111
35. Nündel K, Green NM, Shaffer AL, et al. Cell-intrinsic expression of TLR9 in autoreactive B cells constrains BCR/TLR7-dependent responses. *J Immunol*. 2015;194(6):2504-2512. doi:10.4049/jimmunol.1402425
36. Sacre K, Criswell LA, McCune JM. Hydroxychloroquine is associated with impaired interferon-alpha and tumor necrosis factor-alpha production by plasmacytoid dendritic cells in systemic lupus erythematosus. *Arthritis Res Ther* 2012;14(3):R155. doi:10.1186/ar3895
37. Langreth BR, Waldholz M. New Era of Personalized Medicine. *Oncologist*. 1999;4(5):426-427. doi:10.1634/theoncologist.4-5-426
38. Slamon DJ, Leyland-Jones B, Shak S, et al. Use of Chemotherapy plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer That Overexpresses HER2. *N Engl J Med*. 2001;344(11):783-792. doi:10.1056/NEJM200103153441101
39. Jørgensen JT. Twenty Years with Personalized Medicine: Past, Present, and Future of Individualized Pharmacotherapy. *Oncologist*. 2019;24(7):e432-e440. doi:10.1634/theoncologist.2019-0054
40. Waarts MR, Stonestrom AJ, Park YC, Levine RL. Targeting mutations in cancer. *J Clin Invest*. 132(8):e154943. doi:10.1172/JCI154943
41. Gianni L, Dafni U, Gelber RD, et al. Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. *Lancet Oncol*. 2011;12(3):236-244. doi:10.1016/S1470-2045(11)70033-X
42. Wolff AC, Somerfield MR, Dowsett M, et al. Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology–College of American Pathologists Guideline Update. *Archf Pathol Lab Med*. 2023;147(9):993-1000. doi:10.5858/arpa.2023-0950-SA
43. Fatemi A, Keivani-Boroujeni E, Smiley A. Predictors of severe lupus flare: a prospective follow-up study. *BMC Rheumatol*. 2023;7(1):10. doi:10.1186/s41927-023-00333-y
44. Gensous N, Marti A, Barnetche T, et al. Predictive biological markers of systemic lupus erythematosus flares: a systematic literature review. *Arthritis Res Ther*. 2017;19(1):238. doi:10.1186/s13075-017-1442-6
45. Andrade S de O, Julio PR, Nunes de Paula Ferreira D, Appenzeller S. Predicting lupus flares: epidemiological and disease related risk factors. *Expert Rev Clin Immunol*. 2021;17(2):143-153. doi:10.1080/1744666X.2020.1865156
46. Jorge AM, Smith D, Wu Z, et al. Exploration of Machine Learning Methods to Predict Systemic Lupus Erythematosus Hospitalizations. *Lupus*. 2022;31(11):1296-1305. doi:10.1177/09612033221114805
47. van Vollenhoven RF, Bertsias G, Doria A, et al. 2021 DORIS definition of remission in SLE: final recommendations from an international task force. *Lupus Sci Med*. 2021;8(1):e000538. doi:10.1136/lupus-2021-000538
48. Franklyn K, Lau CS, Navarra SV, et al. Definition and initial validation of a Lupus Low Disease Activity State (LLDAS). *Ann Rheum Dis*. 2016;75(9):1615-1621. doi:10.1136/annrheumdis-2015-207726
49. Vital EM, Merrill JT, Morand EF, et al. Anifrolumab efficacy and safety by type I interferon gene signature and clinical subgroups in patients with SLE: post hoc analysis of pooled

data from two phase III trials. *Ann Rheum Dis*. 2022;81(7):951-961. doi:10.1136/annrheumdis-2021-221425

50. Stohl W, Hilbert DM. The discovery and development of belimumab: the anti-BLyS–lupus connection. *Nat Biotechnol*. 2012;30(1):69-77. doi:10.1038/nbt.2076
51. Merrill JT, Tanaka Y, D'cruz D, et al. Op0139 Efficacy and Safety of Abbv-599 High Dose (elsubrutinib 60 Mg and Upadacitinib 30 Mg) and Upadacitinib Monotherapy for the Treatment of Systemic Lupus Erythematosus: A Phase 2, Double-Blind, Placebo-Controlled Trial. *Ann Rheum Dis*. 2023;82(Suppl 1):91-92. doi:10.1136/annrheumdis-2023-eular.2402
52. Fanton C, Furie R, Chindalore V, et al. Selective expansion of regulatory T cells by NKTR-358 in healthy volunteers and patients with systemic lupus erythematosus. *J Transl Autoimmun*. 2022;5:100152. doi:10.1016/j.jtauto.2022.100152
53. Higgs BW, Liu Z, White B, et al. Patients with systemic lupus erythematosus, myositis, rheumatoid arthritis and scleroderma share activation of a common type I interferon pathway. *Ann Rheum Dis*. 2011;70(11):2029-2036. doi:10.1136/ard.2011.150326
54. Furie RA, Morand EF, Bruce IN, et al. Type I interferon inhibitor anifrolumab in active systemic lupus erythematosus (TULIP-1): a randomised, controlled, phase 3 trial. *Lancet Rheumatol*. 2019;1(4):e208-e219. doi:10.1016/S2665-9913(19)30076-1
55. Morand EF, Furie R, Tanaka Y, et al. Trial of Anifrolumab in Active Systemic Lupus Erythematosus. *N Engl J Med*. 2020;382(3):211-221. doi:10.1056/NEJMoa1912196
56. Dima A, Jurcut C, Baicus C. The impact of anti-U1-RNP positivity: systemic lupus erythematosus versus mixed connective tissue disease. *Rheumatol Int*. 2018;38(7):1169-1178. doi:10.1007/s00296-018-4059-4
57. Stojan G, Petri M. Anti-C1q in systemic lupus erythematosus. *Lupus*. 2016;25(8):873-877. doi:10.1177/0961203316645205
58. Poulsen TBG, Karamehmedovic A, Aboo C, et al. Protein array-based companion diagnostics in precision medicine. *Expert Rev Mol Diagn*. 2020;20(12):1183-1198. doi:10.1080/14737159.2020.1857734
59. Rai R, Chauhan SK, Singh VV, Rai M, Rai G. RNA-seq Analysis Reveals Unique Transcriptome Signatures in Systemic Lupus Erythematosus Patients with Distinct Autoantibody Specificities. Crispin JC, ed. *PLoS ONE*. 2016;11(11):e0166312. doi:10.1371/journal.pone.0166312
60. Lewis MJ, McAndrew MB, Wheeler C, et al. Autoantibodies targeting TLR and SMAD pathways define new subgroups in systemic lupus erythematosus. *J Autoimmun*. 2018;91:1-12. doi:10.1016/j.jaut.2018.02.009
61. Luo H, Wang L, Bao D, et al. Novel Autoantibodies Related to Cell Death and DNA Repair Pathways in Systemic Lupus Erythematosus. *Genomics, Proteomics & Bioinformatics*. 2019;17(3):248-259. doi:10.1016/j.gpb.2018.11.004
62. Kegerreis B, Catalina MD, Bachali P, et al. Machine learning approaches to predict lupus disease activity from gene expression data. *Sci Rep*. 2019;9(1):1-12. doi:10.1038/s41598-019-45989-0

NEDERLANDSE SAMENVATTING

Achtergrond

Systemische Lupus Erythematosus (SLE) is een zeldzame systemische auto-immuunziekte met een prevalentie van 43.7 per 100.000 personen. De klachten die erdoor kunnen ontstaan variëren van relatief milde symptomen, zoals spier- en gewrichtsklachten of een karakteristieke huiduitslag, tot invaliderende of zelfs levensbedreigende symptomen, zoals psychoses en ontstekingen van de inwendige organen. Door de grote variatie aan symptomen die door SLE veroorzaakt kunnen worden, is het stellen van de diagnose vaak lastig. Veel van de symptomen die door SLE kunnen worden veroorzaakt, zijn niet specifiek voor SLE en hebben ook vele andere, vaker voorkomende mogelijke oorzaken. Dit leidt vaak tot een diagnostisch delay, waarbij de periode tussen het ontstaan van de eerste klachten en het stellen van de diagnose regelmatig maanden tot enkele jaren duurt. In deze periode kan de onbehandelde ziekte orgaanschade veroorzaken en hiernaast patiënten een zware ziektelast bezorgen. Hierom is het belangrijk om het diagnostische proces voor SLE te verbeteren en dit diagnostische delay te verkleinen.

Deel 1

Omdat SLE een weinig voorkomend ziektebeeld is, is het lastig om goede data te verzamelen om onderzoek mee te verrichten. Het opzetten van gestructureerde trials om deze data te verzamelen kost veel tijd en geld. Hierom wordt er ook gekeken naar andere bronnen van data die voor onderzoeksdoeleinden te verzamelen, zoals het Elektronisch Patiënten Dossier (EPD). Een van de belangrijkste functies van het EPD is het vastleggen van gegevens die voor de begeleiding van het ziekteproces van de patiënten belangrijk zijn, waaronder symptomen en diagnoses. Door deze functie bevat het EPD een grote hoeveelheid data die potentieel zeer relevant is voor het beantwoorden van onderzoeksvragen. Een grote uitdaging bij het verzamelen van deze data voor onderzoeksdoeleinden is dat het EPD veel vrije tekst of ongestructureerde data bevat, wat lastig te verwerken is.

In de meeste gevallen wordt relevante data handmatig uit het EPD verzameld door het uitvoeren van dossieronderzoek, wat tijdrovend is en een aanzienlijke inter- en intraobserver variatie heeft. In **Hoofdstuk 2** beschrijven we hoe wij een algoritme hebben gemaakt om relevante informatie uit data uit het EPD te kunnen filteren. Voor ons onderzoek ging dit met name om het al dan niet hebben van een diagnose van SLE. Met ons algoritme tonen we aan dat de ongestructureerde data uit het EPD gebruikt kan worden om op een geautomatiseerde, en hiermee repliceerbare, manier relevante informatie te verkrijgen, welke voor onderzoeksdoeleinden hergebruikt kan worden. Dit doet ons algoritme door te zoeken naar het voorkomen van specifieke kernwoorden en hierbij de context van deze kernwoorden te evalueren, om bijvoorbeeld negaties te herkennen. Dergelijke hulpmiddelen kunnen helpen grootschalig dossieronderzoek te faciliteren.

Deel 2

De diagnose van SLE is een klinische diagnose die gesteld wordt op basis van een combinatie van klinische en immunologische kenmerken. Een van de belangrijkste immunologische kenmerken van SLE is de aanwezigheid van autoantistoffen. Over de jaren zijn er meer dan 100 verschillende autoantistoffen ontdekt die geassocieerd worden met SLE. Veel van deze autoantistoffen zijn gericht tegen antigenen uit de celkern en vallen hierdoor in de groep van de antinuclear antibodies (ANAs). Deze autoantistoffen spelen een belangrijke rol in de pathofysiologie van SLE. Van een aantal autoantistoffen is bekend dat ze zelfs jaren voor de diagnose van SLE voor kunnen komen in het bloed van toekomstige patiënten, wat potentie biedt voor het gebruik van autoantistoffen als biomarkers voor vroege diagnostiek. Veel van deze antistoffen komen echter ook voor bij een klein deel van de gezonde populatie, wat de potentie voor een screeningstest in de algemene populatie beperkt. De aanwezigheid van deze autoantistoffen in gezonde mensen zou gezien kunnen worden als een vorm van benigne auto-immuniteit, zonder dat dit tot het ontstaan van een auto-immuunziekte leidt.

In het ziekenhuis komt het regelmatig voor dat patiënten een positieve uitslag hebben voor een ANA-test, zonder dat deze patiënten een diagnose van een auto-immuunziekte krijgen. In **Hoofdstuk 5** hebben we onderzocht of deze positieve testen gezien moeten worden als een teken van een mogelijke toekomstige auto-immuunziekte of als klinisch irrelevant. We tonen aan dat slechts een klein deel van de patiënten met klachten die gerelateerd kunnen worden aan een auto-immuunziekte uiteindelijk daadwerkelijk een dergelijke diagnose krijgen. Een positieve ANA-test was op zichzelf een slechte voorspeller voor het later krijgen van een diagnose van een auto-immuunziekte. Het hebben van zowel symptomen als de aanwezigheid van specifieke ANAs gerelateerd aan de verdachte auto-immuunziekte, was wel een sterkere voorspeller van progressie naar ziekte, alhoewel dit nog steeds weinig voor komt.

Hoewel de voorspellende waarde van individuele antistoffen ogenschijnlijk beperkt is, laten we in **Hoofdstuk 4** zien dat het maken van profielen van autoantistoffen aanvullende kansen biedt. In dit onderzoek hebben we een groot cohort met SLE-patiënten en meerdere relevante controlegroepen gebruikt om profielen van autoantistoffen te creëren. Vervolgens tonen we aan dat toekomstige SLE-patiënten vaak al voor de diagnose een profiel aan autoantistoffen vertonen dat gebruikt kan worden om patiënten met een hoog risico op het ontwikkelen van SLE in een vroege fase te identificeren.

In **Hoofdstuk 3** tonen we aan dat antistoffen gericht tegen CpG oligodeoxynucleotides (ODNs) een hoge prevalentie hebben in het bloed van patiënten met SLE, en hiermee ook een belangrijke bijdrage leveren aan het vormen van de autoantistof profielen. Eerdere onderzoeken hebben aangetoond dat CpG ODNs onder andere plasmacytoide dendritische cellen (pDCs) kunnen stimuleren via TLR9. pDCs zijn belangrijke producenten van

inflammatoire type I interferons (IFN), en zijn hiermee waarschijnlijk voor een groot deel verantwoordelijk voor de toename van de zogeheten IFN signature in patiënten met SLE. In **Hoofdstuk 6** hebben we de relatie tussen autoantistoffen en de IFN signature verder onderzocht. Hierbij tonen we aan dat antistoffen gericht tegen EphB2, RipP1, en RNP70 geassocieerd zijn met een hoge IFN-score.

In **Hoofdstuk 7** hebben we een overzicht gemaakt van recente ontwikkelingen op het gebied van diagnostiek naar SLE. Hierbij hebben we zowel naar autoantistoffen als naar andere biomarkers, zoals chemokines, cytokines, RNAs en genexpressie gekeken. Hoewel dit overzicht geen perfecte biomarker identificeerde, geeft het wel meer inzicht in de pathofysiologie van SLE. We beargumenteren dat de kwaliteit van diagnostische studies significant verbeterd zou kunnen worden door het opzetten van grote population-based cohorten. Deze cohorten kunnen voor veel ziektebeelden interessante inzichten opleveren, en specifiek voor SLE het mogelijk maken om onderzoek te doen naar de immunologische veranderingen in patiënten, voorafgaand aan het stellen van de diagnose.

Conclusie

In dit proefschrift laat ik zien hoe we diverse innovatieve manieren hebben gebruikt om zowel klinische als immunologische data te verzamelen. We laten zien dat technische hulpmiddelen ontworpen kunnen worden om aanvullende databronnen voor onderzoek aan te boren. We hebben de aanwezigheid van diverse autoantistoffen en hun relatie met andere immunologische verschijnselen uitgediept, en laten zien dat profielen van autoantistoffen gebruikt kunnen worden om toekomstige patiënten met SLE in een vroeg stadium te herkennen, met als doel de diagnostische delay hiermee te verkorten.

ENGLISH SUMMARY

Background

Systemic Lupus Erythematosus (SLE) is a rare systemic autoimmune disease with a prevalence of 43.7 per 100.000 people. Symptoms range from relatively mild complaints, such as muscle and joint pains or a characteristic skin rash, to disabling or potentially life-threatening symptoms, such as psychosis and inflammation of internal organs. Because SLE can cause a wide variety of symptoms, establishing the diagnosis is often difficult. Many symptoms caused by SLE are not specific to the disease and have many other, more common possible causes. This often leads to a diagnostic delay, where the time between the onset of first symptoms and the diagnosis regularly spans months to several years. During this period, untreated disease can cause organ damage and imposes a high disease burden on patients. It is therefore important to improve the diagnostic process for SLE and to reduce this diagnostic delay.

Part 1

Because SLE is an uncommon condition, collecting high-quality data for research is challenging. Setting up structured trials to gather such data is time consuming and expensive. Therefore, other data sources are also considered for research purposes, such as the Electronic Health Records (EHR). One of the EHR's main functions is to record information relevant to managing patients' disease, including symptoms and diagnoses. As a result, the EHR contains large amounts of data that may be highly relevant for answering research questions. A major challenge when using EHR data for research is that it contains a lot of free text or unstructured data, which is difficult to process.

In most cases, relevant data are manually extracted from the EHR by conducting chart reviews, which is time consuming and subject to substantial inter- and intraobserver variation. In **Chapter 2** we describe how we developed an algorithm to filter relevant information from EHR data. For our research this focused mainly on whether or not a diagnosis of SLE was present. With our algorithm we show that unstructured EHR data can be used to obtain relevant information in an automated and thus reproducible way that can be reused for research. The algorithm searches for occurrences of specific keywords and evaluates the context of those keywords to recognize, for example, negations. Tools like our algorithm can help facilitate large-scale chart review.

Part 2

The diagnosis of SLE is a clinical diagnosis made on the basis of a combination of clinical and immunological features. One of the most important immunological features of SLE is the presence of autoantibodies. Over the years more than 100 different autoantibodies associated with SLE have been discovered. Many of these autoantibodies target antigens from the cell nucleus and thus fall into the group of antinuclear antibodies (ANAs). These

autoantibodies play an important role in the pathophysiology of SLE. Some autoantibodies are known to appear in the blood of future patients years before the diagnosis of SLE, which offers potential for using autoantibodies as biomarkers for early diagnosis. Many of these antibodies, however, also occur in a small proportion of the healthy population, which limits the potential for a screening test in the general population. The presence of these autoantibodies in healthy individuals could be seen as a form of benign autoimmunity that does not lead to clinical autoimmune disease.

In clinical practice it is common for patients to have a positive ANA test result without ultimately receiving a diagnosis of an autoimmune disease. In **Chapter 5** we investigated whether these positive tests should be interpreted as a sign of a possible future autoimmune disease or as clinically irrelevant. We show that only a small proportion of patients with complaints that could be related to an autoimmune disease eventually receive such a diagnosis. A positive ANA test by itself was a poor predictor of later receiving a diagnosis of an autoimmune disease. Having both symptoms and the presence of specific ANAs related to the suspected autoimmune disease was a stronger predictor of progression to disease, although this still occurred infrequently.

Although the predictive value of individual autoantibodies appears limited, in **Chapter 4** we demonstrate that constructing autoantibody profiles offers additional opportunities. In this study we used a large cohort of SLE patients and multiple relevant control groups to create autoantibody profiles. We then show that future SLE patients often already exhibit an autoantibody profile before diagnosis that can be used to identify patients at high risk of developing SLE at an early stage.

In **Chapter 3** we show that antibodies directed against CpG oligodeoxynucleotides (ODNs) have a high prevalence in the blood of SLE patients and therefore make an important contribution to forming the autoantibody profiles. Previous studies have shown that CpG ODNs can stimulate plasmacytoid dendritic cells (pDCs) via TLR9. pDCs are important producers of inflammatory type I interferons (IFN) and are likely largely responsible for the increase of the so-called IFN signature in SLE patients. In **Chapter 6** we further investigated the relationship between autoantibodies and the IFN signature. We show that antibodies directed against EphB2, RipP1, and RNP70 are associated with a high IFN score.

In **Chapter 7** we provide an overview of recent developments in diagnostics for SLE. We considered both autoantibodies and other biomarkers, such as chemokines, cytokines, RNAs, and gene expression. Although this review did not identify a perfect biomarker, it does provide greater insight into the pathophysiology of SLE. We argue that the quality of diagnostic studies could be significantly improved by establishing large population-based

cohorts. Such cohorts can yield valuable insights for many diseases and, specifically for SLE, enable research into immunological changes in patients prior to diagnosis.

Conclusion

In this thesis I demonstrate how we used several innovative approaches to collect both clinical and immunological data. We show that technical tools can be designed to unlock additional data sources for research. We explored the presence of various autoantibodies and their relationship with other immunological phenomena, and we show that autoantibody profiles can be used to recognize future SLE patients at an early stage, with the aim of reducing diagnostic delay.

ACKNOWLEDGEMENTS (DANKWOORD)

Nadat mijn proefschrift geruime tijd "bijna af" was, is het dan nu eindelijk zo ver, mijn proefschrift is klaar. Ik kijk met veel plezier terug op een periode waarin ik vele uitdagingen aan ben gegaan, obstakels heb overwonnen en ontzettend veel heb geleerd. Ik wil graag een aantal mensen bedanken die een belangrijke rol hebben gespeeld in dit hele traject.

Prof. dr. Van Laar, beste Jaap. Ik denk dat wij meerdere voortgangsgesprekken hebben gehad waar jij eerst rustig mijn ideeën voor nieuwe projecten aanhoorde, en vervolgens vriendelijk de nadruk legde op de lopende projecten die ik misschien beter eerst nog maar kon afronden. Jij wist altijd de grote lijn weer in focus te krijgen als ik deze dreigde te verliezen. Ik heb veel gehad aan jouw adviezen en heb veel van je kunnen leren, bedankt hiervoor.

Dr. Limper, beste Maarten. Vanaf ons eerste gesprek in Madrid heb je mij weten te inspireren met jouw enthousiasme. Jij gaf mij een enorm gevoel van vertrouwen en was altijd laagdrempelig beschikbaar wanneer ik zat met vragen. Ik heb genoten van onze inhoudelijke discussies en hoop dat mijn vragen niet te vaak reden zijn geweest voor een sessie met de boksbal.

Dr. Otten, beste Henny. Tijdens mijn traject als promovendus stond jouw deur altijd open, iets waar ik dan ook graag gebruik van maakte. Ik heb veel bewondering voor jouw vermogen om de link tussen het lab en de kliniek te leggen en heb hier zelf dan ook ontzettend veel aan gehad. Ik kijk met veel plezier terug op onze samenwerking en hoop dat je geniet van je pensioen.

Beste Hein, zonder jou had dit proefschrift waarschijnlijk niet bestaan. Met jouw begeleiding zette ik tijdens mijn M2 in het ZGT de eerste stappen in het onderzoek, wat resulteerde in een presentatie op de EULAR in Madrid, waar de eerste contacten voor dit proefschrift werden gelegd. En ook met jouw voortrekkersrol bij ARCH, wat ook een rode draad in mijn tijd als promovendus is geweest, heb jij onmiskenbaar een belangrijke rol gespeeld in deze fase van mijn carrière.

Leon, ik heb genoten van onze samenwerking bij de data-analyse. Ik had nog nooit van een XGBoost- of Lasso-model gehoord voordat jij dit aan mij introduceerde, dus ik denk dat ik met recht kan zeggen dat jouw hulp onmisbaar is geweest voor het tot stand komen van dit proefschrift. Rowena, toen ik net begon had ik geen idee hoe alles eraan toe ging op het lab. Niet erg handig voor een project wat grotendeels draait om het analyseren van honderden bloedsamples… Heel erg bedankt voor het mij hierbij op sleeptouw nemen, en voor al het werk dat je op het lab voor deze projecten hebt verricht. Hiernaast was het met jouw aanstekelijke goede humeur altijd een zeer prettige samenwerking.

Saskia, jij en de rest van UPOD toonden mij de georganiseerde chaos die de achterkant van het EPD is. Bedankt voor het helpen met het ontwarren van deze chaos. Het heeft me niet alleen enorm geholpen voor het onderzoek, maar ook nog een hoop kennis geleerd over hoe EPD's in elkaar zitten, iets wat ook goed van pas komt bij mijn huidige werk in de kliniek. Suzanne, onze samenwerking was kort, maar krachtig. Ik heb veel plezier gehad met het samen uitzoeken hoe we uit de gigantische hoeveelheid data de juiste informatie konden filteren.

Collega's van de H3 gang; Ik weet nog goed dat op mijn eerste dag bleek dat er al veel voor mij was geregeld, maar een werkplek voor mij was men even vergeten. Maar ik hoefde gelukkig niet lang te zoeken, want ondanks dat ik tussen al de RA- en artroseonderzoekers een beetje een vreemde eend in de bijt was, werd ik gelijk opgenomen in de groep. Het koffiemomentje in de ochtend en discussies over Wie is de Mol en "Nieuwsuur" bij de lunch maakten mijn dag vaak een stuk leuker. Ook het samen delen van successen en tegenslagen die bij onderzoek horen, heeft enorm bijgedragen aan hoe ik terugkijk naar deze tijd. Simon, als aanvoerder van de H3 gang, bedankt dat je altijd klaar stond om met een grote lach alle vragen aan te horen als ik ergens op vast liep. Indien dit (ondanks een uitgebreide statistiekcursus) op statistiek- of methodologie-gebied was, kon ik altijd bij Paco terecht. Anne-Karien, ik wil niet weten met hoeveel versies van biobank- of studieprotocollen ik bij jou langs ben geweest, maar ik ben trots op de projecten die we daarmee op hebben kunnen zetten, zoals de PROFILE studie. En ik was onder de indruk van hoe soepel deze voorstellen vaak langs de METC of biobankcommissie kwamen nadat ze jouw stempel van goedkeuring hadden gekregen. Arno, de lab-man. Ik was iets minder fan van de mij toebedeelde taak van het 's ochtends vroeg sjouwen met de stikstofvaten, maar de grote lach en vrolijke praatjes als ik ze kwam afleveren maakten veel goed.

Ellen, ik denk dat wij samen, ondanks alle hindernissen van COVID, samen een heel mooi project hebben afgeleverd van de raakvlakken van onze onderzoeken, en met de PROFILE studie ook hebben gebouwd aan iets moois voor verder onderzoek. Ik ben blij dat we dit in goede handen achter hebben kunnen laten en ben heel benieuwd naar de verdere resultaten die Bettina ons uiteindelijk hierover zal gaan tonen. Voor het werk aan de PROFILE studie, en voor de samenwerking met de BLISS-Believe studie, wil ik natuurlijk ook de researchverpleegkundigen Karin, Anneloes, Joke en later ook Ria bedanken. Ik was en ben nog steeds onder de indruk van hoe jullie alles wisten te organiseren en denk dat veel onderzoek zonder jullie nooit zo tot hun recht waren gekomen. Daarnaast voelde ik me altijd ontzettend welkom bij jullie op de poli voor een kopje thee of Engels dropje, waarvoor dank.

Rolf, ondanks dat er uiteindelijk geen onderdeel over APS in mijn boekje is gekomen, heb ik ontzettend veel geleerd van jou tijdens de APS-overleggen, wat inmiddels ook in de kliniek nog af en toe van pas komt.

Hiernaast wil ik ook graag Linda en Jorge van Thermo Fisher bedanken, zonder wiens hulp meerdere projecten van dit proefschrift niet van de grond waren gekomen.

Naast mijn taken in het UMCU heeft ook ARCH een belangrijke rol gespeeld tijdens mijn tijd als promovendus. Alexandre, ik ben trost op het werk dat ik onder jouw begeleiding heb kunnen verrichten en heb veel geleerd van de bijbehorende discussies over hoe knelpunten aangepakt konden worden. Ik voelde veel vertrouwen om met de ARCH MDOs en het vernieuwen van het SLE-register aan de slag te gaan. Hiernaast was het ook erg leuk om in het Amsterdam UMC de kans te krijgen iets meer patiënten te zien in plaats van voornamelijk mijn eigen computerscherm, met een patiënt wiens lab-buis uit meer vet dan bloed bestond als bijzondere herinnering. Ook wil ik Irene bedanken, die altijd klaar stond voor vragen (zoals in het geval van hiervoor genoemde patiënt) en wiens toewijding aan het stimuleren van bewegen in patiënten met bindweefselziekten veel indruk op mij heeft gemaakt. De aandacht voor het mensen om te leren gaan met klachten van hun chronische ziekte is iets wat ik hoop ook in mijn eigen klinische werk terug te laten komen.

Helaas was voor mij de tijd in het UMCU niet lang genoeg om alles af te kunnen ronden en is het een significante uitdaging gebleken om het hierna, naast een klinische baan, af te ronden (getuige ook de tijd dat dit heeft geduurd). Wie had kunnen denken dat het combineren van een full-time baan als ANIOS, en later als AIOS, lastig te combineren was met het afronden van een proefschrift? Ik wil dan ook graag mijn oud-collega's van het Sint Antonius ziekenhuis, later het Martini ziekenhuis en momenteel het Treant Scheper ziekenhuis bedanken. Op elk van deze locaties heb ik mij ontzettend op mijn plek gevoeld, wat enorm heeft bijgedragen aan mijn verdere ontwikkeling als clinicus. Ik heb veel gehad aan jullie motiverende woorden om na een lange dag op de afdeling of de SEH toch achter mijn laptop te gaan zitten om stug door te werken en de open eindjes van dit proefschrift af te ronden. Ook heb ik veel gehad aan jullie woorden van steun toen het pad naar de opleiding tot internist niet zo soepel verliep, wat mij zeker heeft geholpen om door te zetten en te komen waar ik nu ben. Ik hoop velen van jullie nog veel vaker terug te zien in de toekomst.

Dan ook een woord van dank aan alle mensen buiten het ziekenhuis die me op de een of andere manier hebben geholpen om dit project tot een mooi einde te brengen.

De heren van Domstad 6/3/5, Jeroen, Stijn, Matts, Arjen, Duco, Edwin, en later ook Jasper. Sporten is voor mij altijd een uitlaatklep geweest, en ik had me hier geen mooiere groep omheen kunnen bedenken. De donderdagavonden en competitiedagen op zaterdag zijn een lange tijd niet weg te denken geweest uit mijn agenda. Ondanks dat dit nu helaas wat

lastiger is omdat ik zonodig naar Groningen moest verhuizen, kan het nog steeds ontzettend waarderen dat nog regelmatig er donderdagavond een belletje komt vanuit de kantine om mijn systeemplafond belachelijk te maken. Ik wil niet weten hoe lang ik jullie heb verteld dat mijn proefschrift "bijna af" was, maar hier dan toch echt: mijn werkstuk is af! Ik ben erg blij om jullie als vrienden te mogen rekenen, ondanks wat eigenaardigheden (zoals Stijn z'n pre-match foto's). Ik kijk uit naar het volgende pokeravondje of wintersportavontuur!

Dan ook mijn Jaarclub, JC Brût. Bedankt voor jullie vriendschap, steun, medeleven, motivatie, etc. etc.. Toen ik als jonge student aan kwam in Groningen had ik niet durven hopen dat ik hier zo'n groep aan vrienden over zou houden. Bedankt voor de vele motiverende woorden om mij tijdens dit lange traject te ondersteunen, en vooral bedankt voor jullie vriendschap en alle dingen die we samen hebben doorgemaakt en nog zullen gaan meemaken. Beste Mark, mijn eeuwige buurman: Ik denk dat het eindelijk tijd is om dat restaurant in Modena maar te gaan boeken. Beste Tom, ik ben ontzettend blij dat jij tijdens mijn periode als promovendus een tijd mijn huisgenoot bent geweest, en dat er in die tijd geen verdere ABBA-incidenten hebben voorgedaan. Ik hoop dat je het ontzettend naar je zin hebt in de Filipijnen, maar hoop ook dat je toch nog weer terugkomt, want Manilla is wel erg ver weg…

Beste Bram, ik ben enorm onder de indruk van hoe jij alle ballen van de SEH, onderzoek, privé en bijzondere sportprestaties in de lucht weet te houden. Jij weet precies wat je moet zeggen om mij te motiveren, wat me enorm heeft geholpen bij dit proefschrift. Ik ben blij dat je mijn paranimf wil zijn en trots om jou een vriend te noemen.

Beste Karmend, bedankt dat je mijn paranimf wil zijn. Ik weet dat je het vooral doet om een excuus te hebben om weer een keer een rokkostuum aan te kunnen doen, maar ik denk graag dat ons gedeelde enthousiasme voor een geslaagde pubquiz, een goed bord pasta (complimenten aan de chef), een mooi hakje, scherpe politieke comedy en vreemde YouTube filmpjes daar ook een rol in hebben gespeeld. Ik ben blij een goede vriend te hebben met wie ik zo veel kan delen.

Lieve Ronne, en natuurlijk ook Michiel en Jinna, bedankt voor alle lieve berichten en steun die ik de afgelopen jaren van jullie heb ontvangen.

Beste Ellie en Jacob, lieve mama en papa. Ten slotte wil ik jullie bedanken voor jullie onvoorwaardelijke steun. Voor een luisterend oor en een kritische vraag als ik mijn gedachten wil delen. Voor de bemoedigende woorden als ik weer met nieuwe ideeen aan kom, en voor de spreekwoordelijke schop onder de kont wanneer ik deze nodig heb. Voor het ondersteunen bij de tegenslagen en het met mij vieren van de successen. Bedankt voor alles. Ik hou van jullie en ben trots op jullie.

ABOUT THE AUTHOR



Tammo Ernst Brunekreef was born on the 25th of July 1993 in the city of Utrecht. He graduated cum laude from the Christelijk Gymnasium Utrecht in 2011. He a bachelor and master in medicine at the University of Groningen in 2011 and 2014 respectively. During his studies, he was involved in various student activities of M.F.V. Panacea, A.S.V. Dizkartes and Lijst Calimero. During the second year of the master's studies, he was an intern at ZGT Almelo, where, under the supervision of Dr. Bernelot Moens, he started a project researching clinical outcomes of patients with Rheumatoid Arthritis, which resulted in an oral presentation at the EULAR congress in 2017. After graduating in 2018, he started his PhD at the Department of Rheumatology & Clinical Immunology and the Centre for Translational Immunology at the UMC Utrecht under the supervision of prof. Jaap van Laar, dr. Maarten Limper and dr. Henny Otten. During his PhD, Tammo performed research into the use of data from the Electronic Health Records for research purposes and various projects aimed at improving the diagnostic process for patients with systemic lupus erythematosus (SLE). During his time as a PhD candidate, he also worked on the ARCH committees for antiphospholipid syndrome (APS) and SLE. In 2022 Tammo left the UMC Utrecht as a full time PhD student and finalized this thesis besides a job as a resident at the department of internal medicine at Sint Antonius Ziekenhuis and Martini Ziekenhuis. In 2025 he started his specialty training to become a specialist of internal medicine at the Treant Scheper Ziekenhuis in Emmen.

