Serum Metabolomic Signatures Can Predict Subclinical Atherosclerosis in Patients With Systemic Lupus Erythematosus

Leda Coelewij, Kirsty E. Waddington,* George A. Robinson,† Elvira Chocano, Thomas McDonnell, Filipa Farinha, Junjie Peng, Pierre Donnes, Edward Smith, Sara Croca, Jyoti Bakshi, Maura Griffin, Andrew Nicolaides, Anisur Rahman, Elizabeth C. Jury,† Ines Pineda-Torra†

OBJECTIVE: Patients with systemic lupus erythematosus (SLE) have an increased risk of developing cardiovascular disease. Standard serum lipid measurements in clinical practice do not predict cardiovascular disease risk in patients with SLE. More detailed analysis of lipoprotein taxonomy could identify better predictors of cardiovascular disease risk in SLE.

APPROACH AND RESULTS: Eighty women with SLE and no history of cardiovascular disease underwent carotid and femoral ultrasound scans; 30 had atherosclerosis plaques (patients with SLE with subclinical plaque) and 50 had no plaques (patients with SLE with no subclinical plaque). Serum samples obtained at the time of the scan were analyzed using a lipoprotein-focused metabolomics platform assessing 228 metabolites by nuclear magnetic resonance spectroscopy. Data were analyzed using logistic regression and 5 binary classification models with 10-fold cross validation. Patients with SLE had global changes in complex lipoprotein profiles compared with healthy controls despite having clinical serum lipid levels within normal ranges. In the SLE cohort, univariate logistic regression identified 4 metabolites associated with subclinical plaque; 3 subclasses of VLDL (very low-density lipoprotein; free cholesterol in medium and large VLDL particles and phospholipids in chylomicrons and extremely large VLDL particles) and leucine. Together with age, these metabolites were also within the top features identified by the lasso logistic regression (with and without interactions) and random forest machine learning models. Logistic regression with interactions differentiated between patients with SLE with subclinical plaque and patients with SLE with no subclinical plaque groups with the greatest accuracy (0.800). Notably, free cholesterol in large VLDL particles and age differentiated between patients with SLE with subclinical plaque and patients with SLE with no subclinical plaque in all models.

CONCLUSIONS: Serum metabolites are promising biomarkers to uncover and predict multimetabolic phenotypes of subclinical atherosclerosis in SLE.

Key Words: atherosclerosis • lipoprotein • lupus erythematosus, systemic • machine learning • metabolomics

Systemic lupus erythematosus (SLE) is a multisystem autoimmune condition that predominantly affects women (women to men ratio of 9:1) with a prevalence of ≈1 in 1000 in the United Kingdom.¹ Patients with SLE have a 5- to 10-fold increased risk of developing cardiovascular disease (CVD) compared with healthy people of the same age and sex.² Strikingly, the presence of SLE in women between the ages of 35 and 44 increases the risk of coronary artery disease by 50 times.² Furthermore, in a large multinational study of...
Nonstandard Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apo</td>
<td>apolipoprotein</td>
</tr>
<tr>
<td>CE</td>
<td>cholesterol ester</td>
</tr>
<tr>
<td>CVD</td>
<td>cardiovascular disease</td>
</tr>
<tr>
<td>FC</td>
<td>free cholesterol</td>
</tr>
<tr>
<td>GSM</td>
<td>Gray Scale Median</td>
</tr>
<tr>
<td>HC</td>
<td>healthy volunteer control</td>
</tr>
<tr>
<td>HDL</td>
<td>high-density lipoprotein</td>
</tr>
<tr>
<td>IDL</td>
<td>intermediate density lipoprotein</td>
</tr>
<tr>
<td>IMT</td>
<td>intima-media thickness</td>
</tr>
<tr>
<td>LDL</td>
<td>low-density lipoprotein</td>
</tr>
<tr>
<td>LR</td>
<td>logistic regression</td>
</tr>
<tr>
<td>LR+I</td>
<td>logistic regression with interactions</td>
</tr>
<tr>
<td>M</td>
<td>medium</td>
</tr>
<tr>
<td>ML</td>
<td>machine learning</td>
</tr>
<tr>
<td>mTOR</td>
<td>mammalian target of rapamycin</td>
</tr>
<tr>
<td>NF-KB</td>
<td>nuclear factor-kappa B</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>PCSK9</td>
<td>protein convertase subtilisin/kexin type 9</td>
</tr>
<tr>
<td>RF</td>
<td>random forest</td>
</tr>
<tr>
<td>SLE</td>
<td>systemic lupus erythematosus</td>
</tr>
<tr>
<td>SLE-NP</td>
<td>patients with SLE with no subclinical plaque</td>
</tr>
<tr>
<td>SLE-P</td>
<td>patients with SLE with subclinical plaque</td>
</tr>
<tr>
<td>VLDL</td>
<td>very low density lipoprotein</td>
</tr>
<tr>
<td>XL</td>
<td>very large</td>
</tr>
<tr>
<td>XS</td>
<td>very small</td>
</tr>
<tr>
<td>XXL</td>
<td>extremely large</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Highlights</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Lipoprotein-based metabolomics identified a significantly disrupted lipoprotein subclass taxonomy in women with systemic lupus erythematosus compared with health volunteers, characterized by increased total lipid, cholesterol, and cholesterol ester content in various VLDL (very low-density lipoprotein) subsets, and reduced HDL (high-density lipoprotein) subsets.</td>
</tr>
<tr>
<td>• Patients with systemic lupus erythematosus with subclinical atherosclerosis plaques had a unique metabolomic profile associated with increased circulating VLDL subsets, leucine and tyrosine and reduced glycine.</td>
</tr>
<tr>
<td>• Interactions between metabolites and patient demographic and treatment features were also important in discriminating between patients with systemic lupus erythematosus with and without subclinical plaque.</td>
</tr>
<tr>
<td>• This study suggests that a composite score of detailed metabolomics with conventional risk factors may be a better predictor of cardiovascular disease risk in patients with systemic lupus erythematosus.</td>
</tr>
</tbody>
</table>

9,547 patients with SLE, a quarter of deaths were attributed to CVD. The precise mechanism of this increased CVD risk is yet to be fully elucidated. While traditional risk factors such as high blood pressure, diabetes, and high cholesterol contribute to the increased risk, they fail to account for it fully. The risk is likely to be multifactorial, resulting from a complex interplay of SLE-driven immunologic dysfunction and traditional CVD risk factors.

Abnormalities in lipid profiles are a traditional risk factor for CVD and can also be affected by chronic inflammatory conditions such as SLE. Lipids are central to driving atherosclerosis, the main pathology underlying CVD. Various fractions of lipoproteins can be distinguished in blood on account of their size and density: HDL (high-density lipoprotein), LDL (low-density lipoprotein), and VLDL (very low-density lipoprotein). Dyslipidemias are present in over 70% of cases of premature coronary heart disease and elevated plasma concentrations of LDL and VLDL can induce the development of atherosclerosis in the absence of other risk factors. In contrast, HDL has antiatherogenic properties that include macrophage cholesterol efflux, anti-oxidation, and protection against thrombosis. Conversely, McMahon et al have demonstrated the existence of a subpopulation of proinflammatory HDL in patients with SLE and rheumatoid arthritis that promotes atherosclerosis and could be a biomarker for increased risk of developing CVD.

Hypercholesterolemia (defined as elevated plasma total cholesterol and/or LDL cholesterol or non-HDL-cholesterol) is found in 34% to 51% of patients with SLE and is characterized by elevated levels of VLDL and triglycerides and low HDL levels. In addition, development of CVD in women with SLE has been found to be associated with smaller subfractions of LDL. Studies in non-SLE patients with CVD suggest that the ratio between serum lipid-associated proteins, ApoB:ApoA1 (apolipoprotein-B:apolipoprotein-A1), is a more effective CVD predictor than routine cholesterol measurements. A higher ApoB:ApoA1 ratio is associated with increased cardiovascular risk, however, the role of this ratio in the prediction of SLE CVD is still being assessed. Overall, dyslipidemia detected in routine lipid screens available in clinical practice fails to fully account for the increased risk of CVD in patients with SLE. Many patients with SLE with normal serum lipid levels on standard assays also go on to have CVD. Therefore, more sensitive and specific lipid profiles need to be delineated to identify high CVD risk patients in SLE cohorts.

Here, we used a nuclear magnetic resonance (NMR) Spectroscopy metabolomics platform and machine learning (ML) analyses to assess the association of lipoprotein subclasses and lipid content and other low
molecular weight serum metabolites with the presence of subclinical atherosclerotic plaque in patients with SLE.

MATERIALS AND METHODS
All data have been made publicly available in Mendeley and can be accessed at http://dx.doi.org/10.17632/fmygdybj2h.1.

Patient Cohort
Serum samples were collected from 80 nonfasting patients with SLE attending a rheumatology clinic at University College London Hospital and fulfilling the American College of Rheumatology classification criteria for lupus (1997).18 Patients had no previous history of CVD (defined as coronary artery disease, stroke, or myocardial infarction with confirmatory evidence from blood tests and/or imaging), and all underwent a vascular ultrasound scan between 2011 and 2013.19 Serum samples were also donated by 39 healthy female volunteers, which were used as controls. Demographic and clinical information for patients were recorded at the time of scan/blood sampling, including sex, age, ethnicity, blood pressure (mean arterial blood pressure: calculated as 2× diastolic pressure+1× systolic pressure divided by 3), routine serology measures, treatment (including hydroxychloroquine, statins, ACE [angiotensin-converting enzyme] inhibitor, immunosuppressives, rituximab—number of treatment cycles and time since last rituximab cycle recorded, prednisolone [and dose], and aspirin), and disease activity assessed by the global British Isles Lupus Assessment Group-2004 index20 and SLE damage index21 (Table II in the Data Supplement). Serum cytokine levels (IL6, IL10, IFN-γ, and TNF-α) were measured in a subset of patients (patients with SLE with no subclinical plaque [SLE-NP] n=19 and patients with SLE with no subclinical plaque [SLE-EP] n=17) from serum taken at the time of the scan; no significant differences were identified between 2 groups (Table II in the Data Supplement). In total, 4 patients were not on any treatment at the time of the scan. All patients gave informed written consent and the study was approved by the combined University College London/University College London Hospital Research Ethics Committee (Reference 06/Q0505/79).

Plaque Detection
A detailed description of the scanning protocol is in the Data Supplement. Briefly, the intima-media thickness (IMT) and the size and nature (stable or unstable) of plaques were measured objectively and noninvasively using vascular ultrasound scans of the common carotid artery, carotid bulb, carotid bifurcation, common femoral artery, and femoral bifurcation. Plaque echogenicity, a measure of plaque stability and lipid content was expressed numerically by Gray Scale Median (GSM) value.22 Lower GSM values signify more echolucent plaque associated with a large lipid core and high inflammatory cell content; whereas plaque with high GSM scores are associated with a small lipid core and high collagen content.23,24,25

Gray Scale Median
Images were normalized using linear scaling with 2 reference points blood (gray scale=0) and adventitia (gray scale=190). After image normalization plaque echogenicity, a measure of plaque stability and lipid content was expressed numerically by Gray Scale Median (GSM) value.22 Lower GSM values signify more echolucent plaque associated with a large lipid core and high inflammatory cell content; whereas plaque with high GSM scores are associated with a small lipid core and high collagen content.23,24,25

Serum Metabolomics
Measures of 228 serum biomarkers were acquired with an established NMR-spectroscopy platform (Nightingale Health).26 These included both absolute concentrations (mmol/L), ratios, and percentages (%) of lipoprotein composition. Serum lipids measured included apolipoproteins and VLDL, LDL, IDL (intermediate density lipoprotein), and HDL particles of different sizes ranging from chylomicrons and extremely large (XXL), very large, large (L), medium (M), small (S), and very small (XS). Lipids within each lipoprotein subclass included total lipid, phospholipids, total cholesterol, cholesterol esters (CE), free cholesterol (FC), and triglycerides. Distribution of these lipids was expressed as a ratio or percentage (%) of total lipid content for each lipoprotein subclass (for list of metabolites see Table III in the Data Supplement).

Data Analysis
Data analysis plan is summarized in Figure 1. Data were analyzed for association using logistic regression (LR) and for classification using 5 different supervised ML algorithms: support vector machine, LR with and without interactions (LR/LR+I), decision trees, and random forest (RF). Ten-fold cross-validation was used to evaluate model performance. Partial Least Squares Discriminant Analysis (sPLS-DA) was used to combine parameter selection and classification into one operation, see Data Supplement for detailed description of the analysis and software/packages used. Demographic, clinical, and treatment variables were adjusted for as appropriate and denoted in figure legends.

Statistical Testing
Statistical tests were performed in Microsoft Excel and GraphPad Prism version 8.3.0 for Windows (GraphPad Software, San Diego). Data was assessed for normality and analyzed with parametric or nonparametric tests as appropriate. Details of statistical tests and parameters accounted for in the analyses are given in the figure legends. P<0.05 were considered statistically significant.
RESULTS

Serum Metabolites Can Differentiate Patients With SLE From Healthy Controls

To establish the taxonomy of lipoproteins in patients with SLE compared with healthy volunteer controls (HCs), detailed lipoprotein-based serum metabolomics was performed (Figure 1, Tables I and III in the Data Supplement). While the routinely available clinical lipid measurements of patients with SLE were within normal ranges (Table IV in the Data Supplement), univariate LRs of the serum metabolites adjusted for age and ethnicity demonstrated a significant difference in the metabolite profile between SLE and HC (Data Files I and II in the Data Supplement). Although age was significantly different between the 2 groups, it did not impact the results; the beta value (which expresses the importance of each variable) ranged between 0.02 and 0.12 (odds ratio 1.05) as opposed to the significant metabolites where the beta coefficient ranged from −5.92 (odds ratio, 0.0027) and 7.88 (odds ratio, 2645; Data Files I and II in the Data Supplement).

Partial Least Squares Discriminant Analysis, a supervised clustering ML model that combines parameter selection and classification into one operation, was then performed using 4 components and 50 metabolite measurements (following model optimization, refer to Methods in the Data Supplement) to rank and validate the metabolite features according to their distribution in SLE and HC. A significant separation between patients with SLE and HCs was observed by plotting principal component (PC)-2 against PC-1 (Figure 2A). The 50 metabolite measurements were ranked by discriminating capability (Figure 2B and 2C). The top 5 ranked metabolites were M-HDL-CE, which was associated with a HC classification, and S-HDL-P, S-VLDL-P, very large-VLDL-FC_%, and IDL-P, which were associated with a SLE classification. Notably, age was not in the top 50 predictors for SLE in the Partial Least Squares Discriminant Analysis plot (Figure 2A through 2C) suggesting that metabolites rather than age differences were driving the separation of patients with SLE from HCs. All 50 metabolites included in the Partial Least Squares Discriminant Analysis clustering (Figure 2B) were also found to be significantly different between SLE and HC in the univariate LRs (Data Files I and II in the Data Supplement) suggesting that patients with SLE have an underlying dyslipidemia that is not detected by routine lipid assessments.

Serum Metabolites Can Predict the Presence of Subclinical Atherosclerotic Plaque in Patients With SLE

To assess whether the disrupted serum metabolite profile identified in patients with SLE could be associated with the presence of subclinical atherosclerosis, patients with SLE...
were stratified based on the presence (SLE-P) or absence (SLE-NP) of plaque(s) detected by vascular ultrasound (Tables II and III in the Data Supplement). Several analysis strategies were applied, and all models were adjusted for ethnicity, age, mean arterial blood pressure, disease duration, disease activity (global British Isles Lupus Assessment Group score), and treatment at the time of the scan (Figure 1). First, univariate LRs identified 4 metabolites which differentiated between SLE-P and SLE-NP patients; Leucine, M-VLDL-FC_%, L-VLDL-FC_% and XXL-VLDL-phospholipids_% which were all increased in serum from SLE-P compared with SLE-NP patients (Figure 3A, Data Files III and IV in the Data Supplement). Of note, these metabolites can be significantly affected by treatment with statins and pro-PCSK9 inhibitors, or body mass index, suggesting that abnormal serum lipid metabolite profiles in SLE-P patients, could be modified using available therapies or interventions (Figure 3A asterisks).

Next, 5 supervised ML models were developed and validated to predict the presence of plaque in patients with SLE; LR, LR+I, support vector machine, RF and Decision tree. Since many of the metabolites measured were biologically interdependent, and therefore highly correlated, homology reduction was applied (see Methods in the Data Supplement). The models were built using the
Figure 3. Identification of important metabolites separating patients with SLE with subclinical plaque (SLE-P) from patients with SLE with no subclinical plaque (SLE-NP).

A. Forest plot depicting statistically significant individual logistic regression results of metabolites in SLE-P (n=30) vs SLE-NP (n=50). Results given in odds ratio (95% CIs). Logistic regressions were adjusted for age, ethnicity, mean arterial blood pressure, global BILAG-2004, disease duration, and treatments at the time of scan (Table II in the Data Supplement). Colored asterisks denote metabolite has previously been shown to be modified by statins, PCSK9 (protein convertase subtilisin/kexin type 9) inhibitors, or body mass index.29–32 (Continued)
homology reduced dataset (124 metabolites) and patient information (age, ethnicity, mean arterial blood pressure, disease duration, global British Isles Lupus Assessment Group index, and treatments at the time of first scan). Sex was not considered as all participants were female (Data File V in the Data Supplement for full lists of the predictors contributing to each model).

The top 3 models, according to classification accuracy, specificity, and F1 scores, were LR, LR+I, and RF (Table). Performance metrics were based on predictions of the models, summarized in confusion matrices (Figure I in the Data Supplement). LR and LR+I had a similar performance, correctly classifying 75% and 80% of SLE-P patients, respectively. The RF model had the best specificity, identifying 45 out of 50 (90%) SLE-NP patients correctly.

The 3 models were further investigated to identify the top features in predicting plaque formation in SLE (Figure 3B through 3E). Four metabolites (XXL-VLDL-phospholipids_%, L-VLDL-FC_%, glycine and tyrosine) and patient age were identified in both LR and RF models as important predictors for SLE-P (Figure 3B). Metabolites, which were significant in individual LR models and also featured in the top 10 metabolites of the LR and RF model, were further investigated for differences between SLE-P and SLE-NP patients (Figure 3C; Figure IIA in the Data Supplement). Receiver operating characteristic curve of the individual metabolites showed an area under the curve of 0.6810 (XXL-VLDL-phospholipids %), 0.7523 (L-VLDL-FC%), 0.7337 (glycine), and 0.6300 (tyrosine; Figure 3D; Figure IIB in the Data Supplement). However, receiver operating characteristic curves based on the classification true positive rate (sensitivity) and false positive rate (1—specificity) of the LR, RF, and LR+I models had an improved area under the curve of 0.78, 0.80, and 0.81, respectively (Figure 3E). This suggested a stronger potential predictive ability when both metabolites and clinical and demographic features (including age and disease duration) were combined and considered together.

The top performing LR+I model (classification accuracy of 0.800) included interactions of each metabolite with all other metabolites and clinical features (assessing over 15,000 possible features; Table). Using the lasso method of shrinkage and selection, only 35 interactions were identified as important and given a beta coefficient to describe the effect size and direction on the model (Figure 3F, Table V in the Data Supplement). Features with larger beta coefficients had the greatest effect on the classification. Notably, L-VLDL-FC_%: age and IDL-triglyceride_%:age, Black/Caribbean ethnicity: hydroxychloroquine treatment and sphingomyelin:leucine were significantly associated with plaque; while S-LDL-CE: rituximab treatment, S+HDL-FC: XXL-VLDL-CE_%, glycine:histidine and statins:ACE inhibitor treatment were associated with the absence of plaques.

Importantly, although age and disease duration were significantly different between the 2 patient groups (Table II in the Data Supplement) and contributed to separation between SLE-P and SLE-NP in the ML models (Data File V in the Data Supplement), these factors did not individually influence metabolite concentrations (Data Files IV and VI in the Data Supplement), thus demonstrating that the metabolite features identified by the models were due to the presence of subclinical plaque.

**Metabolite Interactions Can Differentiate Between SLE-P and SLE-NP Patients**

Using the metabolite interaction features that were selected for the LR+I model (Figure 3F, Table V in the Data Supplement), a Partial Least Squares Discriminant Analysis was performed to rank and validate the metabolite features by their distribution in patients with SLE-P and SLE-NP. By assessing the overall estimation error rate in 10-fold cross-validation, models with 4 components and a subset of 28 metabolite features were chosen for optimal model performance (Figure 4A). This analysis identified a significant separation between SLE-P and SLE-NP patients by plotting PC-2 against PC-1 (Figure 4B). The 28 selected metabolite interaction features were ranked by discriminating capability (Figure 4C and 4D). The 2 highest weighted features were L-VLDL-FC_%:Age and IDL-triglyceride_%:Age, which were also the highest ranked features in the LR+I model (Figure 3F). L-VLDL-FC_% and age were both also included in the top 10 features shared by LR and RF (Figure 3B) and were differentially expressed between SLE-P and SLE-NP patients (Figure 3A) suggesting their influential role in SLE-P patients.

**Figure 3 Continued. B.** Best performing models were determined based on performance statistics (Table). The top 10 features of these logistic regression (LR) and random forest (RF) models are listed to identify common features. C and D, Metabolites which were significant in the individual logistic regression and featured in the top 10 of the LR and RF models were further analyzed using (C) bar charts showing mean, and P-value (see Figure IIA in the Data Supplement) and (D) ROC plots (see Figure IIB in the Data Supplement). E, ROC analysis of the metabolites and features contributing to the LR, RF, and LR+I models, which utilize a combination of metabolomic and clinical features. F, Beta coefficients from the logistic regressions with interactions are plotted for the SLE-P vs SLE-NP analysis. The sign indicates the direction of the effect of the predictor; a positive sign indicates an increased likelihood of a SLE-P prediction, while a negative sign indicates an increased likelihood of a SLE-NP prediction. BMI indicates body mass index; PCSK9, protein convertase subtilisin/kexin type 9; TG, triglyceride; VLDL, very low-density lipoprotein; XS, very small; and XXL, extremely large.
Differential Metabolites Correlated With Clinical Features of SLE-P Patients

Finally, the top 10 metabolites from LR, RF, and all metabolite interactions included in the LR+I models were correlated with clinical and plaque features (Figure 5, Table VI in the Data Supplement). Significant correlations include GSM (measure of plaque stability and lipid content) correlated positively with S-HDL-triglyceride, XS-VLDL-triglyceride, very large-HDL-triglyceride; glycerol and IDL-CE;glycerol and negatively with LDL-D, histidine;XS-VLDL-CE_%; and Glycine:XS-VLDL-CE_%; Plaque number and plaque thickness negatively correlated with histidine;XS-VLDL-CE_%; total plaque area positively correlated with L-VLDL-triglyceride; lactate; and disease activity (British Isles Lupus Assessment Group-2004) positively correlated with M-VLDL-FC_%; and M-VLDL-FC_%;L-LDL-FC_%; and negatively with DHA-FA(22:6, docosahexaenoic acids to total fatty acids); age. The strongest correlation was between disease duration and tyrosinodisease duration.

DISCUSSION

CVD risk in patients with SLE is an important cause of mortality in a cohort of mostly female and relatively young patients compared with the general population. Despite the long-established link between SLE and increased CVD risk, SLE specialists still lack accurate methods of predicting the risk in an individual patient. Traditional cardiovascular risk factors encapsulated in the Framingham equations underestimate the true risk in patients with SLE and fail to predict which patients will have cardiovascular events. A more accurate model of CVD risk in SLE would be useful since it is the main carrier of triglycerides, which are an independent risk factor for CVD, and VLDL particle concentrations have been positively associated with the risk of myocardial infarction. In the JUPITER trial of CVD risk in nearly 12,000 patients, risk among placebo-allocated participants was associated with total VLDL particles, as well as ApoB, total cholesterol, and triglycerides. Furthermore, pharmacological lowering of ApoB-containing lipoproteins, including VLDL, earlier in life is proposed to eliminate high risk of atherosclerotic-CVD in individuals with image-documented subclinical atherosclerosis. Analysis using ML models identified an association between multiple VLDL subsets, amino acids leucine, glycine and tyrosine and clinical features including age with the presence of subclinical plaque, suggesting that more detailed lipoprotein and metabolomic measurements together with demographic information could help to better predict those patients at greatest CVD risk.

While age, disease duration, and treatment are known contributors to CVD risk in SLE, other metabolic factors also contribute to the accelerated risk. Our findings support this showing that the combination of both metabolites and age most strongly distinguished SLE-P from SLE-NP. VLDL subclasses featured predominantly in all the analysis models used suggesting its potential importance in predicting which patients with SLE go on to develop atherosclerosis. VLDL is known to be associated with increased CVD risk. The main carrier of triglycerides, which are an independent risk factor for CVD, and VLDL particle concentrations have been positively associated with the risk of myocardial infarction. In the JUPITER trial of CVD risk in nearly 12,000 patients, risk among placebo-allocated participants was associated with total VLDL particles, as well as ApoB, total cholesterol, and triglycerides. Furthermore, pharmacological lowering of ApoB-containing lipoproteins, including VLDL, earlier in life is proposed to eliminate high risk of atherosclerotic-CVD in individuals with image-documented subclinical atherosclerosis, such as the SLE women in this study.

Mechanistically, the interaction between lipoprotein profile, lupus, and atherogenesis has remained elusive and likely involves a complex cross-regulation between lipoproteins, lupus specific factors (including female sex hormones and treatment) and an activated and dysregulated immune response. Lipoproteins play an important role in maintaining lipid homeostasis, both systemically and at a cellular level via lipid uptake (LDL/VLDL) and efflux (HDL). Lupus-related nontraditional risk factors likely contribute to dyslipidemia, for instance, chronic inflammation can induce reduced serum HDL and increased triglycerides due to increased hepatic VLDL production and reduced clearance of triglyceride-rich

<table>
<thead>
<tr>
<th>Model</th>
<th>F1</th>
<th>Precision</th>
<th>Recall</th>
<th>Specificity</th>
<th>CA</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>0.630</td>
<td>0.708</td>
<td>0.567</td>
<td>0.860</td>
<td>0.750</td>
<td>0.802</td>
</tr>
<tr>
<td>LR+I</td>
<td>0.714</td>
<td>0.769</td>
<td>0.667</td>
<td>0.880</td>
<td>0.800</td>
<td>0.812</td>
</tr>
<tr>
<td>SVM</td>
<td>0.480</td>
<td>0.600</td>
<td>0.400</td>
<td>0.840</td>
<td>0.675</td>
<td>0.707</td>
</tr>
<tr>
<td>RF</td>
<td>0.542</td>
<td>0.722</td>
<td>0.433</td>
<td>0.900</td>
<td>0.725</td>
<td>0.779</td>
</tr>
<tr>
<td>Tree</td>
<td>0.464</td>
<td>0.500</td>
<td>0.433</td>
<td>0.740</td>
<td>0.625</td>
<td>0.630</td>
</tr>
</tbody>
</table>

Performance statistics for 5 predictive models based on serum metabolites at the time of the first scan. The models used were LR with and without interactions (I), SVM, RF, and decision tree (Tree). The (CA) represents the proportion of correctly identified cases, in contrast to specificity, which is the true negative rate. F1 is the weighted average of the precision and recall (see Methods). Statistics are rounded to 3 decimal places. CA indicates classification accuracy; LR, logistic regression; RF, random forest; and SVM, support vector machine.
lipoproteins\(^{44}\); lupus disease activity/damage and inflammatory mediators (including IL-6 and TNF-\(\alpha\)) are independently associated with a proatherogenic lipid profile (elevated triglycerides and low HDL)\(^{45,46}\); and lower LDL activity leading to the accumulation of triglyceride-rich particles.\(^{47,48}\) Small dense LDL is also increased and more easily oxidized as the ability of HDL to prevent the oxidation of LDL is diminished.\(^{49}\) Conversely, subclinical changes in circulating lipoprotein subclasses and their lipid content as described here between patients with SLE with and without plaque are likely to impact immune cell metabolism and function.\(^{50}\) Cell-mediated cholesterol efflux is impaired in patients with SLE\(^{51}\) and LDL composition rather than total LDL levels promoted macrophage infiltration, aortic foam cell formation, and vascular aging in experimental lupus models.\(^{52}\) Furthermore, larger, typically triglyceride-rich, ApoB-containing lipoproteins such as VLDL, may have difficulty leaving the intima because of their larger size or because they get entrapped by components in the subendothelial space.\(^{53}\) Here, these lipoproteins undergo enzymatic modifications that accelerate accumulation and promote aggregation, which is influenced by lipoprotein quantity and composition.\(^{54}\) Notably, the larger triglyceride and cholesterol-rich lipoproteins seem to be more potent than LDL, the most common atherogenic lipoprotein, for provoking greater maladaptive immune activation.\(^{55}\) This supports our previous work showing that VLDL from SLE-P patients could influence the phenotype of \(\text{iNKT cells and monocytes, via altered lipid-antigen presentation.}\(^{19}\) Other mechanisms

**Figure 4.** Partial least squares discriminate analysis validated metabolites identified by logistic regression with interactions to predict patients with systemic lupus erythematosus (SLE) with subclinical plaque (SLE-P).

**A.** Model optimization—model with different components and features kept in the analysis were analyzed, with each color representing a different number of components (Comp), number of features kept in the analysis on the \(x\) axis, and the overall error on the \(y\) axis. **B.** Sparse partial least squares discriminant analysis (sPLS-DA) plot to validate top hits from the logistic regression with interactions. sPLS-DA is a supervised clustering method which separates SLE-P from patients with SLE with no subclinical plaque. **C.** Features included in the sPLS-DA plotted with their factor loading value. **D.** Visualization of the weighting and correlation of each metabolite in component 1 and 2 on the sPLS-DA model. HDL indicates high-density lipoprotein; IDL, intermediate density lipoprotein; LDL, low-density lipoprotein; TG, triglyceride; VLDL, very low density lipoprotein; XL, very large; XS, very small; and XXL, extremely large.
could include disrupted immune cell signaling because of changes in plasma membrane lipid rafts.\textsuperscript{56} In addition, SLE autoantibodies induce endothelial injury, inflammation, and cell-adhesion, all mechanisms promoting atherosclerotic plaque formation.

Previous metabolomics studies in patients with SLE have not focused on cardiovascular risk but rather compared metabolomics profiles between healthy donors and patients with SLE.\textsuperscript{57} One study used mass spectrometry rather than NMR to compare 20 patients with SLE and 9 healthy controls, identifying >100 differentially expressed metabolites but did not assess lipoprotein particles and only one patient had CVD.\textsuperscript{58} Another study using NMR identified raised VLDL and LDL and reduced HDL in patients with SLE.\textsuperscript{59} However, they did not report on lipoprotein subclasses and vascular ultrasound imaging was not performed. Guleria et al\textsuperscript{60} used metabolomics to identify metabolic signatures in different clinical subgroups within a cohort of lupus patients, as we have done here for SLE-P versus SLE-NP patients. Another NMR metabolomics study compared patients with and without lupus nephritis, and healthy controls. Compared with healthy donors, this study reported lower VLDL and LDL in patients with SLE, although higher in the nephritis patients.\textsuperscript{59} The study used patients from India, so ethnicity and lifestyle factors such as diet may have played a role in these results, which do not support other reports. In our more in-depth lipoprotein analysis, we show that...
patients with SLE have a significant alteration in the concentration and content of all lipoprotein subsets compared with healthy controls, with a specific increase of total lipid, cholesterol, and CE content in smaller VLDL subsets but a reduction in HDL subsets. We found a significant increase in the ApoB:ApoA1 ratio in patients with SLE compared with HCs; specifically, a decrease in ApoA1, but no difference in ApoB. This suggests that the increased ApoB:ApoA1 ratio in SLE could be due mainly to a reduction of HDL. About these observations, other studies have reported increased ApoB concentrations\(^\text{43}\) and decreased ApoA\(^\text{1}\),\(^\text{61,62}\) while some studies have associated these differences with autoantibodies against apolipoproteins in SLE.\(^\text{63,64}\)

We also found associations between metabolites and ultrasound scanning measurements. Interestingly, triglycerides in VLDL showed significant correlations with TPA and GSM. Measurements of TPA and echolucency have been shown to have good predictive value for coronary artery disease in women\(^\text{65}\) and thus pertinent in our all-female SLE cohort.

In addition to lipoproteins, leucine, glycine, and tyrosine were also identified by the ML models to be important predictors of SLE-P. Leucine is a branched chain essential amino acid with a potential important role in regulating protein, glucose, and lipid metabolism, in part via activation of the mTOR (mammalian target of rapamycin) protein kinase and promotion of leptin synthesis.\(^\text{66}\) Studies in mice show that leucine supplementation improves diet-induced obesity, insulin resistance, and atherosclerosis outcomes.\(^\text{67,68}\) Increased leucine could be associated with an early response to subclinical plaque development, as we have shown previously.\(^\text{18}\) Glycine, shown here to be reduced in SLE-P, is also reported to have a potential anti-inflammatory role via reducing NF-κB (nuclear factor-κappa B) activation in vascular endothelial cells.\(^\text{69}\) Plasma glycine levels are inversely correlated with acute myocardial infarction in patients undergoing coronary angiography.\(^\text{70}\) Tyrosine, in the form of 3-nitrotyrosine, is associated with oxidized HDL in the human artery wall and circulation in atherosclerosis, and may promote atherosclerosis.\(^\text{71}\) Further work is needed to confirm these findings and to understand fully the complex role of these metabolites in atherogenesis in SLE.

This study has some limitations. While we accounted for most important patient associated factors including treatment, age, and disease duration, we did not have access to sex hormone levels which change with age and could also influence lipoprotein levels.\(^\text{72}\) An important overall aim of this research was to define a clinical algorithm using serum metabolomics and clinical data to help identify patients with SLE at elevated cardiovascular risk who would benefit most from therapeutic or lifestyle intervention. Efforts are ongoing to validate these metabolic signatures in additional larger, multicenter cohorts which are beyond the scope of the current study. It will be important to establish this before clinical algorithms can be defined. Furthermore, while we have identified a number of metabolites associated with subclinical plaque, which have previously been associated with various features of atherosclerosis formation (see above), understanding the specific role these metabolites play in plaque development was beyond the scope of the current study.

In conclusion, the interrogation of lipid subclasses may hold the key to providing insights on how to better stratify CVD risk in SLE. Analysis of lipoproteins using NMR spectroscopy is of particular interest given this high throughput metabolomics analysis is rapid, can be carried out on serum, gives a larger amount of information from each sample and is potentially cost-effective\(^\text{73}\) depending on how many high-risk patients are identified and how the risk is managed. It is possible that a composite score of metabolomics with conventional risk factors may be the best way to assess CVD risk in patients with SLE.\(^\text{74}\)

**REFERENCES**


**ARTICLE INFORMATION**

Received October 1, 2020; accepted January 12, 2021.

**Affiliations**


**Sponsors of Funding**

This work was supported by NIHR University College London Hospital Biomedical Research Centre grant reference BRC531/1/IPT/101350 (IPT/ECJ). L. Coelewij is supported by UCL & Birchbeck MRC Doctoral Training Programme. G.A. Robinson was supported by a PhD studentship from Lupus UK and The Rosettes Trust (M409). K.E. Wardington was funded by a British Heart Foundation PhD Studentship (FS/13/59/30649). J. Peng is supported by Versus Arthritis grant reference 21226.

**Disclosures**

None.


