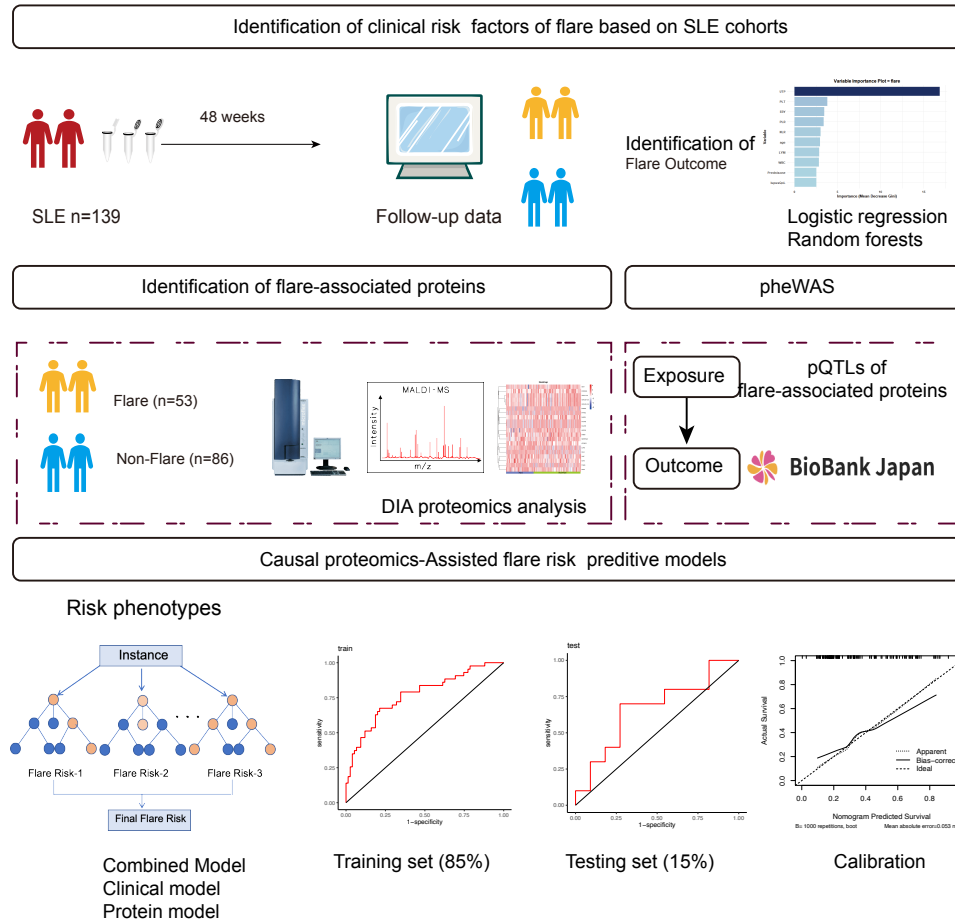


Graphical Abstract



Phenome-wide causal proteomics enhance systemic lupus erythematosus flare prediction: A study in Asian populations

Liying Chen^a, Ou Deng^c, Ting Fang^a, Mei Chen^a, Xvfeng Zhang^a, Ruichen Cong^c, Dingqi Lu^b, Runrun Zhang^b, Qun Jin^c, Xinchang Wang^b,

^a*Second Clinical Medical College, Zhejiang Chinese Medical University, Hangzhou, 310053, China*

^b*Department of Rheumatology, The Second Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, 310053, China*

^c*Graduate School of Human Sciences, Waseda University, Mikajima, Tokorozawa, 359-1192, Japan*

Abstract

Objective: Systemic lupus erythematosus (SLE) is a complex autoimmune disease characterized by unpredictable flares. This study aimed to develop a novel proteomics-based risk prediction model specifically for Asian SLE populations to enhance personalized disease management and early intervention.

Methods: A longitudinal cohort study was conducted over 48 weeks, including 139 SLE patients monitored every 12 weeks. Patients were classified into flare (n = 53) and non-flare (n = 86) groups. Baseline plasma samples underwent data-independent acquisition (DIA) proteomics analysis, and phenome-wide Mendelian randomization (PheWAS) was performed to evaluate causal relationships between proteins and clinical predictors. Logistic regression (LR) and random forest (RF) models were used to integrate proteomic and clinical data for flare risk prediction.

Results: Five proteins (SAA1, B4GALT5, GIT2, NAA15, and RPIA) were significantly associated with SLE Disease Activity Index-2K (SLEDAI-2K) scores and 1-year flare risk, implicating key pathways such as B-cell receptor

Corresponding author: Xinchang Wang, Department of Rheumatology, The Second Affiliated Hospital of Zhejiang Chinese Medical University, No.318, Chaowang Road, Gongshu District, Hangzhou, Zhejiang Province, China. Email: 20054007@zcmu.edu.cn; ossani@126.com

signaling and platelet degranulation. SAA1 demonstrated causal effects on flare-related clinical markers, including hemoglobin and red blood cell counts. A combined model integrating clinical and proteomic data achieved the highest predictive accuracy (AUC = 0.769), surpassing individual models. SAA1 was highlighted as a priority biomarker for rapid flare discrimination.

Conclusion: The integration of proteomic and clinical data significantly improves flare prediction in Asian SLE patients. The identification of key proteins and their causal relationships with flare-related clinical markers provides valuable insights for proactive SLE management and personalized therapeutic approaches.

Keywords: Systemic lupus erythematosus, Flare, Causal proteomics, Phenome-wide mendelian randomization, Longitudinal cohort study

1 Introduction

2 Systemic lupus erythematosus (SLE) is a heterogeneous autoimmune disorder
3 characterized by the production of multiple autoantibodies and involvement
4 of various organ systems. Its clinical course is unpredictable, with patients
5 experiencing remissions and flares sudden increases in disease activity that
6 pose significant management challenges[1, 2]. Although early disease control
7 and adherence to pharmacological interventions are foundational in managing
8 SLE, conventional predictive models relying on clinical observations and ge-
9 netic markers are limited in capturing the disease’s dynamic and multifactorial
10 nature[3, 4].

11 The complexity of SLE flares, often arising unpredictably from multi-
12 faceted and poorly understood triggers, necessitates a nuanced, data-driven
13 predictive approach. Accurate flare prediction requires longitudinal tracking
14 of disease activity, identification of modifiable risk factors, and comprehensive
15 evaluation of clinical, serological, and molecular markers. Recent studies in
16 Asian SLE populations have linked risk factors such as thrombocytopenia,
17 hypocomplementemia, elevated neutrophil-to-lymphocyte ratio (NLR), and
18 high platelet-to-lymphocyte ratio (PLR) to increased flare risk[5, 6]. Addi-
19 tionally, the presence of specific autoantibodies like anti-ribosomal P and
20 anti-phospholipid antibodies is associated with heightened flare susceptibil-
21 ity in Chinese patients[7]. Non-biological factors, including quality of life,
22 psychological stress, and overexertion, also significantly contribute to flare

23 incidence, highlighting the need to consider both biological and psychosocial
24 variables in predictive models[8, 9, 10].

25 Advancements in genome-wide association studies (GWAS) and PheWAS
26 have deepened our understanding of the genetic underpinnings and phenotypic
27 expressions of autoimmune diseases. Mendelian randomization (MR) and
28 PheWAS methodologies enable exploration of causal relationships between
29 genetic variations, proteins, and disease outcomes, offering more robust
30 insights than traditional observational studies[11, 12]. Given SLE's diverse
31 nature, a multi-omics approach that integrates proteomic data with genomic
32 and clinical information is essential for identifying dynamic biomarkers capable
33 of improving flare prediction[13]. In this context, proteomics-based MR holds
34 significant promise for discovering novel, druggable protein targets, thereby
35 enhancing our ability to classify and stratify patients by flare risk.

36 This study aims to address the limitations of current predictive models
37 for SLE flares through a comprehensive, integrative approach. We first
38 conduct an exhaustive analysis of known clinical predictors to build a robust
39 baseline model. Building on this foundation, we incorporate cutting-edge
40 plasma proteomics techniques to gain deeper insights into molecular changes
41 associated with flare events. We explore causal relationships between identified
42 plasma proteins and clinical outcomes through phenome-wide Mendelian
43 randomization analyses, identifying proteins with a causal impact on flare
44 risk. Finally, we leverage advanced machine learning algorithms to integrate
45 proteomic and clinical data, constructing a predictive model that offers more
46 precise risk stratification for SLE flares. This integrative methodology aims
47 to significantly improve early detection of organ damage, guiding timely and
48 personalized therapeutic interventions.

49 **Methods**

50 *Patient Recruitment and Follow-up*

51 This prospective multicenter study recruited 139 SLE patients between August
52 2020 and January 2023. Eligible participants were aged 18-65 and met the 2012
53 SLICC classification criteria[14]. Patients with active infections, malignancies,
54 or other connective tissue diseases were excluded. Disease activity was assessed
55 using the SLEDAI-2K at baseline and at 3, 6, and 12 months to identify flare
56 events.

57 The study was approved by the ethics committee of the Second Affiliated
58 Hospital of Zhejiang Chinese Medical University (approval NO.2020-KL-002-

59 IH01). Informed consent was obtained from all participants. Additional
60 details are in Supplementary Table 1.

61 *Outcome Measurement*

62 The primary outcome was the occurrence of an SLE flare within 12 months,
63 defined as an increase in SLEDAI-2K score by ≥ 3 points from baseline and
64 previous assessments[15]. This widely accepted criterion standardizes flare
65 identification.

66 *Clinical Variables: Definitions and Selection*

67 Baseline data included: (1) Complement levels (C3, C4); (2) Inflammatory
68 markers (NLR, PLR); (3) Medication use (glucocorticoids, immunosuppres-
69 sants); (4) Serological markers (anti-ribosomal P antibodies); (5) Disease
70 activity (SLEDAI-2K score); and (6) Quality of life (LupusQoL questionnaire).
71 LR and RF algorithms identified key clinical predictors by ranking features
72 based on predictive importance. All variables were dichotomized (1 or 0) for
73 simplicity and clinical relevance, facilitating risk stratification. This approach
74 aims to develop a robust predictive model integrating clinical parameters
75 with proteomic biomarkers.

76 *Proteomic Analysis*

77 DIA data were processed into spectral libraries using SpectraST and analyzed
78 with DIA-NN (v1.7.0), ensuring accurate results. Detailed workflows are
79 available at <https://www.iprox.cn/page/HMV006.html>.

80 *Differential Protein and SLEDAI-2K Correlation Analysis*

81 Proteins expressed in at least 25% of samples were analyzed. Student's
82 t-test identified plasma proteins differentiating flare and non-flare patients,
83 considering proteins with $|\log_2(\text{Fold Change})| \geq 1$ and $p < 0.05$ as significant.
84 Spearman's correlation assessed relationships between these proteins and
85 SLEDAI-2K scores, with significance at $p < 0.05$.

86 *Pathway Enrichment Analysis*

87 We performed pathway enrichment analysis on differentially expressed proteins
88 and those correlated with SLEDAI-2K scores. We analyzed Reactome and
89 KEGG pathways, considering those significantly enriched at $p < 0.05$ using
90 DAVID (<https://david.ncifcrf.gov/>).

91 *Causal Proteomics Analysis Using PheWAS*

92 Genetic instruments for identified pQTLs were obtained from 2,958 Han
93 Chinese participants[16]. Outcome data for SLE phenotypes were sourced
94 from BioBank Japan ($n = 179,000$) and supplemented by UK Biobank and
95 FinnGen data ($n_{\text{total}} = 628,000$)[17]. The GWAS dataset included predictors
96 like chronic glomerulonephritis, blood cell counts, and medication use. We
97 selected pQTLs with $p < 5 \times 10^{-8}$, independence ($R^2 < 0.001$ with clumping
98 distance $> 10,000$ kb), and F statistic > 10 to ensure robust instruments.

99 The primary method was inverse-variance weighted Mendelian Random-
100 ization. To address potential violations, we performed sensitivity analyses
101 using MR-Egger regression, weighted median, and mode-based estimators.
102 Heterogeneity was assessed with Cochran’s Q test, and leave-one-out analyses
103 evaluated SNP influence on causal estimates. This approach allowed us to
104 explore causal relationships between serum amyloid A1 (SAA1) pQTLs and
105 SLE flare risk.

106 *Machine Learning-Based Biomarker Selection*

107 We developed predictive models for SLE flares using LR and RF algorithms.
108 Three models were created: (1) Clinical models using validated risk param-
109 eters; (2) Protein models using identified biomarkers; and (3) Combined
110 models integrating clinical and proteomic data. Datasets were split into
111 training (85%) and testing (15%) sets. Model performance was evaluated
112 using Receiver Operating Characteristic curves and calibration curves to
113 assess discriminative ability and reliability.

114 *Statistical Analysis*

115 Statistical analyses were conducted using R (v4.3.2) and GraphPad Prism
116 8. Clinical data are presented as medians with interquartile ranges. Group
117 comparisons used Student’s t-test. Correlation analyses and linear regression
118 assessed variable relationships. PheWAS analyses employed the `TwoSampleMR`
119 R package. Predictive models used LR. Model accuracy was assessed using
120 ROC curves and Area Under the Curve metrics via the `pROC` R package.

121 **Results**

122 *Clinical Data Characteristics and Flare Outcome*

123 As shown in Fig.1, 139 SLE patients were monitored over 48 weeks and
124 classified into flare ($n = 53$) and non-flare ($n = 86$) groups; 38% experienced

125 at least one flare. The flare group had a high female predominance (94%),
126 aligning with SLE's gender disparity. Approximately 47% of flare patients
127 were positive for anti-dsDNA antibodies. Significantly, the flare group had a
128 higher prevalence of rRNP antibodies than the non-flare group ($p = 0.019$),
129 suggesting their potential role in predicting flares.

130 Significant differences in erythrocyte sedimentation rate (ERY, $p = 0.008$),
131 platelet count (PLT, $p = 0.023$), and platelet-to-lymphocyte ratio (PLR,
132 $p = 0.035$) were observed, highlighting their potential as flare risk markers.
133 Medication use showed no significant difference in prednisone use ($p = 0.284$).
134 Hydroxychloroquine use was slightly lower in the flare group (by 11%), though
135 not statistically significant ($p = 0.617$), warranting further investigation into
136 its protective effects. No significant differences in organ involvement were
137 observed ($p > 0.05$). Baseline characteristics are detailed in Table1.

138 *Importance Ranking of Clinical Predictors*

139 We identified key clinical risk factors for SLE flares using univariate LR
140 and RF based on 39 baseline features from 139 patients. In the LR model
141 (Fig.2A), top predictors were: (1) 24-hour urinary total protein (UTP), (2)
142 urinary erythrocyte count (ERY), (3) platelet count (PLT), (4) platelet-to-
143 lymphocyte ratio (PLR), and (5) anti-ribonucleoprotein (rRNP) antibodies.
144 The RF model (Fig.2B) yielded a similar ranking, with NLR as the fifth
145 predictor.

146 *Identification of Differentially Expressed Biomarkers Associated with Flare*

147 Proteomic analysis identified 102 significantly differentially expressed proteins
148 (73 upregulated, 23 downregulated) associated with flares (Fig.3A), based on
149 proteins detected in over 25% of samples. A heatmap (Fig.3B) illustrates these
150 findings. Pathway enrichment analysis using Reactome revealed involvement
151 in key pathways: (1) B-cell receptor (BCR) downstream signaling, (2) response
152 to elevated platelet cytosolic Ca^{2+} , and (3) platelet degranulation (Fig.3C),
153 offering insights into molecular mechanisms underlying flares.

154 *Proteins Significantly Associated with SLEDAI-2K*

155 We identified plasma proteins significantly associated with SLEDAI-2K
156 scores. KEGG pathway analysis indicated their involvement in neurode-
157 generation, lipid metabolism, atherosclerosis, and neurotrophin signaling
158 (Fig.3D). Protein-protein interaction analysis using STRING (Fig.3E) showed

159 functional relationships among these proteins. Among 13 proteins cor-
160 relating with SLEDAI-2K, five were significantly upregulated and posi-
161 tively associated with worse outcomes: (1) Serum Amyloid A-1 (SAA1),
162 (2) β -1,4-galactosyltransferase 5 (B4GALT5), (3) Ribose 5-phosphate Iso-
163 merase A (RPIA), (4) GTPase-activating Protein 2 (GIT2), and (5) N-
164 alpha-acetyltransferase 15 (NAA15). Correlation analysis (Fig.3F) showed
165 significant positive associations between SAA1 and B4GALT5, RPIA, and
166 NAA15, suggesting potential functional relationships in SLE activity and
167 flare risk.

168 *Causal Effects of Flare-Associated Proteins on SLE and Risk Factors*

169 Using PheWAS, we investigated causal mechanisms between plasma proteins
170 and SLE flares, focusing on pQTLs for SAA1. We examined 220 SLE-related
171 outcomes using data from BioBank Japan. Six SNPs associated with SAA1
172 levels were identified as instrumental variables. Inverse-variance weighted
173 (IVW) analysis revealed significant causal effects of SAA1 on four flare-related
174 outcomes (OR = 1.071, 95% CI: 1.004 - 1.143, $p = 0.040$), hemoglobin (OR=
175 0.971, 95% CI: 0.947 - 0.996, $p = 0.023$), red blood cell count (OR = 0.971,
176 95% CI: 0.947 - 0.996, $p = 0.021$), hematocrit (OR =0.967, 95% CI: 0.937
177 - 0.997, $p = 0.031$) (Fig.4A-D, Table 2). All of them exhibited absence of
178 heterogeneity by IVW (Cochran's Q = 0.419; Cochran's Q = 0.263; Cochran's
179 Q = 0.291 ; Cochran's Q =0.104, respectively) and MR-Egger (Cochran's
180 Q= 0.334; Cochran's Q = 0.169; Cochran's Q = 0.195; Cochran's Q =0.068,
181 respectively), horizontal pleiotropy by MR-Egger (intercept = -0.015, $p =$
182 0.585; intercept = -0.001, $p = 0.891$; intercept = -0.003, $p = 0.808$; intercept
183 = -0.005, $p = 0.694$, respectively). The leave-one-out method suggested that
184 the MR analysis results were reliable (Fig.5A-D).

185 *Construction and Internal Validation of Flare Risk Prediction Models*

186 We developed predictive models for SLE flare risk by dividing patients into
187 training and test sets. Using LR and random forest algorithms, and following
188 TRIPOD guidelines[18], we selected six key clinical variables: UTP, ERY,
189 PLT, PLR, anti-rRNP antibodies, and NLR.

190 Three models were developed: a clinical model, a protein-based model
191 using five key proteins (SAA1, B4GALT5, GIT2, NAA15, RPIA), and a
192 combined model integrating clinical and proteomic variables. ROC curve
193 analysis showed the protein-based model had superior predictive accuracy
194 (AUC = 0.744, 95% CI: 0.646 - 0.842) compared to the clinical model (AUC =

195 0.643, 95%CI: 0.541 - 0.745). The combined model achieved the highest AUC
196 (AUC = 0.769, 95% CI : 0.678 - 0.860), indicating the benefit of integrating
197 clinical and proteomic data.

198 Calibration analysis demonstrated good predictive performance for all
199 models (Fig.6A-C). The protein-based model showed particular strength
200 in AUC and Mean Absolute Error (MAE), suggesting its suitability for
201 independent flare prediction. The combined model may be more appropriate
202 in complex clinical scenarios requiring multiple variables for accurate risk
203 assessment.

204 Discussion

205 This is the first study to develop an SLE flare prediction model based on
206 proteomic analysis in an East Asian population. We identified novel molecular
207 pathways and biomarkers associated with SLE flares, offering valuable insights
208 for improved risk stratification and personalized management.

209 Pathway enrichment analysis showed significant involvement of B cell
210 receptor signaling, elevated platelet cytosolic Ca²⁺ responses, and platelet
211 degranulation in SLE flare pathogenesis. KEGG analysis revealed enrichment
212 of neurodegeneration, lipid metabolism, atherosclerosis, and neurotrophin
213 signaling pathways among proteins linked to disease activity. These find-
214 ings highlight the complex interplay between immune activation, vascular
215 responses, and systemic inflammation in SLE flares, particularly affecting
216 hematological, cardiovascular, and neurological systems in this East Asian
217 cohort.

218 We identified five key proteins, SAA1, B4GALT5, GIT2, NAA15, and
219 RPIA, strongly associated with increased SLE flare risk over 48 weeks, in-
220 dependent of conventional risk factors. Our analysis demonstrated a causal
221 effect of elevated SAA1 levels on flare risk factors, underscoring its critical
222 role in disease progression.

223 Integrating these proteomic biomarkers with clinical indicators significantly
224 enhanced our model's predictive accuracy. These results confirm the prognos-
225 tic potential of causal proteomics in SLE flare risk stratification, paving the
226 way for personalized therapies and earlier interventions for high-risk patients.

227 Previous studies support Serum Amyloid A1 (SAA1) as a key biomarker
228 for assessing SLE flare risk, reinforcing its predictive utility. SAA1-associated
229 pQTLs have been causally linked to hematological parameters and depressive

230 disorders in multi-GWAS PheWAS[19], enhancing our understanding of pro-
231 teins in the psychological aspects of SLE flares[20]. As an acute-phase protein,
232 SAA1 is markedly upregulated during inflammation, contributing to organ
233 dysfunction[21]. Our study showed that elevated SAA1 levels causally influ-
234 ence hematological involvement and antirheumatic medication use, consistent
235 with prior findings[22].

236 Elevated SAA1 levels correlate with SLEDAI-2K scores and nervous sys-
237 tem involvement severity[23]. SAA1 is linked to the Th17 cell differentiation
238 pathway, amplifying inflammatory responses[24]. Given its pivotal role in mod-
239 ulating inflammatory mediators and immune cells in SLE pathophysiology[25],
240 investigating SAA1 can advance our understanding of flare mechanisms.

241 Identified as an early immunological diagnostic biomarker with high sensi-
242 tivity and specificity for SLE[26], SAA1 may serve as a valuable biomarker for
243 identifying high-risk patients based on our East Asian PheWAS results. These
244 findings suggest potential therapeutic targets for personalized treatments.
245 However, since infections also trigger flares and SAA1 is associated with white
246 blood cell count and NLR, further research is needed to distinguish SLE flares
247 from infection-related events[27].

248 β -1,4-Galactosyltransferase V (B4GALT5) is crucial in carbohydrate
249 metabolism, specifically in lactosylceramide synthesis [28]. Unexpectedly,
250 our logistic regression showed a strong positive correlation between B4GALT5
251 expression and SLE flares, making it a significant protein associated with
252 flare risk. Given B4GALT5's role in antiviral immunity, overexpression leads
253 to upregulation of inflammatory cytokines and glycosylated surface proteins
254 involved in antigen presentation, cell adhesion, and migration[29].

255 However, we did not establish a causal relationship between B4GALT5 and
256 SLE, possibly due to the low prevalence of B4GALT5 variants in the East Asian
257 population studied. This underscores the need for diverse population studies
258 to evaluate B4GALT5's role in SLE pathogenesis and its predictive value. The
259 unexpected correlation suggests a complex interplay between glycosylation
260 processes, immune responses, and SLE activity. Further investigation is needed
261 to elucidate how B4GALT5 might influence flare risk, given its functions
262 in carbohydrate metabolism and immune regulation. These findings open
263 avenues for understanding the glycosylation-immune axis in SLE and potential
264 therapeutic interventions.

265 GIT2 and NAA15 emerged as potential contributors to SLE flares. GIT2
266 is a scaffold protein regulating aging processes affecting multiple tissues
267 and linked to neurodegeneration and cardiovascular disorders[30]. NAA15

268 is a susceptibility gene for neurodevelopmental disorders[31], implicated in
269 congenital cardiac anomalies, plaque stability in atherosclerosis, and seizure
270 pathophysiology[32, 33]. Although no direct evidence links GIT2 and NAA15
271 to SLE, their identification suggests novel pathways in SLE pathogenesis.
272 Further research is needed to elucidate their specific roles in SLE flares,
273 particularly considering disease duration and age, which merit attention in
274 risk prediction models.

275 SLE progression involves distinct risk factors at different disease stages.
276 Early organ damage is driven by active disease (reflected in SLEDAI-2K), while
277 later damage results from long-term medication effects, especially prolonged
278 glucocorticoid use and withdrawal[34, 35, 36]. We identified key clinical
279 risk factors for flares, including hematological and renal involvement and
280 medication use, aligning with traditional findings. Risk factors may shift
281 during disease progression, with early damage from disease activity and later
282 damage from drug side effects, notably glucocorticoid therapy.

283 Traditional correlation studies offer insights but do not establish causality.
284 We employed proteomic Mendelian randomization to develop robust flare
285 prediction models. By integrating correlational methods with advanced
286 machine learning and causal analyses, we aimed to enhance prediction accuracy.
287 Identifying proteomic biomarkers with direct causal effects provides deeper
288 insights into SLE flare mechanisms, highlighting potential causal pathways
289 and therapeutic targets for future interventions.

290 This study has several limitations. The small sample size and focus on
291 an East Asian population may limit statistical power and generalizability to
292 broader populations. The 12-month follow-up may be insufficient to capture
293 long-term fluctuations in SLE disease activity. Moreover, reliance on advanced
294 proteomics techniques, while valuable, may hinder clinical accessibility and
295 routine implementation. Although our PheWAS analysis revealed associations
296 between genetic protein levels and SLE, establishing definitive causality
297 remains challenging due to the complexities of SLE and potential violations
298 of Mendelian randomization assumptions. Therefore, these results should be
299 interpreted with caution. Future research should prioritize external validation
300 of our predictive model in diverse cohorts, adhering to the TRIPOD guidelines.
301 Studies with larger, heterogeneous populations and extended follow-up periods
302 are essential to address current limitations. Developing methodologies that
303 account for time-varying exposures and non-linear associations will enhance
304 model robustness and accuracy. Further investigation is needed to verify
305 causal relationships between identified proteins and SLE risk, improving

306 clinical applicability. Additionally, comprehensive proteomic analyses across
307 multiple organs are necessary to assess organ-specific protein effects, providing
308 deeper insights into early flare mechanisms.

309 **Conclusion**

310 This study demonstrates the potential of integrating causal proteomics with
311 clinical risk factors to improve SLE flare prediction. Our findings provide
312 significant advancements in understanding the molecular mechanisms under-
313 lying SLE flares and offer a foundation for more precise and personalized
314 approaches to SLE management. The identified proteomic signatures and
315 causal pathways represent promising avenues for future therapeutic interven-
316 tions and risk stratification strategies, with the potential to enhance patient
317 care and outcomes in SLE.

318 **Abbreviations:**

319 SLE: systemic lupus erythematosus;
320 PheWAS: phenome-wide mendelian randomization;
321 GWAS: genome-wide association studies;
322 MR: mendelian randomization;
323 DIA: data-independent acquisition;
324 LR: logistic regression;
325 RF: random forest;
326 UTP: 24-hour urinary total protein;
327 ERY: urinary erythrocyte count;
328 PLT: platelet count;
329 NLR: neutrophil-to-lymphocyte ratio;
330 PLR: platelet-to-lymphocyte ratio;
331 rRNP: anti-ribosomal P antibodies;
332 LupusQoL: lupus quality of life;
333 SLEDAI-2K: Systemic Lupus Erythematosus Disease Activity Index 2000;
334 BCR: B-cell receptor;
335 SLICC: systemic lupus international collaborating clinics;
336 KEGG: Kyoto Encyclopedia of Genes and Genomes;
337 SAA1: serum amyloid A1;
338 B4GALT5: β - 1,4-galactosyltransferase 5;
339 RPIA: Ribose 5-phosphate isomerase A;
340 GIT2: GTPase-activating protein 2;

341 NAA15: N-alpha-acetyltransferase 15;
342 IVW: inverse-variance weighted;
343 TRIPOD: transparent reporting of a multivariable prediction model for individual prognosis
344 or diagnosis.

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

Table 1: Clinical and demographic characteristics of SLE flare.

Number of patients	Total ($n = 139$)	Primary outcome: 1-year flare		P
		Non-flare ($n = 86$)	Flare ($n = 53$)	
Sex, n (%)	130 (94%)	80 (93%)	50 (94%)	1
Age, Median (Q1, Q3)	36 (28.5, 48.5)	34 (31, 42.75)	40 (28, 51)	0.357
Course, Median (Q1, Q3)	6 (3, 12)	6 (3, 12)	6 (3, 11)	0.617
SLEDAI, Median (Q1, Q3)	6 (3, 8)	6 (3, 8)	6 (4, 12)	0.337
dsDNA, n (%)	66 (47%)	41 (48%)	25 (47%)	1
Sm, n (%)	35 (25%)	19 (22%)	16 (30%)	0.386
SSA, n (%)	62 (45%)	37 (43%)	25 (47%)	0.763
Ro52, n (%)	52 (37%)	28 (33%)	24 (45%)	0.185
SSB, n (%)	22 (16%)	12 (14%)	10 (19%)	0.595
rRNP, n (%)	38 (27%)	17 (20%)	21 (40%)	0.019
Scl70, n (%)	1 (1%)	0 (0%)	1 (2%)	0.381
AHA, n (%)	27 (19%)	17 (20%)	10 (19%)	1
AnuA, n (%)	39 (28%)	24 (28%)	15 (28%)	1
CENP, n (%)	2 (1%)	0 (0%)	2 (4%)	0.144
AMA.M2, n (%)	2 (1%)	1 (1%)	1 (2%)	1
U10RNP, n (%)	25 (18%)	15 (17%)	10 (19%)	1
pro, n (%)	48 (34%)	27 (31%)	21 (40%)	0.492
ERY, Median (M25, M75)	3 (0, 13)	2.4 (0, 10.6)	5.5 (1, 20)	0.008
LEU, Median (M25, M75)	6 (2, 24.8)	6 (1, 24.45)	8 (2.3, 26)	0.333
UTP, Median (M25, M75)	102.45 (102.45, 180)	102.45 (102.45, 102.45)	180 (180, 180)	< 0.001
WBC, Median (M25, M75)	5.2 (4.3, 6.86)	5.2 (4.27, 6.9)	5.1 (4.4, 6.4)	0.873
NEUT, Median (M25, M75)	3.43 (2.46, 4.82)	3.46 (2.5, 4.78)	3.33 (2.28, 4.88)	0.714
LYM, Median (M25, M75)	1.36 (1.04, 1.97)	1.33 (1.03, 1.88)	1.43 (1.04, 2.23)	0.534
PLT, Median (M25, M75)	208 (157, 248.5)	217 (172.5, 258.75)	175.5 (138, 225)	0.023
PLR, Median (M25, M75)	145.37 (102.39, 190.67)	156.1 (111.63, 200.63)	134.18 (83.08, 179.71)	0.035
NLR, Median (M25, M75)	2.36 (1.62, 3.51)	2.37 (1.79, 3.48)	2.25 (1.51, 3.54)	0.411
C3, Median (M25, M75)	0.76 (0.66, 0.9)	0.74 (0.66, 0.9)	0.78 (0.67, 0.93)	0.606
C4, Median (M25, M75)	0.15 (0.11, 0.2)	0.15 (0.11, 0.2)	0.15 (0.1, 0.2)	0.776
lupusQoL, Median (M25, M75)	114 (101, 121)	114 (103.25, 120.75)	110 (97, 122)	0.529
Prednisone, Median (M25, M75)	10 (5, 15)	7.5 (5, 15)	10 (5, 15)	0.284
Hydroxychloroquine, n (%)	86 (62%)	57 (66%)	47 (55%)	0.617
Immunosuppressant, n (%)	90 (65%)	57 (66%)	33 (62%)	0.765
Noninvolved, n (%)	31 (22%)	17 (20%)	14 (26%)	0.481
Multisystem, n (%)	40 (29%)	27 (31%)	13 (25%)	0.499
Hematologic, n (%)	22 (16%)	15 (17%)	7 (13%)	0.671
Neuropsychiatric, n (%)	7 (5%)	3 (3%)	4 (8%)	0.427
Mucocutaneous, n (%)	62 (45%)	36 (42%)	26 (49%)	0.514
Renal, n (%)	57 (41%)	39 (45%)	18 (34%)	0.251

Table 2: Mendelian Randomization Results - Causal Effects Between pQTLs and SLE Flare Risk (IVW Approach)

Outcome	N SNP	β	Standard Error	p	OR (95% CI)
Hemoglobin	6	-0.029	0.013	0.023	0.971 (0.947, 0.996)
Hematocrit	6	-0.034	0.016	0.031	0.967 (0.937, 0.997)
Red blood cell count	6	-0.029	0.013	0.021	0.971 (0.947, 0.996)
Anti-inflammatory medication	6	0.069	0.033	0.040	1.071 (1.004, 1.143)

Note: Source by Zheng Lab.

345

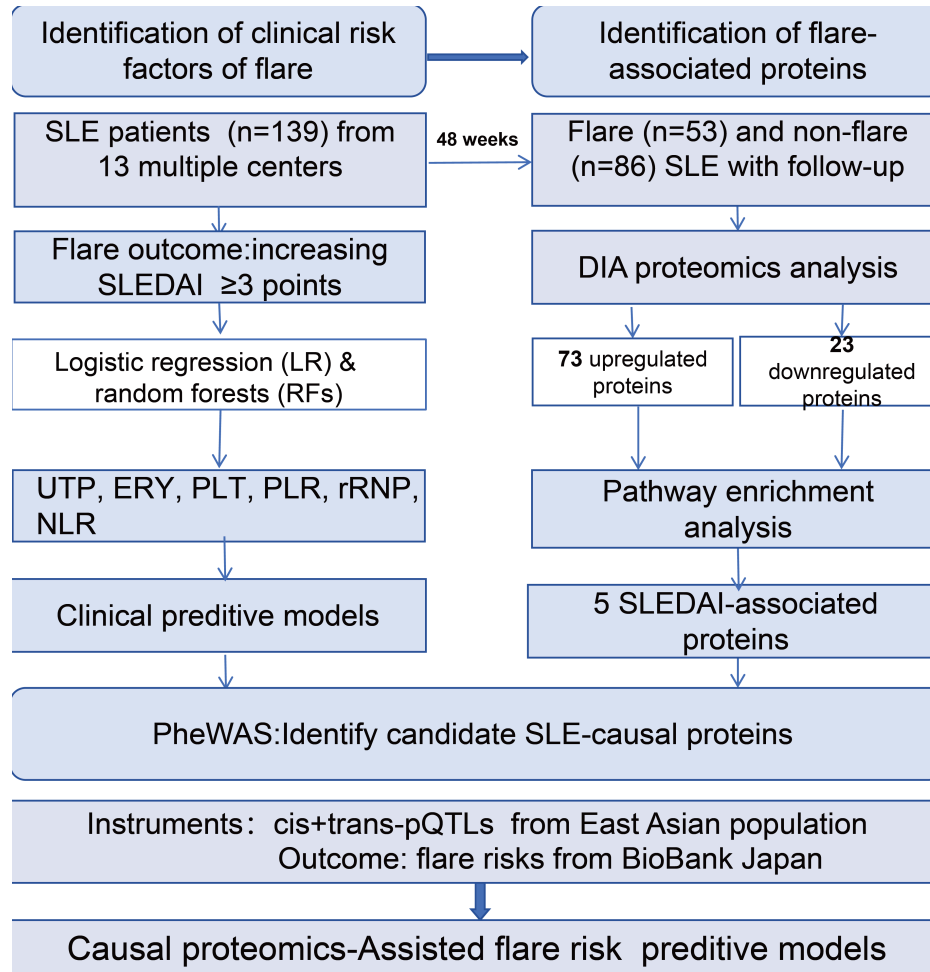


Figure 1: Study design and patient follow-up schema. Schematic representation of the recruitment process, inclusion and exclusion criteria, and longitudinal assessment timeline for SLE patients over the 48-week study period.

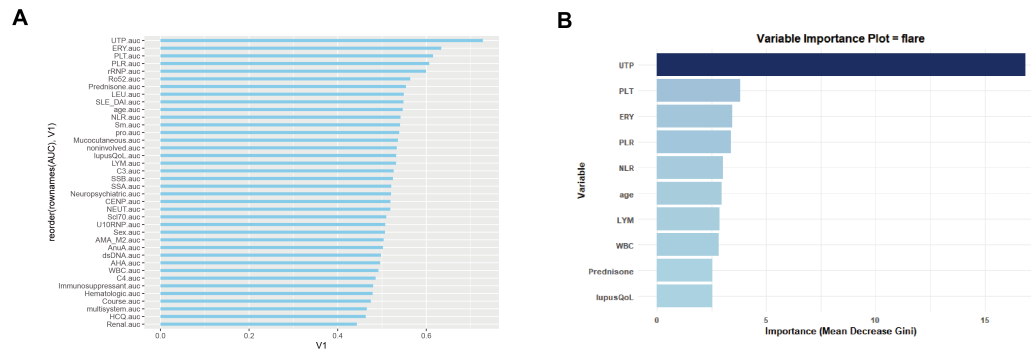


Figure 2: Comparative analysis of clinical variable importance in predicting SLE flares. (A) Ranking of clinical variables based on LR coefficients. (B) Relative importance of clinical variables determined by the RF algorithm, measured by the mean decrease in Gini impurity.

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

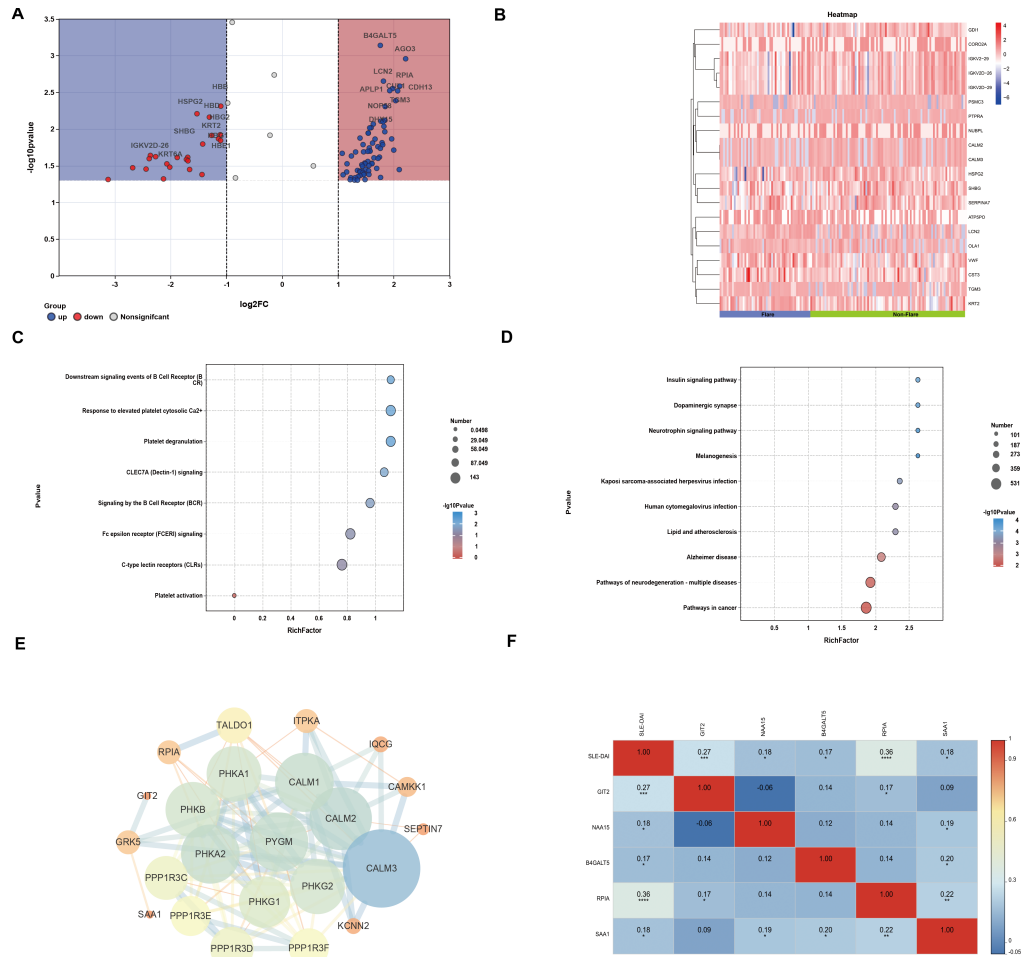


Figure 3: Multi-dimensional proteomic profiling of SLE flares. (A) Volcano plot illustrating the distribution and statistical significance of 102 differentially expressed proteins between flare and non-flare groups (‘the absolute value of the base-2 logarithm of the fold change’ ≥ 1 , $p < 0.05$). (B) Hierarchical clustering heatmap displaying normalized plasma protein expression levels, with a color gradient (blue to red) representing relative protein abundance. (C) Reactome pathway enrichment analysis of flare-associated proteins, with statistical significance indicated by $-\log_{10}(p\text{-value})$. (D) KEGG pathway analysis of proteins correlated with SLEDAI-2K scores. (E) STRING-based protein-protein interaction network of SLEDAI-2K-associated proteins, where node size reflects the degree of interaction. (F) Correlation matrix showing the relationships between upregulated proteins positively associated with SLEDAI-2K scores.

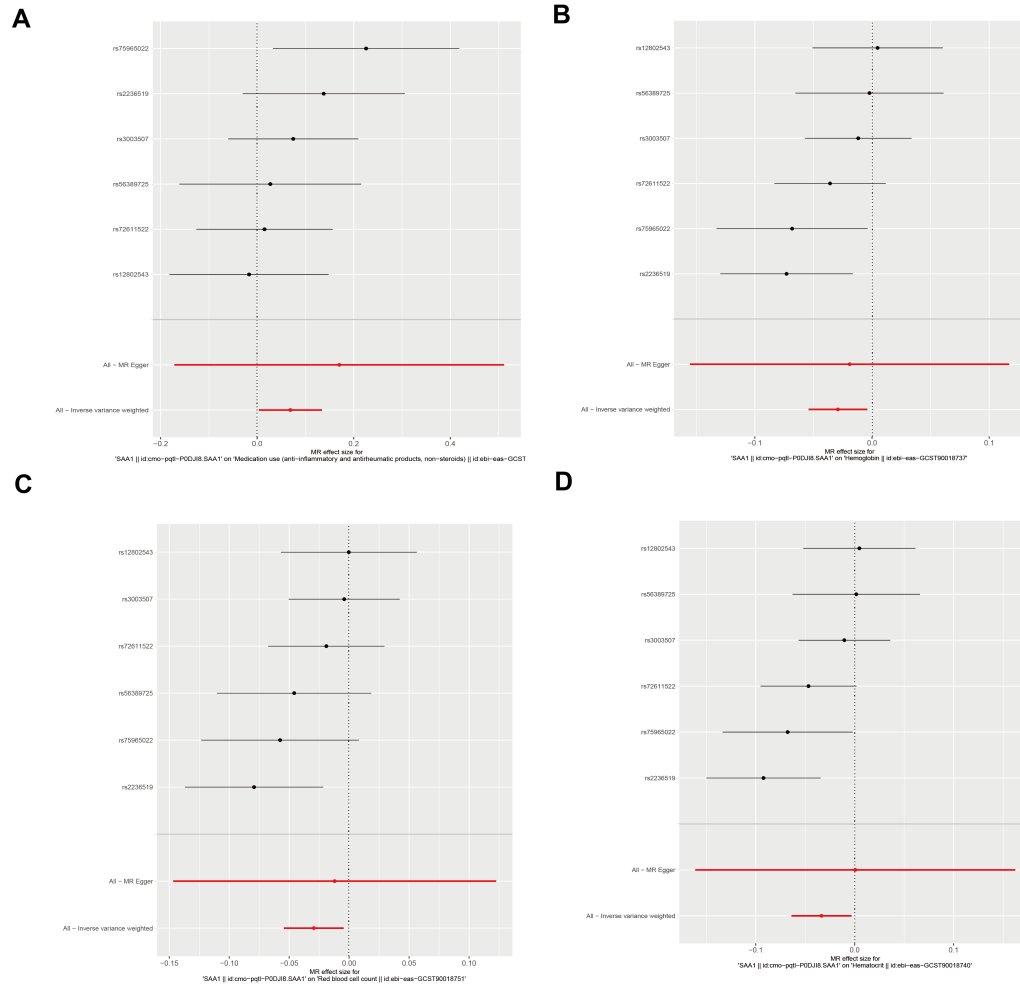


Figure 4: Mendelian randomization analysis of SAA1's causal effects on SLE flare risk factors. Forest plots from random-effects inverse-variance weighted (IVW) analyses depicting the genetic causal relationships between SAA1 and (A) anti-inflammatory and antirheumatic medication use, (B) hemoglobin levels, (C) red blood cell count, and (D) hematocrit. Odds ratios and 95% confidence intervals are shown.

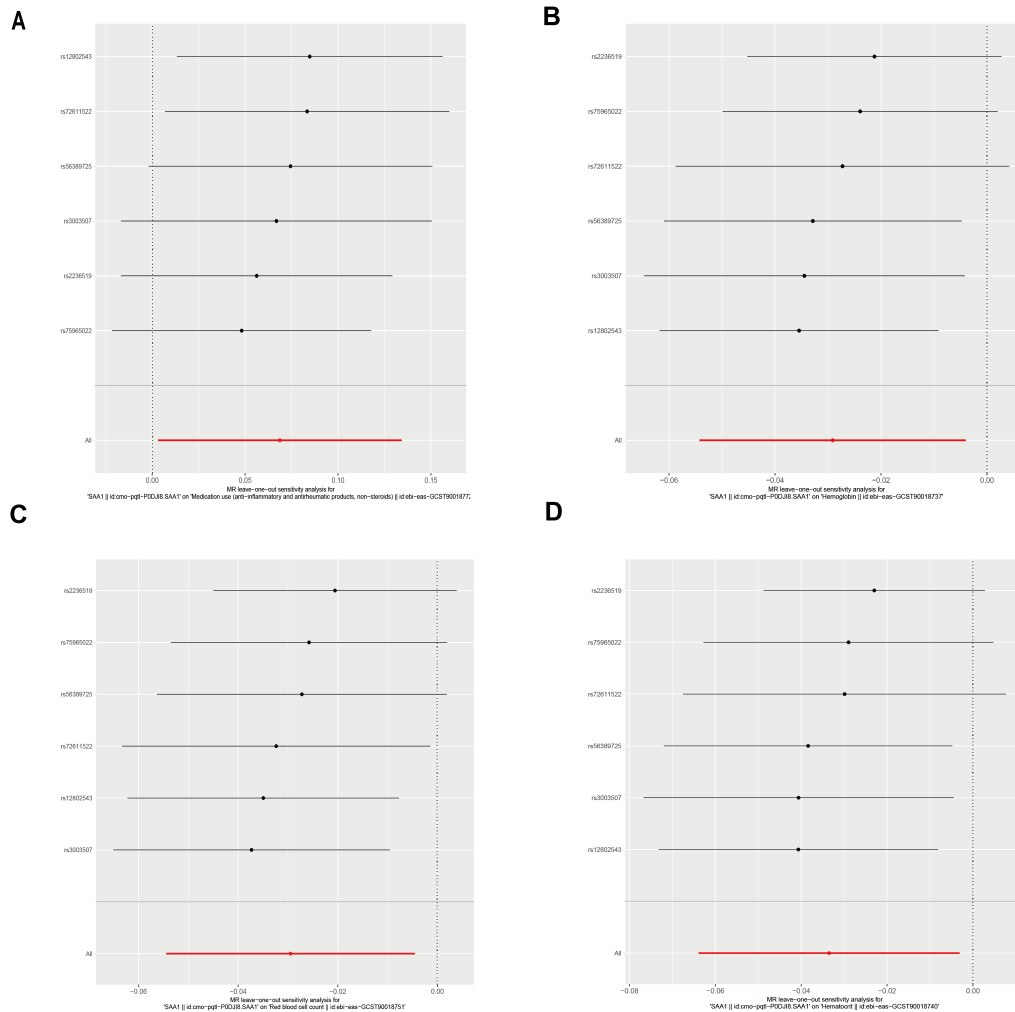


Figure 5: Sensitivity analysis of Mendelian randomization results using the leave-one-out method. (A) Anti-inflammatory and antirheumatic medication use. (B) Hemoglobin levels. (C) Red blood cell count. (D) Hematocrit. This analysis evaluates the robustness of the causal relationships between SAA1 and four key outcomes: anti-inflammatory and antirheumatic medication use, hemoglobin levels, red blood cell count, and hematocrit. Each subplot (A-D) represents one of these outcomes, with individual points showing the Mendelian randomization estimate after excluding one SNP from the analysis, while the vertical line indicates the estimate using all SNPs. This approach helps identify potential outlier SNPs that may disproportionately influence the overall results.

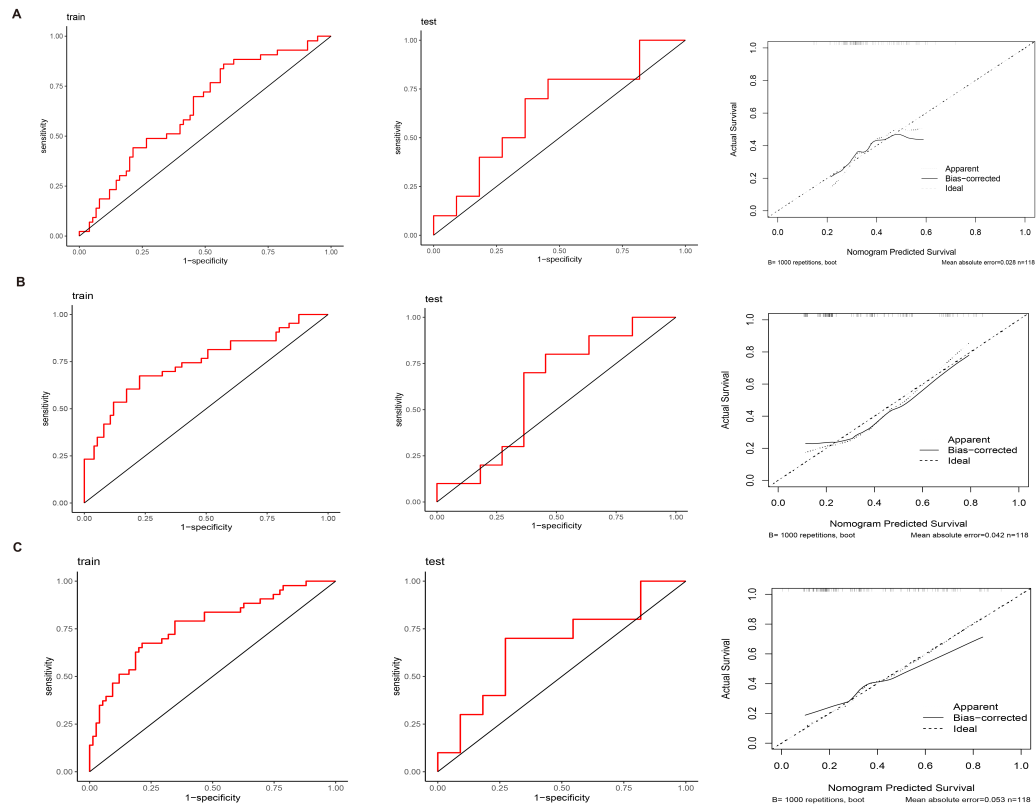


Figure 6: Comparative performance evaluation of SLE flare risk prediction models. (A-C) Receiver Operating Characteristic (ROC) curves (left) and calibration plots (right) for (A) clinical, (B) proteomic, and (C) combined prediction models. ROC curves display sensitivity versus 1-specificity, with Area Under the Curve (AUC) values indicated. Calibration plots show predicted versus observed probabilities, generated using 1,000 bootstrap resamples. Hosmer-Lemeshow goodness-of-fit test P-values are provided to evaluate model calibration.

346 **Supplementary Tables**

347 *1. Name of multi-center hospitals*

Table S1. Name of multi-center hospitals
The Second Affiliated Hospital of Zhejiang Chinese Medical University
Guang'anmen Hospital China Academy of Chinese Medical Sciences
Chinese People's Liberation Army General Hospital
The First Affiliated Hospital of Guangzhou University of Chinese Medicine
Shenzhen Futian Hospital for Rheumatic Diseases
The First Affiliated Hospital of Henan University of CM
First Teaching Hospital of Tianjin University of Traditional Chinese Medicine
The First School of Clinical Medicine Yunnan University of Chinese Medicine
Department of Rheumatology Second Affiliated Hospital of Zhejiang University
The Affiliated Hospital of Medical School Ningbo University
The Second Affiliated Hospital of Jiaxing University
Department of Rheumatology Huzhou Hospital of Traditional Chinese Medicine Affiliated to Zhejiang Chinese Medical University
Shaoxing Hospital of Traditional Chinese Medicine, Shaoxing TCM Hospital Affiliated to Zhejiang Chinese Medical University

348 *Acknowledgments*

349 We thank members of the Genomics Platform (Human Phenome Institute,
350 Fudan University) for library preparation experiments and providing the
351 proteomics analysis. We also appreciate the medical background advice
352 provided by Professor Atsushi Ogiwara of Waseda University.

353 *Authors' Contributions*

354 LC conceptualized the study, developed the methodology, conducted formal
355 analysis, and led the writing of the manuscript. OD contributed to the
356 development of the methodology, analysis, supervision, and revision of the
357 manuscript. TF, MC, and XZ were involved in data collection, curation, and
358 validation. RC, DL, and RZ provided assistance in manuscript preparation
359 and revision. QJ contributed to conceptualization, supervision, and project
360 administration. XW oversaw the project, providing leadership in conceptual-
361 ization, supervision, funding acquisition, and manuscript review. All authors
362 have reviewed and approved the final version of the manuscript.

363 *Funding*

364 This research was supported by the Zhejiang Traditional Chinese Medical
365 Modernization Special Project (grant number 2020ZX008).

366 *Data Availability*

367 The mass spectrometry proteomics data generated in this study have been
368 deposited in the Integrated Proteome Resources (iProX) database under the
369 dataset identifier IPX0007978000. The data are publicly accessible through
370 the iProX website (<https://www.iprox.cn/page/>). Additional datasets that
371 support the findings of this study are available from the corresponding author
372 upon reasonable request.

373 **Declarations**

374 *Ethics approval and consent to participate*

375 The study was approved by the ethics committee of the Second Affiliated 58
376 Hospital of Zhejiang Chinese Medical University (approval NO.2020-KL-002-
377 59 IH01). Informed consent was obtained from all participants.

378 *Consent for publication*

379 Not applicable. Our manuscript does not contain any individual person's
380 data.

381 *Competing Interests*

382 The authors declare no competing interests.

383 References

- 384 [1] Guidelines for referral and management of systemic lupus erythematosus
385 in adults. American College of Rheumatology Ad Hoc Committee on
386 Systemic Lupus Erythematosus Guidelines. *Arthritis and rheumatism*.
387 1999;42(9):1785-96.
- 388 [2] Fanouriakis A, Kostopoulou M, Alunno A, Aringer M, Bajema I, Boletis
389 JN, et al. 2019 update of the EULAR recommendations for the manage-
390 ment of systemic lupus erythematosus. *Annals of the rheumatic diseases*.
391 2019;78(6):736-45.
- 392 [3] Barturen G, Babaei S, Català-Moll F, Martínez-Bueno M, Makowska
393 Z, Martorell-Marugán J, et al. Integrative Analysis Reveals a Molecular
394 Stratification of Systemic Autoimmune Diseases. *Arthritis & rheumatol-
395 ogy (Hoboken, NJ)*. 2021;73(6):1073-85.
- 396 [4] Goetz I, Choong C, Winnie J, Nelson DR, Birt J, Noxon V, et al. Develop-
397 ment of a claims-based flare algorithm for systemic lupus erythematosus.
398 *Current medical research and opinion*. 2022;38(9):1641-9.
- 399 [5] Cho J, Lahiri M, Teoh LK, Dhanasekaran P, Cheung PP, Lateef A.
400 Predicting flares in patients with stable systemic lupus erythematosus.
401 *Seminars in arthritis and rheumatism*. 2019;49(1):91-7.
- 402 [6] Cho J, Liang S, Lim SHH, Lateef A, Tay SH, Mak A. Neutrophil to
403 lymphocyte ratio and platelet to lymphocyte ratio reflect disease activity
404 and flares in patients with systemic lupus erythematosus-A prospective
405 study. *Joint bone spine*. 2022;89(4):105342.
- 406 [7] Sun F, Zhao L, Wang H, Zhang D, Chen J, Wang X, et al. Risk factors
407 of disease flares in a Chinese lupus cohort with low-grade disease activity.
408 *Lupus science & medicine*. 2022;9(1).
- 409 [8] Goldschen L, Ellrodt J, Amonoo HL, Feldman CH, Case SM, Koenen
410 KC, et al. The link between post-traumatic stress disorder and systemic
411 lupus erythematosus. *Brain, behavior, and immunity*. 2023;108:292-301.
- 412 [9] Yelin E, Trupin L, Bunde J, Yazdany J. Poverty, Neighborhoods, Per-
413 sistent Stress, and Systemic Lupus Erythematosus Outcomes: A Qual-
414 itative Study of the Patients' Perspective. *Arthritis care & research*.
415 2019;71(3):398-405.

- 416 [10] Katz P, Wan GJ, Daly P, Topf L, Connolly-Strong E, Bostic R, et al.
417 Patient-reported flare frequency is associated with diminished quality of
418 life and family role functioning in systemic lupus erythematosus. *Quality*
419 *of life research : an international journal of quality of life aspects of*
420 *treatment, care and rehabilitation*. 2020;29(12):3251-61.
- 421 [11] Yuan S, Jiang F, Chen J, Lebwohl B, Green PHR, Leffler D, et al.
422 Phenome-wide Mendelian randomization analysis reveals multiple health
423 comorbidities of coeliac disease. *EBioMedicine*. 2024;101:105033.
- 424 [12] Yuan S, Wang L, Zhang H, Xu F, Zhou X, Yu L, et al. Mendelian ran-
425 domization and clinical trial evidence supports TYK2 inhibition as a ther-
426 apeutic target for autoimmune diseases. *EBioMedicine*. 2023;89:104488.
- 427 [13] Shen X, Kellogg R, Panyard DJ, Bararpour N, Castillo KE, Lee-McMullen
428 B, et al. Multi-omics microsampling for the profiling of lifestyle-associated
429 changes in health. *Nature biomedical engineering*. 2024;8(1):11-29.
- 430 [14] Petri M, Orbai AM, Alarcón GS, Gordon C, Merrill JT, Fortin PR, et al.
431 Derivation and validation of the Systemic Lupus International Collab-
432 orating Clinics classification criteria for systemic lupus erythematosus.
433 *Arthritis and rheumatism*. 2012;64(8):2677-86.
- 434 [15] Fanouriakis A, Tziolos N, Bertsias G, Boumpas DT. Update on the
435 diagnosis and management of systemic lupus erythematosus. *Annals of*
436 *the rheumatic diseases*. 2021;80(1):14-25.
- 437 [16] Xu F, Yu EY, Cai X, Yue L, Jing LP, Liang X, et al. Genome-wide
438 genotype-serum proteome mapping provides insights into the cross-
439 ancestry differences in cardiometabolic disease susceptibility. *Nature*
440 *communications*. 2023;14(1):896.
- 441 [17] Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshihara S,
442 et al. A cross-population atlas of genetic associations for 220 human
443 phenotypes. *Nature genetics*. 2021;53(10):1415-24.
- 444 [18] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting
445 of a multivariable prediction model for individual prognosis or diagno-
446 sis (TRIPOD): the TRIPOD statement. *BMJ (Clinical research ed)*.
447 2015;350:g7594.

- 448 [19] Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, et
449 al. Phenome-wide Mendelian randomization mapping the influence of the
450 plasma proteome on complex diseases. *Nature genetics*. 2020;52(10):1122-
451 31.
- 452 [20] Moroni L, Mazzetti M, Ramirez GA, Farina N, Bozzolo EP, Guerrieri S,
453 et al. Beyond Neuropsychiatric Manifestations of Systemic Lupus Ery-
454 thematosus: Focus on Post-traumatic Stress Disorder and Alexithymia.
455 *Current rheumatology reports*. 2021;23(7):52.
- 456 [21] Rosenau BJ, Schur PH. Antibody to serum amyloid A. *Journal of au-
457 toimmunity*. 2004;23(2):179-82.
- 458 [22] Márquez-Macedo SE, Perez-Arias AA, Pena-Vizcarra Ó R, Zavala-
459 Miranda MF, Juárez-Cuevas B, Navarro-Gerrard MA, et al. Predictors of
460 treatment outcomes in lupus nephritis with severe acute kidney injury and
461 requirement of dialytic support. *Clinical rheumatology*. 2023;42(8):2115-
462 23.
- 463 [23] He J, Tang D, Liu D, Hong X, Ma C, Zheng F, et al. Serum proteome
464 and metabolome uncover novel biomarkers for the assessment of dis-
465 ease activity and diagnosing of systemic lupus erythematosus. *Clinical
466 immunology (Orlando, Fla)*. 2023;251:109330.
- 467 [24] Lee JY, Hall JA, Kroehling L, Wu L, Najjar T, Nguyen HH, et al.
468 Serum Amyloid A Proteins Induce Pathogenic Th17 Cells and Promote
469 Inflammatory Disease. *Cell*. 2020;180(1):79-91.e16.
- 470 [25] Lu X, Huang Y, Zhou M, Guo Y, Zhou Y, Wang R, et al. Artesunate
471 attenuates serum amyloid A-induced M1 macrophage differentiation
472 through the promotion of PHGDH. *International immunopharmacology*.
473 2024;127:111462.
- 474 [26] Zhou G, Wei P, Lan J, He Q, Guo F, Guo Y, et al. TMT-based quanti-
475 tative proteomics analysis and potential serum protein biomarkers for
476 systemic lupus erythematosus. *Clinica chimica acta; international journal
477 of clinical chemistry*. 2022;534:43-9.
- 478 [27] Luo KL, Yang YH, Lin YT, Hu YC, Yu HH, Wang LC, et al. Differential
479 parameters between activity flare and acute infection in pediatric patients
480 with systemic lupus erythematosus. *Scientific reports*. 2020;10(1):19913.

- 481 [28] Yoshihara T, Satake H, Nishie T, Okino N, Hatta T, Otani H, et al.
482 Lactosylceramide synthases encoded by B4galt5 and 6 genes are pivotal
483 for neuronal generation and myelin formation in mice. *PLoS genetics*.
484 2018;14(8):e1007545.
- 485 [29] Zhang L, Ren J, Shi P, Lu D, Zhao C, Su Y, et al. The Immunological
486 Regulation Roles of Porcine β -1, 4 Galactosyltransferase V (B4GALT5)
487 in PRRSV Infection. *Frontiers in cellular and infection microbiology*.
488 2018;8:48.
- 489 [30] van Gastel J, Boddaert J, Jushaj A, Premont RT, Luttrell LM, Janssens
490 J, et al. GIT2-A keystone in ageing and age-related disease. *Ageing
491 research reviews*. 2018;43:46-63.
- 492 [31] Stessman HA, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M, et
493 al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk
494 genes with autism and developmental-disability biases. *Nature genetics*.
495 2017;49(4):515-26.
- 496 [32] Cheng H, Dharmadhikari AV, Varland S, Ma N, Domingo D, Kleyner R,
497 et al. Truncating Variants in RPIA Are Associated with Variable Levels
498 of Intellectual Disability, Autism Spectrum Disorder, and Congenital
499 Anomalies. *American journal of human genetics*. 2018;102(5):985-94.
- 500 [33] Qun L, Wenda X, Weihong S, Jianyang M, Wei C, Fangzhou L, et al.
501 miRNA-27b modulates endothelial cell angiogenesis by directly targeting
502 Naa15 in atherogenesis. *Atherosclerosis*. 2016;254:184-92.
- 503 [34] Ugarte-Gil MF, Alarcón GS. Low Disease Activity Early in the Course
504 of Systemic Lupus Erythematosus. *The Journal of rheumatology*. 2024.
- 505 [35] Ji L, Xie W, Fasano S, Zhang Z. Risk factors of flare in patients with sys-
506 temic lupus erythematosus after glucocorticoids withdrawal. A systematic
507 review and meta-analysis. *Lupus science & medicine*. 2022;9(1).
- 508 [36] Nakai T, Fukui S, Ikeda Y, Suda M, Tamaki H, Okada M. Glucocorticoid
509 discontinuation in patients with SLE with prior severe organ involve-
510 ment: a single-center retrospective analysis. *Lupus science & medicine*.
511 2022;9(1).