REVIEW

Rheumatology & Autoimmunity

Understand SLE heterogeneity in the era of omics, big data, and artificial intelligence

Prianka Puri[1,2] | Simon H. Jiang[3,4,5] | Yang Yang[2] | Fabienne Mackay[6] | Di Yu[2]

[1]Kidney Health Service, Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia

[2]The University of Queensland Diamantina Institute, Faculty of Medicine, The University of Queensland, Woolloongabba, Queensland, Australia

[3]Department of Immunology and Infectious Disease, John Curtin School of Medical Research, Acton, Australian Capital Territory, Australia

[4]Centre for Personalised Immunology, NHMRC Centre for Research Excellence, Acton, Australian Capital Territory, Australia

[5]Department of Renal Medicine, The Canberra Hospital, Garran, Australian Capital Territory, Australia

[6]QIMR Berghofer Medical Research Institute in Brisbane QLD, Herston, Queensland, Australia

Correspondence
Di Yu, The University of Queensland Diamantina Institute, Faculty of Medicine, The University of Queensland, Woolloongabba, QLD 4102, Australia.
Email: di.yu@uq.edu.au

Funding information
Bellberry Limited and The Viertel Charitable Foundation, Grant/Award Number: Bellberry Limited and The Viertel Charitable Foundation

Abstract

Systemic lupus erythematosus (SLE) is a systemic autoimmune disease characterized by extraordinary heterogeneity, due to the complex pathogenesis and diverse manifestations. Stratification of patients for therapy and prognosis represents a major challenge to manage SLE. Conventional biomarkers for disease diagnosis and activity assessment provide very limited insight into immunological pathogenesis and therapeutic response rates. The advancement of "omics" technologies including genomics, transcriptomics, proteomics, and metabolomics has constituted an unprecedented opportunity to characterize the immunopathological landscape in individual patients with SLE. Indeed, genomic studies reveal a subset of SLE patients carrying one or more functional single nucleotide polymorphisms (SNPs) underlying immune dysregulation while transcriptomic studies have revealed subgroups in SLE patients showing distinct signatures for Type I interferon (TI-IFN) pathway activation or aberrant differentiation of B cells into plasma cells. This review will summarize results from the latest studies using omics technology to understand SLE heterogeneity. In addition, we propose that the application of artificial intelligence, such as by machine learning-based nonlinear dimensionality reduction method uniform manifold approximation and projection (UMAP) can further strengthen the analysis of omics big data. The combination of new technology and novel analysis pipeline can lead to breakthroughs in stratifying SLE patients for a better monitoring of disease activity and more precise design of treatment regime, not only for conventional immunosuppression but also novel immunotherapies targeting B-cell activating factor (BAFF), TI-IFN, and interleukin 2 (IL-2).

KEYWORDS
classification, genomics, metabolomics, proteomics, transcriptomics

1 | SLE IS CHARACTERIZED BY EXTRAORDINARY CLINICAL DIVERSITY AND DRIVEN BY MULTIPLE PATHOGENIC PATHWAYS

Systemic lupus erythematosus (SLE, or lupus) is the archetypal multisystem autoimmune disease affecting 20–150 people per 100,000, with female predominance (7-9:1).[1] It is characterized by extraordinary clinical diversity, wherein patients develop different organ manifestations with variable disease activities (DA) and individuals show distinct responses to specific immunosuppressive or immunomodulatory therapies.[2] Such clinical and biological diversity of SLE leads to the argument that SLE as a single disease can be fundamentally flawed.[3] The heterogeneous nature of SLE is itself a reflection of complex and diverse pathogenesis.

The pathogenesis of SLE involves the hyperactivation in both innate and adaptive arms of the immune system, which eventually leads to loss of immune tolerance to autoantigens.[1,4,5] The impaired clearance of DNA or RNA from dead cells including neutrophil extracellular traps (NETs) can stimulate plasmacytoid dendritic cells (pDCs) to produce type I interferon (TI-IFN), which in turn enhances

antigen presentation and promotes both innate and adaptive responses.[6,7] In SLE, there is an expansion of activated B cells with aberrant differentiation in plasma cells to produce autoantibodies,[8] which is supported by excessive B-helper function from CD4[+] T cells.[9] CD4[+] T cells also show enhanced production of proinflammatory cytokines such as interleukin 17 (IL-17) and impaired function for the regulatory T (Treg) subset.[10] CD8[+] T cells in SLE also contribute to organ damage with their activity shown to predict prognosis.[11] The end-organ damage in SLE is mediated by proinflammatory innate and adaptive immune cells as well as the pathogenic antigen-autoantibody immune complex that activate FcγR and complement pathways[7] (Figure 1).

The complex and diverse pathways for SLE pathogenesis are clearly demonstrated by different mouse models for lupus, which are individually driven by distinct pathways of immune dysregulation. Impaired uptake of apoptotic cells in *MFG-E8*[−/−] mice causes splenomegaly and spontaneous formation of germinal centers, the production of autoantibody and the development of glomerulonephritis.[12] The mice with the D18N mutation in *Trex1*, a major cytoplasmic exonuclease that degrades dsDNA and ssDNA, exhibit lupus-like symptoms including systemic inflammation, lymphoid hyperplasia, vasculitis, and kidney disease.[13] Overexpression of B-cell activating factor (BAFF) in mice models results in SLE-like disease including production of autoantibodies and the development of glomerulonephritis.[8,14,15] Aberrant differentiation of follicular helper T (Tfh) cells leads to autoreactive germinal center B-cell formation and autoantibody production in Roquin[san/san] mice.[16–19] Although an individual immune dysregulation is sufficient to drive lupus-like disease mice, the question is whether SLE patients could be divided into subgroups featured by distinct types of immune signatures, which could help to tailor patients with specific immunotherapies.

2 | CONVENTIONAL CRITERIA FOR SLE CLASSIFICATION

Given the heterogeneity in SLE, the current international consensus for diagnosis and classification of SLE is based on a series of clinical, biochemical, and immunological parameters.[20] These criteria are designed to diagnose disease based on an additive score but do not predict disease progression, severity, or risk of relapses. Earlier criteria by the Systemic Lupus International Collaborating Clinics (SLICC) addressed issues with initial diagnostic criteria developed by the American College of Rheumatology (ACR) which omitted features such a mucocutaneous, neuropsychiatric disease, and immunological parameters.[20,21] SLICC criteria considered biopsy-proven lupus nephritis (LN) along with a lupus antibody positivity (antinuclear antibodies [ANAs], double-stranded DNA [dsDNA], and anti-smith antibodies [anti-sm]) sufficient for the diagnosis of SLE.[21] More recently European League Against Rheumatism (EULAR) and the ACR have developed a classification criterion, to create a more unbiased patient-specific approach compared with SLICC, to encompass an extended list of clinical and immunological features.[20,21] The EULAR/ACR criteria have a weighted additive score >10, from either clinical or immunological features and a screening anti-nuclear antibody (ANA) test being positive (Figure 2). There is some data to suggest that EULAR/ACR baseline score >20 may indicate more severe disease at 5 years.[23] However, others believe 5 years is not sufficient to provide diseases accrual, and this classification may miss early organ-specific diseases.[20–22] An unbiased patient-specific approach is required to improve early detection, disease prognosis, and patient-specific therapy.
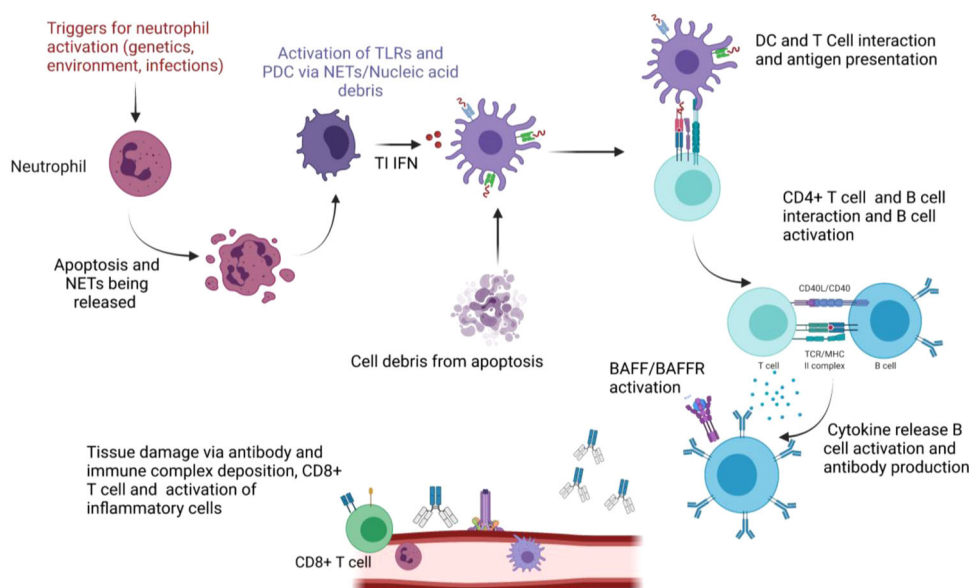

FIGURE 1 Simplified depiction of Immunopathogenesis of SLE. Inciting events or triggers in predisposed individuals leads to poor clarences of nucleic acids and increased neutrophil extracellular traps (NETs). Activation of plasmacytoid dendritic cells (pDCs) leads to Type I IFN (TI-IFN) production and activation of dendritic cells (DCs). Primed DCs present antigens and stimulate T-cell activation and differentiation. CD4[+] T cells promote B-cell activation and differentiation into plasma cells for autoantibody production. Excessive B-cell activating factor (BAFF) support self-reactive B-cell survival and function. Proinflammatory immune cells, autoantigen-autoantibody immune complex (IC), and activated complements are recruited into organs to cause tissue damages. SLE, systemic lupus erythematosus
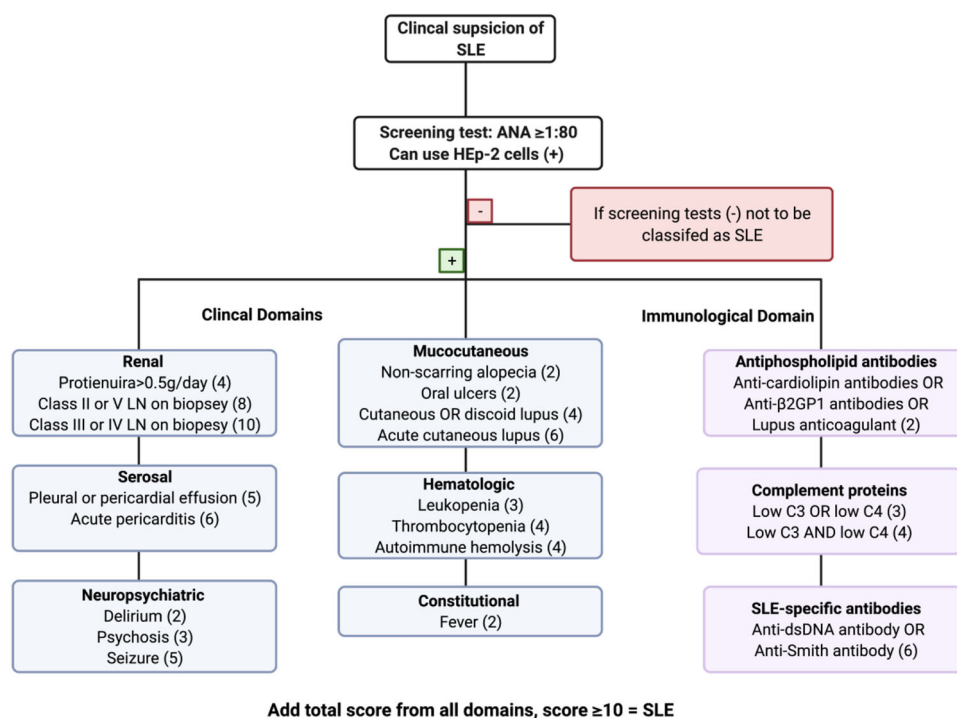
FIGURE 2 Conventional clinical features and immunological biomarkers to classify SLE patients, adapted from American College of Rheumatology/European Alliance of Associations for Rheumatology classification criteria 2019.[22] SLE, systemic lupus erythematosus

A critical complication of SLE is LN. LN affects 29% of Caucasian and up to 80% of Asian patients and is associated with substantial adverse outcomes including mortality and progressive kidney disease.[5] Up to 10% of patients develop kidney failure at 5 years.[1,5] The relapsing and remitting nature of LN, compounded by limited noninvasive biomarkers of DA (such as proteinuria and complement levels), makes it difficult to treat. Consequently, certain patients are left with low levels of DA while others are exposed to the adverse effects of potent immunosuppression.[5] Immunosuppression currently used for LN has significant side effects, and patient-specific factors including tolerance of medications, cost, and compliance make treatment a further challenge.[24]

LN can be suspected in patients with biochemical changes in kidney function, onset of hematuria, proteinuria, and autoantibody positivity. However, definitive diagnosis still requires an invasive kidney biopsy which carries the risk of complications (minor complication rates reported being 8.1%–15%, and major complications rates between 1.5% and 6.6%).[5,25–27] LN is classified into six subtypes based on a broad spectrum of clinicopathologic features (Table 1).[28] The histological grading of the disease provides insights into the activity and chronicity of kidney damage to guide the degree of immunosuppression (with Class I LN being an early disease not requiring aggressive immunosuppression, and Class VI being advanced diseases with significant fibrosis).[5,25] This classification is often insufficient in predicting prognosis and guiding therapy given the variable response to treatment.[5] Ineffective therapies to treat LN in randomized controlled trials are shown to be effective in individual cases.[29–31]

There is an unmet need for further research into SLE and LN stratification by assessing these genetic and immunological changes in patients with the disease. A noninvasive precision medicine approach is required to not only mitigate biopsy-related risk but also to provide patient-specific therapy.

3 | THE APPLICATION OF "OMICS" TECHNOLOGIES AT THE FRONTLINE OF PRECISION MEDICINE FOR SLE

With an increasing understanding of the pathogenesis of SLE by internationally recognized immunologists and clinical scientists, new immunotherapies for SLE have been developed to suppress the key pathway for immune activation by blocking TI-IFN,[32] to reduce excessive B-cell survival and activation by blocking BAFF[31,33] and to reinstate the balance between regulatory and effector T cells by low-dose IL-2.[34–36] It is now more important than ever to classify SLE patients based on systemic immune signatures. This new classification will not only help predict prognosis but essentially guide therapy for precision medicine to improve outcomes. Furthermore, identification of potential biomarkers and immunological pathways associated with specific classification may provide new targets for future therapy development, for overall or specific SLE subgroups. By applying cutting-edge "omics" technologies and artificial intelligence-based analysis, a "systems immunology" approach has been pioneered in stratifying patients with SLE (Figure 3). We will summarize representative studies of this kind using genomics, transcriptomics, metabolomics, and proteomics.

TABLE 1 Biopsy features of Lupus nephritis adapted from the Insertional society of nephrology and Renal pathology society classifications 2018.[28]

LN biopsy class	Features	Modified National Institute of Health activity and chronicity index
LN class I	Minimal changes on light microscopy. Mesangial deposition of IC by IF and EM	**Activity score** Endocapillary Hypercellularity (0–3) Neutrophil infiltration (0–3) Fibrinoid necrosis (0–3) x2 Hyaline deposits (0–3) Cellular or fibrocellular Crescents (0–3) x2 Interstitial inflammation (0-3) *Score total 0–24* **Score** 0. absent 1. <25% glomeruli/cortex or interstium effected 2. 25–50% effected 3. >50% effected
LN class II	Mesangial proliferative LN (any degree of hypercellularity >3 mesangial cells per 3-µm-thick section) on LM. Isolated subepithelial or subendothelial deposits (IF/EM)	
LN class III	Focal LN < 50% glomeruli effected by LM. Active or inactive (endocapillary or extracapillary hypercellularity).	
LN class IV	Diffuse LN > 50% of all glomeruli effect on LM. May have diffuse wireloops and subendothelial deposits	
LN class V	Lupus membranous nephropathy, diffuse thickening of capillary walls on LM. Diffuse subepithelial deposits on EM.	**Chronicity score** Glomerulosclerosis score (0–3) Fibrous crescents (0–3) Tubular atrophy (0–3) Interstitial fibrosis (0–3) *Score total 0–12*
LN class VI	Advanced sclerosing glomerulonephritis, with >90% of glomeruli sclerosed on LM.	

Abbreviations: EM, electron microscopy; IC immune complex; IF, immunofluorescence; LN, lupus nephritis; LM, light miscopy.
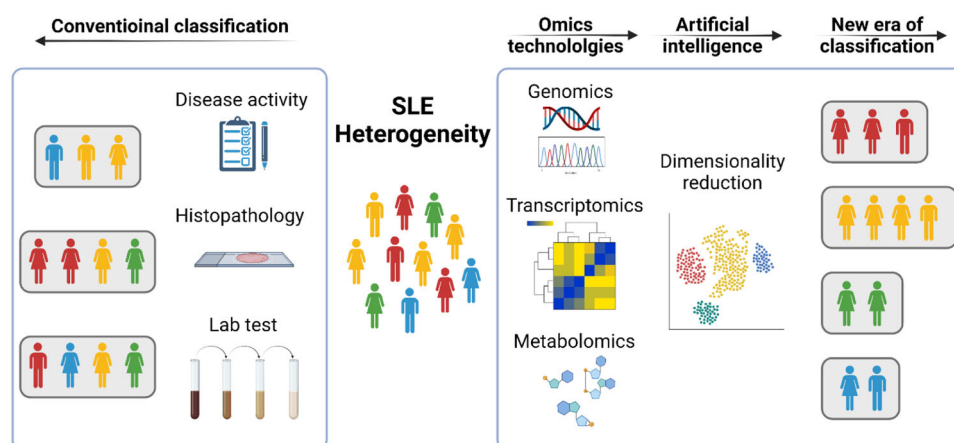
FIGURE 3 A proposed new classification of SLE patients by omics data and artificial intelligence. Compared with the conventional classification of SLE patients by disease activities, histopathological features, and lab test results, the integration of multiomics data by artificial intelligence, such as machine learning-based dimensionality reduction can more efficiently classify SLE patients for precision medicine. SLE, systemic lupus erythematosus

3.1 | Genomics

Genomics are recognized as one of the most potent risk factors underpinning the development of SLE. Twin concordance studies suggest a range from 25% to 54% of genetically identical twins will develop SLE compared with 2%–5% of genetically distinct twins.[37] Symptom heterogeneity exists among these twins and in part is thought to be related to epigenetic modifications of the genome.[38,39] This recognition accelerated genetic studies in SLE patients, with almost 100 loci now associated with SLE through genome-wide association studies (GWAS).[40–44] Although gene loci from different studies differed between cohorts due to distinct ethnicities and sample sizes, common major regulatory pathways and mechanisms in SLE pathogenesis have been elucidated by such studies.[41,45,46] More recently, whole exome or genome sequencing has discovered one or more rare or novel single nucleotide variants (SNVs) contributing to the development of SLE, identifying unique monogenic mutations contributing to heterogeneous phenotypes in SLE patients[46] (Table 2).

One major advance from genome studies is the recognition of common immune pathways implicated in disease. These loci include genes with functions in DNA/RNA metabolism, classic complement pathway regulation, IFN regulation, and T- and B-cell function. Indeed, the risk of developing SLE attributed to these genes tends to reflect the biological pathway involved. As an example, genes involved in RNA or DNA metabolism often cause Mendelian SLE, whereas genes involved in IFN regulation increase risk by a relatively modest 20%–40%. Recently, studies have identified the mechanisms through which these loci contribute to SLE with SNVs in GWAS-associated genes disrupting regulation of TI-IFN resulting in enhanced TI-IFN

TABLE 2 Representative SLE genomics studies

Studies	Patient criteria & Sample size	Profiling methods	Analytic methods	Major discoveries in stratification
Harley et al.[41]	European female SLE, $N = 720$ HC $N = 2337$ Extra data set SLE $N = 1846$ HC $N = 1825$	GWAS of SNPs.	PCA and multiple logistic regression model	Multiple genetic variants implicated in SLE development in European women were identified. These included variants in complement related genes, ITGAM, IFN alpha gene IRF7, coding DNA elongation factor KIAA1542 and DNA splicing gene PXK. Enrichment of these SNVs are thought to contribute to SLE pathogenesis.
Kozyrev et al.[40]	SLE $N = 279$ HC $N = 515$	Genotyping using 100 K Affymetrix array plus peripheral blood mononuclear cell (PBMC) RT-PCR,	Logistic regression analysis and conditional multiple logistic regression analysis	Associations between B-cell scaffold protein with ankyrin repeats gene, (BANK1) mutations, and SLE pathogenesis were seen. A nonsynonymous substitution and alternative splicing in the IP3R regions may lead to increase affinity binding of BANK1 and overactivation in SLE patients leading to disease.
Almlof et al.[45]	SLE $N = 71$ HC $N = 142$	Whole-genome sequencing (WGS) of parent-offspring trios	Random forest (RF)	RF scores allow for the prediction of rare SNPs causing variants in 22 monogenic genes contributing to SLE. These variants were thought to be inherited from a single parent, with RF scores grouping parents into high versus low risk based on WGS data.
Jiang et al.[42]	SLE $N = 69$ (SLE1) replication cohort of SLE 2 $N = 64$ HC, $N = 97$	WGS and WES with GWAS undertaken	ADMIXTURE algorithm	Patients were divided into three groups HC, SLE1, and SLE2. Using a list of 76 GWAS SLE genes groups based on the proportion of rare SNVs noted HC $<$SLE1 $<$SLE2. A substantial number of SLE patients harbor multiple SNV variants in BANK1 and/or BLK. SNV found in BLK1 impair kinase function, related to suppression IFN production. Loss of function variants in BANK1 impair scaffold protein function leading to T1IFN activity. These SNV may account for disease development and IFN production in carriers with SLE.

Abbreviations: GWAS, genome-wide association studies; IFN, interferon; PCA, principal component analysis; RT-PCR, real-time polymerase chain reaction; SNV, single nucleotide variant; SLE, systemic lupus erythematosus.

production.[42] It is also increasingly apparent that many SNVs that contribute to SLE are under significant selection pressure with enrichment of pathogenic variants among rare and novel SNVs. Monogenic mutations such as hereditary C1q deficiency have been implicated in SLE development along with mutations in other complement proteins[45] and rare mutations in the TI-IFN associated Toll-like receptor (TLR) pathway being enriched in SLE patients.[47] GWAS utilizing WES have permitted greater insight into these low-frequency mutations' association with SLE development (Table 2). While the heterogeneity of mutations across distinct arms of the immune system in SLE make treatment challenging in the absence of knowing patient-specific genes and pathways, it also represents an opportunity to individualize therapy where patient-specific signatures are known. Indeed, it has been suggested that certain GWAS loci are associated with specific unique transcriptomic signatures.[48] With further machine learning analysis, genomics can be utilized as a tool to stratify patients with SLE and provide insights into disease pathogenesis.

3.2 | Transcriptomics

Although there is a high genetic contribution to SLE, non-genetic risk factors are also important, which can be better examined by transcriptomics. Transcriptional changes often precede clinical manifestations and may provide patient stratification.[11,48–50] These studies demonstrate that the variations in IFN gene signatures (IGS) and the dysregulation of innate immune cells such as neutrophils or adaptive immune cells such as plasma cells or CD8[+] T cells might stratify patients into different groups (Table 3).

Current transcriptional and genomic data highlight a strong IGS elevation in SLE.[48–50,52,54–56] These studies suggest discordant results surround IGS expression with DA, however, all data suggest TI-IFN is involved in the pathogenesis of SLE. Whole exome sequencing (WES) has highlighted various SVNs in TI-IFN genes found in SLE patients.[42,46] Further analysis of the pediatric population has highlighted different cell types correlate to disease activities based on subgroup stratification of disease among patients, although the relationship between subgroup stratification and conventional classification has not been investigated in detail.[48] In this article, a subgroup analysis of disease activities based on response to drug therapy revealed Cyclophosphamide had a greater suppressive effect on DA modules as compared with steroids, however, steroids were more efficacious than Plaquenil in DA module suppression. Other studies have suggested that IGS are unreliable as a marker of activity due to the monocyte retention during quiesces and flare.[57] Given the potent correlation between transcriptomics and DA, recent randomized controlled trials (RCTs) of anti-interferon therapy show greater efficacy in subgroups with pronounced TI-IFN signatures suggesting targeted anti-interferon therapy may be more efficacious for this subset of patients.[27] Using transcriptional gene panels which correlate to disease could be considered as a cost-effective and noninvasive biomarker of SLE/LN activity.

Transcriptional profiling of both kidney tissue and urine samples may further provide insight into disease pathogenesis severity and response.[53,58–60] Der et al. undertook single-cell RNA-sequencing (scRNA-seq) of renal tubular cells and skin keratinocytes. The found nine major cell clusters, differentially expressed gene (DEG) enrichment was noted for TI-IFN signatures. TI-IFN response scores correlated with renal disease treatment response. Each subclass of LN itself had the differential expression of inflammatory pathways. Patients with Class III/IV LN had higher TI-IFN gene signatures than patients with Class V LN.[58] Arazi et al. undertook the transcriptomic assessment of kidney tissue and urine samples. They found a high correlation of urine immune cells corresponding to infiltrating leukocytes seen in kidney tissue. This suggesting transcript assessment in both kidney tissues and urine samples may provide insights into disease response over traditional renal biopsies alone.[60]

3.3 | Proteomics

High-resolution mass spectrometry can efficiently analyze protein abundance, modifications, and interactions. Proteomic analysis of kidney tissue and showed an immune activation gradient in urinary proteomics which was higher in proliferative disease. Kidney tissues analysis revealed chemokine secretion in the urine was related to intrarenal chemokine production within tubular cells. While transcriptomic analysis found IFN-γ, was found to be the most abundant cytokine produced by infiltrating CD8[+] T cells in the kidney tissue across all classes of LN.[53] Proteomic analysis of neutrophil subsets has shown differential expression of interferon regulated within SLE patients.[61] Bashant et al.[61] demonstrated that low density proinflammatory granulocytes (neutrophils) expressed an abundance of TI-IFN proteins compared with normal density neutrophils within SLE patients. Cytoskeleton changes were noted in these low-density granulocytes thought to contribute to increased trafficking and adherence of these neutrophils to microvasculature causing tissue damage during disease flare.[61] These studies suggest proteomic analysis may provide additional information on cell regulation and immunoactivity in serum, urine, and kidney tissue samples in patients with SLE.

3.4 | Metabolomics

Metabolomics comprehensively measure the repertoire of metabolites and (or small molecules) present in cells, tissues, and body fluids. Serum/plasma, and to a lesser extent, urine are most commonly body fluids from patients to be used to profile metabolomes, which have been applied to understand disease mechanisms, stratify patients and predict prognosis or response to therapy.[62] Although the metabolomic study to understand the heterogeneity in SLE patients is yet to be performed, cohort studies analyzing SLE patient's sera have suggested metabolic alternations in these patients compared with healthy controls (HCs).[63]

TABLE 3 Representative SLE transcriptomics studies

Studies	Patient criteria & Sample size	Profiling methods	Analytic methods	Major discoveries in stratification
Chiche et al.[50]	SLE $N = 62$ with 157 samples collected HC $N = 20$ adult patients	Microarray-based transcript-omics of whole blood	Second generation of a modular framework	Assessment of IFN signatures revealed that 87% of SLE patient samples had upregulated IFN patterns. The upregulation of not only IFN alpha was noted but also IFN beta and gamma were seen the SLE patients. Modular analysis of IFN signatures revealed three distinct modules which had variable activation thresholds (M1.2 < M3.4 < M5.12) and were useful for stratifying patients within IFN modules. Variability with the longitudinal assessment of IFN gene signature (IGS) was seen within patients.
Banchereau et al.[48]	SLE $N = 158$ pediatric (profiled 924 longitudinal blood samples from 158 pediatric patients). HC $N = 48$ pediatric patients	Microarray-based transcript-omics of whole blood	Unsupervised hierarchical clustering. Conducted modular analysis	By clustering of the interindividual SLEDAI correlation matrix, seven patient groups were identified (PG1–7). Each group had different immune signatures related to activity, containing erythropoiesis; IFN, myeloid lineage/neutrophils, plasmablasts, lymphoid lineage. PG1 and 6 mainly had erythropoiesis correlating to activity and had the lowest rates of LN. PG 2 and 3: plasmablasts and/or lymphoid lineage correlated to activity. PG4 and 5: IFN correlated with DA. PG 7 Myeloid and plasma blast signatures correlating to DA. It was also noted plasma blast signature was the most robust marker of disease activity, particularly for African American patients. Increase neutrophil transcripts were noted in prior and during disease flare.
Figgett et al.,[49]	SLE $N = 161$ HC $N = 57$	Transcript-omics: RNA-seq of whole blood	Unsupervised k-means clustering	Four groups in SLE patients: G1: similar to HC, low anti-Ro$^+$ and, the highest rate of photosensitivity; G2: the highest rate of serositis, high expression of TLR7, high ISGs; G3: increased neutrophil signature, decreased PC signature, more flares than G1 & G2. G4: the highest disease activities and autoantibody titers, the highest rate of renal disorder and discoid rash, more flares than G1 & G2, high expression of BAFF, high ISG.
Hong et al.[51]	SLE pregnant Women $N = 92$ SLE women undergoing assisted reproduction $N = 25$ SLE nonpregnant $N = 20$ HC pregnant $N = 43$ HC nonpregnant $N = 34$	Whole blood Microarray Transcript-omics	Hierarchical clustering to assess differentially expressed transcripts (DET) PCAQ-Gen assessment of transcriptional patterns	Four gene expression modules were identified in SLE patients (M1.2, IFN, M 4.11 Plasma cells, M2.3 erythropoiesis M4.2 inflammation and M5.15 neutrophils) SLE pregnancies with complications, e.g., pre-eclampsia (PET) had higher fold expression of the neutrophil signature versus noncomplicated SLE pregnancy. Upregulation of IFN and plasma cell signature were seen in fatal complications other than PET in SLE patients.

TABLE 3 (Continued)

Studies	Patient criteria & Sample size	Profiling methods	Analytic methods	Major discoveries in stratification
Nehar-Belaid et al.[52]	Cohort 1: children SLE (N = 33) and HC (N = 11), PBMCs (n = ~276,000) Cohort 2: adult SLE (N = 8) and HC (N = 6), PBMCs (n = ~82,000)	Transcriptomics: scRNA-seq of PBMCs	Hierarchical clustering on combined child and adult matrices	Four groups in SLE patients: G1: the expansion of subclusters with a high signature of ISGs, lower disease activities than G2, MMF treatment in the majority. G2: the highest disease activities, the expansion of subclusters with the high signature of ISGs, MMF treatment in minority; G3 and G4: mixed with HC, the expansion of "memory" CD4 + and CD8 + T-cell subclusters.
Fava et al.[53]	SLE N = 30	scRNA-seq and proteomics of kidney tissue and urine	PCA	Analysis showed an immune activation gradient in urinary proteomics which was higher in proliferative disease. Kidney tissues analysis revealed chemokine secretion in the urine was related to intrarenal chemokine production. IFN-γ was found to be the most abundant cytokine produced by infiltrating CD8 + T cells in the kidney tissue across all classes of LN.

Abbreviations: IGS, IFN gene signature; LN, lupus nephritis; PCA, principal component analysis; SLEDAI, SLE diseases activity index; SLE, systemic lupus erythematosus.

Metabolomic profiling of SLE patients has demonstrated increased oxidative stress, fatty acid oxidation, and changes in glycolysis related to DA, these changes may precede clinical flares and are potential biomarkers for DA.[63,64] Pearl et al. analyzed 36 SLE patients versus 42 HC using peripheral blood lymphocytes (PBLs) by Liquid chromatography-tandem mass spectrometry (LC-MS). The most profound changes they noted in SLE patients' PBLs versus HC samples were in the pentose phosphate pathway (PPP).[65] Enhanced oxidative stress was thought to reduce cysteine levels while cystine, kynurenine, cytosine were all increased. Kynurenine accumulation was reversed via the administration of N-acetyl cysteine (NAC) versus placebo in SLE patients. In vitro analysis of Kynurenine showed it was an activator of the mammalian target of rapamycin (mTOR). This provides mechanistic insights into disease development in SLE via metabolic changes.[36] Analysis of fecal metabolites by Zhang et al. also found increased levels of fecal xanthurenic acid and kynurenic acid elevated in SLE patients versus HC. This suggests that metabolomic analysis may provide insights into disease mechanisms in SLE. Yan et al. performed Gas chromatography–mass spectrometry (GC-MS) in 30 SLE patients versus 29 HC.[66] They noted L-valine was elevated at the time of SLE diagnosis, while L-tryptophan detection was correlated with DA. More recently Zhang et al. noted significant changes in glycerophospholipid metabolism in SLE patients' sera.[67] Elevations in L-pyroglutamic acid levels were markedly increased in patients with active disease and correlated with disease activities, this was validated using area under curve (AUC) and receiver-operating characteristic curves (ROC). These data suggest that disease-specific metabolites may be considered noninvasive biomarkers for diagnosis and disease activities. However further analysis is required to use key metabolite changes to understand the heterogeneity of SLE patients and dynamics of flare-remission cycles of the disease.

4 | IMPROVEMENT OF THE MINING OF OMICS DATA BY NEW ANALYTIC METHODS INCLUDING MACHINE LEARNING

A major challenge in the stratification of SLE patients based on omics data is how to interpret data that significantly exceed the scale of conventional clinical and immunological measures. Compared with clinical data are usually composed of up to dozens of variables, data generated by omics technologies can range from hundreds (such as metabolites) to tens of thousands (such as transcripts). The linear dimensionality reduction method principal component analysis (PCA) is a machine learning algorithm that has been widely used to handle data with high dimensionality.[68] PCA reduces data into low-dimensional space and clusters samples. The method PCA aims to project the data into the most variant directions of the feature space, thus compressing the data and maintaining global information. The first principal component (PC1) corresponds to the direction with the largest contribution in feature space, followed by the second principal component (PC2) with the
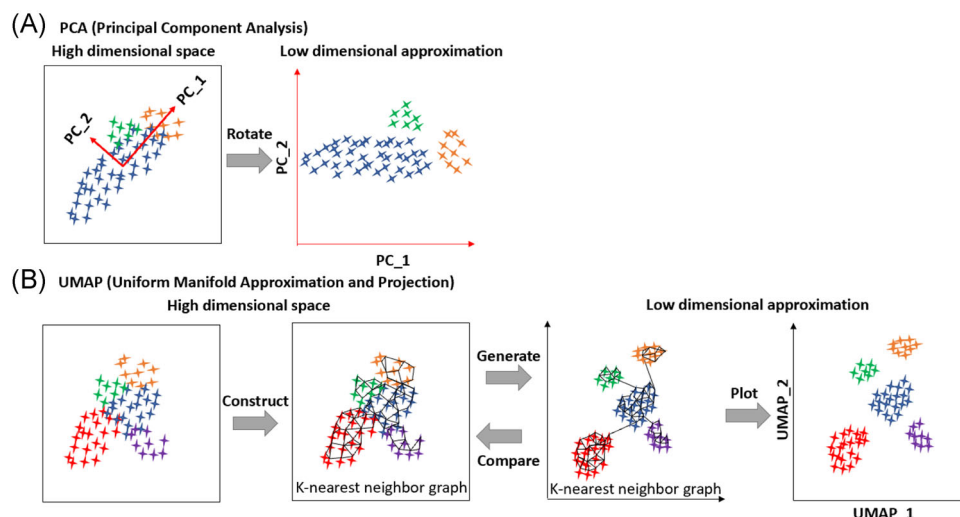
FIGURE 4 Basics for dimensionality reduction. (A) Principal component analysis (PCA) aims to project the data into the most variant directions with the first principal component (PC1) corresponds to the direction with the largest contribution in feature space, followed by the second principal component (PC2) with second-largest contribution. (B) Uniform manifold approximation and projection (UMAP) constructs a k-nearest neighbor (kNN) graph to approximate the manifold structure of data, then embeds the kNN graph to two-dimensional space to visualize the data

second largest contribution (Figure 4A). Based on the embedding data by PCA, clustering methods like k-means and hierarchical clustering can be applied to generating clusters of samples, which can be further investigated with the association with patients' clinical and immunological features. PCA was applied by several studies to stratify SLE patients which have been summarized above.[49]

Although PCA can retain the information of most variables, it often fails to preserve the local clustering structure since the local relationship between samples is less dependent on PC1 or PC2. Therefore, new methods including uniform manifold approximation and projection (UMAP) have been developed. UMAP relies on persistent homology to maintain the manifold topological structure.[69] Specifically, UMAP first constructs a k-nearest neighbour (kNN) graph to approximate the manifold structure of data, then embeds the kNN graph to two-dimensional space to visualize the data (Figure 4B). UMAP has been shown to have a better capacity at maintaining local information as well as global structure in analyzing datasets for scRNA-seq.[70]

A recent study systemically compared UMAP with other mainstream methods for dimensionality reduction including PCA and demonstrated that UMAP was superior to PCA for clustering accuracy, neighbour information preserving and feature separating.[70] In this study, a data set consisting of longitudinal transcriptome profiles of 65 SLE patients and 20 HCs was analyzed. UMAP, but not PCA, clearly demonstrated the separation of SLE patients from HCs. Furthermore, the embedding space by UMAP revealed new clustering structures, that is, subgroups in SLE patients' samples. Such clustering structures were found associated with patient visiting date and the trends of disease improvement or deterioration.[70] This is an example of the capacity of new methods that can effectively analyze omics data and strengthen the detection of the heterogeneity within patients for disease diagnosis and therapy choices.

5 | FUTURE PERSPECTIVE

Several recent reviews have touched on the topic of how omics technologies can pave the way for precision medicine in SLE, including the promise of transcriptomic profiling and machine learning integration may provide an insight into SLE immune pathways and the application of multi-omics approaches including genomics, immunophenotyping, proteomic, and transcriptomic assessments.[38,71–73] Compared with other reviewers also suggesting the application of omics technologies, we propose that the further integration of these "multi-omics" technologies using artificial intelligence including machine learning has the potential to identify non-invasive biomarkers of DA, novel prognostic markers for stratification and the utilization of precision medicine to tailor the choice of target therapies and duration of immunosuppression. Clinical translation of this study into standardized bioassays may prove to be superior to current markers of DA in SLE/LN and deliver improved patient-specific outcomes.

CONFLICT OF INTERESTS
The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS
Di Yu: Conceptualization; Writing; Supervision. **Prianka Puri:** Writing. **Simon H. Jiang:** Writing. **Yang Yang:** Writing. **Fabienne Mackay:** Supervision.

ORCID

Prianka Puri http://orcid.org/0000-0003-0752-4651
Simon H. Jiang http://orcid.org/0000-0002-3973-5114
Fabienne Mackay http://orcid.org/0000-0002-6074-2693
Di Yu http://orcid.org/0000-0003-1721-8922

REFERENCES

1. Tsokos GC. Systemic lupus erythematosus. *N Engl J Med.* 2011; 365(22):2110-2121.
2. Allen ME, Rus V, Szeto GL. Leveraging heterogeneity in systemic lupus erythematosus for new therapies. *Trends Mol Med.* 2021; 27(2):152-171.
3. Agmon-Levin N, Mosca M, Petri M, Shoenfeld Y. Systemic lupus erythematosus one disease or many? *Autoimmun Rev.* 2012;11(8): 593-595.
4. Choi J, Kim ST, Craft J. The pathogenesis of systemic lupus erythematosus-an update. *Curr Opin Immunol.* 2012;24(6):651-657.
5. Almaani S, Meara A, Rovin BH. Update on lupus nephritis. *Clin J Am Soc Nephrol.* 2017;12(5):825-835.
6. 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.
7. Tsokos GC. Autoimmunity and organ damage in systemic lupus erythematosus. *Nat Immunol.* 2020;21(6):605-614.
8. Vincent FB, Morand EF, Schneider P, Mackay F. The BAFF/APRIL system in SLE pathogenesis. *Nat Rev Rheumatol.* 2014;10(6): 365-373.
9. Linterman MA, Rigby RJ, Wong RK, et al. Follicular helper T cells are required for systemic autoimmunity. *J Exp Med.* 2009;206(3): 561-576.
10. Suarez-Fueyo A, Bradley SJ, Tsokos GC. T cells in systemic lupus erythematosus. *Curr Opin Immunol.* 2016;43:32-38.
11. McKinney EF, Lyons PA, Carr EJ, et al. A CD8+ T cell transcription signature predicts prognosis in autoimmune disease. *Nat Med.* 2010;16(5):586-591.
12. Hanayama R, Tanaka M, Miyasaka K, et al. Autoimmune disease and impaired uptake of apoptotic cells in MFG-E8-deficient mice. *Science.* 2004;304(5674):1147-1150.
13. Grieves JL, Fye JM, Harvey S, Grayson JM, Hollis T, Perrino FW. Exonuclease TREX1 degrades double-stranded DNA to prevent spontaneous lupus-like inflammatory disease. *Proc Natl Acad Sci USA.* 2015;112(16):5117-5122.
14. Mackay F, Woodcock SA, Lawton P, et al. Mice transgenic for BAFF develop lymphocytic disorders along with autoimmune manifestations. *J Exp Med.* 1999;190(11):1697-1710.
15. Vincent FB, Northcott M, Hoi A, Mackay F, Morand EF. Association of serum B cell activating factor from the tumour necrosis factor family (BAFF) and a proliferation-inducing ligand (APRIL) with central nervous system and renal disease in systemic lupus erythematosus. *Lupus.* 2013;22(9):873-884.
16. Vinuesa CG, Cook MC, Angelucci C, et al. A RING-type ubiquitin ligase family member required to repress follicular helper T cells and autoimmunity. *Nature.* 2005;435(7041):452-458.
17. He J, Tsai LM, Leong YA, et al. Circulating precursor CCR7(lo)PD-1(hi) CXCR5(+) CD4(+) T cells indicate Tfh cell activity and promote antibody responses upon antigen reexposure. *Immunity.* 2013;39(4): 770-781.
18. Yu D, Tan AH, Hu X, et al. Roquin represses autoimmunity by limiting inducible T-cell co-stimulator messenger RNA. *Nature.* 2007;450(7167):299-303.
19. Deng J, Wei Y, Fonseca VR, Graca L, Yu D. T follicular helper cells and T follicular regulatory cells in rheumatic diseases. *Nat Rev Rheumatol.* 2019;15(8):475-490.
20. 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.
21. Petri M, Orbai AM, Alarcón 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.
22. Ugarte-Gil MF, Pons-Estel GJ, Griffin R, Vilá LM, Reveille JD, Alarcón GS. Achievement of the 2019 European Alliance of Associations for Rheumatology/American College of Rheumatology Criteria for Systemic Lupus Erythematosus and Amount of Damage Accrual: results from a multiethnic multicenter cohort. *Arthritis Care Res.* 2021;73(7):1038-1040.
23. Whittall Garcia LP, Gladman DD, Urowitz M, Touma Z, Su J, Johnson SR. New EULAR/ACR 2019 SLE Classification Criteria: defining ominosity in SLE. *Ann Rheum Dis.* 2021.80(6):767–774.
24. Chambers SA, Rahman A, Isenberg DA. Treatment adherence and clinical outcome in systemic lupus erythematosus. *Rheumatology.* 2007;46(6):895-898.
25. Yu F, Haas M, Glassock R, Zhao MH. Redefining lupus nephritis: clinical implications of pathophysiologic subtypes. *Nat Rev Nephrol.* 2017;13(8):483-495.
26. Korbet SM, Volpini KC, Whittier WL. Percutaneous renal biopsy of native kidneys: a single-center experience of 1,055 biopsies. *Am J Nephrol.* 2014;39(2):153-162.
27. Whittier WL, Korbet SM. Timing of complications in percutaneous renal biopsy. *J Am Soc Nephrol.* 2004;15(1):142-147.
28. Bajema IM, Wilhelmus S, Alpers CE, et al. Revision of the International Society of Nephrology/Renal Pathology Society classification for lupus nephritis: clarification of definitions, and modified National Institutes of Health activity and chronicity indices. *Kidney Int.* 2018;93(4):789-796.
29. Contis A, Vanquaethem H, Truchetet ME, et al. Analysis of the effectiveness and safety of rituximab in patients with refractory lupus nephritis: a chart review. *Clin Rheumatol.* 2016;35(2): 517-522.
30. Ramos-Casals M, Soto MJ, Cuadrado MJ, Khamashta MA. Rituximab in systemic lupus erythematosus: a systematic review of off-label use in 188 cases. *Lupus.* 2009;18(9):767-776.
31. Fontana F, Alfano G, Leonelli M, et al. Efficacy of Belimumab for active lupus nephritis in a young Hispanic woman intolerant to standard treatment: a case report. *BMC Nephrol.* 2018;19(1):276.
32. Felten R, Scher F, Sagez F, Chasset F, Arnaud L. Spotlight on anifrolumab and its potential for the treatment of moderate-to-severe systemic lupus erythematosus: evidence to date. *Drug Des Devel Ther.* 2019;13:1535-1543.
33. Furie R, Rovin BH, Houssiau F, et al. Two-year, randomized, controlled trial of belimumab in lupus nephritis. *N Engl J Med.* 2020;383(12):1117-1128.
34. He J, Zhang X, Wei Y, et al. Low-dose interleukin-2 treatment selectively modulates CD4(+) T cell subsets in patients with systemic lupus erythematosus. *Nat Med.* 2016;22(9):991-993.
35. He J, Zhang R, Shao M, et al. Efficacy and safety of low-dose IL-2 in the treatment of systemic lupus erythematosus: a randomised, double-blind, placebo-controlled trial. *Ann Rheum Dis.* 2020; 79(1):141-149.
36. Humrich Jens Y, von Spee-Mayer Caroline, Siegert Elise, et al. Low-dose interleukin-2 therapy in refractory systemic lupus erythematosus: an investigator-initiated, single-centre phase 1 and 2a clinical trial. *Lancet.* 2019;1(1):44-49.
37. Deapen D, Escalante A, Weinrib L, et al. A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis Rheum.* 1992;35(3):311-318.
38. Teruel M, Chamberlain C, Alarcon-Riquelme ME. Omics studies: their use in diagnosis and reclassification of SLE and other systemic autoimmune diseases. *Rheumatology.* 2017;56(suppl_1):i78-i87.
39. Alarcon-Riquelme ME. New attempts to define and clarify lupus. *Curr Rheumatol Rep.* 2019;21(4):11.
40. Kozyrev SV, Abelson AK, Wojcik J, et al. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nat Genet.* 2008;40(2):211-216.
41. Harley JB, Alarcón-Riquelme ME, Criswell LA, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat Genet.* 2008;40(2):204-210.
42. Jiang SH, Athanasopoulos V, Ellyard JI, et al. Functional rare and low frequency variants in BLK and BANK1 contribute to human lupus. *Nat Commun.* 2019;10(1):2201.
43. Panousis NI, Bertsias GK, Ongen H, et al. Combined genetic and transcriptome analysis of patients with SLE: distinct, targetable
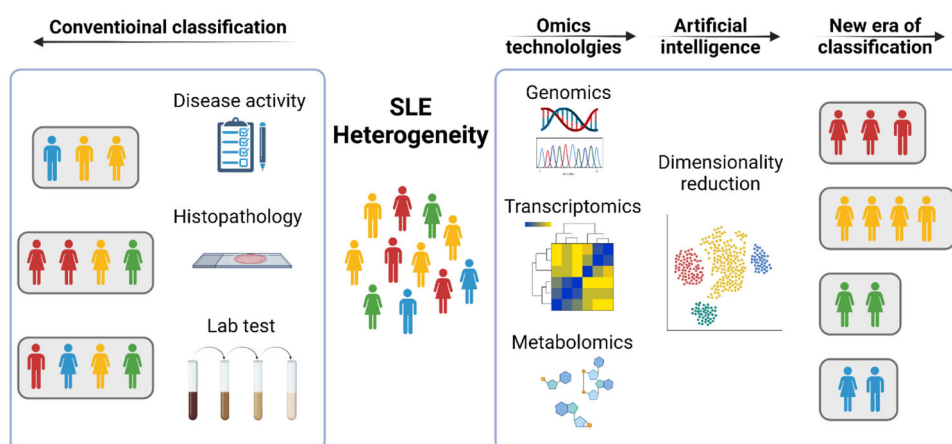
signatures for susceptibility and severity. *Ann Rheum Dis*. 2019; 78(8):1079-1089.

44. Connaughton DM, Hildebrandt F. Personalized medicine in chronic kidney disease by detection of monogenic mutations. *Nephrol Dial Transplant*. 2020;35(3):390-397.

45. Almlöf JC, Nystedt S, Leonard D, et al. Whole-genome sequencing identifies complex contributions to genetic risk by variants in genes causing monogenic systemic lupus erythematosus. *Hum Genet*. 2019;138(2):141-150.

46. Jiang SH, Stanley M, Vinuesa CG. Rare genetic variants in systemic autoimmunity. *Immunol Cell Biol*. 2020;98(6):490-499.

47. Lea-Henry TN, Chuah A, Stanley M, et al. Increased burden of rare variants in genes of the endosomal Toll-like receptor pathway in patients with systemic lupus erythematosus. *Lupus*. 2021. doi:10.1177/09612033211033979

48. Banchereau R, Hong S, Cantarel B, et al. Personalized immunomonitoring uncovers molecular networks that stratify lupus patients. *Cell*. 2016;165(3):551-565.

49. Figgett WA, Monaghan K, Ng M, et al. Machine learning applied to whole-blood RNA-sequencing data uncovers distinct subsets of patients with systemic lupus erythematosus. *Clin Transl Immunology*. 2019;8(12):e01093.

50. Chiche L, Jourde-Chiche N, Whalen E, et al. Modular transcriptional repertoire analyses of adults with systemic lupus erythematosus reveal distinct type I and type II interferon signatures. *Arthritis Rheumatol*. 2014;66(6):1583-1595.

51. Hong S, Banchereau R, Maslow BL, et al. Longitudinal profiling of human blood transcriptome in healthy and lupus pregnancy. *J Exp Med*. 2019;216(5):1154-1169.

52. Nehar-Belaid D, Hong S, Marches R, et al. Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nat Immunol*. 2020;21(9):1094-1106.

53. Fava A, Buyon J, Mohan C, et al. Integrated urine proteomics and renal single-cell genomics identify an IFN-gamma response gradient in lupus nephritis. *JCI Insight*. 2020;5(12):e138345.

54. 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. *PLOS One*. 2016;11(11):e0166312.

55. El-Sherbiny YM, Psarras A, Md Yusof MY, et al. A novel two-score system for interferon status segregates autoimmune diseases and correlates with clinical features. *Sci Rep*. 2018;8(1):5793.

56. Mistry P, Nakabo S, O'Neil L, et al. Transcriptomic, epigenetic, and functional analyses implicate neutrophil diversity in the pathogenesis of systemic lupus erythematosus. *Proc Natl Acad Sci USA*. 2019;116(50):25222-25228.

57. Catalina MD, Bachali P, Geraci NS, Grammer AC, Lipsky PE. Gene expression analysis delineates the potential roles of multiple interferons in systemic lupus erythematosus. *Commun Biol*. 2019;2:140.

58. Der E, Ranabothu S, Suryawanshi H, et al. Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis. *JCI Insight*. 2017;2(9).

59. Der E, Suryawanshi H, Morozov P, et al. Tubular cell and keratinocyte single-cell transcriptomics applied to lupus nephritis

reveal type I IFN and fibrosis relevant pathways. *Nat Immunol*. 2019;20(7):915-927.

60. Arazi A, Rao DA, Berthier CC, et al. The immune cell landscape in kidneys of patients with lupus nephritis. *Nat Immunol*. 2019; 20(7):902-914.

61. Bashant KR, Aponte AM, Randazzo D, et al. Proteomic, biomechanical and functional analyses define neutrophil heterogeneity in systemic lupus erythematosus. *Ann Rheum Dis*. 2021;80(2):209-218.

62. Beger RD, Dunn W, Schmidt MA, et al. Metabolomics enables precision medicine: "A White Paper, Community Perspective". *Metabolomics*. 2016;12(10):149.

63. Yan B, Huang J, Zhang C, et al. Serum metabolomic profiling in patients with systemic lupus erythematosus by GC/MS. *Mod Rheumatol*. 2016;26(6):914-922.

64. De Livera AM, Dias DA, De Souza D, et al. Normalizing and integrating metabolomics data. *Anal Chem*. 2012;84(24):10768-10776.

65. Perl A, Hanczko R, Lai ZW, et al. Comprehensive metabolome analyses reveal N-acetylcysteine-responsive accumulation of kynurenine in systemic lupus erythematosus: implications for activation of the mechanistic target of rapamycin. *Metabolomics*. 2015;11(5):1157-1174.

66. Yan R, Jiang H, Gu S, et al. Fecal metabolites were altered, identified as biomarkers and correlated with disease activity in patients with systemic lupus erythematosus in a GC-MS-based metabolomics study. *Front Immunol*. 2020;11:2138.

67. 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*. 2021;60(2):598-606.

68. Svante Wold KE, Paul Geladi. Principal component analysis. *Chemom Intell Lab Syst*. 1987;2(1-3).

69. Leland McInnes, John Healy JM. UMAP: uniform manifold approximation and projection for dimension reduction. *ArXiv e-prints*. 2018. doi:1802.03426v3

70. Yang Y, Sun H, Zhang Y, et al. Dimensionality reduction by UMAP reinforces sample heterogeneity analysis in bulk transcriptomic data. *Cell Rep*. 2021;36(4):109442.

71. Fujio K, Takeshima Y, Nakano M, Iwasaki Y. Review: transcriptome and trans-omics analysis of systemic lupus erythematosus. *Inflamm Regen*. 2020;40:11.

72. Nagafuchi Y, Shoda H, Fujio K. Immune profiling and precision medicine in systemic lupus erythematosus. *Cells*. 2019;8(2):140.

73. Kwon YC, Chun S, Kim K, Mak A. Update on the genetics of systemic lupus erythematosus: genome-wide association studies and beyond. *Cells*. 2019;8(10):1180.

Graphical Abstract

This graphical abstract will be a part of HTML, Online and Print versions.

Classification of SLE patients by omics data and artificial intelligence.