Integrative analysis of transcriptomic profiles reveals shared gene signatures of Rheumatoid Arthritis and Systemic Lupus Erythromatosus

Neetu Tyagi
Regional Centre for Biotechnology
Dinesh Gupta (dinesh@icgeb.res.in)
International Centre for Genetic Engineering and Biotechnology https://orcid.org/0000-0002-7548-8835

Research article

Keywords: Meta-analysis, rheumatoid arthritis, systemic lupus erythromatosus, Multi-Cohort Analysis, gene expression, Microarray, Autoimmunity

Posted Date: September 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-869926/v1

Abstract

Background

Autoimmune diseases develop when a person's immune system starts developing immune response against its own healthy cells, tissues, or any other cell constituents. Rheumatoid Arthritis (RA) and Systemic Lupus Erythromatosus (SLE) are the two most common systemic inflammatory autoimmune diseases, sharing various clinical as well as pathological signatures. Although multiple studies have been conducted to date, very little is known about molecular pathogenesis and overlapping molecular signatures of the two diseases. Motivated to explore the common molecular disease features, we conducted a meta-analysis of the publicly available microarray gene expression datasets of RA and SLE.

Methods

Common and unique gene signatures of RA and SLE were identified based on analysis of microarray gene-expression datasets. Hub genes were identified by performing network analysis of protein-protein interaction (PPI) networks of the identified genes. Gene ontology functional enrichment and integrative pathway analysis was also performed to understand the underlying molecular mechanisms in the diseases.

Results

Intriguingly, out of the identified signature genes, 9 are upregulated and 24 are downregulated. Many of the common gene signatures identified in this study provide clues to the shared pathological mechanisms of RA and SLE. Amongst the identified signatures, MMP8, NFIL3, B4GALT5, HIST1H1C, NMT2, PTGDS and DUSP14, are the robust gene signatures shared by all the RA and SLE datasets. Functional analysis revealed that the common signatures are involved in the pathways such as mTOR signaling pathways, virus infection-related pathways, bone remodeling, activation of matrix metalloproteinase pathway, immune and inflammatory response-related pathways.

Conclusions

The common gene signatures and related pathways identified in this study substantiate the shared pathological mechanism involved in both diseases. Furthermore, our analysis of multi-cohort and multiple microarray datasets allow discovery of novel leads for clinical diagnosis and potential novel drug targets.

Background

Autoimmunity refers to a condition of an organism's immune response against its own cells or tissues and the other normal body constituents. More than 80 autoimmune diseases have been characterized; few of them are well known, including Rheumatoid Arthritis (RA), Systemic Lupus Erythromatosus (SLE), type 1 diabetes, multiple sclerosis, and psoriatic arthritis. In contrast, others are rare and difficult to diagnose [1]. Two of the common systemic inflammatory autoimmune diseases are RA and SLE, are the focus of this study.

RA is a multisystem chronic systemic inflammatory autoimmune disease, primarily affects the joints and can lead to abnormality and functional limitation in the absence of suitable treatment. The patients have inflammation, synovial hyperplasia, autoantibody production (rheumatoid factor, RF), bone deformities, cartilage destruction, and systemic involvement of cardiovascular, pulmonary, neurological and endocrine systems [2]. SLE is another complex, heterogeneous, chronic autoimmune disease which affects multiple organs and tissues, including skin, joints, lungs,

kidneys, blood and CNS. In SLE, there is an interplay between various immunological, environmental, hormonal and genetic factors [3, 4].

The aetiology and pathophysiology of the two diseases are interconnected in genetic, immunological and environmental aspects. Both the diseases share several common mechanisms; however, they are different in their immunological mechanism, reflecting in their clinical presentation and implications. Although the first appearances of these two autoimmune diseases have different phenotypes, both the disorders are heterogeneous, multifactorial disorders that share molecular mechanisms, which are further elicited by similar clinical and pathophysiological features. The two autoimmune diseases share specific clinical and pathogenic features related to B and T cell hyperactivity, many susceptibility loci, and a pathogenic array of autoantibodies [5, 6]. The treatments are very similar for the autoimmune disorders, except in organ damage or in cases where the features of one disease dominate over the other. To design a personalized treatment, it is crucial to discover the molecular level similarities and differences in the diseases linked with respective patient groups, apart from clinical features.

Identifying specific gene expression signatures by gene expression patterns analysis can also provide pivotal information that help to understand the molecular mechanisms that underlie the diseases. With the advancement of high-throughput technologies like microarray and next-generation sequencing, these types of studies are becoming feasible and common. This development of high-throughput technologies provides us with the opportunity to analyze the gene expression on a genome-wide scale and in fact widely used in context to autoimmune diseases during the past decade [7–9].

Meta-analysis techniques offer the opportunity to integrate and combine the data from different sources, representing multiple studies. Several conventional tools exist for meta-analysis, including MAMA [10], MetaOmics [11], INMEX [12], Elastic-Net [13], MetaIntegrator [14] etc. We have used MetaIntegrator in this study. The MetaIntegrator package offers a straightforward, custom framework in automatically downloading data from GEO, integrating the MetaSignature database, and better visualization of the results. The package has demonstrated its application in various areas like discovering diagnostics, prognostics, novel drug targets identification and drug repurposing too. It uses a random effect model for meta-analysis, embraces heterogeneity, and avoids the batch effect correction limitations by computing effect size for each dataset.

Previous studies focused on integrated data analysis of different studies for a single disease have been successfully conducted to gain more rational and decisive results. For example, Song et al. [15] conducted a meta-analysis by integrating three publicly available datasets for SjS. A study was conducted to identify 37 gene signatures associated with SLE, where Arasappan et al. [16] performed a pathway-based meta-analysis using four SLE datasets. For RA datasets, several Interferon (IFN) related apoptotic process and cell cycle pathways were identified by Olsen et al. [17, 18]. Afroz et al. [19] conducted a study to identify RA-specific unique gene signatures via meta-profiling that might provide mechanistic insights of this disease. A recent analysis was performed on publicly available SLE datasets divided into discovery and validation to conduct a multi-cohort meta-analysis using the MetaIntegrator framework to reveal unified signatures related to SLE [3]. Kroger et al. [20] performed a study to pick the differentially regulated genetic pathways associated with the PBMC SLE patient compared to the unaffected healthy controls.

Apart from the analysis of single disease multiple source datasets, the meta-analysis techniques offer tremendous opportunity to integrate data from different diseases to reveal novel common gene signatures, which may be missed in single studies data analysis. For example, Tuller et al. [8] analyzed the publicly available data from six different autoimmune diseases (SLE, multiple sclerosis, RA, juvenile RA, type 1 diabetes, Crohn's disease and ulcerative colitis) taking the PBMC samples. The study aimed to understand the intra-regulatory mechanism in PBMC, which can be common to all autoimmune diseases or specific to any few of them. They found certain chemokines and interleukin genes were differentially expressed with the analyzed autoimmune diseases. To uncover the co-expression patterns, Silva et al. [21] integrated the SLE and RA expression datasets and profiling modules for specifically induced or repressed and

co-modulated genes. Higgs et al. [22] conducted a study to analyze type 1 IFN related common signatures by integrating data from SLE, myositis, RA and scleroderma. Toro-Dominguez et al. [23] uncovered the common signatures from SLE, RA and SjS PBMC patients and conducted the gene expression meta-analysis using the publicly available gene expression datasets. A recent study was reported by Badr and Häcker [24], to find the linker gene signature and pathways of RA and TB. Their study identified the co-regulated genes and pathways associated with the diseases.

With an aim to identify the robust gene signatures using meta-analysis approach, we performed a meta-analysis of publicly available multi-cohort gene expression microarray datasets of PBMC and whole blood samples from SLE and RA patients with respect to controls. To the best of our knowledge, this is the first integrative study of the two systemic autoimmune diseases, based on large-scale gene expression datasets downloaded from GEO database to explore the shared gene expression signatures in the patients with respect to the healthy controls for RA and SLE.

Methods

Data collection

Data collection was carried out by searching the Gene Expression Omnibus (GEO). For RA, the search terms used to retrieve the datasets was “rheumatoid arthritis and microarray”, for SLE, the term was “Systemic lupus erythromatosus and microarray”, with the filter organism as *Homo sapiens*, study type expression profiling and entry type Dataset/Series. Initially 536 records were recovered. The datasets were downloaded from the GEO database using the GEOquery R package, integrated in the MetaIntegrator R [64]. The downloaded datasets were filtered for making the final dataset suitable for the study, as described in results section.

Data Pre-processing and Meta-analysis

We processed the selected GEO accessions using the MetaIntegrator, which downloaded and pre-processed the datasets corresponding to the GEO accessions. Subsequently, the individual meta-analysis of each disease datasets was carried out using the MetaIntegrator framework [14]. The framework detects the significantly deregulated genes between cases (RA or SLE) and healthy controls. It computes Hedges g effect size for each gene in each dataset and pools these effect sizes across the datasets.

$$\text{Hedges g effect size } (g) = J \frac{\overline{X}_1 - \overline{X}_0}{\sqrt{\frac{(n_1-1)S_1^2 + (n_0-1)S_0^2}{n_1+n_0-2}}}$$

J is Hedges g correction factor, $_1$ and $_0$ are the average expression; S_1 and S_0 are the standard deviation deviations of cases and controls. The n_1 and n_2 are the number of cases and controls. The summary effect size g_s was computed using a random effect model.

$$\text{Pooled effect size } (g_s) = \frac{\sum_i^n W_i g_i}{\sum_i^n W_i}$$

n is the no. of studies, g_i is the hedges’ g of the gene within dataset i, W_i is the weight calculated by $1/(V_i + T^2)$, V_i is the variance of the gene within a given dataset i and T^2 is the inter-dataset variation estimated by DerSimonian-laired method.

The random effect model usually gives more conservative results by extracting fewer differentially expressed genes (DEGs), however with more confidence. MetaIntegrator also calculates Cochrane's Q value for evaluating heterogeneity of effect size estimates between studies. It also uses Fisher's method for combining p-values across studies. After meta-analysis, the most significant DEGs were selected for the downstream analysis, using the following filtering criteria: FDR < 0.05, Effect size > 0.40 and expressed in at least three studies. A heatmap of DEGs for both RA and SLE was created in R. Subsequently, we checked for the common gene signatures among the two autoimmune diseases.

Hub genes and network analysis

To generate the protein-protein interaction (PPI) network, NetworkAnalyst was used to create PPI networks for the common gene signatures and the RA and SLE specific DEGs, using the option to use innateDB interactome database [65]. The common gene signatures and top 50 DEGs from individual disease meta-analysis were used to construct their respective networks for identifying hub genes.

Gene ontology and Integrative pathway analysis

To identify the over-represented biological terms and enriched deregulated pathways, we used Enrichr, a web-based tool [66]. The differentially expressed genes obtained from individual Meta analysis of RA and SLE, and the common gene signatures revealed in our study were used as input in Enrichr tool. Default settings were used for the functional annotation and the p-value was computed using Fisher’s exact test. The key biological terms were ranked based on the significance threshold of p-value < 0.05.

Results

Data download and filtering

The gene expression datasets downloaded from the GEO were manually checked to remove duplicates and irrelevant studies, after which 20 studies remained. Out of the 20 studies, we selected the studies that reported gene expression in whole blood, PBMC or blood cell components, and excluded studies regarding other tissues such as synovial fluid or other tissues to ensure comparable gene expression. This filtering also ensured the removal of any potential bias due to tissue-specific gene expression. We also excluded studies based on drug-treated samples. The filtering, as mentioned above, led to 14 definitive studies to be included for our analysis.

After a thorough search and excluding datasets as specified above, seven SLE datasets (GSE11909, GSE50772, GSE22098, GSE8650, GSE4588, GSE61635, and GSE10325) and seven RA datasets (GSE56649, GSE15573, GSE4588, GSE17755, GSE11827, GSE57383, and GSE68689) were selected for further analysis. A total of 904 samples were considered for downstream analysis, representing data from 366 SLE patients, 168 healthy controls for SLE, 222 RA patients and 148 healthy controls for RA.

Data division and meta-analysis

To achieve an extensive, unbiased study of the shared signatures underlying RA and SLE, we identified and downloaded the publicly available GEO gene expression datasets. From the initial datasets obtained from public databases, we randomly selected 14 studies from both the diseases that matched the study inclusion criteria (see methods). The seven datasets for RA (5 PBMC, 1 WB, 1 CD4 T cells and B cells) and 7 of SLE (5 PBMC, 1 WB and 1 CD4 T cells and B cells) were processed for further analysis, contained samples from 222 RA patients with 148 controls, and 366 SLE patients with 168 controls, the complete workflow is shown in **Fig. 1a and Fig. 1b**. A detailed summary of the included datasets and samples can be found in **Table 1**. The previous findings suggest that discovery datasets with a sample size of 250-300 are enough to identify robust gene signatures for any disease using the MetaIntegrator framework. The Discovery dataset includes samples from PBMC, whole blood and CD4 T and B cells of RA and SLE patients with the respective

controls. Based on the meta-analysis using the MetaIntegrator R package, using Random Effect Model (REM), 160 RA genes and 1630 SLE genes were identified as significantly differentially expressed in the RA and SLE datasets versus healthy controls (FDR<0.05 in the REM). A list of 50 most significantly up or downregulated genes for RA and SLE are shown in **Additional file 1: Supplementary Table S1.**

Identification of common gene signatures in RA and SLE

For meta-analysis of the study dataset, the MetaIntegrator R package was used, with a selected cut-off for effect size (ES) and false discovery rate (FDR) to determine the DEGs for both the diseases, among healthy controls vs. diseased. ES is the ratio of the mean difference between groups to the standard deviation denoted as d, the convention values for small, medium, and high effect size were 0.20, 0.40 and 0.60, respectively. We identified 162 (56 upregulated, 106 downregulated) and 1630 (574 upregulated, 1056 downregulated) DEGs (ES>0.4, FDR<0.05) of RA and SLE, respectively. As per our hypothesis, we identified the 33 common gene signatures (9 upregulated and 24 downregulated) in both RA and SLE with a false discovery rate (FDR<0.05), effect size >0.4 and a minimum of 3 studies should support the same (see **Additional file 2: Supplementary Table S2**). Venn diagram highlights the unique and common gene signatures of RA and SLE (**Fig. 2**). In discovery datasets of RA and SLE Meta-scores distinguish patient samples from the healthy controls with an Area Under the Receiver Operating Characteristics (AUROC) of 0.871 (95% confidence interval (CI): 0.70-1) and 0.935 (95% CI: 0.73-1) respectively (**Fig. 3a, Fig. 3b**). Amongst the 33 common gene signatures, 20 are already reported to be associated with the diseases. However, to the best of our knowledge, 13 are not reported to be associated with these diseases. Particularly striking was the highly significant up-regulation of the genes NEXN, IRF9, B4GALT5, MMP8, HIT1H1C, NFIL3, HIST2H2AA4, DDX60L and VMP1. Out of these, IRF9 is an interferon regulatory factor involved in the IFN alpha signaling pathway and bone remodeling pathways [25,26]. NEXN was recently found deregulated between RA and TB [24]. Similarly, MMP8 is a matrix metalloproteinase associated with many autoimmune diseases, including RA [27]. Another important gene, NFIL3, is a nuclear factor interleukin 3 regulated gene involved in many autoimmune disorders, including SLE [28]. In brief, Interferons (IFN), Interleukins (IL), lymphokine and Tumor Necrosis Factor (TNF) are the types of cytokines that regulate the development of different cells of the immune system. In agreement with the previous studies, we find that these genes play a regulatory role in the pathogenesis of both diseases. A heatmap of the highly differentially regulated genes of RA, SLE and common gene signatures is shown in **Additional file 3: Supplementary Figure S1**. The complete description of the common signatures is listed in **Additional file 4: Supplementary Table S3**. In our study, we have also evaluated the gained and lost genes, gained are those DEGs identified in the meta-analysis, not present in the individual analysis and lost are those DEGs found in any individual study, but not in the meta-analysis, they may show low signals but show consistent expression pattern throughout the datasets. In the case of RA, we obtained 82 gained and 1313 lost genes, whereas, for SLE, the gained genes were 903 and 1631 lost genes. A list is shown in **Additional file 5: Supplementary Table S4** for all the gained and lost genes via this meta-analysis.

Hub genes network analysis

Based on the methodology, we generated three interaction networks as described in the methods section. The interaction network for the identified common gene signatures comprise of 33 seeds with 630 connecting nodes and of 980 edges representing the interaction between these proteins. To arrive at more biologically relevant information, this analysis identified key hub genes among the common signatures, and the top differentially regulated genes for RA and SLE. The protein-protein interaction network for the common gene signatures is shown in **Fig. 4**. We analyzed the protein-protein interaction network for the common gene signatures and found 15 hub genes with a high degree of centrality and betweenness. The top hub genes were RPS2 (degree: 156, betweenness: 34190), PABPC1 (degree: 152, betweenness: 63040), RPL5 (degree: 146, betweenness: 49501), RPS8 (degree: 145, betweenness: 31190), and EEF1G (degree: 122, betweenness: 55075).

From the interaction network of differentially regulated genes in RA, it was found that for the upregulated genes, the hub genes with the highest degree and betweenness are PKM (degree: 99, betweenness: 50885), IRF9 (degree: 47, betweenness: 17423), and OS9 (degree: 42, betweenness: 15352) for downregulated genes- DHX9 (degree: 185, betweenness: 165422), RPL5 (degree: 146, betweenness: 60548) and RPS11 (degree: 117, betweenness: 38181).

For the SLE interaction network, the notable hub genes include STAT1 (degree: 223, betweenness: 183938), ISG15 (degree: 188, betweenness: 167205), and MYD88 (degree: 87, betweenness: 60392) were the topmost differentially regulated hub genes. For the downregulated genes, the significant hub genes are DDX5 (degree: 156, betweenness: 85071), NAP1L1 (degree: 88, betweenness: 47262), CD44 (degree: 56, betweenness: 31368).

All the details of hub genes for the common gene signatures and disease-specific genes (RA and SLE) are listed in **Additional file 6: Supplementary Table S5**. The networks of RA and SLE for the top 50 DEGs are shown in **Additional file 3: Supplementary Figure S3** and **Supplementary Figure S4**. The common gene signatures showed expression in all the studies, shown by forest plots in **Fig. 5**. The forest plots represent very similar expression for the common (upregulated and downregulated) genes of RA and SLE across studies suggests their important role in both diseases. Many of these genes such as NFIL3, NMT2, EIF4B, PTGDS etc. identified as hub genes in our network analysis too, which further proves their importance in underlying mechanism of RA and SLE.

Identification of overrepresented biological pathways and Gene Ontology terms

Pathway enrichment analysis was performed using the Enrichr web-based tool by uploading the list of RA and SLE differentially regulated genes obtained by meta-analysis. For RA upregulated genes, TNF signaling pathways, Interleukin related pathways, osteoclast differentiation, bone remodeling, haemoglobin's chaperon-related pathways, TNFR1 signaling, and virus infection-related pathways were enriched. For the top downregulated genes, legionellosis, mTOR signaling pathway, skeletal muscular hypertrophy regulation related pathway, SODD/TNFR1 signaling related pathways were enriched. For SLE upregulated genes, interferon-alpha/beta signaling, immune system signaling by interferons, cytokine signaling in the immune system, cell cycle checkpoints related pathway, NOD-like receptor signaling, toll-like receptor signaling pathways and many viral infections related signaling pathways were found enriched. For the downregulated genes, T cell receptor signaling pathway, hematopoietic cell lineage, mTOR signaling pathway, multi-drug resistance factor-related pathways, and skeletal muscular hypertrophy regulation related pathway and various gene expression and transcription factors related pathways were enriched. The common gene signatures identified by us include the enriched pathways related to the mTOR signaling pathway, necroptosis, bone remodeling, Regulation of eIF4e and p70 S6 kinase pathway, translational silencing related and virus infection-related pathways. The enriched pathways complete information for common and disease-specific are listed in **Additional file 7: Supplementary Table S6**. For common genes, the enriched Gene Ontology (GO) term for GO: biological processes, nuclear-transcribed mRNA catabolic processes (GO: 0000956), viral gene expression (GO: 0019080), viral transcription (GO: 0019083), and the regulation of cytoskeleton organization (GO: 0051493). For GO: Molecular function, mRNA binding (GO: 0003729), ubiquitin-protein transferase inhibitor activity (GO: 0055105), beta-N-acetyl glucosaminylglycopeptide beta-1, 4-galactosyltransferase activity (GO: 0003831), and translation elongation factor activity (GO: 0003746). The details about the GO-terms for common and specific to RA and SLE are listed in **Additional file 8: Supplementary Table S7**.

Discussion

Previous RA and SLE gene expression meta-analyses have been limited to a few studies and did not focus on shared gene signatures [23][21]. The MetaIntegrator framework used in our study leverages biological and technical heterogeneity to discover robust gene signatures ranges from cancer to autoimmune diseases and also for other infectious diseases [29–37]. Here we performed a multi-cohort analysis of more than 900 samples from 14 datasets representing the real-world heterogeneity in terms of platform, age, sex, cell type and genetic background to identify the

common gene signatures between RA and SLE. The result of our study highlights the importance of global gene expression meta-analysis to achieve the information which cannot be discovered in the analysis of the individual datasets. MetaIntegrator framework allows us to integrate multiple datasets from different platforms in various diseases consistently. Using MetaIntegrator, we found 162 and 1630 differentially regulated genes in RA and SLE individual disease meta-analysis. We have performed a meta-analysis of publicly available RA and SLE diseases datasets and identified 33 common gene signatures with 9 upregulated and 24 downregulated genes. We found significantly enriched pathways related to the shared signature genes like mTOR signaling pathways, bone remodeling, various virus infection related pathways, PI3K-AKT signaling pathway, matrix metalloproteinase pathways, osteoclast differentiation, TNF signaling pathways, and Jak-STAT signaling pathways.

The experimental exploration of functional role of the 33 common gene signatures in the shared pathogenesis of RA and SLE will enhance our understanding of the molecular mechanisms of the disease. Previous studies conducted on individual RA and SLE datasets have shown the over-expression of type-1 IFN related genes. In this study, we identified Interferon regulatory factor 9 (IRF9) as a common gene signature for RA and SLE, which is a type-I IFN related gene with important role in interferon signaling and bone remodeling pathways. It is an integral transcription factor that mediates the type I interferon antiviral response as part of interferon-stimulated gene factor 3 (IRGF3) [38]. Many of the previous studies confirm its role in rheumatoid arthritis as it affects the fibroblast-like synoviocytes (FLS) in TNF induced RA via the SIRT-1/NF-kB signaling pathway [25] and its elevated expression is also explored in SLE [39]. Among other upregulated genes, the particularly striking gene such as MMP8 (matrix metalloproteinases 8) is a known emerging marker for several autoimmune diseases (Rheumatoid arthritis, Dermatomyositis, Inflammatory bowel diseases, Insulin-dependent diabetes mellitus, type 1 diabetes, Systemic lupus erythromatosus, Osteonecrosis of the femoral head (ONFH)) [40]. MMP8 is believed to be one of the key regulators of tissue degradation and remodeling, which plays a central role in various important processes such as morphogenesis, tissue repair, wound healing and remodeling, in response to injury [41]. Various MMPs like MMP-2, MMP-3, and MMP-9 elevated level is reported for RA and SLE [27, 42, 43]. NFIL3 (E4BP4) is an important human transcription factor in autoimmune response, the over-expression of which leads to the downregulation of autoimmune responses in SLE patients via inhibiting the expression of CD40L [44] mutations in this gene can alter the immune hypothesis and sensitize for arthritis pathology [28]. Intriguingly, we have identified two histone genes, H1ST1H1C (linker histone) and HIST2H2AA4 (core histone). The Linker histones act as a trigger in certain autoimmune diseases, as the immune system on extracellular exposure can easily target H1, resulting in activation of mast cell or B cells and ultimately developing the disease [45, 46]. It is speculated that the linker histones are crucial in the pathogenesis of various diseases. Also, it is known that linker histones can repress the transcriptional activity as these binds to transcription factors and cofactors binding site on DNA. Here in this study, we report the elevated expression of linker histone (H1ST1H1C) across RA and SLE datasets, suggesting its role in pathogenesis of the two diseases. Linker histones play a pathophysiological role in mast cell- mediated type I hypersensitivity and help in defense against virus replication indirectly. In general linker histones provides a promising way to regulate innate immunity, thus providing potential target against certain infections.

Another important gene is B4GALT5, a glycosyltranferase which catalyzes the synthesis of lactosylceramide (LacCer) by transferring galactose from UDP-galactose to glucosylceramide (GlcCer). This is a biomarker for multiple sclerosis autoimmune disease [47]. The altered glycosylation of proteins is frequently related to the abnormal expression of glycosyltranferase [48]. B4GALT5 is a glycosyltranferase present on the Golgi cell surface and involved in certain functions like cell adhesion, matrix interactions, and signaling cascades. Its expression level is significantly related with the immune infiltration level of CD4 + and macrophages cells especially [49–51]. VMP1 is another over-expressed gene, It is an ER-associated multi-spanning membrane protein, important for autophagosome formation, in turn, found central importance in immunity [52]. Another important identified gene is NEXN, a F-actin binding protein and previously identified as a disease gene causing cardiomyopathies [53]. The gene expression level of these genes has been found

elevated in many pathological diseases, corroborating, and validating our results. This further suggests their role in the shared pathogenesis of RA and SLE (**see** Fig. 5).

This study also found the downregulation of a few genes, for example, DDX60L, an interferon-stimulated gene product that restricts certain virus infections in many autoimmune diseases [54]. We found the downregulation of specific transcription factors like EIF4B, EEF1G, and NACA, which are important genes related to autoimmune regulation [55, 56]. Other important genes including NMT2, PTGDS, DUSP1, and CLIC3. Among these, NMT (N-Myristoyltransferase) gene has two isoforms NMT1 and NMT2 both are required for T-cell development and TCR signaling [57]. Wen et al. reported the decreased expression of NMT1 in the CD4 + T-cells of RA patients [58]. PTGDS (prostaglandin G2 synthase) is another important gene that might be involved in important processes like adaptive immune response, regulation of cell adhesion molecules, leukocyte migration and chemokine signaling [59]. Interestingly we found the decreased expression of Dual-specificity phosphatase 14 (DUSP14 or MKP6), it is a MAPK phosphatase, which negatively regulates the T-cell receptor signaling and immune responses via inhibiting the TAB1 activation. Thus DUSP14 mediates the inhibition of the T-cell activation signaling process [60]. Chloride intracellular channel 3 (CLIC3) is another important gene in various immune or inflammatory responses, providing a novel therapeutic target for autoimmune diseases [61].

Among all the identified common signatures, several genes have been reported as biomarkers for one or more autoimmune diseases, such as MMP8, NFIL3, B4GALT5, HIST1H1C, NMT2, PTGDS, and DUSP14. The results suggested that our findings are consistent with the previous discoveries; moreover, for the first time via our meta-analysis, we are reporting shared genetic signatures of RA and SLE.

Summarily, the etiology and pathogenesis of RA and SLE is interconnected in several ways, including elevated production of cytokines and other inflammatory mediators, dysregulation of innate and adaptive immune system, clonal expansion of autoreactive lymphocytes, and the dysregulated neutrophills activation [62][63]. The role of the identified gene signatures in this study suggests their involvement in innate as well as adaptive immune responses and other immune regulations indicating the shared pathogenesis of the two diseases. Furthermore, many of the immune cells and system regulated by these gene signatures are itself the major effectors of the inflammatory response in both the diseases.

Among the novel putative biomarkers, the particularly striking genes are DDX60L, DUSP14, HIST2H2AA4, NMT2, RPS2, RPS8, TTC3, ZMYND19, and NARS2. In these, NARS2 is a putative member of the class II family of aminoacyl-tRNA synthetase, plays a crucial role in protein biosynthesis, TTC34 is a known marker for SLE. Here we report a new member i.e., TTC3, shared by the two diseases. Various ribosomal protein members are known for different autoimmune diseases, we have identified RPS2 and RPS8 common in both diseases.

These gene signatures that are highly expressed across RA and SLE patients' datasets, conclusively suggests the involvement of these genes and associated pathways in the pathogenesis of the two diseases. However, their functional relevance regarding gene expression and the disease outcome needs to be further explored and validated in a large external validation dataset. Nevertheless, the shared gene signatures identified in our study provide novel perspectives and clues to understand the pathophysiological mechanism of RA and SLE. These signatures can help to determine the novel potential drug targets for the treatment against both diseases.

Conclusions

Exploring the precise causative factors of RA and SLE remains a challenge. Our study provides alternate biological insights regarding the common regulatory mechanism to investigate the pathological and altered gene expression signatures related to the two diseases. Moreover, the identified common signature genes may also open exploration for new potential targets for understanding the molecular mechanism and the novel treatment options for RA and SLE. The

shortlisted genes may provide the basis for the common mechanisms of pathogenesis and pave the way for exploration of novel leads for clinical diagnosis, novel biomarkers and common drug targets for the two diseases.

Abbreviations

IFN: Interferon

RA: Rheumatoid arthritis

SLE: Systemic lupus erythromatosus

PBMC: Peripheral blood mononuclear cell

SjS: Sjogren's syndrome

PPI: protein-protein interaction

DEGs: Differentially expressed genes

ES: Effect Size

AUROC: Area under the receiver operating characteristic

TNF: Tumor Necrosis Factor

IL: Interleukins

Declarations

Availability of data and materials

The datasets used and analyzed during the current study are available in NCBI Gene expression Omnibus (GEO).

Acknowledgments

We acknowledge ICGEB for providing the necessary infrastructure and facilities for the research. Junior Research Fellowship awarded to NT by GlaxoSmithKline (GSK, India), is duly acknowledged (no. RCB/PhD-BI/2019/1017).

Funding

This work was supported by the bioinformatics infrastructure grant from the Department of Biotechnology, Government of India (no. BT/PR40151/BTIS/137/5/2021).

Author information

Affiliations

Translational Bioinformatics Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India, and Regional Centre for Biotechnology, Faridabad.

Neetu Tyagi

Translational Bioinformatics Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India.

Dr. Dinesh Gupta

Contributions

NT contributes to perform the experiment and analyzed the data. NT wrote the first version of the manuscript. DG provide the conception and design of the study with the guidance at each step of analysis. DG revised the manuscript for important intellectual content and approved the final script.

Corresponding author

Correspondence to Dinesh Gupta

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Karopka T, Fluck J, Mevissen HT, Glass Ä. The Autoimmune disease database: A dynamically compiled literature-derived database. BMC Bioinformatics. 2006;7:325. doi:10.1186/1471-2105-7-325.
2. Mitchell DM, Spitz PW, Young DY, Bloch DA, McShane DJ, Fries JF. Survival, prognosis, and causes of death in rheumatoid arthritis. Arthritis Rheum. 1986;29:706–14. doi:10.1002/art.1780290602.
3. Haynes WA, Haddon DJ, Diep VK, Khatri A, Bongen E, Yiu G, Balboni I, Bolen CR, Mao R, Utz PJ, et al. Integrated, multicohort analysis reveals unified signature of systemic lupus erythematosus. JCI Insight 2020, 5, doi:10.1172/jci.insight.122312.
4. Pabón-Porras MA, Molina-Ríos S, Flórez-Suárez JB, Coral-Alvarado PX, Méndez-Patarroyo P, Quintana-López G. Rheumatoid arthritis and systemic lupus erythematosus: Pathophysiological mechanisms related to innate immune system. SAGE Open Med. 2019;7:205031211987614. doi:10.1177/2050312119876146.
5. Tang S, Lui SL, Lai KN. Pathogenesis of lupus nephritis: An update. Nephrology. 2005;10:174–9.
6. Lynn AH, Kwoh CK, Venglish CM, Aston CE, Chakravarti A. Genetic epidemiology of rheumatoid arthritis. Am J Hum Genet. 1995;57:150–9.
7. Burska AN, Roget K, Blits M, Soto Gomez L, Van De Loo F, Hazelwood LD, Verweij CL, Rowe A, Goulielmos GN, Van Baarsen LGM, et al. Gene expression analysis in RA: Towards personalized medicine. Pharmacogenomics J. 2014;14:93–106. doi:10.1038/tpj.2013.48.
8. *Genes Immun.* **2013**, *14*, 67–82, doi:10.1038/gene.2012.55.

9. Van ’t Veer LJ, Bernards R. Enabling personalized cancer medicine through analysis of gene-expression patterns. Nature. 2008;452:564–70.
10. Zhang Z, Fenstermacher D An introduction to MAMA (meta-analysis of micro array data) system. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings; 2005; Vol. 7 VOLS, pp. 7730–7733.
11. Ma T, Huo Z, Kuo A, Zhu L, Fang Z, Zeng X, Lin C-W, Liu S, Wang L, Liu P, et al. MetaOmics: analysis pipeline and browser-based software suite for transcriptomic meta-analysis. Bioinformatics. 2019;35:1597–9. doi:10.1093/bioinformatics/bty825.
12. Xia J, Fjell CD, Mayer ML, Pena OM, Wishart DS, Hancock REW. INMEX — a web-based tool for integrative meta-analysis of expression data. 2013, *41*, 63–70, doi:10.1093/nar/gkt338.
13. Hughey JJ, Butte AJ. Robust meta-analysis of gene expression using the elastic net. 2015, *43*, 1–11, doi:10.1093/nar/gkv229.
14. Vallania F, Tomczak A, Andres-terre M. Empowering Multi-Cohort Gene Expression Analysis to Increase Reproducibility Empowering Multi-Cohort Gene Expression Analysis to Increase Reproducibility. 2016, doi:10.1101/071514.
15. Song GG, Kim JH, Seo YH, Choi SJ, Ji JD, Lee YH. Meta-analysis of differentially expressed genes in primary Sjogren’s syndrome by using microarray. Hum Immunol. 2014;75:98–104. doi:10.1016/j.humimm.2013.09.012.
16. Arasappan D, Tong W, Mummaneni P, Fang H, Amur S. Meta-analysis of microarray data using a pathway-based approach identifies a 37-gene expression signature for systemic lupus erythematosus in human peripheral blood mononuclear cells. BMC Med 2011, 9, doi:10.1186/1741-7015-9-65.
17. Olsen NJ, Sokka T, Seehorn CL, Kraft B, Maas K, Moore J, Aune TM. A gene expression signature for recent onset rheumatoid arthritis in peripheral blood mononuclear cells. Ann Rheum Dis. 2004;63:1387–92. doi:10.1136/ard.2003.017194.
18. Olsen NJ, Moore JH, Aune TM. Gene expression signatures for autoimmune disease in peripheral blood mononuclear cells. Arthritis Res Ther. 2004;6:120–8. doi:10.1186/ar1190.
19. Afroz S, Giddaluru J, Vishwakarma S, Naz S, Khan AA, Khan. N. A comprehensive gene expression meta-analysis identifies novel immune signatures in rheumatoid arthritis patients. Front Immunol 2017, 8, doi:10.3389/fimmu.2017.00074.
20. Kröger W, Mapiye D, Entfellner JBD, Tiffin N. A meta-analysis of public microarray data identifies gene regulatory pathways deregulated in peripheral blood mononuclear cells from individuals with Systemic Lupus Erythematosus compared to those without. BMC Med Genomics. 2016;9:1–11. doi:10.1186/s12920-016-0227-0.
21. Silva GL, Junta CM, Mello SS, Garcia PS, Rassi DM, Sakamoto-Hojo ET, Donadi EA, Passos GAS. Profiling meta-analysis reveals primarily gene coexpression concordance between systemic lupus erythematosus and rheumatoid arthritis. Ann N Y Acad Sci. 2007;1110:33–46. doi:10.1196/annals.1423.005.
22. Higgs BW, Liu Z, White B, Zhu W, White WI, Morehouse C, Brohawn P, Kiener PA, Richman L, Fiorentino D, et al. Patients with systemic lupus erythematosus, myositis, rheumatoid arthritis and scleroderma share activation of a common type I interferon pathway. Ann Rheum Dis. 2011;70:2029–36. doi:10.1136/ard.2011.150326.
23. Toro-Domínguez D, Carmona-Sáez P, Alarcón-Riquelme ME. Shared signatures between rheumatoid arthritis, systemic lupus erythematosus and Sjögren’s syndrome uncovered through gene expression meta-analysis. Arthritis Res Ther. 2014;16:1–8. doi:10.1186/s13075-014-0489-x.
24. Badr MT, Häcker G. Gene expression profiling meta-analysis reveals novel gene signatures and pathways shared between tuberculosis and rheumatoid arthritis. PLoS One. 2019;14:1–16. doi:10.1371/journal.pone.0213470.

25. Jiang F, Zhou HY, Zhou LF, Zeng W, Zhao LH. IRF9 Affects the TNF-Induced Phenotype of Rheumatoid-Arthritis Fibroblast-Like Synoviocytes via Regulation of the SIRT-1/NF-κB Signaling Pathway. Cells Tissues Organs. 2020;209:110–9. doi:10.1159/000508405.
26. Zhao GN, Jiang DS, Li H. Interferon regulatory factors: At the crossroads of immunity, metabolism, and disease. Biochim Biophys Acta - Mol Basis Dis. 2015;1852:365–78. doi:10.1016/j.bbadis.2014.04.030.
27. Cox JH, Starr AE, Kappelhoff R, Yan R, Roberts CR, Overall CM. Matrix metalloproteinase 8 deficiency in mice exacerbates inflammatory arthritis through delayed neutrophil apoptosis and reduced caspase 11 expression. Arthritis Care Res. 2010;62:3645–55. doi:10.1002/art.27757.
28. Schlenner S, Pasciuto E, Lagou V, Burton O, Prezzemolo T, Junius S, Roca CP, Seillet C, Louis C, Dooley J, et al. NFIL3 mutations alter immune homeostasis and sensitise for arthritis pathology. Ann Rheum Dis. 2019;78:342–9. doi:10.1136/annrheumdis-2018-213764.
29. Haynes WA, James Haddon D, Diep VK, Khatri A, Bongen E, Yiu G, Balboni I, Bolen CR, Mao R, Utz PJ, et al. A unified molecular signature of systemic lupus erythematosus revealed by integrated, multi-cohort transcriptomic analysis. bioRxiv. 2019;94305:1–73. doi:10.1101/834093.
30. Robinson M, Sweeney TE, Barouch-Bentov R, Sahoo MK, Kalesinskas L, Vallania F, Sanz AM, Ortiz-Lasso E, Albornoz LL, Rosso F, et al. A 20-Gene Set Predictive of Progression to Severe Dengue. Cell Rep. 2019;26:1104–11.e4. doi:10.1016/j.celrep.2019.01.033.
31. Bongen E, Lucian H, Khatri A, Fragiadakis GK, Bjornson ZB, Nolan GP, Utz PJ, Khatri P. Sex Differences in the Blood Transcriptome Identify Robust Changes in Immune Cell Proportions with Aging and Influenza Infection. Cell Rep. 2019;29:1961–73.e4. doi:10.1016/j.celrep.2019.10.019.
32. Chen R, Khatri P, Mazur PK, Polin M, Zheng Y, Vaka D, Hoang CD, Shrager J, Xu Y, Vicent S, et al. A meta-Analysis of lung cancer gene expression identifies PTK7 as a survival gene in lung adenocarcinoma. Cancer Res. 2014;74:2892–902. doi:10.1158/0008-5472.CAN-13-2775.
33. Khatri P, Roedder S, Kimura N, Vusser K, De; Morgan AA, Gong Y, Fischbein MP, Robbins RC, Naesens M, Butte AJ, et al. A common rejection module (CRM) for acute rejection across multiple organs identifies novel therapeutics for organ transplantation. 2013, *210*, 2205–2221, doi:10.1084/jem.20122709.
34. Sweeney TE, Shidham A, Wong HR, Khatri P. A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. Sci Transl Med. 2015;7:1–16. doi:10.1126/scitranslmed.aaa5993.
35. Andres-terre M, Mcguire HM, Pouliot Y, Bongen E, Sweeney TE, Tato CM, Andres-terre M, Mcguire HM, Pouliot Y, Bongen E, et al. Integrated, Multi-cohort Analysis Identifies Conserved Transcriptional Signatures across Resource Integrated, Multi-cohort Analysis Identifies Conserved Transcriptional Signatures across Multiple Respiratory Viruses. Immunity. 2015;43:1199–211. doi:10.1016/j.immuni.2015.11.003.
36. Li MD, Burns TC, Morgan AA, Khatri P Integrated multi-cohort transcriptional meta-analysis of neurodegenerative diseases. 2014, 1–23.
37. Lofgren S, Hinchcliff M, Carns M, Wood T, Aren K, Arroyo E, Cheung P, Kuo A, Valenzuela A, Haemel A, et al. Integrated, multicohort analysis of systemic sclerosis identifies robust transcriptional signature of disease severity. Ref Inf JCI Insight. 2016;1:89073. doi:10.1172/jci.insight.89073.
38. Paul A, Tang TH, Ng SK. Interferon regulatory factor 9 structure and regulation. Front Immunol. 2018;9:1–9. doi:10.3389/fimmu.2018.01831.
39. Thibault DL, Chu AD, Graham KL, Balboni I, Lee LY, Kohlmoos C, Landrigan A, Higgins JP, Tibshirani R, Utz PJ. IRF9 and STAT1 are required for IgG autoantibody production and B cell expression of TLR7 in mice. J Clin Invest 2008, 118, doi:10.1172/JCI30065.

40. An F, Du J, Cao Y, Shi J, Guo Y, Jin T, Li J, Chen J, Li P, Dong M, et al. *MMP8 polymorphism is associated with susceptibility to osteonecrosis of the femoral head in a Chinese Han population*.
41. Nagase H, Visse R, Murphy G. Structure and function of matrix metalloproteinases and TIMPs. 2006, doi:10.1016/j.cardiores.2005.12.002.
42. Chang YH, Lin IL, Tsay GJ, Yang SC, Yang TP, Ho KT, Hsu TC, Shiau MY. Elevated circulatory MMP-2 and MMP-9 levels and activities in patients with rheumatoid arthritis and systemic lupus erythematosus. Clin Biochem. 2008;41:955–9. doi:10.1016/j.clinbiochem.2008.04.012.
43. Lee JM, Kronbichler A, Park SJ, Kim SH, Han KH, Kang HG, Ha IS, Cheong H, Il; Kim KH, Kim G, et al. Association between Serum Matrix Metalloproteinase- (MMP-) 3 Levels and Systemic Lupus Erythematosus: A Meta-analysis. *Dis. Markers* 2019, *2019*, doi:10.1155/2019/9796735.
44. Kim HS, Sohn H, Jang SW, Lee GR. The transcription factor NFIL3 controls regulatory T-cell function and stability. Exp Mol Med 2019, 51, doi:10.1038/s12276-019-0280-9.
45. Araki Y, Mimura T The Histone Modification Code in the Pathogenesis of Autoimmune Diseases. *Mediators Inflamm.* 2017, *2017*, doi:10.1155/2017/2608605.
46. Nemtsova MV, Zaletaev DV, Bure IV, Mikhaylenko DS, Kuznetsova EB, Alekseeva EA, Beloukhova MI, Deviatkin AA, Lukashev AN, Zamyatnin AA. Epigenetic changes in the pathogenesis of rheumatoid arthritis. Front Genet. 2019;10:1–13. doi:10.3389/fgene.2019.00570.
47. Mayo L, Trauger SA, Blain M, Nadeau M, Patel B, Jorge I, Mascanfroni ID, Yeste A, Kivisäkk P, Kallas K, et al. B4GALT6 regulates astrocyte activation during CNS inflammation. Nat Med. 2014;20:1147–56. doi:10.1038/nm.3681.B4GALT6.
48. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. Nat Rev Cancer. 2015;2015 159:*15*, 540–55. doi:10.1038/nrc3982.
49. Han Y, Li Z, Wu Q, Liu H, Sun Z, Wu Y, Luo J. B4GALT5 High Expression Associated With Poor Prognosis of Hepatocellular Carcinoma., doi:10.21203/rs.3.rs-659248/v1.
50. C R, BD S. Targeted mutations in beta1,4-galactosyltransferase I reveal its multiple cellular functions. Biochim Biophys Acta. 2002;1573:258–70. doi:10.1016/S0304-4165(02)00392-6.
51. Youakim A, Dubois DH, Shur BD Localization of the long form of beta-1,4-galactosyltransferase to the plasma membrane and Golgi complex of 3T3 and F9 cells by immunofluorescence confocal microscopy. *Proc. Natl. Acad. Sci. U. S. A.* 1994, *91*, 10913, doi:10.1073/PNAS.91.23.10913.
52. Wang P, Kou D, Le W. Roles of VMP1 in Autophagy and ER–Membrane Contact: Potential Implications in Neurodegenerative Disorders. Front Mol Neurosci 2020, 13, doi:10.3389/fnmol.2020.00042.
53. Wang H, Li Z, Wang J, Sun K, Cui Q, Song L, Zou Y, Wang X, Liu X, Hui R, et al. Mutations in NEXN, a Z-Disc gene, are associated with hypertrophic cardiomyopathy. Am J Hum Genet. 2010;87:687–93. doi:10.1016/j.ajhg.2010.10.002.
54. Grünvogel O, Esser-Nobis K, Reustle A, Schult P, Müller B, Metz P, Trippler M, Windisch MP, Frese M, Binder M, et al. DDX60L Is an Interferon-Stimulated Gene Product Restricting Hepatitis C Virus Replication in Cell Culture. J Virol. 2015;89:10548–68. doi:10.1128/jvi.01297-15.
55. Kroczynska B, Mehrotra S, Arslan AD, Kaur S, Platanias LC. Regulation of interferon-dependent mRNA translation of target genes. J Interf Cytokine Res. 2014;34:289–96. doi:10.1089/jir.2013.0148.
56. Luan M, Shang Z, Teng Y, Chen X, Zhang M, Lv H, Zhang R. The shared and specific mechanism of four autoimmune diseases. Oncotarget. 2017;8:108355–74. doi:10.18632/oncotarget.19383.
57. Finlay DK. N-myristoylation of AMPK controls T cell inflammatory function. Nat Immunol. 2019;20:252–4.
58. Wen Z, Jin K, Shen Y, Yang Z, Li Y, Wu B, Tian L, Shoor S, Roche NE, Goronzy JJ, et al. N-myristoyltransferase deficiency impairs activation of kinase AMPK and promotes synovial tissue inflammation. Nat Immunol.

2019;20:313–25. doi:10.1038/s41590-018-0296-7.
59. Zou R, Zheng M, Tan M, Xu H, Luan N, Zhu L. Decreased ptgds expression predicting poor survival of endometrial cancer by integrating weighted gene co-expression network analysis and immunohistochemical validation. Cancer Manag Res. 2020;12:5057–75. doi:10.2147/CMAR.S255753.
60. Tan T-H, Chen D-Y, Chuang H-C, Huang C-Y, Yang C-Y, Li J-P, Chiu L-L, Lan J-L. Signaling by Inhibiting TAB1 Activation (DUSP14/MKP6) Negatively Regulates TCR Dual-Specificity Phosphatase 14. 2014, doi:10.4049/jimmunol.1300989.
61. Salao K. *The role of chloride intracellular channel 1 in immune/inflammatory responses*; 2016.
62. Fresneda Alarcon M, McLaren Z, Wright HL. Neutrophils in the Pathogenesis of Rheumatoid Arthritis and Systemic Lupus Erythematosus: Same Foe Different M.O. Front Immunol. 2021;0:570. doi:10.3389/FIMMU.2021.649693.
63. Pabón-Porras MA, Molina-Ríos S, Flórez-Suárez JB, Coral-Alvarado PX, Méndez-Patarroyo P, Quintana-López G. Rheumatoid arthritis and systemic lupus erythematosus: Pathophysiological mechanisms related to innate immune system: https://doi*org*/10.1177/2050312119876146 2019, 7, 205031211987614, doi:10.1177/2050312119876146.
64. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. 2007, *23*, 1846–1847, doi:10.1093/bioinformatics/btm254.
65. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Web Serv issue Publ online 2019, 47, doi:10.1093/nar/gkz240.
66. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016, 44, doi:10.1093/nar/gkw377.

Tables

Table 1 GEO datasets and samples summary for RA and SLE.

SLE datasets

S.No.	GSE_ids	Source	Case	Control	Used Samples	Total Samples	PMID	Centre
1	GSE11909	PBMC	63 SLE	12H	75	175	18631455	Baylor University, Texas, USA
2	GSE50772	PBMC	61	20	81	81	25861459	ITGR Diagnostics Discovery, South San Francisco
3	GSE22098	Whole blood	28 adults + 12 pediatric SLE	42 H	83	274	20725040	Baylor University, Texas, USA
4	GSE8650	PBMC	38	21H	59	246	17724127	Baylor University, TX, Texas, USA
5	GSE4588	CD4 T and B cells	26	15	41	49	NA	Université catholique de Louvain, Institut de Recherches Expérimentales et Cliniques, Brussels, Belgium
6	GSE61635	PBMC	99	30	129	129	NA	Eli Lilly and Company, USA
7	GSE10325	PBMC	39	28	67	67	23826184	UT Southwestern Medical Center , USA
RA datasets								
1	GSE56649	peripheral blood	13 RA	9(H)	22	22	25880754	Peking University, People's Hospital, China
2	GSE15573	PBMC	18(RA)	15(C)	33	33	19710928	CEA IG, Gene Expression Platform, Evry, France
3	GSE4588	CD4 B and T cells	15(RA)	19 (C)	34	49	NA	Université catholique de Louvain, Institut de Recherches Expérimentales et Cliniques,

								Brussels, Belgium
4	GSE17755	Peripheral blood	RA(112)	HI (45), HC(8)	165	165	21496236	Wakayama Medical University, Laboratory of Immune
5	GSE11827	PBMC	27(RA),	22(control)	49	49	21059672	Faculty of Medicine and Pharmacy, Inserm Unit 519, ROUEN, France
6	GSE57383	Peripheral blood	RA(27)	HC(19)	46	112	25333715	University of Rochester Medical Center, School of Medicine and Dentistry, Rochester, USA
7	GSE68689	whole blood	10(RA)	5 (C) (6 biological replicates)	21	21	NA	Selventa, Cambridge, USA

Figures

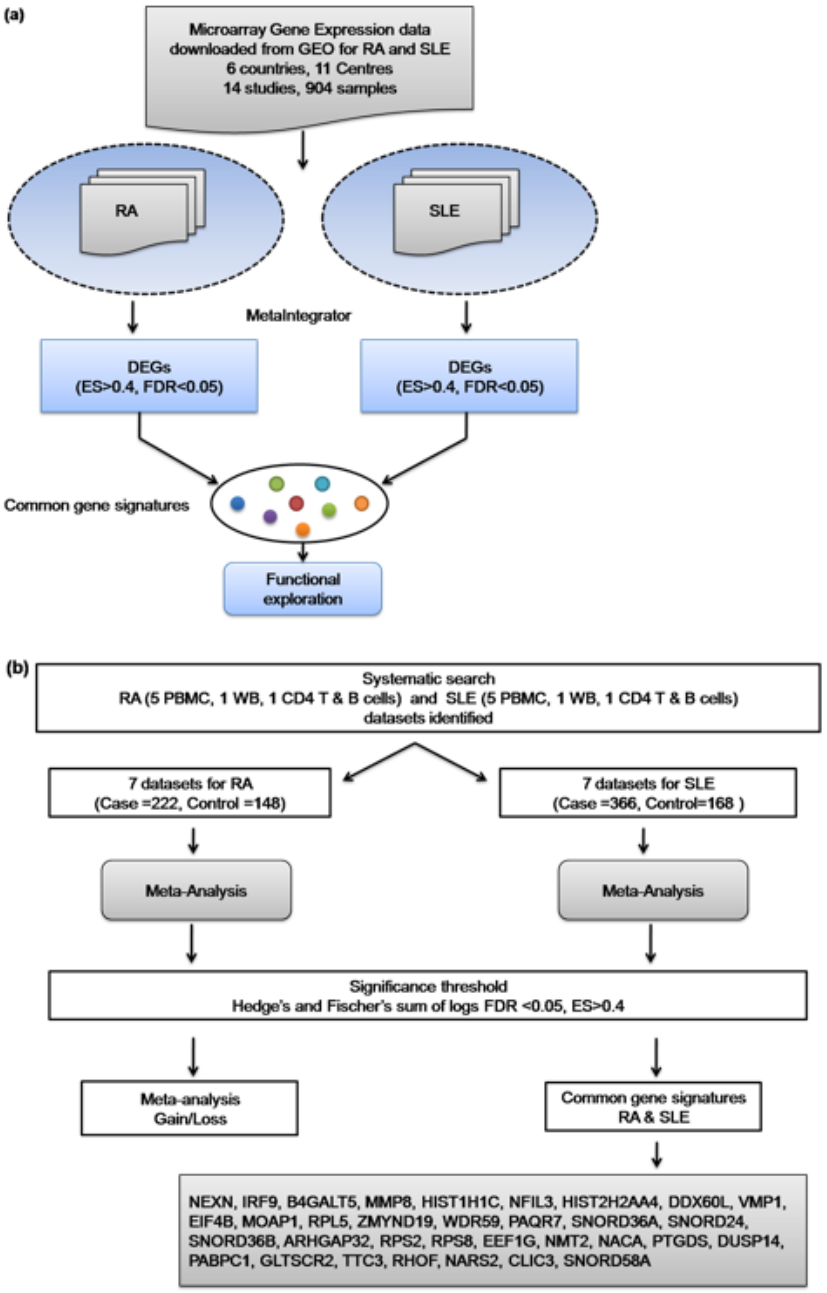

Figure 1

Meta-analysis workflow of the study: (a) A brief outline of the study about countries or centres and the overall idea. (b) Detailed workflow of datasets, meta-analysis, and signatures information.

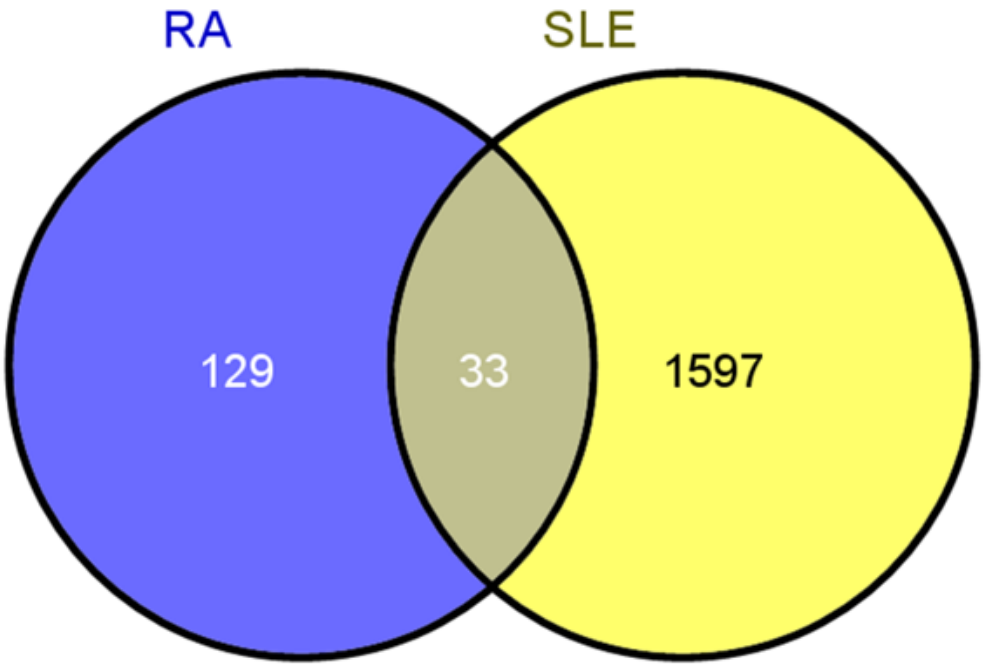

Figure 2

A Venn diagram of DEGs obtained via individual meta-analysis for both diseases. The intersection showed the common gene signatures of RA and SLE.

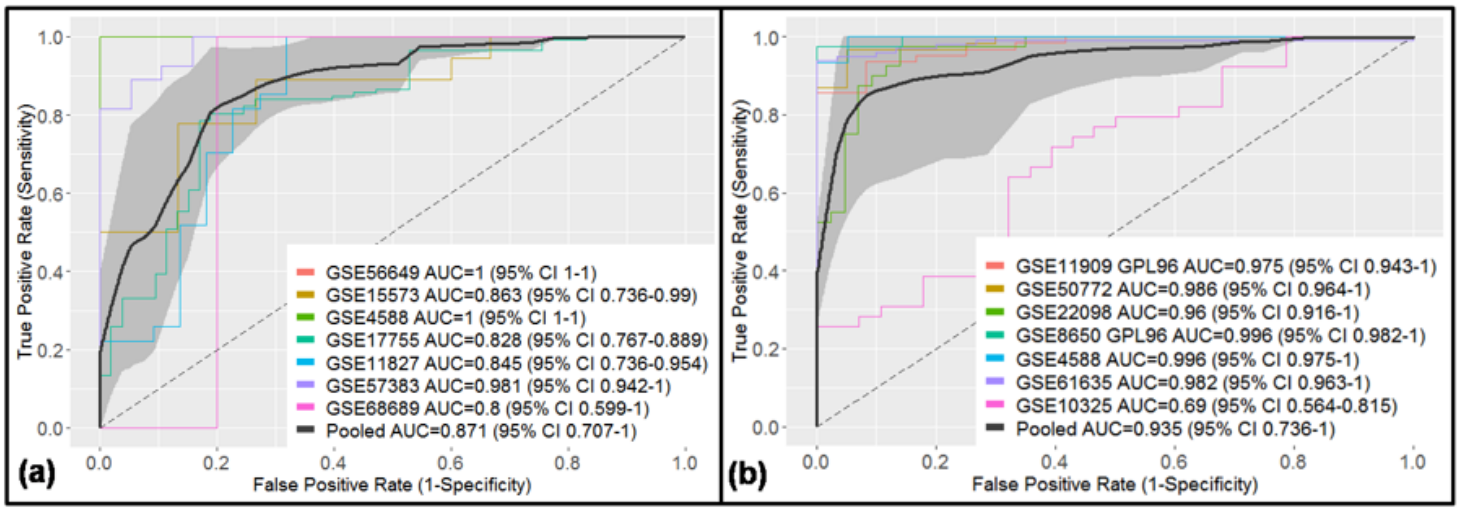

Figure 3

Receiver operating characteristic curves of RA and SLE. (a) For RA datasets and (b) for SLE datasets. A perfect classifier must have an AUROC of 1, while a random classifier has an AUC of 0.5. Here the summary curve is a composite of the individual studies from PBMC, WB, and CD4 T and B Cells samples. The precision curve for RA and SLE datasets is shown in Figure.S2.

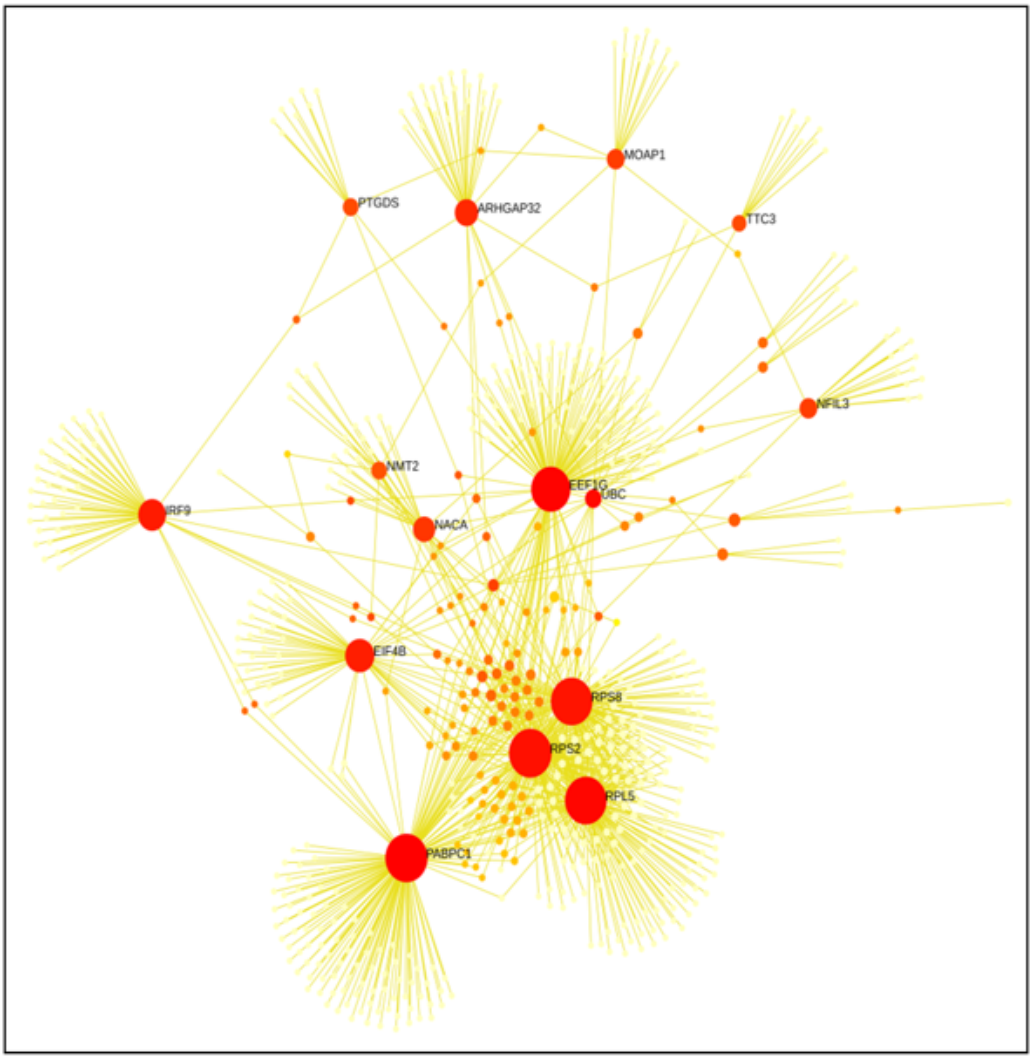

Figure 4

Receiver operating characteristic curves of RA and SLE. (a) For RA datasets and (b) for SLE datasets. A perfect classifier must have an AUROC of 1, while a random classifier has an AUC of 0.5. Here the summary curve is a composite of the individual studies from PBMC, WB, and CD4 T and B Cells samples. The precision curve for RA and SLE datasets is shown in Figure.S2.

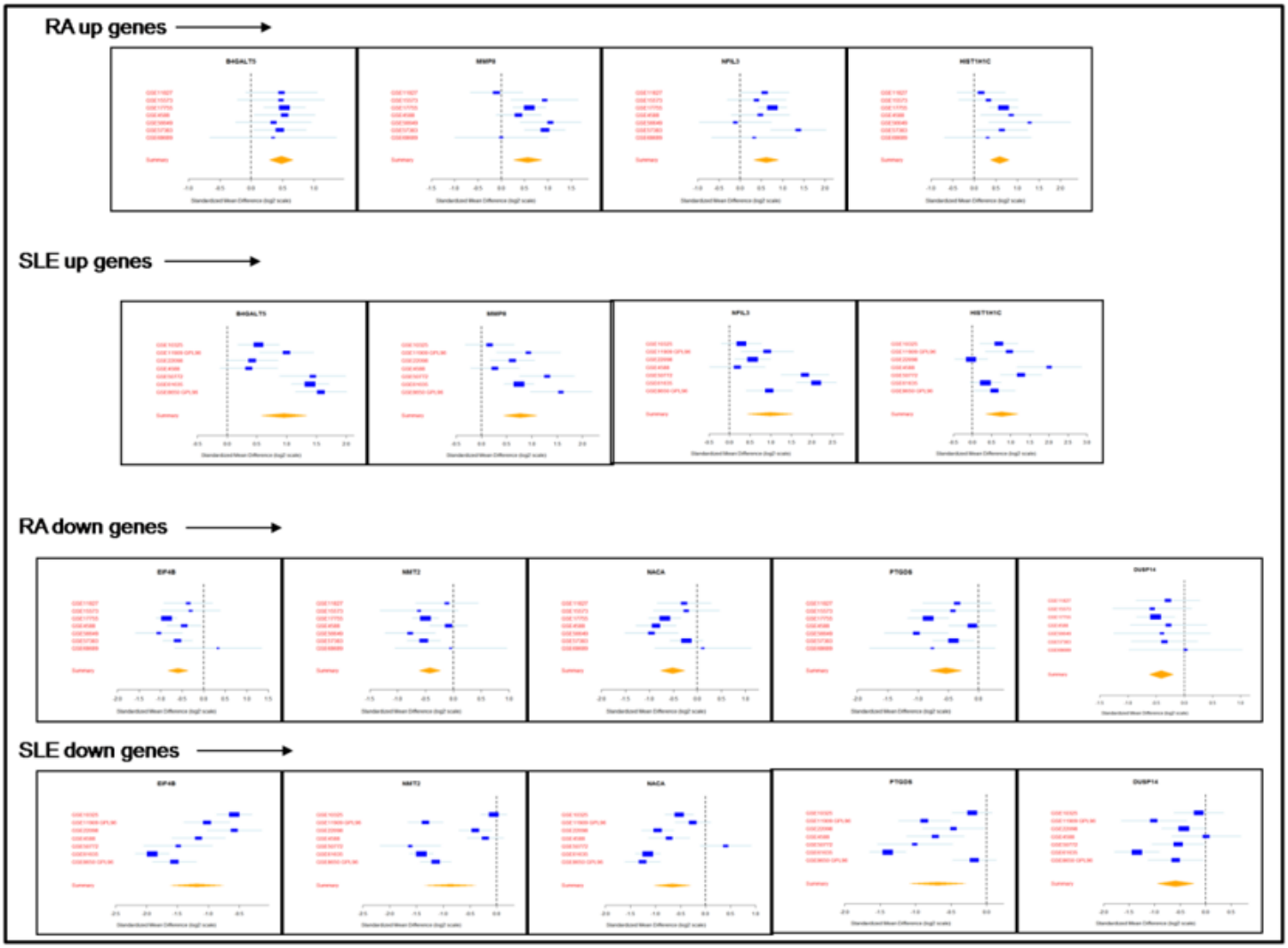

Figure 5

Forest plot generated from MetaIntegrator for the shared gene signatures of RA and SLE, expression in all 7 studies for each disease. Forest plot x-axis shows the standardized mean difference (log2 scale) for genes in multiple studies. The size of the blue box is inversely proportional to the standardarized mean difference of the gene in each study. Whiskers represent 95 % confidence interval. The yellow diamond represents the combined mean difference for each gene and its width denotes the 95% confidence interval.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableS1.doc
- SupplementaryTableS2.doc
- SupplementaryTableS3.xls
- SupplementaryTableS4.xls
- SupplementaryTableS5.xls
- SupplementaryTableS6.xls

- SupplementaryTableS7.xls
- Supplementaryfigures.doc