Transcription Factor Activity Inference in Systemic Lupus Erythematosus

Raul Lopez-Dominguez [1], **Daniel Toro-Dominguez** [1], **Jordi Martorell-Marugan** [1,2], **Adrian Garcia-Moreno** [1], **Christian H. Holland** [3,4], **Julio Saez-Rodriguez** [3,4], **Daniel Goldman** [5], **Michelle A. Petri** [5], **Marta E. Alarcon-Riquelme** [1,6,*] and **Pedro Carmona-Saez** [1,4,7,*]

1. Centre for Genomics and Oncological Research, Pfizer-University of Granada-Andalusian Regional Government, 18016 Granada, Spain; raul.lopez@genyo.es (R.L.-D.); daniel.toro@genyo.es (D.T.-D.); jordi.martorell@genyo.es (J.M.-M.); adrian.garcia@genyo.es (A.G.-M.)
2. Atrys Health S.A., 08025 Barcelona, Spain
3. Faculty of Medicine, Institute of Computational Biomedicine, Heidelberg University, 69120 Heidelberg, Germany; christian.holland@bioquant.uni-heidelberg.de (C.H.H.); julio.saez@bioquant.uni-heidelberg.de (J.S.-R.)
4. Joint Research Center for Computational Biomedicine, Faculty of Medicine, RWTH Aachen University, 52062 Aachen, Germany
5. Division of Rheumatology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; dgoldm11@jhmi.edu (D.G.); mpetri@jhmi.edu (M.A.P.)
6. Unit of Chronic Inflammatory Diseases, Institute of Environmental Medicine, Karolinska Institutet, 17177 Solna, Sweden
7. Biostatistics, Department of Statistics, Faculty of Medicine, University of Granada, 18071 Granada, Spain
* Correspondence: marta.alarcon@genyo.es (M.E.A.-R.); pcarmona@ugr.es (P.C.-S.)

Abstract: Background: Systemic Lupus Erythematosus (SLE) is a systemic autoimmune disease with diverse clinical manifestations. Although most of the SLE-associated loci are located in regulatory regions, there is a lack of global information about transcription factor (TFs) activities, the mode of regulation of the TFs, or the cell or sample-specific regulatory circuits. The aim of this work is to decipher TFs implicated in SLE. Methods: In order to decipher regulatory mechanisms in SLE, we have inferred TF activities from transcriptomic data for almost all human TFs, defined clusters of SLE patients based on the estimated TF activities and analyzed the differential activity patterns among SLE and healthy samples in two different cohorts. The Transcription Factor activity matrix was used to stratify SLE patients and define sets of TFs with statistically significant differential activity among the disease and control samples. Results: TF activities were able to identify two main subgroups of patients characterized by distinct neutrophil-to-lymphocyte ratio (NLR), with consistent patterns in two independent datasets—one from pediatric patients and other from adults. Furthermore, after contrasting all subgroups of patients and controls, we obtained a significant and robust list of 14 TFs implicated in the dysregulation of SLE by different mechanisms and pathways. Among them, well-known regulators of SLE, such as STAT or IRF, were found, but others suggest new pathways that might have important roles in SLE. Conclusions: These results provide a foundation to comprehend the regulatory mechanism underlying SLE and the established regulatory factors behind SLE heterogeneity that could be potential therapeutic targets.

Keywords: transcription factor activity inference; clustering analysis; systemic lupus erythematosus; disease classification

1. Introduction

Systemic Lupus Erythematosus (SLE) is a complex and multisystemic autoimmune disease characterized by the production of autoantibodies leading to chronic inflammation and organ damage. SLE is very heterogeneous, with possible affliction in almost any organ and diverse clinical manifestations including skin rashes, arthritis and renal failure [1].

SLE is known to be much more prevalent in females than in males and when it does occur in men tends to run a more severe course [2]. SLE patients usually suffer uncertain courses of flares, improvement and remission of disease activity, which makes this disease unpredictable. Disease activity can be measured by several indices, but one of the most accepted is the SLE Disease Activity Index (SLEDAI). It is measured by 24 laboratory and weighted clinical variables in nine organ systems including renal, skin and arthritis [3]. SLE is a challenge for researchers because of the difficulty of making an early diagnosis and the fact that the available drugs do not have a therapeutic effect for all patients [4].

With the development of genomics technologies, our knowledge about the pathogenesis of SLE and the underlying molecular mechanisms has significantly increased. In this context, Genome Wide Association Studies (GWAS) have been widely used to identify novel susceptibility genes in SLE, with more than 60 loci found to be associated with the disease [5], including, for example, alleles of genes located in the human leukocyte antigen (HLA) region [6]. Gene expression signatures have been also largely analyzed across different cohorts, leading to the uncovering of abnormal expression of interferon type I inducible genes in SLE patients, which has been termed the interferon gene signature [7,8]. Recent works have attempted to establish a molecular classification of SLE using omics data [9–11]. Banchereau et al. [9] established seven groups of SLE based on gene expression correlation with disease activity. Toro-Dominguez et al. [10] described longitudinal SLE subgroups that differ in clinical features, with two groups that showed high correlation with the percentage of neutrophils and lymphocytes. Finally, an integrative analysis of transcriptome and methylation data stratified different systemic autoimmune diseases, including SLE, into four groups defined by genetic, clinical, serological, and cellular features [11]. Despite all these significant discoveries, we are still at an early stage in the understanding of the mechanistic molecular networks that drive SLE. Besides, different studies have reported that autoimmune disease-associated loci are enriched in transcription factor binding sites and gene regulatory regions [12,13]. These findings are mainly based on the analysis of sequence motifs in SLE-loci or physical interactions with transcription factors (TFs), which do not provide information about TFs activities, the mode of regulation of the TFs, or cell or sample-specific regulatory circuits.

Deciphering TF activity can provide models of regulatory mechanisms to explain the observed changes in gene expression patterns. During the last few years, different studies have shown that TF activity can be inferred from the expression levels of its targeted genes, known as TF-regulon [14]. This approach has been successfully used for the analysis of several types of cancer [15] or to evaluate drug sensitivity [16], among other applications.

In this work, we have systematically analyzed the TFs activity patterns of almost all known human TF-regulons in two gene expression cohorts of SLE patients and control samples. We first estimated the activity of TFs for each sample using DoRothEA [17], a curated target gene database, and analyzed TFs that showed a marked differential activity in SLE with respect to healthy samples. We also explored the stratification of SLE patients based on TFs activity profiles in both SLE cohorts. We obtained consistent results from the analysis of the two independent datasets, finding two main groups of SLE patients based on TFs activity patterns, which were in agreement with previous evidence about SLE stratification [10]. Moreover, we defined the set of TFs with differential activity in SLE samples that provides evidence about the regulatory circuits associated with SLE and disease manifestations.

2. Materials and Methods

2.1. Data Selection and Preprocessing

We used two independent gene expression datasets, one from adult and another from pediatric SLE patients, as previously described [10]. The pediatric set is composed of 158 SLE patients and 46 healthy controls and the adult dataset comprises 301 SLE patients and 20 healthy samples.

Raw data files were downloaded from Gene Expression Omnibus (GEO) (GSE65391 and GSE121239). For the adult dataset, Affymetrix Cel files were used and background correction and normalization were performed with rma function from affy R package. Normalized values of the pediatric dataset were downloaded from GEO series GSE65391 using the R package GEOquery. Sample processing pipeline is described in the original publication [9]. Array probes were annotated with gene symbols and duplicated genes were merged to their median expression value. SLE samples with SLEDAI greater than 5 were selected first because a SLEDAI score > 5 is associated with a probability of initiating therapy in > 50% of cases [18]. Then, for each patient only the sample with the highest score was used for further analysis. Table 1 shows a summary of the dataset including clinical characteristics that were used in our analysis, such as total white blood cell counts (WBC), erythrocyte sedimentation rate (ESR), levels of C3 and C4 in serum, and percentages of lymphocytes, neutrophils and monocytes. Pediatric data was obtained following the protocols approved by the Institutional Review Boards at the University of Texas Southwestern Medical Center (092010-067) and Baylor University Medical Center (011-200) and informed consent was obtained from adults and the parents or guardians of those younger than 18 years of age according to [9]. Adult data from the SPARE [19] were obtained following the protocol approved by Johns Hopkins University School of Medicine Institutional Review Board. Patients were enrolled from the Hopkins Lupus Cohort after informed consent was obtained and the studies were carried out in accordance with the Helsinki Declaration.

Table 1. Clinical characterization of Systemic Lupus Erythematosus (SLE) patients.

	Adult	Pediatric
Gender	67 female and 2 male	102 female and 14 male
SLEDAI	8.493 ± 2.5	13.371 ± 6.6
% Neutrophil	67.289 ± 15.4	63.963 ± 14.8
% Lymphocyte	22.641 ± 12.9	24.808 ± 12.4
% Monocyte	7.68 ± 4.0	7.415 ± 3.8
C3 (mg/dL)	99.667 ± 39.0	78.645 ± 37.5
C4 (mg/dL)	17.87 ± 10.2	12.495 ± 9.5
WBC (K/cu mm)	6.396 ± 3.3	6.507 ± 2.9
ESR (mm/h)	41.765 ± 31.2	50.038 ± 37.0
Proteinuria	20 patients with proteinuria and 49 without	69 patients with proteinuria and 47 without
Pyuria	11 patients with pyuria and 58 without	41 patients with pyuria and 75 without

2.2. Inferring Transcription Factor Activities

Transcription factor activity from the two independent cohorts was estimated using DoRothEA, a curated database of transcription factor-target gene interactions. The methodology is described in detail in [16,17]. Briefly, the normalized gene expression levels were scaled and recentered and the transcription factor activity was calculated using aREA (analytic rank-based enrichment analysis) algorithm from VIPER R package. Briefly, this algorithm performs an enrichment analysis of the ranked gene expression signature for each TF-regulon, inferring the TF activity of each TF using the expression of its target genes in a sample-to-sample approach. We used the TF regulons from DoRotheA with the highest confidence in order to avoid false positives. The final consensus transcription factor regulons (CTFRs) consisted of 168 TFs and 2602 unique targets. This method allowed us to build a transcription factor activity matrix with TFs in rows and samples in columns and each entry of the matrix representing a normalized enriched score of transcription factor activity in each sample.

2.3. Subgroup Identification

Hierarchical clustering analysis (using Euclidean distance and average agglomerative method) and principal component analysis were applied to the TFs activity matrix in

order to establish sets of samples with similar TF activity patterns. To define the number of clusters for each dataset we used Calinski and Harabasz index implemented in the R package NbClust.

2.4. Differential Activity and Statistical Analysis

The TFs activity matrix was also used to analyze differential activity between SLE and healthy controls in each dataset. We performed this analysis using linear models implemented in limma R package. Transcription factors with False Discovery Rate (FDR) < 0.05 were considered statistically significant for further analysis.

Gene Set Enrichment Analysis (GSEA) was applied with the list ranked by the t value of the differentially expressed genes between SLE and healthy controls and the target genes of each TF in order to obtain the leading-edge subsets for each TF. These subsets are the gene lists that contribute the most to the enrichment score, and they are the most differentially expressed target genes for every TF. Functional enrichment analysis was performed using the Enrichr webtool. In order to identify drugs that target the significant TFs we queried the CLUE database [20]. We downloaded the full drug information from the Repurposing tool and selected those drugs whose targets are significant TFs.

For statistical analysis among clinical variables (SLEDAI, % Neutrophil, % Lymphocyte, % Monocyte, ESR, WBC, C3 and C4) between the SLE groups obtained by subgroup identification the Mann–Whitney U test was used.

3. Results

3.1. Analysis of TF Activity Revealed Two Main Groups of SLE Patients

SLE is a very heterogeneous disease and there is consistent evidence about differences in global gene expression programs among different SLE patients. In order to evaluate if TFs activity patterns can reveal SLE subgroups, we first performed clustering analysis of SLE samples in both datasets.

Unsupervised clustering analysis of SLE patients identified two main groups in both datasets (Figure 1A). Cluster 1 in the adult cohort is composed of 47 samples (68%) while cluster 2 included 22 samples (32%). In the pediatric cohort, cluster 1 consisted of 62 samples (53%), whereas cluster 2 is composed of 54 samples (47%). The analysis of clinical variables of samples from each cluster revealed a significant difference in the proportion of samples that were enriched in lymphocyte (Adults: $p = 1.76 \times 10^{-4}$ and Pediatric: $p < 0.0001$ 1.33×10^{-12}) and neutrophil percentages (Adults: $p = 1.96 \times 10^{-4}$ and Pediatric: $p < 0.0001$) in each group (Figure 2A). While cluster 1 was enriched in samples with higher neutrophil percentage, cluster 2 contained the most samples with a higher percentage of lymphocytes. We also analyzed the distribution of the other cellular proportions using cibersort, a deconvolution algorithm [21] and although differences in other cell types were observed in one of the cohorts, we did not find significant differences across clusters that were consistent in both datasets (See Figure S1). These findings were consistent in both independent datasets, and they were in agreement with our previous observations reported in [10] where we described a stratification of SLE patients into three main groups, mainly associated with differences in the percentages of these cell populations when disease activity increases.

Interestingly, principal component analysis of the SLE and healthy samples revealed that healthy individuals were grouped together with the subset of SLE samples enriched in lymphocytes (Figure 1B). Although information about the cell populations of healthy samples was not available, we used a previously published dataset that evaluated the normal value of the neutrophil-to-lymphocyte ratio (NLR) from 413 healthy samples [22] to have a reference of the NLR values for healthy samples. Therefore, we compared the NLR values from each cluster with the NLR values from the healthy set. In both datasets, the NLR values of SLE samples in cluster 1 were significantly higher than in the healthy samples (Adults: $p < 0.0001$ and Pediatric: $p < 0.0001$). On the other hand, the cluster

2 samples showed similar NLR values to the healthy individuals (Adults: p = 0.6441 and Pediatric: p = 0.8143) (Figure 2B).
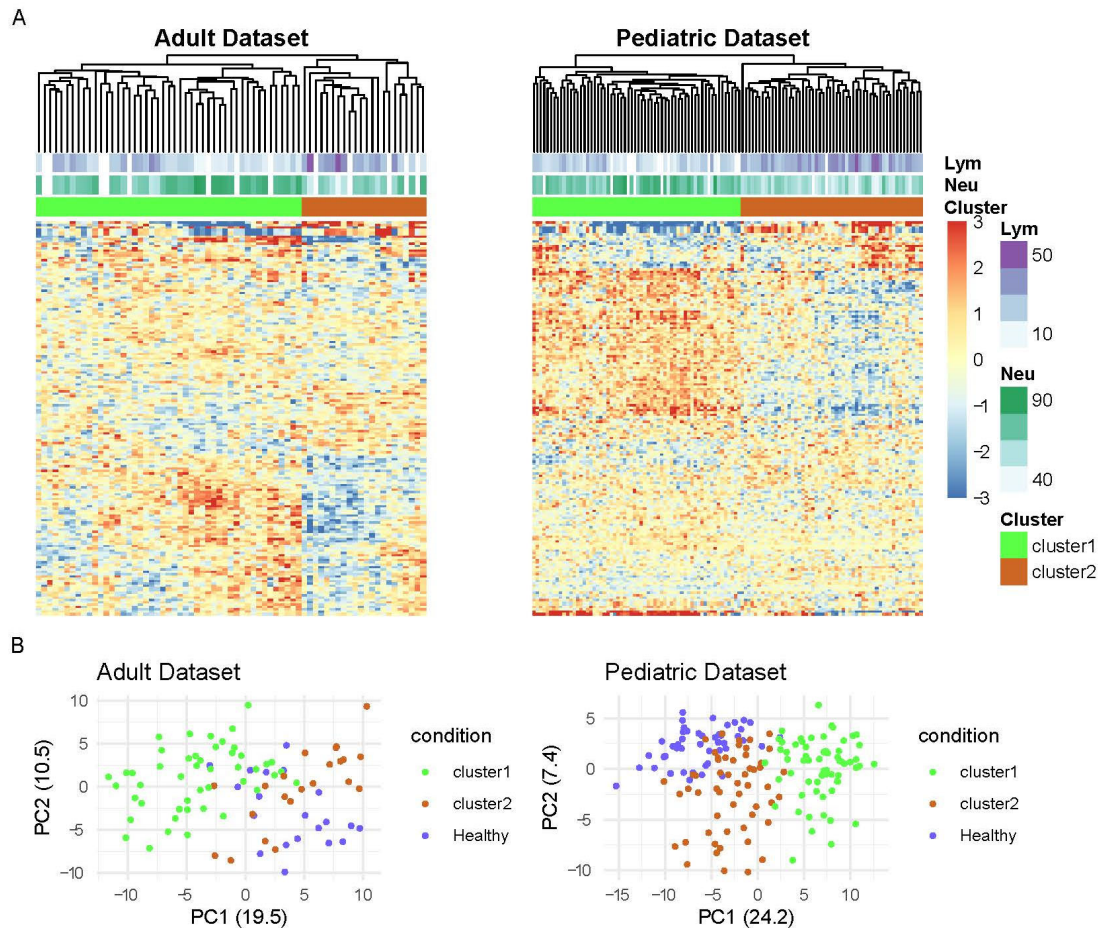

Figure 1. Clustering analysis of transcription factors (TFs) activity profiles in SLE. (**A**) Unsupervised hierarchical clustering of lupus samples based on TFs activity patterns. Top color bars represent percentage of lymphocytes and neutrophils and cluster group. (**B**) Plot showing projection in the two principal components of SLE samples from two clusters together with healthy samples.

Taken together, these results show that activity patterns of TF-regulons can distinguish between two main groups of SLE patients that are characterized by differences in neutrophil/lymphocyte proportions. In addition, there is clear evidence of the shared transcription activity patterns between SLE patients enriched in lymphocyte proportions and the healthy controls.

3.2. TFs with Differential Activity in SLE and Healthy Samples

In order to establish the set of TFs operating in SLE we first performed a differential activity analysis of all SLE against healthy samples. We found a set of 49 TFs that showed significant differential activity between SLE and healthy samples in both datasets (Figure S2). Nevertheless, there was intragroup heterogeneity in the activity value of these TFs. For example, activity patterns for MYC, RFX5, RFXAP and RFXANK were clearly different across the two clusters of SLE samples described previously. Interestingly, the expression level of most of the target genes of these TFs varied within different cell types collected by Expression Atlas (Figure S2B). For example, the expression of HLA genes, regulated by RFXANK, RFXAP and RFX5, was much higher in B cell types and dendritic with respect to the rest. On the other hand, there are some genes, such as BBC3, CXCL2,

ICAM1, PTEN, EGR3, HBA2 or IMPA2, regulated by MYC, whose expression was higher in neutrophil cell types. This fact might reflect the differences in cell proportions found in the two SLE clusters.
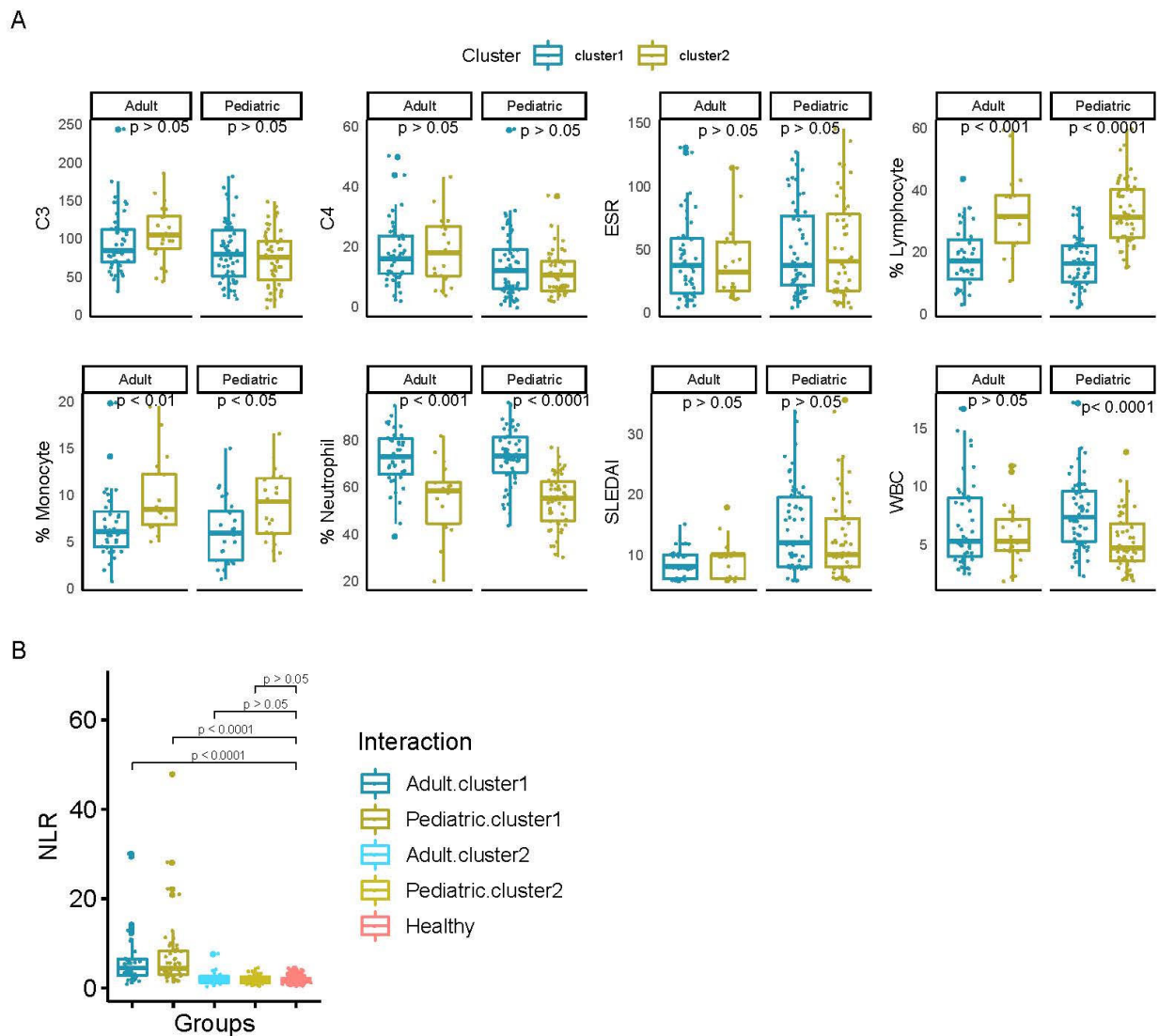

Figure 2. Clinical characterization of SLE clusters. (**A**) Representation of the differences between each SLE cluster in both datasets for each of the clinical variables. Asterisks represent the significance and 'ns' the non-significance. (**B**) Depiction of the changes of NLR between each cluster respect to healthy. Numbers represent the significance levels. Mann–Whitney U test was used to compare among groups.

Therefore, we compared each SLE cluster independently against healthy samples in order to identify TFs that consistently showed differential activity with respect to healthy individuals. This analysis revealed a set of 96 and 60 TFs in cluster 1 and cluster 2, respectively, for the adult dataset, and 135 and 57 TFs in cluster 1 and 2, respectively, for the pediatric set. Sixty-nine TFs in cluster 1 overlapped in both sets (Figure S3A) while 21 TFs in cluster 2 had significant differential activity between healthy samples and both datasets (Figure S3B). From these results, 14 TFs were consistently activated or repressed across all SLE samples, regardless of SLE subgroup (Figure 3A). Specifically, SMAD1, ARNTL, WT1, RELB, SPIB and TCF7L2 had lower activity in SLE while GATA4, NFATC1, E2F2, PPARD, IRF3, STAT2, IRF1, STAT1 were identified as TFs with greater activity.
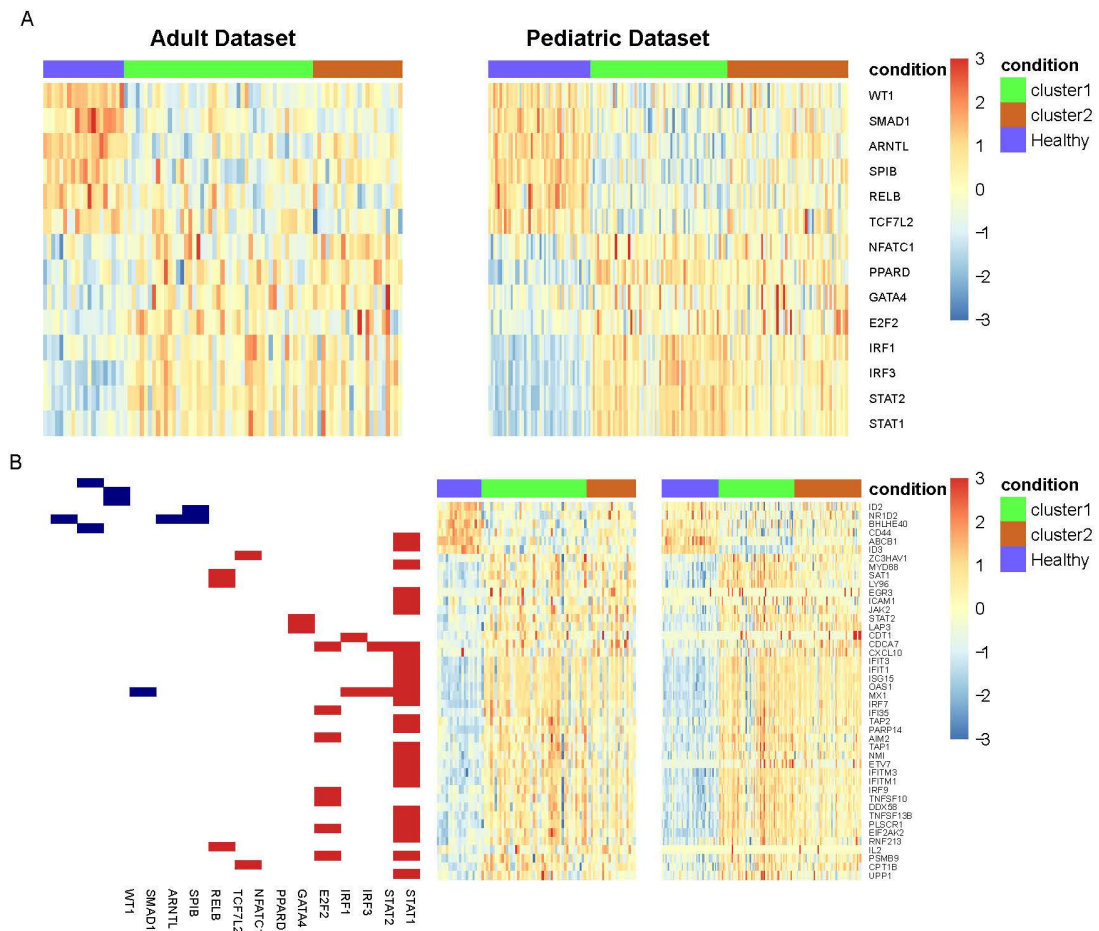
Figure 3. Robust Differential Activity Analysis across SLE and healthy samples. (**A**) Heatmaps show normalized activity values of 14 transcription factors that were found consistently significant across SLE patient and healthy samples in the two independent cohorts (**B**) Gene expression values of leading-edge subsets genes. On the left-hand side there is a graph showing target genes for each TF. Blue and red colors indicate which TFs had lower activity and higher activity in SLE with respect to healthy samples, respectively. Heatmaps on the right-hand side show gene expression values of the corresponding genes (found in the leading-edge subset). Top bars indicate the cluster assignation for each SLE patient as well as healthy samples.

We next evaluated which was the set of target genes whose expression contributed more to the activity signal of the 14 TFs defined in the analysis. To this end, a GSEA was carried out to identify the leading-edge subset of target genes most differentially expressed between SLE clusters and healthy samples for each TF. This analysis yielded a total of 44 target genes (Figure 3B). In this figure it is shown that transcription factors with lower activity in SLE with respect to the healthy controls were involved in the regulation of the underexpressed genes, whilst the overexpressed genes were regulated mainly by the transcription factors with higher activity. Not unexpectedly, most overexpressed genes were involved in interferon signaling pathways and susceptibility to viral infection, which are mainly the targets of STAT1. On the other hand, the few underexpressed genes are related to photoperiodism and circadian clock activity. These biological processes have been previously reported in SLE and other autoimmune diseases but the association has not been well-studied [23,24]. To complete the analysis, we evaluated drugs that can potentially target the set of significant TFs obtained using information from the CLUE database. To this end, we defined the set of drugs that target directly these TFs (Table 2). From the 15 drugs obtained, PPARD is the TF that is associated with 12 of these drugs, most of them related to Perixosome proliferator receptor (PPAR), which has been described to be associated with

Lupus-like autoimmunity development (see discussion). In addition, we found drugs that target NFATC1, IRF3 and STAT1. Interestingly, all the discovered drugs' target TFs had higher activity in SLE compared to healthy controls.

Table 2. Information about significant TFs that are drugs-targets from the CLUE database.

Drug	Mechanisms of Action (MoA)	TF Target	Indication	Phase
Bezafibrate	PPAR receptor agonist	PPARD	Cholesterol	Launched
DG-172	PPAR receptor inverse agonist	PPARD		Preclinical
Elafibranor	PPAR receptor agonist	PPARD		Phase 3
FH-535	PPAR receptor antagonist, WNT signaling inhibitor	PPARD		Preclinical
GSK-0660	PPAR receptor antagonist	PPARD		Preclinical
GSK3787	PPAR receptor antagonist	PPARD		Preclinical
GW-0742	PPAR receptor agonist	PPARD		Preclinical
GW-501516	PPAR receptor agonist	PPARD		Phase 2
Icosapent	Platelet aggregation inhibitor	PPARD	Hypertriglyceridemia	Launched
L-165041	PPAR receptor agonist	PPARD		Preclinical
Sulindac	Cyclooxygenase inhibitor	PPARD	Osteoarthritis, rheumatoid arthritis, ankylosing spondylitis	Launched
Tretinoin	Retinoid receptor agonist, retinoid receptor ligand	PPARD	Leukemia	Launched
INCA-6	Calcineurin inhibitor	NFATC1		Preclinical
piceatannol	SYK inhibitor	IRF3		Preclinical
CKD-712	NFkB pathway inhibitor	STAT1		Phase 1

4. Discussion

Systemic lupus erythematosus is a complex and heterogeneous disease with limited diagnosis and treatment options. Despite research efforts and clinical trials that have been conducted to establish new treatments, only belimumab, a monoclonal antibody against B cell-activating factor (BAFF), has been approved in the last few decades [25]. Therefore, there is an urgent need to decipher the main biological mechanisms underlying SLE to expand our knowledge about the disease and define new biomarkers, classification schemes and treatment options. In the last few years, a large number of SLE biomarkers have been described by GWAS and whole transcriptomics analysis. Interestingly, the positional analysis of these SLE associated loci has revealed that most of them are located in regulatory regions, but there is a lack of a global analysis of TFs activity patterns associated with SLE disease.

In this work we have inferred the TF activities for most human TFs using gene expression levels of SLE patients with active status of the disease in two different cohorts. First, the unsupervised clustering analysis of the activity matrix revealed a two-cluster structure in both datasets, characterized mainly by differences in the neutrophil and lymphocyte proportions. Our molecular classification and clinical characterization are coherent with previous works that stratified SLE or systemic autoimmune diseases in which different proportions of neutrophils and lymphocytes were associated with different groups [9–11]. In this context, there is recent evidence about the potential role of neutrophils' and lymphocytes' proportions as potential markers to stratify SLE patients into clinically separate groups [26,27]. Indeed, the potential of the neutrophil-lymphocyte ratio (NLR) as a cheap and effective biomarker of the activity or response to treatment in autoimmune pathologies is being analyzed by different groups [28–30] and it has also been described that alterations in the balance of neutrophils, monocytes and lymphocytes explain resistance to treatments in patients with rheumatoid arthritis [31], although the molecular mechanisms are not totally characterized

After comparing TF activities from SLE and healthy samples, we defined 49 TFs with significant differential activity. A detailed analysis revealed that some of them could be biased by the heterogeneity of cell population in SLE samples, which is remarkable for the

TFs MYC, RFX5, RFXAP and RFXANK. RFX5, RFXANK and RFXAP, that act as enhancers of the gene expression of the Major Histocompatibility Class II genes (MHC II) [32]. These are expressed in professional antigen-presenting cells (APCs) [33] such as monocytes, B cells and dendritic cells. This could explain the low activity of these TFs in cluster 1, representing neutrophils. On the other hand, MYC regulates a large set of genes, but the presence of ribosomal genes such as RPS15, RPL19, RPS19, RPS6, RPL3, RPL22, RPL6, RPL32, RPL27A, RPL23 and RPS16 stands out. The expression of these genes in different types of blood cells is very heterogeneous [34].

Due to the observed heterogeneity in TF activities of SLE patients, we decided to compare TF activities of each cluster with healthy samples in order to obtain consistent and unbiased regulation patterns across all SLE patients. From these analyses we established 14 TFs that were consistently activated or repressed in SLE.

Those with greater activity in SLE were STAT1, STAT2, IRF1, IRF3, NFATC1, PPARD, E2F2 and GATA4. As expected, this set contains transcription factors that are well-known in the context of SLE pathogenesis including STAT1, STAT2, IRF1 and IRF3 which are activators of interferon genes [35,36]. The analysis of drug–TFs association revealed that STAT1 is the target of CKD-712, which is an inhibitor of NF-κB pathway, a proinflammatory mediator [37]. On the other hand, a SYK inhibitor acts on IRF3. This mechanism has been shown to be effective in rheumatoid arthritis and lupus-prone MRL/lpr mice [38]. In fact, the overexpression of Syk in healthy T cells leads to a SLE-like T cell phenotype, suggesting that the inhibition of Syk gives the opposite effect. Syk has been proposed as a therapeutic target [39].

NFATC1 is overexpressed in lupus-prone MRL/lpr mice activating the calcium/NF-AT pathway [40]. Moreover, this TF is the target of a preclinical calcineurin inhibitor [41], the inhibitory mechanism through which cyclosporine and tracolimus exert their effects when used in SLE patients [42].

Mice deficient in Ppard develop Lupus-like autoimmunity with increased production of autoantibodies and abnormal apoptotic cell clearance [43]. There are some drugs whose target is PPARD (Table 2). One of the most noteworthy drugs that target PPARD, tretinoin, is related to the mechanism of action (MoA) retinoid receptor agonist. The improvement of the inflammatory symptoms of SLE has been reported with retinoic acid treatment in murine models and human disease [44]. SLE amelioration could be achieved via retinoic acid treatment through three mechanisms [45]. One of these is by reversing microbial dysbiosis [46]; secondly, by inhibiting the activity of Pin-1, which activates the TLR-7/TLR-9/IRAK-1/IRF-7 signal that contributes to SLE phenotype [47]; and, thirdly, by reestablishing the vitamin A levels in SLE patients, which improves the T helper 17 (Th17) and regulatory T cell (Treg) balance [44]. Additionally, PPARD is a target of sulindac, which is a cyclooxygenase inhibitor launched in clinical trials of rheumatoid arthritis or spondylitis.

The rest of the significant TFs had lower activity in SLE than in controls: SMAD1, ARNTL, WT1, RELB, SPIB and TCF7L2. SMAD1, along with other genes involved in the BMP/Smad signaling pathway, is repressed through NF-κB signaling pathway in bone marrow-derived mesenchymal stem cells (BMMSCs) from SLE patients [48]. On the other hand, RELB is a subunit of NF-κB and takes part in the development of dendritic cells [49]. In the murine lupus model, Relb-modified dendritic cells decreased the interferon-γ expression [50]. ARNTL is a TF that forms a core component of the circadian clock. This system regulates the gene expression of genes involved in several biological processes according to circadian rhythms. As we have described previously, there are studies that correlated circadian clock dysregulation and SLE pathogenesis [23,24].

Lupus nephritis is one of the most severe manifestations of SLE, characterized by the inflammation of the kidneys and the loss of podocytes [51]. WT1 is a well-known podocyte marker [52] and in murine model its expression is decreased [53].

SPIB belongs to the ETS family of TFs and promotes the development of plasmacytoid dendritic cells (pDC), the major producers of type I interferon and is involved in

the development of germinal center B cells [54]. However, SPIB has been shown to be underexpressed in the B cells of SLE patients [55]. SPIB, as well as E2F2, GATA4 and TCF7L2 have been associated with other autoimmune diseases through differential gene expression and genetic polymorphisms, respectively [56–58].

Although many of the TFs associated with SLE patients have been previously described, in this work we described the stratification of SLE patients into two subgroups based on global TF activity profiles, which are characterized by differences in the neutrophil and lymphocyte proportions. In addition, we identified 14 significant and robust TFs across SLE patients. These results reveal regulation mechanisms regarding SLE heterogeneity, which might be possible therapeutic targets. The groups observed here are consistent with previous findings [10] and can link molecular heterogeneity to clinical manifestations or response to therapies, providing opportunities for novel therapeutic developments or better disease diagnosis.

References

1. Moulton, V.R.; Suarez-Fueyo, A.; Meidan, E.; Li, H.; Mizui, M.; Tsokos, G.C. Pathogenesis of Human Systemic Lupus Erythematosus: A Cellular Perspective. *Trends Mol. Med.* **2017**, *23*, 615–635. [CrossRef] [PubMed]
2. Wasef, S.Z.Y. Gender differences in systemic lupus erythematosus. *Gend. Med.* **2004**, *1*, 12–17. [CrossRef]
3. Mikdashi, J.; Nived, O. Measuring disease activity in adults with systemic lupus erythematosus: The challenges of administrative burden and responsiveness to patient concerns in clinical research. *Arthritis Res. Ther.* **2015**, *17*, 1–10. [CrossRef] [PubMed]
4. Thong, B.; Olsen, N.J. Systemic lupus erythematosus diagnosis and management. *Rheumatology* **2016**, *56*, i3–i13. [CrossRef] [PubMed]
5. Teruel, M.; Alarcón-Riquelme, M.E. The genetic basis of systemic lupus erythematosus: What are the risk factors and what have we learned. *J. Autoimmun.* **2016**, *74*, 161–175. [CrossRef] [PubMed]
6. Gambino, C.M.; Di Bona, D.; Aiello, A.; Carru, C.; Duro, G.; Guggino, G.; Ferrante, A.; Zinellu, A.; Caruso, C.; Candore, G.; et al. HLA-C1 ligands are associated with increased susceptibility to systemic lupus erythematosus. *Hum. Immunol.* **2018**, *79*, 172–177. [CrossRef]
7. Baechler, E.C.; Batliwalla, F.M.; Karypis, G.; Gaffney, P.M.; Ortmann, W.A.; Espe, K.J.; Shark, K.B.; Grande, W.J.; Hughes, K.M.; Kapur, V.; et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2610–2615. [CrossRef]
8. Bennett, L.; Palucka, A.K.; Arce, E.; Cantrell, V.; Borvak, J.; Banchereau, J.; Pascual, V. Interferon and Granulopoiesis Signatures in Systemic Lupus Erythematosus Blood. *J. Exp. Med.* **2003**, *197*, 711–723. [CrossRef]
9. Banchereau, R.; Hong, S.; Cantarel, B.; Baldwin, N.; Baisch, J.; Edens, M.; Cepika, A.-M.; Acs, P.; Turner, J.; Anguiano, E.; et al. Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients. *Cell* **2016**, *165*, 551–565. [CrossRef]

10.	Toro-Domínguez, D.; Martorell-Marugán, J.; Goldman, D.; Petri, M.; Carmona-Sáez, P.; Alarcón-Riquelme, M.E. Stratification of Systemic Lupus Erythematosus Patients into Three Groups of Disease Activity Progression According to Longitudinal Gene Expression. *Arthritis Rheumatol.* **2018**, *70*, 2025–2035. [CrossRef] [PubMed]

11.	Barturen, G.; Babaei, S.; Català-Moll, F.; Martínez-Bueno, M.; Makowska, Z.; Martorell-Marugán, J.; Carmona-Sáez, P.; Toro-Domínguez, D.; Carnero-Montoro, E.; Teruel, M.; et al. Integrative Analysis Reveals a Molecular Stratification of Systemic Autoimmune Diseases. *Arthritis Rheumatol.* **2020**, 2326–5205. [CrossRef]

12.	Dozmorov, M.G.; Wren, J.D.; E Alarcón-Riquelme, M. Epigenomic elements enriched in the promoters of autoimmunity susceptibility genes. *Epigenetics* **2014**, *9*, 276–285. [CrossRef]

13.	Harley, J.B.; Chen, X.; Pujato, M.; Miller, D.; Maddox, A.; Forney, C.; Magnusen, A.F.; Lynch, A.; Chetal, K.; Yukawa, M.; et al. Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat. Genet.* **2018**, *50*, 699–707. [CrossRef]

14.	Essaghir, A.; Toffalini, F.; Knoops, L.; Kallin, A.; Van Helden, J.; Demoulin, J.-B. Transcription factor regulation can be accurately predicted from the presence of target gene signatures in microarray gene expression data. *Nucleic Acids Res.* **2010**, *38*, e120. [CrossRef]

15.	Falco, M.M.; Bleda, M.; Carbonell-Caballero, J.; Dopazo, J. The pan-cancer pathological regulatory landscape. *Sci. Rep.* **2016**, *6*, 39709. [CrossRef]

16.	Garcia-Alonso, L.M.; Iorio, F.; Matchan, A.; A Fonseca, N.; Jaaks, P.; Peat, G.; Pignatelli, M.; Falcone, F.; Benes, C.H.; Dunham, I.; et al. Transcription Factor Activities Enhance Markers of Drug Sensitivity in Cancer. *Cancer Res.* **2018**, *78*, 769–780. [CrossRef] [PubMed]

17.	Garcia-Alonso, L.; Holland, C.H.; Ibrahim, M.M.; Turei, D.; Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **2019**, *29*, 1363–1375. [CrossRef] [PubMed]

18.	Romero-Diaz, J.; Isenberg, D.; Ramsey-Goldman, R. Measures of adult systemic lupus erythematosus: Updated Version of British Isles Lupus Assessment Group (BILAG 2004), European Consensus Lupus Activity Measurements (ECLAM), Systemic Lupus Activity Measure, Revised (SLAM-R), Systemic Lupus Activity Questi. *Arthritis Rheum.* **2011**, *63*, S37–S46. [CrossRef] [PubMed]

19.	Zollars, E.; Courtney, S.M.; Wolf, B.J.; Allaire, N.; Ranger, A.; Hardiman, G.; Petri, M. Clinical Application of a Modular Genomics Technique in Systemic Lupus Erythematosus: Progress towards Precision Medicine. *Int. J. Genom.* **2016**, *2016*, 1–7. [CrossRef] [PubMed]

20.	Subramanian, A.; Narayan, R.; Corsello, S.M.; Peck, D.D.; Natoli, T.E.; Lu, X.; Gould, J.; Davis, J.F.; Tubelli, A.A.; Asiedu, J.K.; et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **2017**, *171*, 1437–1452. [CrossRef]

21.	Newman, A.M.; Liu, C.L.; Green, M.R.; Gentles, A.J.; Feng, W.; Xu, Y.; Hoang, C.D.; Diehn, M.; Alizadeh, A.A. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **2015**, *12*, 453–457. [CrossRef] [PubMed]

22.	Forget, P.; Khalifa, C.; Defour, J.-P.; Latinne, D.; Van Pel, M.-C.; De Kock, M. What is the normal value of the neutrophil-to-lymphocyte ratio? *BMC Res. Notes* **2017**, *10*, 1–4. [CrossRef] [PubMed]

23.	Cao, Q.; Zhao, X.; Bai, J.; Gery, S.; Sun, H.; Lin, D.-C.; Chen, Q.; Chen, Z.; Mack, L.; Yang, H.; et al. Circadian clock cryptochrome proteins regulate autoimmunity. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 12548–12553. [CrossRef]

24.	Lin, G.-J.; Huang, S.-H.; Chen, S.-J.; Wang, C.-H.; Chang, D.-M.; Sytwu, H.-K. Modulation by Melatonin of the Pathogenesis of Inflammatory Autoimmune Diseases. *Int. J. Mol. Sci.* **2013**, *14*, 11742–11766. [CrossRef] [PubMed]

25.	Dolgin, E. Lupus in crisis: As failures pile up, clinicians call for new tools. *Nat. Biotechnol.* **2019**, *37*, 7–8. [CrossRef] [PubMed]

26.	Alarcón-Riquelme, M.E. New Attempts to Define and Clarify Lupus. *Curr. Rheumatol. Rep.* **2019**, *21*, 1–7. [CrossRef]

27.	Smith, C.K.; Kaplan, M.J. The role of neutrophils in the pathogenesis of systemic lupus erythematosus. *Curr. Opin. Rheumatol.* **2015**, *27*, 448–453. [CrossRef] [PubMed]

28.	Boulos, D.; Proudman, S.M.; Metcalf, R.G.; McWilliams, L.; Hall, C.; Wicks, I.P. The neutrophil-lymphocyte ratio in early rheumatoid arthritis and its ability to predict subsequent failure of triple therapy. *Semin. Arthritis Rheum.* **2019**, *49*, 373–376. [CrossRef]

29.	D'Amico, E.; Zanghì, A.; Romano, A.; Sciandra, M.; Palumbo, G.A.M.; Patti, F. The Neutrophil-to-Lymphocyte Ratio is Related to Disease Activity in Relapsing Remitting Multiple Sclerosis. *Cells* **2019**, *8*, 1114. [CrossRef]

30.	Erre, G.L.; Paliogiannis, P.; Castagna, F.; Mangoni, A.A.; Carru, C.; Passiu, G.; Zinellu, A. Meta-analysis of neutrophil-to-lymphocyte and platelet-to-lymphocyte ratio in rheumatoid arthritis. *Eur. J. Clin. Investig.* **2019**, *49*, e13037. [CrossRef]

31.	Tasaki, S.; Suzuki, K.; Kassai, Y.; Takeshita, M.; Murota, A.; Kondo, Y.; Ando, T.; Nakayama, Y.; Okuzono, Y.; Takiguchi, M.; et al. Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission. *Nat. Commun.* **2018**, *9*, 1–12. [CrossRef]

32.	Handunnetthi, L.; Ramagopalan, S.V.; Ebers, G.C.; Knight, J.C. Regulation of major histocompatibility complex class II gene expression, genetic variation and disease. *Genes Immun.* **2009**, *11*, 99–112. [CrossRef]

33.	Kambayashi, T.; Laufer, T.M. Atypical MHC class II-expressing antigen-presenting cells: Can anything replace a dendritic cell? *Nat. Rev. Immunol.* **2014**, *14*, 719–730. [CrossRef]

34.	Guimaraes, J.C.; Zavolan, M. Patterns of ribosomal protein expression specify normal and malignant human cells. *Genome Biol.* **2016**, *17*, 1–13. [CrossRef]

35.	Goropevšek, A.; Holcar, M.; Avčin, T. The Role of STAT Signaling Pathways in the Pathogenesis of Systemic Lupus Erythematosus. *Clin. Rev. Allergy Immunol.* **2016**, *52*, 164–181. [CrossRef]

36. Zhang, Z.; Shi, L.; Song, L.; Ephrem, E.; Petri, M.; Sullivan, K.E. Interferon regulatory factor 1 marks activated genes and can induce target gene expression in systemic lupus erythematosus. *Arthritis Rheumatol.* **2014**, *67*, 785–96. [CrossRef] [PubMed]

37. Sun, S.-C.; Chang, J.-H.; Jin, J. Regulation of nuclear factor-κB in autoimmunity. *Trends Immunol.* **2013**, *34*, 282–289. [CrossRef] [PubMed]

38. Deng, G.-M.; Kyttaris, V.C.; Tsokos, G.C. Targeting Syk in Autoimmune Rheumatic Diseases. *Front. Immunol.* **2016**, *7*. [CrossRef] [PubMed]

39. Grammatikos, A.P.; Ghosh, D.; Devlin, A.; Kyttaris, V.C.; Tsokos, G.C. Spleen Tyrosine Kinase (Syk) Regulates Systemic Lupus Erythematosus (SLE) T Cell Signaling. *PLoS ONE* **2013**, *8*, e74550. [CrossRef]

40. Kyttaris, V.C.; Zhang, Z.; Kampagianni, O.; Tsokos, G.C. Calcium signaling in systemic lupus erythematosus T cells: A treatment target. *Arthritis Rheum.* **2011**, *63*, 2058–2066. [CrossRef]

41. Frischbutter, S.; Schultheis, K.; Pätzel, M.; Radbruch, A.; Baumgrass, R. Evaluation of calcineurin/NFAT inhibitor selectivity in primary human Th cells using bar-coding and phospho-flow cytometry. *Cytom. Part A* **2012**, *81*, 1005–1011. [CrossRef]

42. Mok, C.C. Calcineurin inhibitors in systemic lupus erythematosus. *Best Pract. Res. Clin. Rheumatol.* **2017**, *31*, 429–438. [CrossRef] [PubMed]

43. Mukundan, L.; Odegaard, J.I.; Morel, C.R.; Heredia, J.E.; Mwangi, J.W.; Ricardo-Gonzalez, R.R.; Goh, Y.S.; Eagle, A.R.; Dunn, S.E.; Awakuni, J.U.; et al. PPAR-δ senses and orchestrates clearance of apoptotic cells to promote tolerance. *Nat. Med.* **2009**, *15*, 1266–1272. [CrossRef] [PubMed]

44. Handono, K.; Firdausi, S.N.; Pratama, M.Z.; Endharti, A.T.; Kalim, H. Vitamin A improve Th17 and Treg regulation in systemic lupus erythematosus. *Clin. Rheumatol.* **2016**, *35*, 631–638. [CrossRef]

45. Abdelhamid, L.; Luo, X.M. Retinoic Acid, Leaky Gut, and Autoimmune Diseases. *Nutrients* **2018**, *10*, 1016. [CrossRef] [PubMed]

46. Zhang, H.; Liao, X.; Sparks, J.B.; Luo, X.M. Dynamics of Gut Microbiota in Autoimmune Lupus. *Appl. Environ. Microbiol.* **2014**, *80*, 7551–7560. [CrossRef]

47. Wei, S.; Yoshida, N.; Finn, G.; Kozono, S.; Nechama, M.; Kyttaris, V.C.; Zhou, X.Z.; Tsokos, G.C.; Lu, K.P. Pin1-Targeted Therapy for Systemic Lupus Erythematosus. *Arthritis Rheumatol.* **2016**, *68*, 2503–2513. [CrossRef]

48. Tang, Y.; Xie, H.; Chen, J.; Geng, L.; Chen, H.; Li, X.; Hou, Y.; Lu, L.; Shi, S.; Zeng, X.; et al. Activated NF-κB in Bone Marrow Mesenchymal Stem Cells from Systemic Lupus Erythematosus Patients Inhibits Osteogenic Differentiation Through Downregulating Smad Signaling. *Stem Cells Dev.* **2013**, *22*, 668–678. [CrossRef] [PubMed]

49. Shih, V.F.-S.; Davis-Turak, J.; Macal, M.; Huang, J.Q.; Ponomarenko, J.; Kearns, J.D.; Yu, T.; Fagerlund, R.; Asagiri, M.; Zuniga, E.I.; et al. Control of RelB during dendritic cell activation integrates canonical and noncanonical NF-κB pathways. *Nat. Immunol.* **2012**, *13*, 1162–1170. [CrossRef]

50. Wu, H.; Lo, Y.; Chan, A.; Law, K.S.; Mok, M.Y. Rel B-modified dendritic cells possess tolerogenic phenotype and functions on lupus splenic lymphocytes in vitro. *Immunology* **2016**, *149*, 48–61. [CrossRef]

51. Nakamura, T.; Ushiyama, C.; Suzuki, S.; Hara, M.; Shimada, N.; Sekizuka, K.; Ebihara, I.; Koide, H. Urinary Podocytes for the Assessment of Disease Activity in Lupus Nephritis. *Am. J. Med Sci.* **2000**, *320*, 112–116. [CrossRef]

52. Kanemoto, K.; Takahashi, S.; Shu, Y.; Usui, J.; Tomari, S.; Yan, K.; Hamazaki, Y.; Nagata, M. Variable expression of podocyte-related markers in the glomeruloid bodies in Wilms tumor. *Pathol. Int.* **2003**, *53*, 596–601. [CrossRef]

53. Liao, R.; Liu, Q.; Zheng, Z.; Fan, J.; Peng, W.; Kong, Q.; He, H.; Yang, S.; Chen, W.; Tang, X.; et al. Tacrolimus Protects Podocytes from Injury in Lupus Nephritis Partly by Stabilizing the Cytoskeleton and Inhibiting Podocyte Apoptosis. *PLoS ONE* **2015**, *10*, e0132724. [CrossRef] [PubMed]

54. Willis, S.N.; Tellier, J.; Liao, Y.; Trezise, S.; Light, A.; O'Donnell, K.; Garrett-Sinha, L.A.; Shi, W.; Tarlinton, D.M.; Nutt, S.L. Environmental sensing by mature B cells is controlled by the transcription factors PU.1 and SpiB. *Nat. Commun.* **2017**, *8*, 1–14. [CrossRef] [PubMed]

55. Becker, A.M.; Dao, K.H.; Han, B.K.; Kornu, R.; Lakhanpal, S.; Mobley, A.B.; Li, Q.-Z.; Lian, Y.; Wu, T.; Reimold, A.M.; et al. SLE Peripheral Blood B Cell, T Cell and Myeloid Cell Transcriptomes Display Unique Profiles and Each Subset Contributes to the Interferon Signature. *PLoS ONE* **2013**, *8*, e67003. [CrossRef] [PubMed]

56. Zhang, R.; Wang, L.; Pan, J.-H.; Han, J. A critical role of E2F transcription factor 2 in proinflammatory cytokines-dependent proliferation and invasiveness of fibroblast-like synoviocytes in rheumatoid Arthritis. *Sci. Rep.* **2018**, *8*, 1–10. [CrossRef]

57. Jia, W.; Wu, W.; Yang, D.; Xiao, C.; Huang, M.; Long, F.; Su, Z.; Qin, M.; Liu, X.; Zhu, Y.Z. GATA4 regulates angiogenesis and persistence of inflammation in rheumatoid arthritis. *Cell Death Dis.* **2018**, *9*, 1–15. [CrossRef]

58. Lyssenko, V.; Lupi, R.; Marchetti, P.; Del Guerra, S.; Orho-Melander, M.; Almgren, P.; Sjögren, M.; Ling, C.; Eriksson, K.-F.; Lethagen, υsa-L.; et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. *J. Clin. Investig.* **2007**, *117*, 2155–2163. [CrossRef] [PubMed]