

An integrative Bayesian network approach to highlight key drivers in systemic lupus erythematosus

Samaneh Maleknia

University of Tehran

Zahra Salehi

Tehran University of Medical Sciences School of Medicine

Ali Sharifi-Zarchi

Sharif University of Technology

Vhid Rezaei-Tabar

Allameh Tabatabai University

Kaveh Kavousi (✉ kkavousi@ut.ac.ir)

University of Tehran <https://orcid.org/0000-0002-1906-3912>

Research article

Keywords: SLE; signaling pathways; Bayesian network; meta-analysis; cross-platform normalization; bioinformatics

Posted Date: January 23rd, 2020

DOI: <https://doi.org/10.21203/rs.2.21746/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on June 23rd, 2020. See the published version at <https://doi.org/10.1186/s13075-020-02239-3>.

Abstract

Background: To perceive a comprehensive illustration of the systemic lupus erythematosus (SLE), as a complex and multifactorial disease, draw a biological challenge. Dealing with this challenge needs employing sophisticated bioinformatics algorithms to discover the unknown aspects. This study aimed to distinguish key molecular characteristics of SLE pathogenesis, which may serve as effective targets for therapeutic intervention. **Methods:** In the present study, six gene expression microarray datasets included SLE patients (n=220) and healthy controls (n=135) were collected. We integrated the datasets by cross-platform normalization. Subsequently, through BNrich method, the structures of Bayesian networks (BNs) were extracted from SLE, TCR and BCR signaling pathways and the parameters of BNs were estimated using integrated datasets. Finally, a meta-analysis was performed to distinguish the signaling pathways alterations in the SLE pathogenesis. **Results:** We identified the most dysregulated genes involved in the clearance mechanism (most inhibition in MACROH2A2 and H4C9, most activation in SNRPD3, MACROH2A1 and RO60). Augmented autoantigens presentation by MHCII (HLA-DQA1, HLA-DOB and HLA-DPB1) and CD80 and CD86 to CD28 biological functions were also observed. The increased expression of FCGR1A, C8G, C7 and C9 genes and decreasing biological functions in edges from C1R and C1S to C2 and from C1R to C4B, all involved in end-organ damage, were detected. On the other hand, many of subnetworks alterations of TCR and BCR reported in previous research (PI3K/AKT, MAPK, AP-1 transcription factor). **Conclusions:** Overall, the BNrich as a hybridized network construction method, which integrates intergenic relations inferred from signaling pathways and gene expression profiling, can be employed to study complex diseases. Here we applied BNrich and highlighted significant genes and intergenic relationships and systemic alterations in SLE, TCR and BCR signaling pathways.

1. Introduction

Called as “the Great Imposter”, systemic lupus erythematosus (SLE) is a chronic and complex autoimmune disease with multi-system manifestations in which many questions about its etiology remain unanswered [1,2]. The disease prevalence rate estimated about 5 million people worldwide and it occurs mostly before age 45 years [3–5]. The loss of central and peripheral tolerance to self is proposed to be a crucial first step in the progress of SLE [6]. The autoimmunity can be provoked in genetically susceptible individuals by some environmental triggers like hormones (eight times higher in women), infections (microbes, virus ...), ultraviolet light, toxins and etc. [7].

Certain lines of evidence suggested that a break mediated by type I interferon (IFN) cytokines in central B cell tolerance, increases numbers of autoreactive clones to reach the periphery [8] which ultimately result in autoantibodies production and immune complex deposition that finally damage the end organ [2,7,9,10]. Since, type I IFN cytokines can enhance immune responses by modulating both T cells and B cells functions, they were considered as the master regulators of SLE pathogenesis for the past several decades. Recently, high rates of type I, type II and type III circulating interferons were associated with distinct clinical manifestation in SLE patients at active phase [11]. Nowadays, genome-wide association studies (GWAS) coupled with gene expression profiling data shed light on the critical role regarding the

dysregulation of genes involved in B-cell receptor (BCR) and T-cell receptor (TCR) mediated signaling in B cells and T cells of SLE patients, respectively [7]. As a validation, an aberrant TCR signaling followed by hyperresponsiveness of T cells has been shown in patients with SLE [12]. Despite numerous strongminded studies that have been done, the pathogenesis is still far from clear.

Towards a better intuition of molecular mechanisms involved in multifactorial immune-related diseases especially SLE, lots of researchers acknowledge the gene network approach, which offers a comprehensive analysis of the interplay between multiple genes and pathways, helps to prioritizes main driver genes and pathways, in order to propose a better drug development [13]. Bayesian network (BN), as a worthwhile powerful gene network construction methods, integrates and models biological data with causal relationships [14–18]. By applying gene expression data of SLE patients, Li et al. recently reconstructed a BN, in which prior edges were randomly sampled based on the text-mining, and demonstrated a significant role of the JAK-STAT pathway in SLE pathogenesis[13]. Remarkably, some pathway enrichment analysis (PEA) methods such as BNrich were also developed based on BN properties [19–21]. In addition, to detect significant pathways, BNrich also identifies significant genes (nodes) and biological relationships (edges). The BNrich method has the key concept; the structure of BN is reconstructed by the signaling pathway structure. Further, the parameters of BN are estimated by gene expression data that is continuous. Since, in BNs, the number of outgoing and incoming edges of a node shows the number of children and number of parents, respectively [22], the expression level of the node is modelled as a regression function of expression levels of its parents. In addition, every edge exhibits a biological relationship between two genes, which reflects a biological function (e.g. activation, inhibition, methylation, etc.).

In the present study, to construct trained BNs and identify significant nodes and edges in SLE, as a complex autoimmune disease, we (1) integrated peripheral blood mononuclear cells (PBMCs) microarray gene expression by cross-platform normalization [23]; (2) employed SLE, TCR and BCR signaling pathways as underlying structures of BNs;(3) estimated the parameters of BNs by integrated datasets [22,24]; (4) merged the obtained significant parameters of BNs in desired pathways as “meta-analysis” data integration [23]; (5) highlighted signaling pathways alterations involved in the SLE pathogenesis. Finally, we (6) demonstrated that some of the genes and intergenic relationships have a vital role and can be used as effective targets for therapeutic intervention in SLE patients. Notably, the proposed computational methodology can be applied to a variety of complex diseases to illuminate new mechanistic insights in their pathogenesis.

2. Methods

At the first step, six PBMCs gene expression of human microarray datasets associated with SLE, which originated from three platforms, were downloaded and each paired data related to the same platform integrated by cross-platform normalization algorithm. Subsequently, SLE, TCR and BCR signaling pathways were employed as structures of BN and trained by three major datasets individually by BNrich

method. Consequently, the significant parameters merged by meta-analysis to achieve the final criteria and detect alterations in considering pathways. The workflow was shown in Figure 1.

Figure 1. Workflow and analysis procedure; from row SLE and healthy control (HC) data in various datasets to identify signaling pathways alterations in SLE patients.

2.1. Gene expression datasets

The PBMCs gene expression of human microarray datasets that contain both SLE patients and healthy control subjects (published or updated in 2010-2019) downloaded from Gene Expression Omnibus (GEO) database: GSE17755 [25], GSE 12374 [26], GSE 50772 [27], GSE 81622, GSE 121239 [28,29] and GSE 126307 [30]. We described the details of the data in Table 1.

Table 1. Description of the datasets used in the study.

NO	GSE-NO	GPL/Platform	No of sample		Cell-type	Update(year)	Race
			SLE patients	control			
1	17755	1291/ Hitachisoft	22	55	PBMC	2010	Japanese
2	12374	1291/ Hitachisoft	11	6	PBMC	2012	Japanese
3	50772	570/Affymetrix	61	20	PBMC	2015	Unknown ¹
4	81622	10558/Illumina	30	25	PBMC	2016	Unknown ²
5	121239	13158/Affymetrix	65	20	PBMC	2018	Unknown ³
6	126307	13369/Illumina	31	9	PBMC	2019	Several races ⁴

¹ The data collected in the USA/South San Francisco, but the race of subjects is unknown.

² The data collected in the USA/Dallas, but the race of subjects is unknown.

³ The data collected in Spain / Granada, but the race of subjects is unknown.

⁴ The data collected in USA/Dallas, but the race of subjects is Australian, Australian (Irish / Scottish descent), born in India-ethnicity unknown, Caucasian, Caucasian/Japanese, English, Filipino, Indian, Iraqi, Latin American, Persian, Spanish, White Australian / Anglo-Celtic.

2.2. Cross-platform normalization

Since the gene expression datasets of PBMCs were related to different platforms, we merged the data through cross-platform normalization algorithms [23] to (1) increase sample sizes and improve genes signature selection [23,31–34]; (2) appraise the heterogeneity of the overall estimate, and (3) decrease the effects of individual study-specific biases [23,34]. Finally, we integrated all datasets related to the same platform and built three major datasets from Hitachisoft, Affymetrix and Illumina.

All procedures for data processing and integration were carried out using the R statistical programming language. Each expression datasets were preprocessed by *normalizeQuantiles* function from R package *limma* and then all datasets derived from the same platform were integrated with *Reduce* function. Afterwards, the empirical Bayes method (ComBat) from the R package *sva* was used for batch effect removal [35].

2.3. The signalling pathways as prior information

The SLE signaling pathway (hsa:05322) and two most important pathways in SLE pathogenesis, TCR (hsa:04660) and BCR (hsa:04662) signaling pathways were implemented in this study. All of the pathways were extracted directly from the KEGG database (Release 90.0, April 1, 2019) [36].

2.4. Bayesian network approach

The three desired pathways were exploited as structures of BNs [21]. In the parameter estimate step, the mean value of gene expression for each node can be modelled as a linear regression of its parents' gene expression because the gene expression data is continuous [22,24]. Thus, when gene has parents in the network and describe the estimations of for the SLE patient and healthy control datasets, respectively; they can be modelled as follows: (see Equations in the Supplementary Files)

The and are the residual values. Estimated parameters of BNs were derived from the following distribution [37]: (see Equations in the Supplementary Files)

The parameters of the structures were estimated by SLEs (patients) and healthy controls of the three major datasets individually. It means for each parameter six values were available; SLE vs controls in three datasets. We compared corresponding β parameters between each pair of trained networks (patient and healthy states) by independent t-test [38] and gained their p-value or false discovery rate (FDR). Consequently, for every parameter related to a node or an edge, for example , after independent t-test between patient and healthy states, we had FDR_A (Affymetrix), FDR_I (Illumina) and FDR_H (Hitachisoft).

2.5. Meta-Analysis

By using the meta-analysis approach [23], we incorporated the results of platforms rather than continue to integrate different platforms data. At the first step, the three major datasets were assessed by BNrich method individually. The FDR was also acquired from independent t-test between all paired parameters associated with the SLE patient () and healthy control ().

Subsequently, we defined the parameter similar to the expression fold-change, but for demonstrating the variations of BNrich parameters, as below: (see Equations in the Supplementary Files)

which refers to the the corresponding parameter in the SLE and healthy control.

Finally, we integrated the acquired significant parameters of BNrich ($FDR < 0.05$ & > 1) in the above-mentioned pathways by the following meta-analysis method to compute the final criteria: (see Equations in the Supplementary Files)

where , and are in Affymetrix, Illumina and Hitachisoft datasets based on parameter. Furthermore, , = 95 and were the number of subjects in Affymetrix, Illumina and Hitachisoft datasets, respectively. If the parameter in each dataset was not significant, the weight of the parameter in that dataset was zero. When the was associated with the node parameters, the sign of indicated up-regulation (+) or down-regulation (-) of the gene and the absolute value of was used to measure the value of activation or

inhibition. Where the was associated with the edge parameters, the sign of indicated increasing (+) or decreasing (-) the biological function and the absolute value of was used to measure the value of these increasing and decreasing functions.

3. Results

3.1 Cross-platform normalization

After downloading six mentioned datasets (Table 1), to reduce the batch effect generated between different arrays, each paired data emanated from the same platform were integrated using the ComBat algorithm. The efficiency of the ComBat process in our integration for batch removal can be verified by the comparative boxplots and principle component analysis (PCA) plots (Figure 2, 3 and Additional file1).

3.2 BNrich results

Initially, the SLE, TCR, BCR signaling pathways were utilized as structures of BN. Subsequently, significant parameters (nodes and edges) were determined using BNrich method. Table 2 presents the number of nodes and edges in three platforms based on FDR and for SLE, TCR and BCR signaling pathways. Finally, the significant nodes and edges were used to estimate the final criteria (equation 4) for meta-analysis.

Table 2. The number of all significant nodes and edges of BNs extracted form SLE, TCR and BCR signaling pathways obtained via BNrich

		FDR < 0.05						Signaling pathways
		Affymetrix		Illumina		Hitachisoft		
		nodes	edges	nodes	edges	nodes	edges	
all		73	3	97	10	47	7	SLE
		70	142	66	106	66	72	TCR
		50	86	53	72	48	46	BCR
> 1		0	3	0	8	33	5	SLE
		12	96	16	81	41	49	TCR
		11	59	14	49	33	31	BCR

The common significant nodes and edges among all platforms in SLE, TCR and BCR signaling pathways were represented in Table 3.

Table 3. The common significant nodes and edges amongst all platforms in SLE, TCR and BCR signaling pathway

		FDR						
	from	To	Affimetrix	Illumina	Hitachisoft	Affimetrix	Illumina	Hitachisoft
SLE	C1QB	C2	5.47E-06	1.3E-07	1.75E-08	1.226995	-4.270952	2.054599
TCR	RASGRP1	KRAS	0.0003886	0.00086	0.00000163	-2.706435	-1.530479	2.225356
	RASGRP1	NRAS	1.22E-11	0.00524	0.01120064	-1.933531	2.260514	1.534244
	NRAS	RHOA	0.0001855	0.00477	0.01303765	-3.039338	-1.9127	-3.809746
	FOS	TNF	2.5E-11	7.6E-19	0.0000199	-1.527266	-4.435995	1.347727
	JUN	IFNG	0.0000203	2.9E-07	1.64E-13	-3.942466	-2.16126	-5.837809
	PIK3R1	AKT2	0.0015904	1.5E-07	0.0000412	2.482333	-3.540513	1.36432
	VAV1	CDC42	0.0011554	0.00344	0.00937873	4.302419	-2.365204	-1.299876
	CD28	GRB2	4.13E-09	2.1E-05	0.0000772	-1.139874	1.06101	-2.332912
		JUN ¹	0.0000179	2.2E-05	1.68E-20	-5.536212	1.80163	2.346542
BCR	LYN	BTK	0.0180325	5.7E-06	1.85E-19	-1.037645	-1.843287	-1.037645
	NRAS	FOS	0.0000358	2E-06	0.00015958	-7.277228	-1.613897	-7.277228
	PIK3CA	AKT3	6.41E-11	0.00064	0.0002779	-2.148301	-3.367104	-2.148301
	PIK3R1	AKT2	0.0015904	1.5E-07	0.00015394	2.482333	-3.540513	2.482333
		JUN ¹	5.58E-12	1.5E-05	1.68E-20	-2.081757	-1.535783	-2.081757

¹ The node JUN is the only significant common node across all platforms in both TCR and BCR signaling pathways.

All significant parameters of different platforms, based on three examined signaling pathways, were presented in Additional file2.

3.3 Meta-analysis and pathways alterations detection

The corresponding significant parameters of major datasets merged by meta-analysis, which followed by determining final criteria (equation 4) to distinguish alterations of signaling pathways in SLE patients compared to healthy controls. The range of final criteria in all nodes and edges were represented in Table 4.

By reason that () was positive or minus across different platforms, the final criteria (equation 4) may be tended to zero.

Table 4. The minimum and maximum final criteria values of nodes and edges in three signaling pathway

		SLE		TCR		BCR	
		nodes	edges	nodes	edges	nodes	edges
Min		-6.13	-5.24	-3.99	-9.86	-7.46	-6.61
	Max	4.20	3.49	3.98	5.38	4.08	5.39

The three nodes and edges that have the extreme values of C_i in each desired pathway were illustrated in Table 5 and the all C_i represented in Additional file3.

Table 5. The top three nodes and edges of the studied pathways.

	nodes		edges	
	Down	Up	increasing biological function	decreasing biological function
SLE	CD40	FCGR1A	CD86@CD28	C1R@C2
	CD86	C8G	C1QC@C2	C1R@C4B
	MACROH2A2	C7	C4A@C3	C1S@C2
TCR	FOS	MAPK14	MAP3K14@IKBKB	PPP3R1@NFATC1
	MAPK9	PAK5	PTPN6@CTLA4	RELA@IL5
	VAV3	NCK1	NCK2@PAK4	MAPK14@NFATC3
BCR	INPPL1	FCGR2B	MAPK3@JUN	NRAS@RAF1
	KRAS	LILRB2	HRAS@JUN	AKT2@CHUK
	VAV3	LILRB3	PIK3R2@AKT2	GRB2@SOS1

While among genes involved in clearance mechanism of SLE pathways, SNRPD3, MACROH2A1, RO60, H2BC7, H2AZ1, ACTN1 and SNRPB were upregulated, TRIM21, H4C8, H4C4, H4C9 and MACROH2A2 were downregulated. In antigen presentation process mediated by MHCII, nodes such as CD40, CD86, IFNG and CD28 and two edges, including CD80 and CD86 to CD28, were found to be dysregulated. In addition, our analysis showed increased expression of FCGR1A, C8G, C7, C9, FCGR3B, C6 and CTSG genes and decreased expression of C3 gene, all of them participate into damaging of the end organ. Particularly, in that part, we also found dysregulation of several important edges like C1R, C1S and C1QB to C2, and etc. (Figure 4).

By integrating BCR and TCR signaling pathways information with gene expression data, we identified many of the altered key driver genes in these pathways among SLE patients. Our results revealed 16 shared dysregulated nodes (eg: AKT1, AKT2, AKT3, MALT1, MAPK3, SOS and etc.) and 40 common dysregulated edges (eg: PIK3CA to AKT1 and AKT2, IKBKB to NFKBIA, NFKBIB and NFKBIE) between these two pathways (Figures 5 and 6).

4. Discussion

In the present study, for the first time, through heterogeneous data aggregation in BNs (by BNrich method), we merged the gene expression data into SLE, TCR and BCR signaling pathways to uncover key genes and intergenic relationships which may serve as effective targets for therapeutic intervention in SLE patients.

We observed that the BNrich parameters associated with the same gene show substantial variation among different pathways. It is well established that as a BNs' property, every node implements a linear regression model based on its parents (upstream components) as regressors and hence, its output

completely depends on the topology of the underlying pathway. For instance, in the Illumina platform, the value of for JUN gene in TCR and BCR signaling pathway were 1.80163 and -1.535783, correspondingly (Additional file2). The main distinction of BNrich method in comparison some other well-known approaches like gene set enrichment analysis (GSEA) method [39] is that the latter assigns the same value to a gene amongst all pathways and determines the significance of genes regardless of intergenic relationships [40].

The uncleaned apoptotic components which recognized by both the innate and the adaptive immune systems can trigger the pathogenesis of SLE. Therefore, the dysregulation of effective genes involved in the clearance of apoptotic components has a fundamental role in SLE [41–45]. Remarkably, we showed upregulation of MACROH2A1, SNRPD3, and Ro60 and downregulation of MACROH2A2, TRIM21 and H4C9 (Figure 4B). In particular, this procedure followed by antigen presentation process aided by MHCII. While HLA-DPB1, HLA-DOB and HLA-DQA1 as key mediators of this pathway were distinctively activated, HLA-DPA1 and HLA-DQA2 showed reduction (Figure 4C). In previous studies, it had been shown that both HLA-DPB1 and HLA-DQA1 genetic variants predispose individuals to SLE [46–48]. Whereas CD40, CD86, IFNG and CD28 exhibited lower expression, both TNF and CD80 were augmented. TNF α is produced by a variety of cells and several studies have shown its elevated level in serum of SLE patients. Using anti-TNF monoclonal antibodies like infliximab and etanercept to treat SLE patients, also confirmed the importance of TNF cytokine as a targeting agent [49]. We reported significant increased biological functions in CD80@CD28 and CD86@CD28 (Figure 4C) for the first time, which activate TCR signaling pathway following MHCII antigen presentation and finally result in growing up the differentiation and proliferation rate of T cells in response to the apoptotic autoantigens. Recently, Katsuyama et al. indicated a T cells hyperresponsiveness following abnormal TCR signaling pathway in SLE patients [12]. At the end of the SLE signaling pathway, end-organ damage reported via several key genes (FCGR1A, C3, C7, C8G and C9). Amongst them, the deficiencies of early complement proteins, including C1, C4 and C2, was strongly associated with the development of SLE [50]. Our results also showed alteration in several biological functions; reduction in C1R@C2, C1R@C4B, C1S@C2 and promotion of C1R@C4A, C4A@C3, C1QB@C2 (Figure 4D). In overall, the findings could provide a valuable clue regarding prevention of CD80 and CD86 to CD28 costimulatory signals besides enhancing clearance mechanisms through activation of C1R to C2 and C4B, as well as C1S to C2 biological function, in order to discover new probable therapeutics targets which, lead to relieve of SLE symptoms.

As outlined earlier, the fundamental importance of distributed BCR and TCR signaling pathways in SLE pathogenesis has been well established via both clinical and experimental evidence [51,52]. In this context, we also observed defects in several genes involved in these signaling pathways (Figures 5B and 6B). For instance, in TCR signaling pathway flux, expression of some genes like CD4, PTPRC, PDCD1, CD28 and ZAP70 (which had several significant edges with their neighbours) were suppressed. Recently, a study by Matsuo et al. revealed that a genetic mutation in ZAP70 resulted in the TCR signaling defect which followed by T follicular helper (Tfh) cells development and the manifestation of lupus-like systemic autoimmunity [53]. The results may propose ZAP70 gene as one hub genes of the TCR signaling pathway implicated in SLE pathogenesis. As shown in Figure 5, all significant genes at the end

stage of the TCR signaling pathway, including TNF, IL5, IL10 and INFG were inhibited. In the BCR signaling pathway, the most up- and down-regulated genes were FCGR2B and INPPL, respectively. Furthermore, we also identified significant higher expression of CD79A and CD79B (also known as Igα and Igβ, respectively). Anti-CD79b mAb, with the ability to induce B cell anergy in collagen-induced arthritis (CIA), stated as a novel target against autoimmunity [54].

Interestingly, our results addressed the dysregulation of several noteworthy nodes and edges shared by both TCR and BCR signaling pathways. Amongst them were PI3K/AKT signaling pathway (Figures 5 and 6), which showed particularly increased expression of AKT1 and AKT2 and decreased expression of VAV gene homologs (VAV1, VAV2 and VAV3). In contrast with our study, heightened expression, as well as phosphorylation level of Akt, have been shown in SLE patients, which further support the major role of PI3K/Akt/TSC/mTOR signaling pathway in the disease [55,56]. According to our results, the MAPK signaling pathway as an effective subnetwork of TCR and BCR signaling in SLE pathogenesis [56], displayed a significant dysregulation. Not only MAPK genes (e.g. MAPK3, MAPK12 and MAPK14) were enhanced, PPP3CA gene acts in the downstream of BCR via MAPK signaling pathway, was also had a higher expression in CaN@NEAF, KEGG module in SLE patients. Similarly, the strong association of PPP3CA with SLE-like disease has been described as newly [57]. The transcription factor AP-1 includes FOS and JUN genes regulates cytokines and chemokines of TCR and BCR pathways in inflammatory and autoimmune disease [55]. In TCR signaling pathways, we observed the hub genes JUN and FOS, which had increased or decreased biological relations with their neighbours, significantly were inhibited. However, in the BCR signaling pathway, the significant edges entered to JUN and FOS from their upstream genes and FOS was inhibited.

Further investigation will be required to examine the efficiency of CD80 and CD86@CD28, C1R and C1S @C2, C1R@C4B, PI3K/AKT signaling pathways, besides JUN gene-targeting therapy in SLE disease. Moreover, analysing of the RNA-Seq datasets could be better for distinguishing expression levels and biological relationships of genes' isoforms in considered signaling pathways. Likewise, it is necessary to perform the BNRich approach among some other autoimmune diseases due to identify SLE signature genes more precisely.

Conclusions

In the present study, we exploited BNRich to hybridize intergenic relations inferred from signaling pathways and gene expression profiling in order to improve the insight about SLE as a complex disease. We figured out the significant genes and intergenic relationships and systemic alterations in SLE, TCR and BCR signaling pathways. Through cross-platform normalization and meta-analysis, large enough SLE patients were analyzed. Since multiple platforms were utilized, the effects of individual study-specific biases were reduced.

Abbreviations

BNs	Bayesian networks
BCR	B-cell receptor
CIA	collagen-induced arthritis
FDR	false discovery rate
GEO	Gene Expression Omnibus
GSEA	gene set enrichment analysis
GWAS	genome-wide association studies
HC	healthy control
IFN	interferon
PEA	pathway enrichment analysis
PBMC	peripheral blood mononuclear cell
PCA	principle component analysis
SLE	systemic lupus erythematosus
Tfh	T follicular helper
TCR	T-cell receptor

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and material: The R package of BNrich is available on [GitHub](#). All gene expression data are available in [GEO](#) database and the raw data of signaling pathways are available in [KEGG](#).

Competing interests: The authors declare that they have no competing interests.

Funding: This work has been supported by the Institute of Biochemistry and Biophysics, University of Tehran, Tehran.

Authors' contributions: SM, ASZ, VRT and KK established the method and developed the R package of BNrich. SM, ASZ and VRT performed data and statistical analysis. SM data collection and evaluation. ZS,

KK and SM contributed to interpretation of data. All authors performed editing and participated in the finalization of the manuscript and approved the final draft.

Acknowledgements: The authors would like to thank Mao Tanabe from Institute of Chemical Research, Kyoto University, for their expert insights in signaling pathways shared via several emails contributing greatly to the quality of this research.

References

1. Yang, F.; Zhai, Z.; Luo, X.; Luo, G.; Zhuang, L.; Zhang, Y.; Li, Y.; Sun, E.; He, Y. Bioinformatics identification of key candidate genes and pathways associated with systemic lupus erythematosus. *Clin. Rheumatol.* **2019**.
2. Javinani, A.; Ashraf-Ganjouei, A.; Aslani, S.; Jamshidi, A.; Mahmoudi, M. Exploring the etiopathogenesis of systemic lupus erythematosus: a genetic perspective. *Immunogenetics* **2019**, *71*, 283–297.
3. Helmick, C.G.; Felson, D.T.; Lawrence, R.C.; Gabriel, S.; Hirsch, R.; Kwoh, C.K.; Liang, M.H.; Kremers, H.M.; Mayes, M.D.; Merkel, P.A.; et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. *Arthritis Rheum.* **2008**, *58*, 15–25.
4. Garris, C.; Shah, M.; Farrelly, E. The prevalence and burden of systemic lupus erythematosus in a medicare population: Retrospective analysis of medicare claims. *Cost Eff. Resour. Alloc.* **2015**, *13*.
5. Meas, R.; Burak, M.J.; Sweasy, J.B. DNA repair and systemic lupus erythematosus. *DNA Repair (Amst)*. **2017**, *56*, 174–182.
6. Zharkova, O.; Celhar, T.; Cravens, P.D.; Satterthwaite, A.B.; Fairhurst, A.M.; Davis, L.S. Pathways leading to an immunological disease: systemic lupus erythematosus. *Rheumatology (Oxford)*. **2017**, *56*, i55–i66.
7. Tsokos, G.C.; Lo, M.S.; Reis, P.C.; Sullivan, K.E. New insights into the immunopathogenesis of systemic lupus erythematosus. *Nat. Rev. Rheumatol.* **2016**, *12*, 716–730.
8. Wardemann, H.; Yurasov, S.; Schaefer, A.; Young, J.W.; Meffre, E.; Nussenzweig, M.C. Predominant autoantibody production by early human B cell precursors. *Science (80-)*. **2003**, *301*, 1374–1377.
9. Swanson, C.L.; Wilson, T.J.; Strauch, P.; Colonna, M.; Pelanda, R.; Torres, R.M. Type I IFN enhances follicular B cell contribution to the T cell-independent antibody response. *J. Exp. Med.* **2010**, *207*, 1485–1500.
10. Xu, H.C.; Grusdat, M.; Pandyra, A.A.; Polz, R.; Huang, J.; Sharma, P.; Deenen, R.; Köhrer, K.; Rahbar, R.; Diefenbach, A.; et al. Type I Interferon Protects Antiviral CD8+ T Cells from NK Cell Cytotoxicity. *Immunity* **2014**, *40*, 949–960.
11. Oke, V.; Gunnarsson, I.; Dorschner, J.; Eketjäll, S.; Zickert, A.; Niewold, T.B.; Svenungsson, E. High levels of circulating interferons type I, type II and type III associate with distinct clinical features of active systemic lupus erythematosus. *Arthritis Res. Ther.* **2019**, *21*.

12. Katsuyama, T.; Tsokos, G.C.; Moulton, V.R. Aberrant T cell signaling and subsets in systemic lupus erythematosus. *Front. Immunol.* **2018**, *9*.
13. Li, Y.; Higgs, R.E.; Hoffman, R.W.; Dow, E.R.; Liu, X.; Petri, M.; Wallace, D.J.; Dörner, T.; Eastwood, B.J.; Miller, B.B.; et al. A Bayesian gene network reveals insight into the JAK-STAT pathway in systemic lupus erythematosus. *PLoS One* **2019**, *14*.
14. Yu, J.; Smith, V.A.; Wang, P.P.; Hartemink, A.J.; Jarvis, E.D. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* **2004**, *20*, 3594–3603.
15. Gendelman, R.; Xing, H.; Mirzoeva, O.K.; Sarde, P.; Curtis, C.; Feiler, H.S.; McDonagh, P.; Gray, J.W.; Khalil, I.; Korn, W.M. Bayesian network inference modeling identifies TRIB1 as a novel regulator of cell-cycle progression and survival in cancer cells. *Cancer Res.* **2017**, *77*, 1575–1585.
16. Luo, Y.; El Naqa, I.; McShan, D.L.; Ray, D.; Lohse, I.; Matuszak, M.M.; Owen, D.; Jolly, S.; Lawrence, T.S.; Kong, F.-M. (Spring); et al. Unraveling biophysical interactions of radiation pneumonitis in non-small-cell lung cancer via Bayesian network analysis. *Radiother. Oncol.* **2017**, *123*, 85–92.
17. Agrahari, R.; Foroushani, A.; Docking, T.R.; Chang, L.; Duns, G.; Hudoba, M.; Karsan, A.; Zare, H. Applications of Bayesian network models in predicting types of hematological malignancies. *Sci. Rep.* **2018**, *8*, 6951.
18. Ramos, J.; Das, J.; Felty, Q.; Yoo, C.; Poppiti, R.; Murrell, D.; Foster, P.J.; Roy, D. NRF1 motif sequence-enriched genes involved in ER/PR -ve HER2 +ve breast cancer signaling pathways. *Breast Cancer Res. Treat.* **2018**, *172*, 469–485.
19. Isci, S.; Ozturk, C.; Jones, J.; Otu, H.H. Pathway analysis of high-throughput biological data within a Bayesian network framework. **2011**, *27*, 1667–1674.
20. Kourou, K.; Papaloukas, C.; Fotiadis, D.I. Integration of Pathway Knowledge and Dynamic Bayesian Networks for the Prediction of Oral Cancer Recurrence. *IEEE J. Biomed. Heal. Informatics* **2017**, *21*, 320–327.
21. Maleknia, S.; Sharifi-Zarchi, A.; Tabar, V.R.; Namazi, M.; Kavousi, K. BNrich: A Bayesian network approach to the pathway enrichment analysis. *bioRxiv* **2020**, 2020.01.13.905448.
22. Neapolitan, R.E. *Learning Bayesian networks*, first.; Pearson Prentice Hall: Chicago, 2004; ISBN 0130125342.
23. Walsh, C.; Hu, P.; Batt, J.; Santos, C. Microarray Meta-Analysis and Cross-Platform Normalization: Integrative Genomics for Robust Biomarker Discovery. *Microarrays* **2015**, *4*, 389–406.
24. Nagarajan, R.; Scutari, M.; Lèbre, S. *Bayesian Networks in R*; Springer New York: New York, NY, 2013; ISBN 978-1-4614-6445-7.
25. Lee, H.M.; Sugino, H.; Aoki, C.; Nishimoto, N. Underexpression of mitochondrial-DNA encoded ATP synthesis-related genes and DNA repair genes in systemic lupus erythematosus. *Arthritis Res. Ther.* **2011**, *13*, R63.
26. Lee, H.M.; Mima, T.; Sugino, H.; Aoki, C.; Adachi, Y.; Yoshio-Hoshino, N.; Matsubara, K.; Nishimoto, N. Interactions among type I and type II interferon, tumor necrosis factor, and β -estradiol in the

- regulation of immune response-related gene expressions in systemic lupus erythematosus. *Arthritis Res. Ther.* **2009**, *11*, R1.
27. Kennedy, W.P.; Maciucă, R.; Wolslegel, K.; Tew, W.; Abbas, A.R.; Chaivorapol, C.; Morimoto, A.; McBride, J.M.; Brunetta, P.; Richardson, B.C.; et al. Association of the interferon signature metric with serological disease manifestations but not global activity scores in multiple cohorts of patients with SLE. *Lupus Sci. Med.* **2015**, *2*.
28. 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.
29. Petri, M.; Fu, W.; Ranger, A.; Allaire, N.; Cullen, P.; Magder, L.S.; Zhang, Y. Association between changes in gene signatures expression and disease activity among patients with systemic lupus erythematosus. *BMC Med. Genomics* **2019**, *12*.
30. Jiang, S.H.; Athanasopoulos, V.; Ellyard, J.I.; Chuah, A.; Cappello, J.; Cook, A.; Prabhu, S.B.; Cardenas, J.; Gu, J.; Stanley, M.; et al. Functional rare and low frequency variants in BLK and BANK1 contribute to human lupus. *Nat. Commun.* **2019**, *10*.
31. Hamid, J.S.; Hu, P.; Roslin, N.M.; Ling, V.; Greenwood, C.M.T.; Beyene, J. Data Integration in Genetics and Genomics: Methods and Challenges. *Hum. Genomics Proteomics* **2009**, *1*.
32. Larsen, M.J.; Thomassen, M.; Tan, Q.; Sørensen, K.P.; Kruse, T.A. Microarray-based RNA profiling of breast cancer: Batch effect removal improves cross-platform consistency. *Biomed Res. Int.* **2014**, *2014*.
33. Irigoyen, A.; Jimenez-Luna, C.; Benavides, M.; Caba, O.; Gallego, J.; Ortuño, F.M.; Guillen-Ponce, C.; Rojas, I.; Aranda, E.; Torres, C.; et al. Integrative multi-platform meta-analysis of gene expression profiles in pancreatic ductal adenocarcinoma patients for identifying novel diagnostic biomarkers. *PLoS One* **2018**, *13*.
34. Taminau, J.; Lazar, C.; Meganck, S.; Nowé, A. Comparison of Merging and Meta-Analysis as Alternative Approaches for Integrative Gene Expression Analysis. *ISRN Bioinforma.* **2014**, *2014*, 1–7.
35. Leek, J.T.; Johnson, W.E.; Parker, H.S.; Jaffe, A.E.; Storey, J.D. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **2012**, *28*, 882–883.
36. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361.
37. Wasserman, L. All of Statistics: A Concise Course in Statistical Inference. In *The American Statistician*; 2004; Vol. 59, pp. 161–163&216–217 ISBN 0387402721.
38. Andrade, J.M.; Estévez-Pérez, M.G. Statistical comparison of the slopes of two regression lines: A tutorial. *Anal. Chim. Acta* **2014**, *838*, 1–12.
39. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: a knowledge-based

- approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15545–50.
40. García-Campos, M.A.; Espinal-Enríquez, J.; Hernández-Lemus, E. Pathway Analysis: State of the Art. *Front. Physiol.* **2015**, *6*, 383.
 41. Asadipour, M.; Hassan-Zadeh, V.; Aryaeian, N.; Shahram, F.; Mahmoudi, M. Histone variants expression in peripheral blood mononuclear cells of patients with rheumatoid arthritis. *Int. J. Rheum. Dis.* **2018**, *21*, 1831–1837.
 42. Hung, T.; Pratt, G.A.; Sundararaman, B.; Townsend, M.J.; Chaivorapol, C.; Bhangale, T.; Graham, R.R.; Ortmann, W.; Criswell, L.A.; Yeo, G.W.; et al. The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science (80-)*. **2015**, *350*, 455–459.
 43. Kamiyama, R.; Yoshimi, R.; Takeno, M.; Iribe, Y.; Tsukahara, T.; Kishimoto, D.; Kunishita, Y.; Sugiyama, Y.; Tsuchida, N.; Nakano, H.; et al. Dysfunction of TRIM21 in interferon signature of systemic lupus erythematosus. *Mod. Rheumatol.* **2018**, *28*, 993–1003.
 44. Espinosa, A.; Dardalhon, V.; Brauner, S.; Ambrosi, A.; Higgs, R.; Quintana, F.J.; Sjöstrand, M.; Eloranta, M.L.; Gabhann, J.N.; Winqvist, O.; et al. Loss of the lupus autoantigen Ro52/Trim21 induces tissue inflammation and systemic autoimmunity by disregulating the IL-23-Th17 pathway. *J. Exp. Med.* **2009**, *206*, 1661–1671.
 45. Shimada-Sugimoto, M.; Otowa, T.; Miyagawa, T.; Khor, S.S.; Kashiwase, K.; Sugaya, N.; Kawamura, Y.; Umekage, T.; Kojima, H.; Saji, H.; et al. Immune-related pathways including HLA-DRB1*13:02 are associated with panic disorder. *Brain. Behav. Immun.* **2015**, *46*, 96–103.
 46. Cruz-Tapias, P.; Castiblanco, J.; Correa, N.E.; Montoya-Ortíz, G. *Autoimmunity - from bench to bedside*; 2013; ISBN 9789587383669.
 47. Prokunina, L.; Castillejo-López, C.; Öberg, F.; Gunnarsson, I.; Berg, L.; Magnusson, V.; Brookes, A.J.; Tentler, D.; Kristjansdóttir, H.; Gróndal, G.; et al. A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nat. Genet.* **2002**, *32*, 666–669.
 48. Georgopoulou, C.; Zintzaras, E.; Papadimitropoulos, M.; Spyropoulou, M.; Stavropoulou, A.; Moutsopoulos, H.; Manoussakis, M. HLA DR-DQ haplotypes and genotypes in patients with systemic lupus erythematosus with co-existing Sjogren's syndrome, systemic lupus erythematosus without Sjogren's syndrome and primary Sjogren's syndrome: clinical and laboratory associations. *Arthritis Res. Ther.* **2005**, *7*, P100.
 49. Jacob, N.; Stohl, W. Cytokine disturbances in systemic lupus erythematosus. *Arthritis Res. Ther.* **2011**, *13*.
 50. Macedo, A.C.L.; Isaac, L. Systemic lupus erythematosus and deficiencies of early components of the complement classical pathway. *Front. Immunol.* **2016**, *7*.
 51. Jakez-Ocampo, J.; Paulín-Vera, C.M.; Gómez-Martín, D.; Lima, G.; Vargas-Rojas, M.I.; Llorente, L.; Rivadeneyra-Espinoza, L.; Pérez-Romano, B.; Calva-Cevenini, G.; Ruiz-Argüelles, A.; et al. Vβ T cell receptor (TCR) genes in circulating cells of patients with systemic lupus erythematosus and their healthy relatives. *Gac. Med. Mex.* **2018**, *154*, 74–79.

52. Ma, K.; Du, W.; Wang, X.; Yuan, S.; Cai, X.; Liu, D.; Li, J.; Lu, L. Multiple functions of B cells in the pathogenesis of systemic lupus erythematosus. *Int. J. Mol. Sci.* **2019**, *20*.
53. Matsuo, T.; Hashimoto, M.; Sakaguchi, S.; Sakaguchi, N.; Ito, Y.; Hikida, M.; Tsuruyama, T.; Sakai, K.; Yokoi, H.; Shirakashi, M.; et al. Strain-Specific Manifestation of Lupus-like Systemic Autoimmunity Caused by Zap70 Mutation . *J. Immunol.* **2019**, *202*, 3161–3172.
54. Hardy, I.R.; Anceriz, N.; Rousseau, F.; Seefeldt, M.B.; Hatterer, E.; Irla, M.; Buatois, V.; Chatel, L.E.; Getahun, A.; Fletcher, A.; et al. Anti-CD79 Antibody Induces B Cell Anergy That Protects against Autoimmunity. *J. Immunol.* **2014**, *192*, 1641–1650.
55. He, J.; Ma, J.; Ren, B.; Liu, A. Advances in systemic lupus erythematosus pathogenesis via mTOR signaling pathway. *Semin. Arthritis Rheum.*
56. Beşliu, A.N.; Pistol, G.; Marica, C.M.; Bănică, L.M.; Chiţonu, C.; Ionescu, R.; Tănăseanu, C.; Tamsulea, I.; Matache, C.; Stefănescu, M. PI3K/Akt signaling in peripheral T lymphocytes from systemic lupus erythematosus patients. *Roum. Arch. Microbiol. Immunol.* **2009**, *68*, 69–79.
57. Wilbe, M.; Kozyrev, S. V.; Farias, F.H.G.; Bremer, H.D.; Hedlund, A.; Pielberg, G.R.; Seppälä, E.H.; Gustafson, U.; Lohi, H.; Carlborg, Ö.; et al. Multiple Changes of Gene Expression and Function Reveal Genomic and Phenotypic Complexity in SLE-like Disease. *PLoS Genet.* **2015**, *11*.

Figures

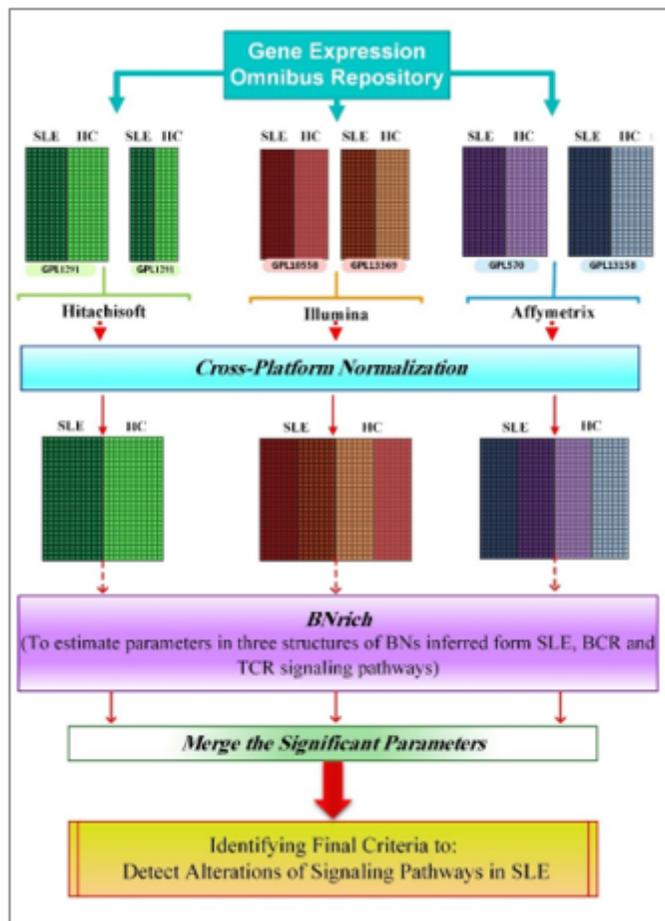


Figure 1

Workflow and analysis procedure; from row SLE and healthy control (HC) data in various datasets to identify signaling pathways alterations in SLE patients.

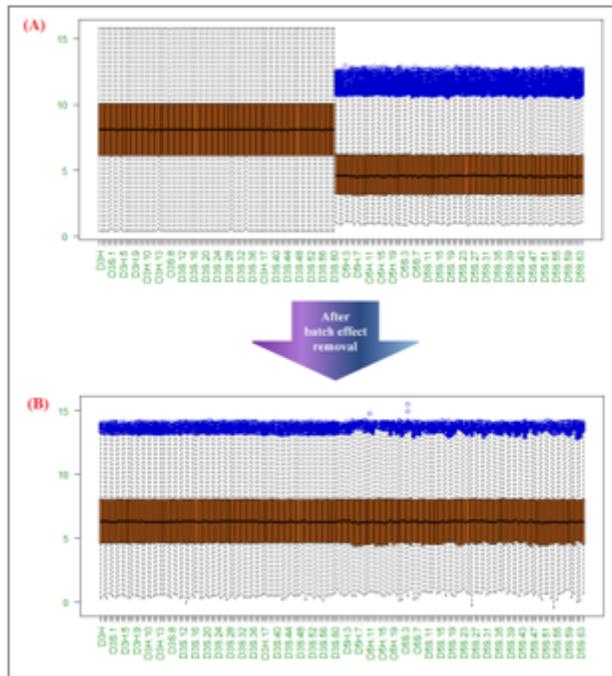


Figure 2

Graphical demonstration of the batch effect removal using ComBat. Boxplots for the gene expression distributions in Affymetrix (A) before and (B) after batch removal. The distributions show the normalization and decreasing technical diversities between datasets. In the horizontal axis, the j th healthy control subjects and the j th SLE patient in i th dataset were illustrated with DiH,j and DiS,j , correspondingly.

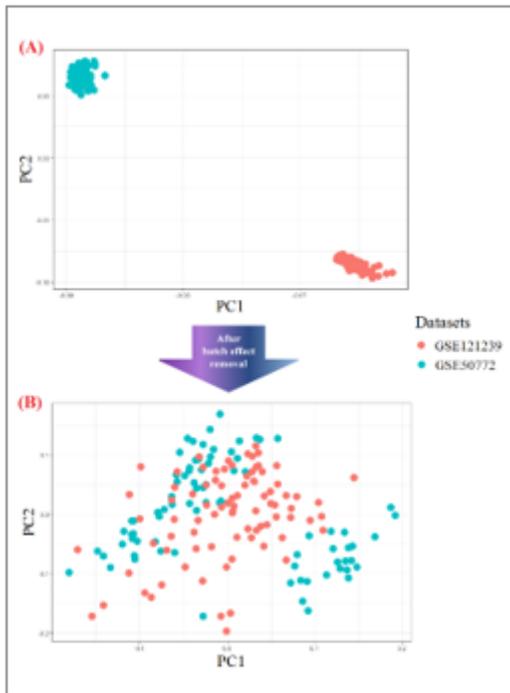


Figure 3

Graphical demonstration of the batch effect removal using ComBat. The PCA plots of the microarray datasets (A) before, and (B) after batch removal for Affymetrix platforms. Each dot represents one sample and the color indicates its dataset.

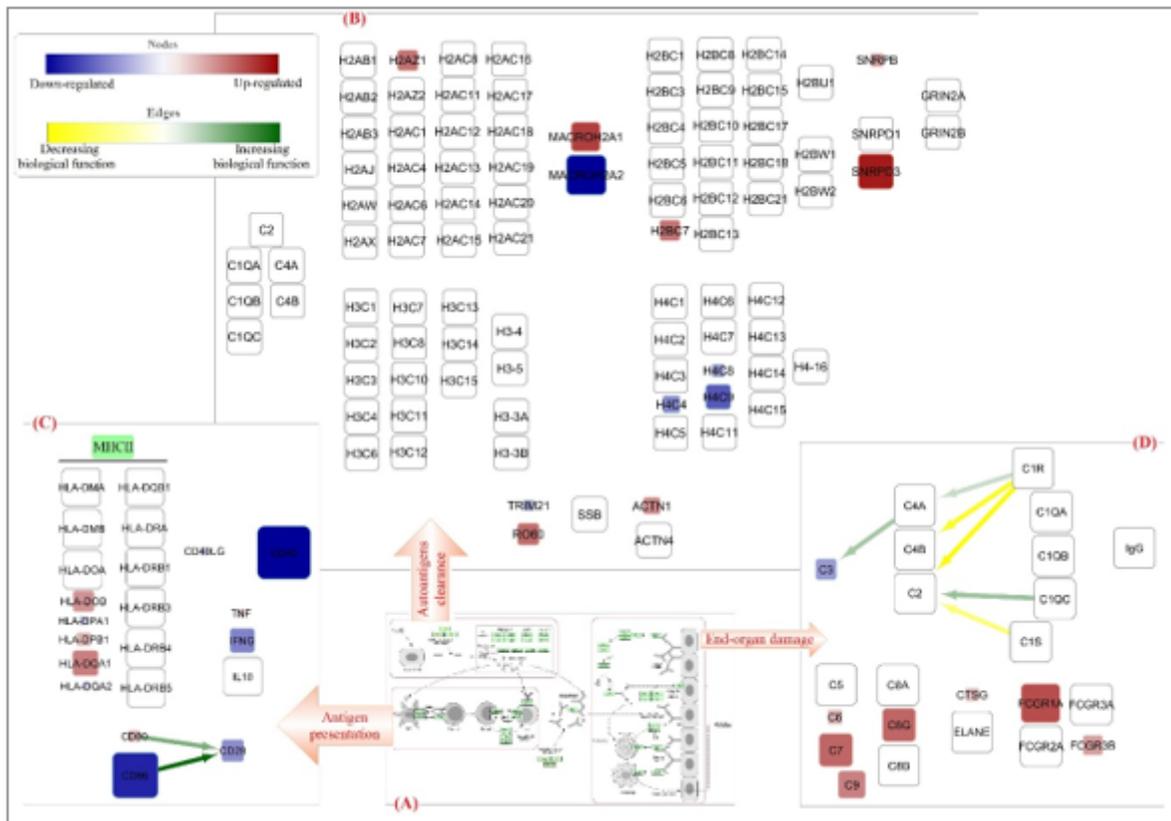


Figure 4

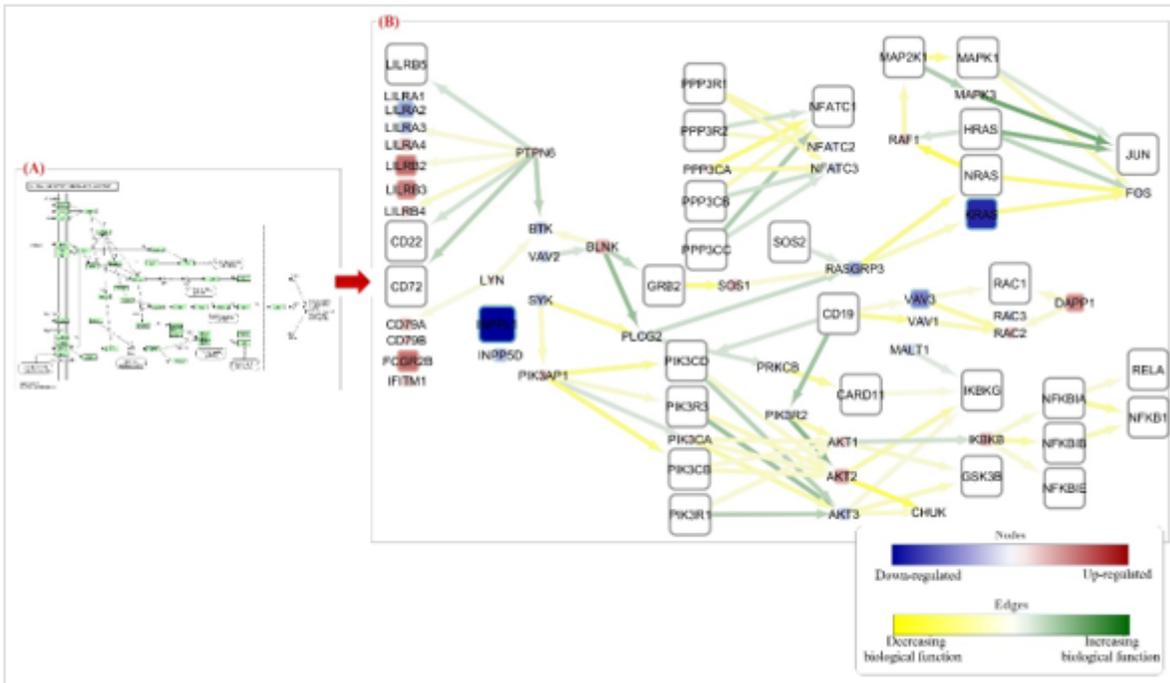


Figure 6

The alterations in BCR signaling pathway; The schematic figure of the BCR signaling pathway in KEGG (A); and the BCR signaling pathway alterations based on calculated final criteria (B). The color and size of nodes reflect the values and the absolute values of the $\log_2(fc^*)$ respectively and it is equivalent to up-regulate or down-regulate of genes. Moreover, the color of the edges reflects the values of the $\log_2(fc^*)$ respectively and it is equivalent to increasing biological function or decreasing biological function.

Supplementary Files

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

- [Additionalfile3.xlsx](#)
- [Additionalfile2.xlsx](#)
- [Additinalfile1.pdf](#)