

Identifying pathogenic variants related to systemic lupus erythematosus by integrating genomic databases and a bioinformatic approach

by Universitas Ahmad Dahlan Yogyakarta 22

Submission date: 12-Jan-2024 08:19AM (UTC+0700)

Submission ID: 2269678753

File name: 19.pdf (3.62M)

Word count: 6741

Character count: 37736



38

Original article

eISSN 2234-0742
Genomics Inform 2023;21(3):e37
<https://doi.org/10.5808/gi.23002>

Received: January 12, 2023
Revised: June 15, 2023
Accepted: August 9, 2023

*Corresponding author:
E-mail: dyonisanp@staff.uns.ac.id

© 2023 Korea Genome Organization
This is an open-access article distributed under the terms of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Identifying pathogenic variants related to systemic lupus erythematosus by integrating genomic databases and a bioinformatic approach

Ratih Dewi Yudhani¹, Dyonisa Nasirochmi Pakha^{1*}, Suyatmi Suyatmi², Lalu Muhammad Irham³

¹Department of Pharmacology, Faculty of Medicine, Universitas Sebelas Maret, Surakarta 57126, Indonesia

²Department of Histology, Faculty of Medicine, Universitas Sebelas Maret, Surakarta 57126, Indonesia

³Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta 55166, Indonesia

Systemic lupus erythematosus (SLE) is an inflammatory-autoimmune disease with a complex multi-organ pathogenesis, and it is known to be associated with significant morbidity and mortality. Various genetic, immunological, endocrine, and environmental factors contribute to SLE. Genomic variants have been identified as potential contributors to SLE susceptibility across multiple continents. However, the specific pathogenic variants that drive SLE remain largely undefined. In this study, we sought to identify these pathogenic variants across various continents using genomic and bioinformatic-based methodologies. We found that the variants rs35677470, rs34536443, rs17849502, and rs13306575 are likely damaging in SLE. Furthermore, these four variants appear to affect the gene expression of *NCF2*, *TYK2*, and *DNASE1L3* in whole blood tissue. Our findings suggest that these genomic variants warrant further research for validation in functional studies and clinical trials involving SLE patients. We conclude that the integration of genomic and bioinformatic-based databases could enhance our understanding of disease susceptibility, including that of SLE.

Keywords: bioinformatic, genetic factors, genomic variants, single-nucleotide polymorphism, systemic lupus erythematosus

Introduction

Systemic lupus erythematosus (SLE) is a chronic inflammatory-autoimmune disease with a complex multi-organ pathogenesis. The incidence of SLE is highest in women of reproductive age, with a female-to-male ratio around 9:1. The clinical signs of SLE are autoantibody secretion, complement activation, and immune complex deposition, all of which can result in extensive tissue and organ damage and a poor prognosis [1]. The pathogenesis of SLE has yet to be fully understood, but it is believed to involve a complex interplay between genetic predisposition and environmental factors [1,2].

This disease remains a significant challenge for researchers and clinicians due to its complex etiology and pathogenesis, heterogeneous clinical manifestations, and unpredictable exacerbations. Therefore, efforts for identifying potential biomarkers are urgent.

ly pursued for several critical reasons. For instance, SLE is often misdiagnosed because no single test is sufficiently sensitive or specific. Additionally, there is no reliable laboratory test for predicting flares and exacerbations, or for identifying specific organ involvement in individual patients with diverse SLE manifestations. Furthermore, the limited availability of biomarkers impedes the discovery of new candidates for SLE [3]. Biomarkers reflect a range of biological processes, including genomic, molecular, histological, and serological markers, that correlate with clinical manifestations or disease pathogenesis [4].

Both twin and familial studies indicate a strong relationship between genetic factors and SLE. Furthermore, over 80 SLE susceptibility loci have been identified through numerous genome-wide and candidate-gene association studies (GWAS) [1]. Through GWAS, it has been documented that several single-nucleotide polymorphisms (SNPs) are associated with the pathogenesis of SLE. However, the disease is not influenced by a single genetic susceptibility alone. Instead, it results from a complex interplay of multiple variants, which could serve as genetic biomarkers. Despite this, the identification of SNPs with missense mutations as the most significant pathogenic variants contributing to SLE remains limited. Consequently, this study aimed to identify genetic variations associated with SLE pathogenesis in a wide range of populations, using genomic databases and bioinformatics approaches. This research is anticipated to provide a comprehensive overview of potential biomarker candidates. These candidates could be used to further investigate the intricate pathogenesis of SLE related to genetic profiles, thereby supporting current targets in SLE drug discovery and development.

Methods

Identifying genomic variants is crucial not only for understanding the structure of the human genome, but also for gaining a deeper understanding of disease biology. In this study, we utilized several bioinformatics-based approaches to integrate variants associated with SLE. The study was conducted using the National Human Genome Research Institute (NHGRI) GWAS Catalog Database (<https://www.ebi.ac.uk/gwas/>) with the term "systemic lupus erythematosus" (EFO ID: EFO_0002690), which was downloaded on October 28, 2022 [5]. This yielded 1,259 variants and risk alleles. We then focused on the missense variant due to its potential to alter protein function. A p-value threshold of $< 10^{-8}$ was established to distinguish true positives from false-positives [6,7]. This resulted in a total of 46 variants. Subsequently, an odds ratio greater than one was used to rank the risk of genes affecting SLE, yield-

ing 19 data points after removing duplicates. The SNPnexus (<https://www.snp-nexus.org>) with PolyPhen-2 databases was employed to determine which variants influenced protein changes in the disease. These were classified as benign, possibly damaging, or probably damaging [8-12], and were extracted on October 28, 2022. A total of four SNPs with 29 variants were predicted as possibly or probably damaging. Population data for the variants were also extracted from SNPnexus using 1000 Genomes Data. The expression profile of the three involved genes was then evaluated using the GTEx Portal (<http://www.gtexportal.org/home/>) to understand gene expression across various tissues. This data was obtained from the GTEx Portal on October 28, 2022.

A summary of the methodology used to screen for SLE-associated variants is illustrated in the various steps of the bioinformatics pipeline (Fig. 1). A similar method was used by Puspitaningrum et al. [13] to distinguish genomic variants in Sjögren syndrome. Additionally, the same approach was used to identify pathogenic variants of genes related to coronavirus disease 2019 and chickenpox [14].

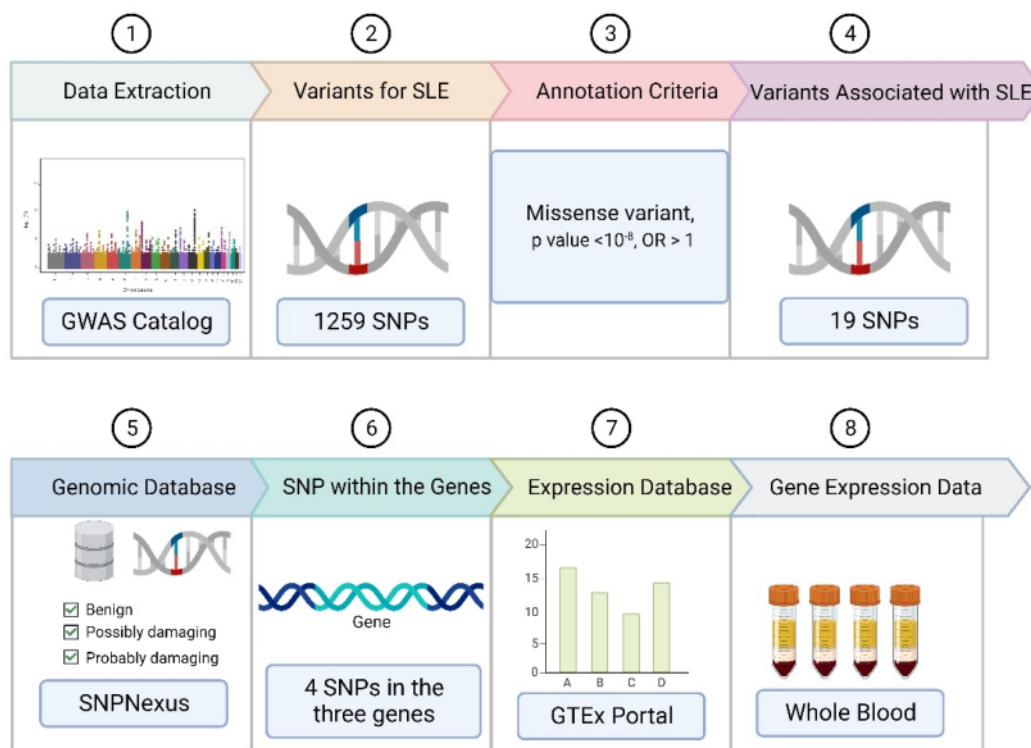
Results

Identification of SLE-associated SNPs

Nineteen SLE-associated SNPs were collected from the GWAS Catalog Database (Table 1). Furthermore, as shown in Table 2, four variant SNPs were identified from the PolyPhen-2 database as contributing to protein-level damage. These SNPs represented three distinct genes: *NCF2*, *TYK2*, and *DNASE1L3*. The *NCF2* gene with the SNP rs13306575 was predicted having possibly damaging by increasing the possibility of having SLE, while the SNP rs17849502 in this gene was predicted to be probably damaging at the protein level. Moreover, both the *TYK2* and *DNASE1L3* genes were predicted to alter the pathogenesis of SLE through a probably damaging mechanism. The strongest contribution was found for the *TYK2* and *DNASE1L3* genes, with the SNPs rs34536443 and SNP rs35677470, respectively, which were predicted to be probably damaging (score 0.999).

Distribution of SLE-associated SNPs

Table 3 displays the distribution of allele frequencies gathered from the 1000 Genome Database. As indicated in Table 3, this study revealed that, with the exception of rs13306575, Africans and East Asians had allele frequencies of less than 1% or none at all for all variants. In contrast, the South Asian population exhibited an allele frequency of 1%–2% for three SNPs. The American population had frequencies exceeding 2% for all variants. For



31 1. Schematic of bioinformatics pipeline to identify systemic lupus erythematosus (SLE) susceptibility genes across multiple continents. GWAS, genome-wide association studies; SNP, single-nucleotide polymorphism.

rs13306575, the highest distribution was observed in East Asia, nearing 7%, while the other SNPs were predominantly distributed in Europe, with frequencies ranging from approximately 3% to 6%.

Identification of the tissue gene expression of SLE-associated SNPs

The distribution of the expression of the three genes in different tissues was obtained through GTEx Portal. Of the three genes, *NCF2* was predominantly expressed in whole blood, followed by spleen, lung, and cells (namely, Epstein-Barr virus [EBV]-transformed lymphocytes) (Fig. 2). Similarly, *TYK2* was mainly expressed in cells (EBV-transformed lymphocytes) and spleen, followed by whole blood and the lung (Fig. 3). Meanwhile, the expression of *DNASE1L3* was primarily found in the spleen. In contrast, this gene had low levels of expression in the lung and whole blood (Fig. 4).

Discussion

Genetic variants of SLE-associated SNPs

NCF2 gene

Tables 1 and 2 indicate that neutrophil cytosolic factor 2 (*NCF2*), which is located on chromosome 1 and had the SNPs rs17849502 and rs13306575, was one of the strongest contributors to the risk of SLE and could be potentially damaging (score: 0.998). *NCF2* is a subunit of the enzyme complex nicotinamide adenine dinucleotide phosphate (NADPH) oxidase, which generates superoxide in phagocytes, including neutrophils and leukocytes. Thus, this enzyme will digest (neutralize) foreign pathogens or remove cell debris [15]. The leukocyte NADPH oxidation complex consists of two membrane proteins (CYBA and CYBB) and three soluble proteins (*NCF1*, *NCF2*, and *NCF4*) alongside a small GTPase (Rac1/2) as an activator [16]. Deficiencies in any of these proteins could lead to the onset of chronic granulomatous disease (CGD), a primary immunodeficiency disorder that typi-

cally present in early childhood. This disease is characterized by low levels of reactive oxygen species (ROS) in the phagosome, which are associated with a deficiency in NADPH oxidase activity. The reduction in NADPH activity leads to abnormal phagocyte digestion and the formation of granulomas in various organs, resulting in recurrent infections, inflammatory disorders, and auto-

immunity [17].

In addition to CGD, many studies have identified that alterations in the *NCF2* gene sequence might be related to the onset of lupus and lupus-like diseases [18]. Most autoimmune diseases are characterized by a reduced immunological tolerance mechanism, which leads to the production of autoantibodies and chronic inflammation, causing damage to various tissues and organs. This pathogenic interaction could potentially explain the link between certain autoimmune diseases and reduced ROS production, which is a result of inadequate activity of the NADPH oxidase complex [19].

The rs17849502 SNP is one of the missense variants of the *NCF2* gene that results in a histidine to glutamine (H389Q) substitution in the *NCF2* protein. This substitution leads to a reduction of *NCF2*'s binding affinity with the Vav1 protein. These processes further inhibit NADPH oxidase activity due to the stimulation of the signaling pathway involved in Vav1 and could be a genetic risk factor for both adult- and juvenile-onset SLE [20]. Both rs17849502 (*NCF2* gene, substitution His389Gln) and rs13306575 (*NCF2* gene, substitution Arg395Trp) were strongly associated with SLE in North Americans of Hispanic descent ($p = 4.91 \times 10^{-9}$ and 1.50×10^{-11} , respectively). However, only rs17849502 was significantly associated with SLE in North Americans of European descent (EA) ($p = 9.47 \times 10^{-14}$). Arg-395 plays a role in stabilizing the interaction between *NCF2* and the C-terminal tail of *NCF4* via hydrogen bond formation with the carboxyl oxygen of *NCF4* residue 339. The Arg-395 → Trp mutation (rs13306575) deteriorates this interaction, which, consequently, will destabilize loop 395–402 of *NCF2* and disrupt the

Table 1. Missense variants of SLE-associated SNPs with $p < 10^{-8}$

SNP	p-value
rs35677470	5×10^{-9}
rs2286672	3×10^{-9}
rs7097397	2×10^{-9}
rs1131476	1×10^{-9}
rs10516487	4×10^{-10}
rs1061502	9×10^{-11}
rs1801274	6×10^{-11}
rs11574637	3×10^{-11}
rs566731348	4×10^{-12}
rs2476601	3×10^{-12}
rs2304256	2×10^{-12}
rs13306575	2×10^{-14}
rs17849502	3×10^{-17}
rs2230926	1×10^{-17}
rs1131665	9×10^{-21}
rs34536443	2×10^{-25}
rs9274384	5×10^{-31}
rs7097397	5×10^{-48}
rs1274384	6×10^{-54}

SLE, systemic lupus erythematosus; SNP, single-nucleotide polymorphism.

Table 2. SLE-associated SNPs and their effects at the protein level

SNP	Chromosome	Gene	Score	Prediction
rs35677470	chr3	<i>DNASE1L3</i>	0.979–0.999	Probably damaging
rs34536443	chr19	<i>TYK2</i>	0.973–0.999	Probably damaging
rs17849502	chr1	<i>NCF2</i>	0.919–0.998	Probably damaging
rs13306575	chr1	<i>NCF2</i>	0.803–0.806	Possibly damaging

SLE, systemic lupus erythematosus; SNP, single-nucleotide polymorphism.

Table 3. The distribution of allele frequencies of four SNPs across multiple continents

SNP	Allele		Allele frequency				
	REF allele	ALT allele	African	American	East Asian	European	South Asian
rs35677470	G	A	0.0030	0.0317	None	0.0527	0.0215
rs34536443	G	C	0.0015	0.0202	None	0.0288	0.0061
rs17849502	G	T	None	0.0245	None	0.0596	0.0143
rs13306575	G	A	0.0008	0.0245	0.069400	None	0.0020

SNP, single-nucleotide polymorphism; REF, reference; ALT, alternative.

this SNP might be functional, as it reduces the pSTAT1-induced level of IFN- α in peripheral blood mononuclear cells (PBMCs) relative to cells derived from patients with the G allele, hence decreasing IFNAR signaling [28]. The C allele of rs34536443 diminished p-STAT3 levels induced by IFN- γ and IL-12. In addition, it decreased STAT2-induced IFN- β in a murine model of multiple sclerosis (MS) and PBMCs of MS patients carrying the C allele compared to the G allele [29].

In line with that meta-analysis, Diogo et al. [30] documented that allele C of rs34536443 serves to protect against several autoimmune diseases, including inflammatory bowel disease, rheumatoid arthritis, and SLE. Contreras-Cubas et al. [31] demonstrated a protective OR for the C allele of rs34536443 (OR, 0.370; $p=0.034$) in childhood-onset SLE as an independent predictor after adjusting for sex and ancestry. The protective effect of rs34536443 variants was also found in adult-onset SLE in Mexican patients (OR, 0.277; $p=0.008$).

The pattern of activated cytokine signaling linked to TYK2 indicated that small molecules that inhibit TYK2 could be potential candidate drugs for SLE patients [32]. A greater understanding of TYK2's molecular and cellular characteristics and studies indicating that TYK2 gene polymorphisms are protective against developing SLE will lead to the discovery of therapeutic approaches for autoimmune diseases such as SLE.

DNASE1L3 gene

Tables 1 and 2 indicate that the DNASE1L3 gene, which is located on chromosome 3 and had the SNP rs35677470, had one of the strongest contributions to the risk of SLE and was predicted to be probably damaging (score 0.999). DNASE1L3 is a Mg²⁺Ca²⁺ dependent endonuclease predominantly expressed in myeloid cells [33,34]. One of the roles of DNASE1L3 is to clear DNA released from cells undergoing apoptosis [35]. Meanwhile, it has been found that the absence of DNASE1L3 leads to the accumulation of circulating apoptotic bodies. This triggers an autoimmune response, producing anti-DNA antibodies [34,35]. Yu et al. [36] showed a higher antibody reactivity to DNASE1L3 in SLE patients than in healthy individuals. These anti-DNASE1L3 antibodies are linked significantly with increased SLE disease activity, alongside elevated anti-dsDNA and low complement levels. Moreover, the anti-DNASE1L3 antibody-positive subgroup had higher upregulation of IFN, myeloid/neutrophil/granulocyte, and inflammation modules, indicating that this antibody is associated with increased immune pathway stimulation. However, the data were insufficient to determine the function of anti-DNASE1L3 antibodies in predicting future flares [36].

Moreover, it has been documented that DNASE1L3 deficiency contributes to pediatric-onset SLE, characterized by anti-dsDNA antibodies, reduced complement, and antineutrophil cytoplasmic antibodies [33-35]. Hence, this also suggests that DNASE1L3 might have a protective role from autoimmunity [34]. In addition, DNASE1L3 $-/-$ mice developed as a model of pediatric-onset SLE had similar symptoms to humans deficient in DNASE1L3. These mice developed a delay in immune activation, early anti-dsDNA, and anti-chromatin antibodies. The disparity in the timing of the activation of anti-dsDNA and other antibody responses is suggested to be due to the impairment in the initial priming steps, indicating a specific immune activation defect [33].

The missense variant rs35677470 at the DNASE1L3 locus is linked to the development of SLE [37,38]. The structural analysis by Zervou et al. [37] revealed a modification of the conserved electrostatic network between the guanidinium group of the Arg206 side chain and the carboxylate group of Glu170, forming a strong salt bridge. An electrostatic salt bridge network is formed in conjunction with the Arg208 to Asp219 charge interaction. Meanwhile, the defective rs35677470 SNP allele, with the substitution of arginine by cysteine, disrupts this network and the molecular architecture, including disordered protein folding. This might affect DNASE1L3 enzymatic activity [37]. Additionally, the 3p14.3 locus containing the DNASE1L3 gene (exon 8) is linked to the risk of SLE, which is a potentially causal missense variant rs35677470 [37,38]. A study by Coke et al. [38] found that the OR for SLE among heterozygous and homozygous carriers of the rs35677470 risk allele was 1.14 (95% CI, 1.05 to 1.24) and 1.68 (95% CI, 1.14 to 2.47), respectively, indicating that both alleles are risk factors. Therefore, SLE patients carrying the rs35677470 risk allele may benefit from DNASE1L3 replacement therapy or other promising therapies to restore DNASE1L3 protein cellular secretion [38]. Furthermore, a previous study found that anti-DNASE1L3 antibodies in SLE were produced by autoreactive VH4-34+ B cells with the 9G4 idiotype, which are intrinsically autoreactive B cells escaping the tolerance checkpoint in SLE. Hence, this previous study indicated that VH4-34+ B cells and 9G4 antibodies are potential targets for SLE treatment [36].

Distribution of genetic variants of SLE-associated SNPs

This study found that most SNPs were primarily distributed in Europeans, except for rs13306575, which was mainly in East Asia (Table 3). These alleles were also distributed among the American and South Asian populations. This distribution aligns with the documented prevalence of SLE, which is notably higher among Asians and African Americans [39]. A systematic review by Stojan

and Petri [40] also showed that the highest estimates of incidence and prevalence of SLE are in North America (23.2/100,000 person-years and 241/100,000 people, respectively). This current study also found that the frequency of alleles of the four SNPs is below 1% in the African population. This finding is supported by a systematic review that found the lowest SLE incidence rates in Africa and Ukraine (0.3/100,000 person-years). Furthermore, Northern Australia has the lowest prevalence of SLE (0 cases among 847 people) [40].

A previous study by Ueki et al. [41] found that rs35677470 (R206C) was accompanied by a homozygous C686 allele in all Asian and African populations, suggesting that the *DNASE1L3* gene exhibits low genetic structure diversity in terms of non-synonymous SNPs since it codes for an enzyme that has been well-conserved throughout human evolution. In contrast, a heterozygous (C686/T686) allele was found in three Caucasian populations (Turkish, German, and Mexican) with a frequency between 3.5% and 15.4%. Compared to Asian and African populations, the genotype distribution of Caucasian populations differed significantly ($p = 1.03 \times 10^{-17}$), suggesting this allele is Caucasian-specific. This variation could be related to the prevalence of autoimmune disease [41].

Furthermore, a previous study by Faezi et al. [27] found that the C allele of rs34536443 in the Iranian population had no impact on SLE susceptibility. This highlighted the importance of genetic divergence in diverse populations and the contribution of different genes to the etiopathogenesis of a multigenic disease like SLE [27]. Meanwhile, the rs34536443 allele was associated with autoimmune disease in the European population. However, the rs34536443 allele is absent or highly uncommon in Asian populations [42].

In addition, numerous studies have indicated that the T allele of the rs17849502 SNP is closely related to the occurrence of adult-onset SLE in various ethnic groups [20,21,43,44]. Furthermore, Bakutenko et al. [15] revealed a strong association between the minor T allele of the rs17849502 SNP and the risk of juvenile-onset SLE in the Belarusian population. The observed frequencies of the T allele in the case groups of juvenile-onset SLE and clinical control were 14.3% and 4.9%, respectively. Moreover, Jacob et al. [20] also observed a significant association of rs17849502 with the European American subgroup for both adult-onset SLE and childhood-onset SLE. However, no association was evident between rs17849502 and SLE in Asian Americans or African Americans [20]. In contrast, Kim-Howard et al. [43] found an association between rs17849502 and SLE in European American, Hispanic, and African American populations. Fur-

thermore, an independent association was observed between a non-synonymous variant rs13306575 and SLE in Hispanic and Korean populations ($p_{HS} = 7.04 \times 10^{-7}$ and $p_{KR} = 3.30 \times 10^{-3}$, respectively). Additionally, a significant interaction between rs13306575 and rs17849502 was observed in the Hispanic population, significantly elevating the risk (OR, 6.55) [43].

The tissue gene expression of SLE-associated SNPs

Gene regulatory networks govern both standard and tissue-specific processes, determining gene expression and its levels. While tissue specificity is often described based on gene expression levels, it's understood that individual genes or gene sets alone cannot adequately define the diverse processes that differentiate various tissues. Instead, a combination of regulatory elements, primarily transcription factors, work alongside other genetic and environmental factors to control gene transcription and protein phenotype. The transcriptomic data provided by the Genotype-Tissue Expression (GTEx) consortium presents an unparalleled opportunity to investigate the intricate regulatory patterns required to maintain the specific functional activity of genes across different human tissues [45].

To determine expression levels in human tissues of the genetic variants associated with SLE, we utilized the GTEx Portal database at <http://www.gtexportal.org/home/>. The database demonstrated that the *NCF2* gene was predominantly expressed in whole blood (Fig. 2). The *TYK2* gene was predominantly expressed in both cells (EBV-transformed lymphocytes) and spleen. In contrast, the expression of *DNASE1L3* was mainly found in the spleen (Figs. 3 and 4).

These *NCF2* and *TYK2* gene expression findings are consistent with the Human Protein Atlas [46,47], indicating that *NCF2* is highly expressed in neutrophils and macrophages. In contrast, the *TYK2* gene is expressed to encode a cytoplasmic and membrane-bound protein in most tissues, particularly the spleen. Slightly different from GTEx, *DNASE1L3* gene expression was highest in the liver, followed by the spleen [48]. In line with the Human Protein Atlas, Chai et al. [49] suggested that *DNASE1L3*, also known as DNAase γ , belongs to the DNase family. It is predominantly expressed in the liver and lymphoid organs, like the spleen, and is primarily expressed by dendritic cells and macrophages.

The high expression of *NCF2* in neutrophils and macrophages is consistent with previous reports of *NCF2* function. Bakutenko et al. [15] reported that *NCF2* is a part of the enzyme NADPH oxidase, which secretes superoxide in neutrophil phagosomes and other phagocytic leukocytes. Superoxide is a crucial component in

removing foreign pathogens or cell debris. TYK2 is a member of the Janus kinases, mediating cytokines' intracellular signaling via STAT activation. Although the JAK family of non-receptor kinases is small, this family is an essential intracellular signaling molecule, serving as a crucial connection in the sequence of events from cytokines to cellular responses [50]. This function may explain why TYK2 is predominantly expressed as a cytoplasmic and membrane-bound protein in numerous tissues. The DNASE1L3 protein possesses DNA hydrolysis activity that can cleave single- or double-stranded DNA, which is essential for human plasma DNA homeostasis [51]. A large population of phagocytic cells, including Kupffer cells and macrophages, is found in the liver. Macrophages are essential immune cells that actively play a role in maintaining homeostasis and integrity [52], including DNA integrity. Moreover, the function of DNASE1L3 is DNA hydrolysis and the elimination of apoptotic bodies. This provides a plausible explanation for why DNASE1L3 is highly expressed in the liver.

GWAS are commonly used to ascertain the statistical relationships between SNPs and numerous significant common diseases, providing fresh insights. Despite the characterization of only a limited number of variants, understanding the functional relationship between these variants and phenotypic traits has proven challenging [53]. This bioinformatics approach has shown that the risk genes identified are linked to SLE pathogenesis. However, given its limitations, additional preclinical (both *in vitro* and *in vivo*) and clinical studies are required to validate and integrate data. This will help to elucidate complex interactions with phenotypes and facilitate the translation of these discoveries into medical practices.

This present bioinformatic study revealed that genetic variants of the *NCF2*, *TYK2*, and *DNASE1L3* genes were associated with the risk of SLE development in individuals and populations. Both *NCF2* and *TYK2* are highly expressed in whole blood, cells (EBV-transformed lymphocytes), and the spleen. Meanwhile, *DNASE1L3* is primarily expressed in the spleen and liver. We found that two variants of the *NCF2* gene (rs17849502 and rs13306575) were strong contributors to the risk of SLE. Contrastingly, the allele C of rs34536443 in the *TYK2* gene protects against several autoimmune diseases, including SLE. As for *DNASE1L3*, the missense variant rs35677470 was identified as a risk factor for SLE. This study also revealed that *NCF2*, *TYK2*, and *DNASE1L3* could be potential targets, which may facilitate the discovery of new therapeutic approaches for SLE because several variants on those genes are strongly associated with the risk of the disease. Further investigations are necessary to identify potential biomarkers for SLE based on the pathogenic variants found in this investigation.

ORCID

Ratih Dewi Yudhani: <https://orcid.org/0000-0001-6781-8251>

Dyonisa Nasirochmi Pakha: <https://orcid.org/0000-0001-7768-6061>

Suyatmi Suyatmi: <https://orcid.org/0000-0002-8223-2862>

Lalu Muhammad Irham: <https://orcid.org/0000-0002-0091-4887>

Authors' Contribution

Conceptualization: RDY, LMI. Data curation: DNP. Formal analysis: RDY, DNP. Methodology: RDY, LMI. Writing – original draft: RDY, DNP. Writing – review & editing: LMI, SS.

18

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

References

1. Shuoshan X, Changjuan X, Honglin Z, Qinghua Z, Shaxi O, Qi W, et al. Genetic variants related to systemic lupus erythematosus revealed using bioinformatics. *Eur J Inflamm* 2022 Mar 4 [Epub]. <https://doi.org/10.1177/20587392211070407>.
2. Rees F, Doherty M, Grainge MJ, Lanyon P, Zhang W. The worldwide incidence and prevalence of systemic lupus erythematosus: a systematic review of epidemiological studies. *Rheumatology (Oxford)* 2017;56:1945-1961.
3. Ahearn JM, Liu CC, Kao AH, Manzi S. Biomarkers for systemic lupus erythematosus. *Transl Res* 2012;159:326-342.
4. Greenan-Barrett J, Doolan G, Shah D, Virdee S, Robinson GA, Choida V, et al. Biomarkers associated with organ-specific involvement in juvenile systemic lupus erythematosus. *Int J Mol Sci* 2021;22:7619.
5. Buniello A, MacArthur JA, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019;47:D1005-D1012.
6. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* 2008;32:381-385.
7. Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet* 2016;24:1202-1205.
8. Oscanoa J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annota-

- tion of human genome sequence variation (2020 update). *Nucleic Acids Res* 2020;48:W185-W192.
9. Dayem Ullah AZ, Oscanoa J, Wang J, Nagano A, Lemoine NR, Chelala C. SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic Acids Res* 2018;46:W109-W113.
 10. Dayem Ullah AZ, Lemoine NR, Chelala C. A practical guide for the functional annotation of genetic variations using SNPnexus. *Brief Bioinform* 2013;14:437-447.
 11. Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). *Nucleic Acids Res* 2012;40:W65-W70.
 12. Chelala C, Khan A, Lemoine NR. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. *Bioinformatics* 2009;25:655-661.
 13. Puspitaningrum AN, Perwitasari DA, Adikusuma W, Djalilah GN, Dania H, Maliza R, et al. Integration of genomic databases and bioinformatic approach to identify genomic variants for sjogren's syndrome on multiple continents. *Media Farm J Ilmu Farm* 2022;19:71-81.
 14. Irham LM, Adikusuma W, Lolita L, Puspitaningrum AN, Afief AR, Sarasmita MA, et al. Investigation of susceptibility genes for chickenpox disease across multiple continents. *Biochem Biophys Rep* 2023;33:101419.
 15. Bakutenko IY, Haurylchyk ID, Nikitchenko NV, Sechko EV, Kozyro IA, Tchitchko AM, et al. Neutrophil cytosolic factor 2 (NCF2) gene polymorphism is associated with juvenile-onset systemic lupus erythematosus, but probably not with other autoimmune rheumatic diseases in children. *Mol Genet Genomic Med* 2022;10:e1859.
 16. Cachat J, Deffert C, Hugues S, Krause KH. Phagocyte NADPH oxidase and specific immunity. *Clin Sci (Lond)* 2015;128:635-648.
 17. Arnold DE, Heimall JR. A review of chronic granulomatous disease. *Adv Ther* 2017;34:2543-2557.
 18. Jordan MA, Baxter AG. Genetic predisposition, humans. In: *The Autoimmune Diseases* (Rose NR, Mackay IR, eds.). 6th ed. San Diego: Academic Press, 2020. pp. 383-418.
 19. O'Neill S, Brault J, Stasia MJ, Knaus UG. Genetic disorders coupled to ROS deficiency. *Redox Biol* 2015;6:135-156.
 20. Jacob CO, Eisenstein M, Dinauer MC, Ming W, Liu Q, John S, et al. Lupus-associated causal mutation in neutrophil cytosolic factor 2 (NCF2) brings unique insights to the structure and function of NADPH oxidase. *Proc Natl Acad Sci U S A* 2012;109:E59-E67.
 21. Armstrong DL, Eisenstein M, Zidovetzki R, Jacob CO. Systemic lupus erythematosus-associated neutrophil cytosolic factor 2 mutation affects the structure of NADPH oxidase complex. *J Biol Chem* 2015;290:12595-12602.
 22. Ohl K, Tenbrock K. Oxidative stress in SLE T cells, is NRF2 really the target to treat? *Front Immunol* 2021;12:633845.
 23. Lee YH, Bae SC. Association between TYK2 polymorphisms and susceptibility to autoimmune rheumatic diseases: a meta-analysis. *Lupus* 2016;25:1307-1314.
 24. Pellenz FM, Dieter C, Lemos NE, Bauer AC, Souza BM, Crispim D. Association of TYK2 polymorphisms with autoimmune diseases: a comprehensive and updated systematic review with meta-analysis. *Genet Mol Biol* 2021;44:e20200425.
 25. Shao WH, Cohen PL. The role of tyrosine kinases in systemic lupus erythematosus and their potential as therapeutic targets. *Expert Rev Clin Immunol* 2014;10:573-582.
 26. Ronnblom L, Alm GV. Systemic lupus erythematosus and the type I interferon system. *Arthritis Res Ther* 2003;5:68-75.
 27. Faezi ST, Soltani S, Akbarian M, Aslani S, Hamzeh E, Jamshidi A, et al. Association of TYK2 rs34536443 polymorphism with susceptibility to systemic lupus erythematosus in the Iranian population. *Rheumatol Res* 2018;3:151-159.
 28. Gorman JA, Hundhausen C, Kinsman M, Arkatkar T, Allenspach EJ, Clough C, et al. The TYK2-P1104A autoimmune protective variant limits coordinate signals required to generate specialized T cell subsets. *Front Immunol* 2019;10:44.
 29. Couturier N, Bucciarelli F, Nurtudin RN, Debouverie M, Lebrun-Frenay C, Defer G, et al. Tyrosine kinase 2 variant influences T lymphocyte polarization and multiple sclerosis susceptibility. *Brain* 2011;134:693-703.
 30. Diogo D, Bastarache L, Liao KP, Graham RR, Fulton RS, Greenberg JD, et al. TYK2 protein-coding variants protect against rheumatoid arthritis and autoimmunity, with no evidence of major pleiotropic effects on non-autoimmune complex traits. *PLoS One* 2015;10:e0122271.
 31. Contreras-Cubas C, Garcia-Ortiz H, Velazquez-Cruz R, Barajas-Olmos F, Baca P, Martinez-Hernandez A, et al. Catalytically impaired TYK2 variants are protective against childhood- and adult-onset systemic lupus erythematosus in Mexicans. *Sci Rep* 2019;9:12165.
 32. Ghoreschi K, Augustin M, Baraliakos X, Kronke G, Schneider M, Schreiber S, et al. TYK2 inhibition and its potential in the treatment of chronic inflammatory immune diseases. *J Dtsch Dermatol Ges* 2021;19:1409-1420.
 33. Sisirak V, Sally B, D'Agati V, Martinez-Ortiz W, Ozcakar ZB, Da-

- vid J, et al. Digestion of chromatin in apoptotic cell mMicroparticles prevents autoimmunity. *Cell* 2016;166:88-101.
34. Shi G, Abbott KN, Wu W, Salter RD, Keyel PA. Dnase1L3 regulates inflammasome-dependent cytokine secretion. *Front Immunol* 2017;8:522.
 35. Al-Mayouf SM, Sunker A, Abdwani R, Abrawi SA, Almurshedi F, Alhashmi N, et al. Loss-of-function variant in *DNASE1L3* causes a familial form of systemic lupus erythematosus. *Nat Genet* 2011;43:1186-1188.
 36. Yu Y, Gomez-Banuelos E, Li J, Cashman KS, Paz M, Trejo-Zambrano MI, et al. Neutralizing anti-DNase1L3 antibodies derive from autoreactive VH4-34+-B cells and associate with the interferon signature in SLE. Preprint at: <https://doi.org/11.1101/2021.06.07.21258180> (2021).
 37. Zervou MI, Andreou A, Matalliotakis M, Spandidos DA, Goulielmos GN, Eliopoulos EE. Association of the *DNASE1L3* rs35677470 polymorphism with systemic lupus erythematosus, rheumatoid arthritis and systemic sclerosis: structural biological insights. *Mol Med Rep* 2020;22:4492-4498.
 38. Coke LN, Wen H, Comeau M, Ghanem MH, Shih A, Metz CN, et al. Arg206Cys substitution in *DNASE1L3* causes a defect in *DNASE1L3* protein secretion that confers risk of systemic lupus erythematosus. *Ann Rheum Dis* 2021;80:782-787.
 39. Yap DY, Chan TM. Lupus nephritis in Asia: clinical features and management. *Kidney Dis (Basel)* 2015;1:100-109.
 40. Stojan G, Petri M. Epidemiology of systemic lupus erythematosus: an update. *Curr Opin Rheumatol* 2018;30:144-150.
 41. Ueki M, Takeshita H, Fujihara J, Iida R, Yuasa I, Kato H, et al. Caucasian-specific allele in non-synonymous single nucleotide polymorphisms of the gene encoding deoxyribonuclease I-like 3, potentially relevant to autoimmunity, produces an inactive enzyme. *Clin Chim Acta* 2009;407:20-24.
 42. Li Z, Rotival M, Patin E, Michel F, Pellegrini S. Two common disease-associated TYK2 variants impact exon splicing and TYK2 dosage. *PLoS One* 2020;15:e0225289.
 43. Kim-Howard X, Sun C, Molineros JE, Maiti AK, Chandru H, Adler A, et al. Allelic heterogeneity in *NCF2* associated with systemic lupus erythematosus (SLE) susceptibility across four ethnic populations. *Hum Mol Genet* 2014;23:1656-1668.
 44. Reid S, Alexsson A, Frodlund M, Morris D, Sandling JK, Bolin K, et al. High genetic risk score is associated with early disease onset, damage accrual and decreased survival in systemic lupus erythematosus. *Ann Rheum Dis* 2020;79:363-369.
 45. Sonawane AR, Platig J, Fagny M, Chen CY, Paulson JN, Lopes-Ramos CM, et al. Understanding tissue-specific gene regulation. *Cell Rep* 2017;21:1077-1088.
 46. Human Protein Atlas. The Human Protein Atlas - *NCF2* 2023. Stockholm: Human Protein Atlas, 2023. Accessed 2023 Jan 12. Available from: <https://www.proteinatlas.org/ENSG00000116701-NCF2/tissue>.
 47. Human Protein Atlas. The Human Protein Atlas - *TYK2* 2023. Stockholm: Human Protein Atlas, 2023. Accessed 2023 Jan 12. Available from: <https://www.proteinatlas.org/ENSG00000105397-TYK2/tissue>.
 48. Human Protein Atlas. The Human Protein Atlas - *DNASE1L3* 2023. Stockholm: Human Protein Atlas, 2023. Accessed 2023 Jan 12. Available from: <https://www.proteinatlas.org/ENSG00000163687-DNASE1L3/tissue>.
 49. Chan RW, Serpas L, Ni M, Volpi S, Hiraki LT, Tam LS, et al. Plasma DNA profile associated with *DNASE1L3* gene mutations: clinical observations, relationships to nuclease substrate preference, and in vivo correction. *Am J Hum Genet* 2020;107:882-894.
 50. Hromadova D, Elewaut D, Inman RD, Strobl B, Gracey E. From science to success? Targeting tyrosine kinase 2 in spondyloarthritis and related chronic inflammatory diseases. *Front Genet* 2021;12:685280.
 51. Li B, Ge YZ, Yan WW, Gong B, Cao K, Zhao R, et al. *DNASE1L3* inhibits proliferation, invasion and metastasis of hepatocellular carcinoma by interacting with beta-catenin to promote its ubiquitin degradation pathway. *Cell Prolif* 2022;55:e13273.
 52. Binatti E, Gerussi A, Barisani D, Invernizzi P. The role of macrophages in liver fibrosis: new therapeutic opportunities. *Int J Mol Sci* 2022;23:6649.
 53. Fernald GH, Capriotti E, Daneshjou R, Karczewski KJ, Altman RB. Bioinformatics challenges for personalized medicine. *Bioinformatics* 2011;27:1741-1748.

Identifying pathogenic variants related to systemic lupus erythematosus by integrating genomic databases and a bioinformatic approach

ORIGINALITY REPORT

15%

SIMILARITY INDEX

9%

INTERNET SOURCES

12%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

- 1** www.biomedcentral.com 1%

Internet Source
- 2** eprints.uad.ac.id 1%

Internet Source
- 3** Yikai Yu, Eduardo Gomez-Banuelos, Jessica Li, Kevin S. Cashman et al. "Neutralizing anti-DNase1L3 antibodies derive from autoreactive VH4-34+ B cells and associate with the interferon signature in SLE", Cold Spring Harbor Laboratory, 2021 1%

Publication
- 4** Don L. Armstrong, Miriam Eisenstein, Raphael Zidovetzki, Chaim O. Jacob. "Systemic Lupus Erythematosus-associated Neutrophil Cytosolic Factor 2 Mutation Affects the Structure of NADPH Oxidase Complex", Journal of Biological Chemistry, 2015 1%

Publication

5

Cecilia Contreras-Cubas, Humberto García-Ortiz, Rafael Velázquez-Cruz, Francisco Barajas-Olmos et al. "Catalytically Impaired TYK2 Variants are Protective Against Childhood- and Adult-Onset Systemic Lupus Erythematosus in Mexicans", Scientific Reports, 2019

Publication

<1 %

6

ouci.dntb.gov.ua

Internet Source

<1 %

7

Submitted to Utah Education Network

Student Paper

<1 %

8

www.omicsdi.org

Internet Source

<1 %

9

Latanya N Coke, Hongxiu Wen, Mary Comeau, Mustafa H Ghanem et al. " Arg206Cys substitution in causes a defect in DNASE1L3 protein secretion that confers risk of systemic lupus erythematosus ", Annals of the Rheumatic Diseases, 2021

Publication

<1 %

10

scholarhub.ui.ac.id

Internet Source

<1 %

11

biblio.ugent.be

Internet Source

<1 %

12

www.biorxiv.org

<1 %

13

www.spandidos-publications.com

Internet Source

<1 %

14

Dominika Hromadová, Dirk Elewaut, Robert D. Inman, Birgit Strobl, Eric Gracey. "From Science to Success? Targeting Tyrosine Kinase 2 in Spondyloarthritis and Related Chronic Inflammatory Diseases", *Frontiers in Genetics*, 2021

Publication

<1 %

15

Submitted to University of the Southern Caribbean

Student Paper

<1 %

16

Jacquelyn A. Gorman, Christian Hundhausen, Mackenzie Kinsman, Tanvi Arkatkar et al. "The TYK2-P1104A Autoimmune Protective Variant Limits Coordinate Signals Required to Generate Specialized T Cell Subsets", *Frontiers in Immunology*, 2019

Publication

<1 %

17

Connie C. Qiu, Roberto Caricchio, Stefania Gallucci. "Triggers of Autoimmunity: The Role of Bacterial Infections in the Extracellular Exposure of Lupus Nuclear Autoantigens", *Frontiers in Immunology*, 2019

Publication

<1 %

18

www.e-jyms.org

Internet Source

<1 %

19

Ali Javinani, Amir Ashraf-Ganjouei, Saeed Aslani, Ahmadsreza Jamshidi, Mahdi Mahmoudi. "Exploring the etiopathogenesis of systemic lupus erythematosus: a genetic perspective", Immunogenetics, 2019

Publication

<1 %

20

link.springer.com

Internet Source

<1 %

21

Bo Li, Yu-Zhen Ge, Wei-Wei Yan, Bin Gong, Kun Cao, Rui Zhao, Chao Li, Ye-Wei Zhang, Yi-Heng Jiang, Shi Zuo. " inhibits proliferation, invasion and metastasis of hepatocellular carcinoma by interacting with β -catenin to promote its ubiquitin degradation pathway ", Cell Proliferation, 2022

Publication

<1 %

22

jmsgtr.tamhsc.edu

Internet Source

<1 %

23

Kuojun Zhang, Ke Ye, He Tang, Zhihao Qi, Tianyu Wang, Jie Mao, Xiangyu Zhang, Sheng Jiang. "Development and Therapeutic Implications of Tyrosine Kinase 2 Inhibitors", Journal of Medicinal Chemistry, 2023

Publication

<1 %

24 Wen-Hai Shao, Philip L Cohen. "The role of tyrosine kinases in systemic lupus erythematosus and their potential as therapeutic targets", Expert Review of Clinical Immunology, 2014
Publication <1 %

25 koreamed.org
Internet Source <1 %

26 scitechdaily.com
Internet Source <1 %

27 text.123doc.net
Internet Source <1 %

28 9dok.org
Internet Source <1 %

29 Helena Idborg, Arash Zandian, Ann-Sofi Sandberg, Bo Nilsson et al. "Two subgroups in systemic lupus erythematosus with features of antiphospholipid or Sjögren's syndrome differ in molecular signatures and treatment perspectives", Arthritis Research & Therapy, 2019
Publication <1 %

30 Luis Alonso González, Manuel Francisco Ugarte-Gil, Graciela S Alarcón. "Systemic lupus erythematosus: The search for the ideal biomarker", Lupus, 2020
Publication <1 %

31

journals.lww.com

Internet Source

<1 %

32

Chih-Wei Chung, Seng-Cho Chou, Tzu-Hung Hsiao, Grace Joyce Zhang, Yu-Fang Chung, Yi-Ming Chen. "Machine learning approaches to identify systemic lupus erythematosus in anti-nuclear antibody-positive patients using genomic data and electronic health records", *BioData Mining*, 2024

Publication

<1 %

33

Guilan Shi, Kennady N. Abbott, Wenbo Wu, Russell D. Salter, Peter A. Keyel. "Dnase1L3 Regulates Inflammasome-Dependent Cytokine Secretion", *Frontiers in Immunology*, 2017

Publication

<1 %

34

Yun Deng, Jian Zhao, Daisuke Sakurai, Andrea L Sestak et al. " Decreased expression associates with lupus-risk variants and elevated antinuclear antibody production ", *Annals of the Rheumatic Diseases*, 2016

Publication

<1 %

35

ebin.pub

Internet Source

<1 %

36

jefftbowles.com

Internet Source

<1 %

37

Internet Source

<1 %

38

Ham, Seokjin, and Tae-Young Roh. "A Follow-up Association Study of Genetic Variants for Bone Mineral Density in a Korean Population", *Genomics & Informatics*, 2014.

Publication

<1 %

39

Suad M.. "Chapter 1 Genetics and Epigenetic in Systemic Lupus Erythematosus", *IntechOpen*, 2012

Publication

<1 %

40

T. J. Vyse. "Association of polymorphisms across the tyrosine kinase gene, TYK2 in UK SLE families", *Rheumatology*, 03/27/2007

Publication

<1 %

41

Xian-Gang Mo, Wei Liu, Yao Yang, Saber Imani et al. "NCF2, MYO1F, S1PR4, and FCN1 as potential noninvasive diagnostic biomarkers in patients with obstructive coronary artery: A weighted gene co-expression network analysis", *Journal of Cellular Biochemistry*, 2019

Publication

<1 %

42

Zeynab Cheraghi, Seyed Ali Ziai, Zahra Fazeli, Abbas Gheisoori. "Unveiling sex-based impact of TYK2 rs2304256 polymorphism on

<1 %

interferon beta-1alpha responsiveness in COVID-19 patients", Gene Reports, 2023

Publication

-
- | | | |
|----|---|------|
| 43 | Zhi Min Song, Leïla Bouchab, Elodie Hudik, Romain Le Bars, Oliver Nüsse, Sophie Dupré-Crochet. "Phosphoinositol 3-phosphate acts as a timer for reactive oxygen species production in the phagosome", Journal of Leukocyte Biology, 2017
Publication | <1 % |
| 44 | aacijournal.biomedcentral.com
Internet Source | <1 % |
| 45 | api.repository.cam.ac.uk
Internet Source | <1 % |
| 46 | arthritis.pure.elsevier.com
Internet Source | <1 % |
| 47 | digitalcommons.wustl.edu
Internet Source | <1 % |
| 48 | dm5migu4zj3pb.cloudfront.net
Internet Source | <1 % |
| 49 | dspace.dtu.ac.in:8080
Internet Source | <1 % |
| 50 | journals.sagepub.com
Internet Source | <1 % |
| 51 | repo.lib.tokushima-u.ac.jp
Internet Source | <1 % |

52

www.austinpublishinggroup.com

Internet Source

<1 %

53

www.researchgate.net

Internet Source

<1 %

54

www.tandfonline.com

Internet Source

<1 %

55

www.wjgnet.com

Internet Source

<1 %

56

"Genetics of Rare Autoimmune Diseases",
Springer Science and Business Media LLC,
2019

Publication

<1 %

57

C. O. Jacob, M. Eisenstein, M. C. Dinauer, W.
Ming et al. "Lupus-associated causal mutation
in neutrophil cytosolic factor 2 (NCF2) brings
unique insights to the structure and function
of NADPH oxidase", Proceedings of the
National Academy of Sciences, 2011

Publication

<1 %

58

Eric Henrique Roma, Juan Pereira Macedo,
Grazielle Ribeiro Goes, Juliana Luar
Gonçalves et al. "Impact of reactive oxygen
species (ROS) on the control of parasite loads
and inflammation in Leishmania amazonensis
infection", Parasites & Vectors, 2016

Publication

<1 %

59

M Postal, NA Sinicato, KO Peliçari, R Marini, LTL Costallat, S Appenzeller. "Clinical and serological manifestations associated with interferon- α levels in childhood-onset systemic lupus erythematosus", Clinics, 2012

Publication

<1 %

60

"The Microbiome in Rheumatic Diseases and Infection", Springer Science and Business Media LLC, 2018

Publication

<1 %

61

Rebecca W.Y. Chan, Lee Serpas, Meng Ni, Stefano Volpi et al. "Plasma DNA Profile Associated with DNASE1L3 Gene Mutations: Clinical Observations, Relationships to Nuclease Substrate Preference, and InVivo Correction", The American Journal of Human Genetics, 2020

Publication

<1 %

62

doi.org
Internet Source

<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On