Medical Sains : Jurnal Ilmiah Kefarmasian
Vol. 9 No. 1, January - March 2024
https://ojs.stfmuhammadiyahcirebon.ac.id/index.php/iojs

MISSENSE MUTATIONS IN THE *IRAK1* GENE ARE ASSOCIATED WITH AN INCREASED RISK OF SYSTEMIC LUPUS ERYTHEMATOSUS

Adnan[1], Dyah Aryani Perwitasari[1]*, Rita Maliza[2], Nanik Sulistyani[1]
[1]*Faculty of Pharmacy, Ahmad Dahlan University*
[2]*Biology Department, Faculty of Mathematics and Natural Sciences, Andalas University*
Email Corresponding: dyah.perwitasari@pharm.uad.ac.id

Submitted: November 13, 2023 Revised: February 1, 2024 Accepted: February 13, 2024

ABSTRACT

Single nucleotide polymorphisms (SNPs) are the most common form of genetic variation in humans. Missense SNPs can change protein structure and function. This study aimed to determine missense SNPs of the *IRAK1* gene that can affect the amino acid sequence and lead to changes in protein structure and function, as well as their relationship as a risk factor for SLE. In this in silico method, several bioinformatics tools have been used to identify missense SNPs, including their properties and impacts, as well as their interaction networks with proteins. The tools used were PolyPhenv2, SIFT, PhD-SNP, PROVEAN, SNAP, Panthers, I-Mutant 3.0, and GeneMania. Four missense SNPs, rs11465830, rs1059702, rs1059703, and 10127175, were obtained from the NCBI SNP database. The SIFT test results showed that all the SNPs were tolerant. In the test results obtained using PolyPhen, the four SNPs were benign. The results of the probe test indicated that the four SNPs were neutral. When tested with SNAP, one SNP was neutral, and three others had an impact. In the PhD-SNP test, all SNPs were neutral. In the panther test, all SNPs were benign. The I-mutant assay showed that the four SNPs could decrease protein stability. Tests with GeneMania have reported that most interactions between genes were between *IRAK1* and *MYD88*, and physical interactions were the most dominant form of interaction. Conclusion. rs10127175, rs11465830, rs1059702, and rs1059703 are missense SNPs in IRAK1, which can disrupt protein stability and be a risk factor for SLE.

Keywords: *IRAK1*, SNP, Missense, Systemic Lupus Erythematosus

INTRODUCTION

SLE is an autoimmune disorder characterized by overproduction of antibodies (Goulielmos et al., 2018). Its pathogenesis is influenced by environmental, hormonal, and genetic factors (Tanzilia et al. 2021). Genetic factors are one of the main causes of SLE (Julià et al., 2018). SNP are genetic variations that appear most frequently in the human genome. Many quantitative studies have been conducted to determine the relationship between genetic variation and complex diseases (Lee et al., 2005).

90% of the genetic variation that occurs in the human genome comes from SNPs. The most common type of DNA sequence variation is single nucleotide polymorphisms (SNPs), which are characterized by a single base pair change in the allele. Approximately 500,000 coding regions of the human genome contain many SNPs (Collins et al., 1999). Missense SNPs, also known as nsSNPs, are essential to note because they alter amino acid residues in human proteins, creating functional diversity (Lander, 1996).

Functional variants can affect protein structure and function, resulting in changes in the sequence of amino acids that make up proteins (Capriotti & Altman, 2011). The adverse effects include changes in gene regulatory protein structure, destabilization, geometry, and hydrophobicity, which affect the protein cargo (Petukh et al., 2019), dynamics, translation,

stability, and intra-protein interactions (Petukh et al., 2019) can impact the cell's structural integrity (Thomas et al., 1999).

Identifying SNPs in a candidate gene is challenging for thousands of SNPs (Zhernakova et al. 2009). Bioinformatic approaches are essential for determining which SNPs can be analyzed based on their functional relevance to distinguish between neutral and SNPs that can affect protein stability (Ilhan & Tezel, 2013; Patnala et al., 2013). In silico methods have been used to determine how genetic sequence differences can affect protein structure and function. In Silico techniques are less costly and time-consuming than experimental tests to identify SNPs, and the results can be used as a preliminary test to analyze the impact of SNPs (Venkata Subbiah et al., 2020).

The chromosome Xq28 region is the location of a gene encoding an interleukin-1 receptor-associated kinase, reported as a genetic location associated with the occurrence of SLE (Kaufman, 2013). IL-1 and *IRAK1* control NF-KB, an important transcription factor in the development of SLE. *IRAK1* also influences the TLR signaling pathway, leading to immune activation. These mechanisms play a role in SLE development (Marshak-rothstein, 2006). SNPs in IRAK1 have been reported to confer susceptibility to SLE in Asian and European populations (Kaufman, 2013).

SNP rs1059703 SNP rs1059703 is susceptible to SLE in Hispanics (HA), Asians (AS), and European-Americans (EA) but not in people of African descent (Kaufman, 2013). This study aimed to identify missense SNPs in the *IRAK1* gene that may cause changes in the amino acid sequence that may result in changes in the structure and function of the protein with respect to its susceptibility to SLE.

METHODS

Data source

Missense SNPs were obtained from NCBI SNP database (https://www.ncbi.nlm.nih.gov/SNP/). Only missense SNPs that are thought to affect amino acid makeup and result in changes to the structure and function of the protein were retrieved. The overall course of this study is shown in **Figure 1**.

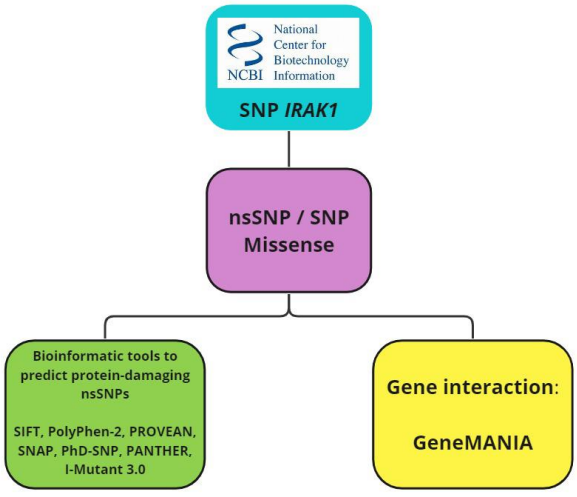

Figure 1. Research Algorithm

Bioinformatic tools to predict protein-damaging nsSNPs

Several software packages were used in this study. The first is Sorting Intolerant from Tolerant (SIFT) (https://sift.bii.a-star.edu.sg/), which determines whether a change in amino acid sequence can have a damaging effect on a protein (Vaser et al., 2016).

PolyPhen-2 (Polymorphism Phenotyping v2) (http://genetics.bwh.harvard.edu/pph2/) was used to identify the potential impact of amino

acid changes on protein stability and function considering structural evolution and to identify the potential presence of missense mutations (Adzhubei et al., 2013).

The Protein Variation Effect Analyzer (PROVEAN; http://provean.jcvi.org) is a tool used to predict whether amino acid changes affect a protein's biological function (Choi et al., 2012). SNAP (https://rostlab.org/services/snap/) distinguishes between effect variants and non-synonymous neutral/SNP variants by considering various sequences and variant features (Hecht et al., 2015).

Predictor of human Deleterious Single Nucleotide Polymorphisms (PhD-SNP; http://snps.biofold.org/phd-snp/phdsnp.html) (Capriotti et al., 2006). The online predictor utilizes a support vector machine algorithm to categorize nsSNPs as disease-related or neutral polymorphisms.

To predict whether a particular SNP has a functional impact on a protein, analyses using Panther, Protein Analysis Through Evolutionary Relationships were conducted (https://www.pantherdb.org/tools) (Thomas et al., 2006).

Protein stability change prediction

Predictions of changes in protein stability were made using the I-Mutant 3.0 tool (http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi). This tool predicts how the thermodynamic stability of a protein can be affected by Single-point variation based on the free energy difference between the variant protein and the free protein (Venkata Subbiah et al., 2022). DDG values were derived from protein sequences with the results: DDG > 0.5 kcal/mol (stable), DDG < -0.5 kcal/mol (unstable), and -0.5 ≤ 0.5 kcal/mol (neutral).

Gene interaction

Identify IRAK1 gene interaction networks with other genes using GeneMania (http://www.genemania.org). This tool determines the function and information related to the expression, location, protein domains, and shared pathways of genes (Warde-Farley et al., 2010).

RESULTS

Retrieval of nsSNPs from NCBI SNP database

IRAK1 missense SNPs were extracted from the NCBI for Biotechnology Information SNP database. A total of 3642 SNPs were identified, and after restriction with only missense and Global MAF value range 0.01-0.9, only ten missense SNPs remained. out of 10 existing snp, Six SNPs were excluded because they were identical to other SNPs, and only four missense SNPs remained, namely rs10127175, rs1059702, rs1059703, and rs11465830 (**Table I**). SNP identification using the Polyphen database showed that the four SNPs were benign. All SNP (four) were identified as neutral based on Provean and PhD-SNP databases. Four SNPs were reported as benign based on analysis using Panthers and, based on I-Mutant examination, reported that all four SNPs reduced protein stability.

Table I. Destructive SNP Missense Predictions

SNP ID	Amino acid change	SIFT score	SIFT prediction	Polyphen score	Polyphen prediction	Proven score	Proven prediction	SNAP prediction	PhD-SNP prediction	Panther
rs1059702	F196S	0.4	Tolerated	0.000	Benign	1.7	Neutral	Effect	Neutral	probably benign
rs1059703	S532L	1	Tolerated	0.000	Benign	0.6	Neutral	Effect	Neutral	probably benign
rs11465830	R194H	0.2	Tolerated	0.000	Benign	0.4	Neutral	Effect	Neutral	probably benign

rs1012 7175	C203S	0.4	Tolera ted	0.014	Benign	0.4	Neutra l	Neutral	Neutral	probabl y benign

Using bioinformatics tools in silico resulted in three missense SNPs that could impact protein structure and function changes. Using the SNAP tool, rs1059702, rs1059703, and rs11465830 were found to potentially change the structure and function of the protein. There are several differences in the respective measurement results of the in silico tool. This can happen because each tool has a different assessment method and parameters that are assessed, and thus, there is the potential for differences in measurement results.

Predicting of protein stability using I-Mutant 3.0

Four SNPs, rs1059702, rs1059703, rs10127175, and rs11465830, had DDG values of less than zero. These values indicated that SNPs are unstable and can reduce protein stability. **Table II** displays the results of this study.

Table II. Prediction of protein stabilization

SNP	Change in Amino acids	Value of DDG	Prediction of effect
rs10127175	C203S	-1.4	decrease
rs1059702	F196S	-0.5	decrease
rs1059703	S532L	-0.4	decrease
rs11465830	R194H	-0.9	decrease

Gene interactions

The GeneMania application aims to obtain information about physical interactions, genetic interactions, protein domains, and the pathways involved (Warde-Farley et al., 2010). An overview of the interactions of IRAK1 with other genes is shown in Figure 2. Figure 2 shows the interactions between IRAK1 and the 20 other genes. It is found that most interactions occur between the *IRAK1* gene and *MYD88*, and physical interactions are the most common forms of interaction.

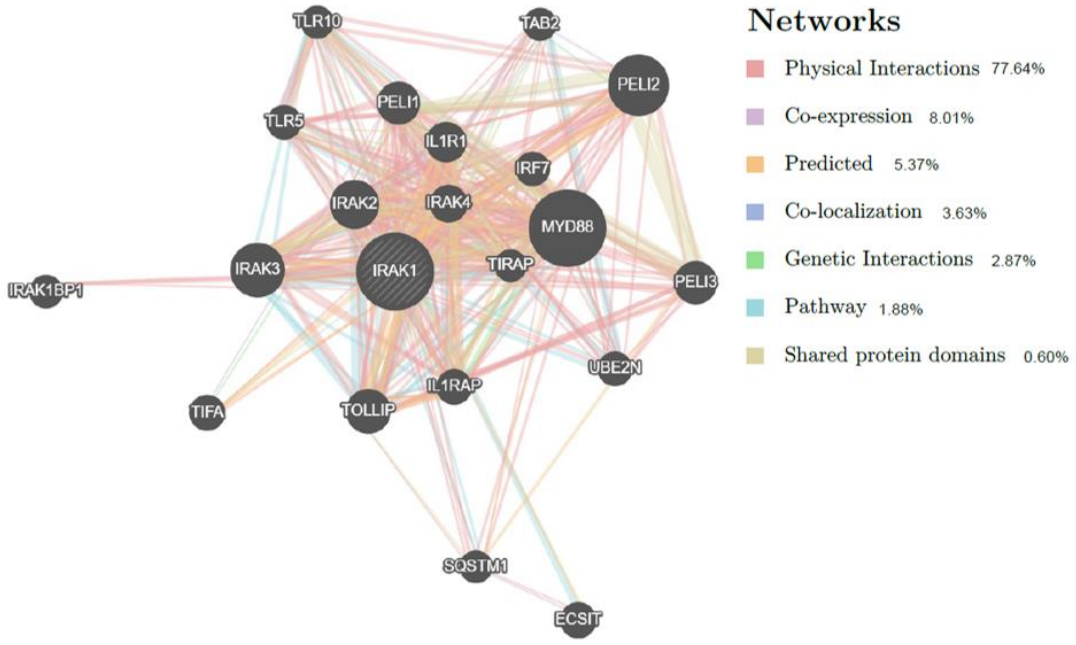

Figure 2. GenMania Was Used For Gene Interaction Analysis.

Expression quantitative trait loci (eQTL)

Examination of eQTLs to obtain information about *IRAK1* gene expression in tissues was performed using GtexPortal (https://gtexportal.org/home/). The most apparent functional impact of genetic variation was included in eQTL annotation (Emilsson, 2008). IRAK1 expression is shown in **Figure 3**.

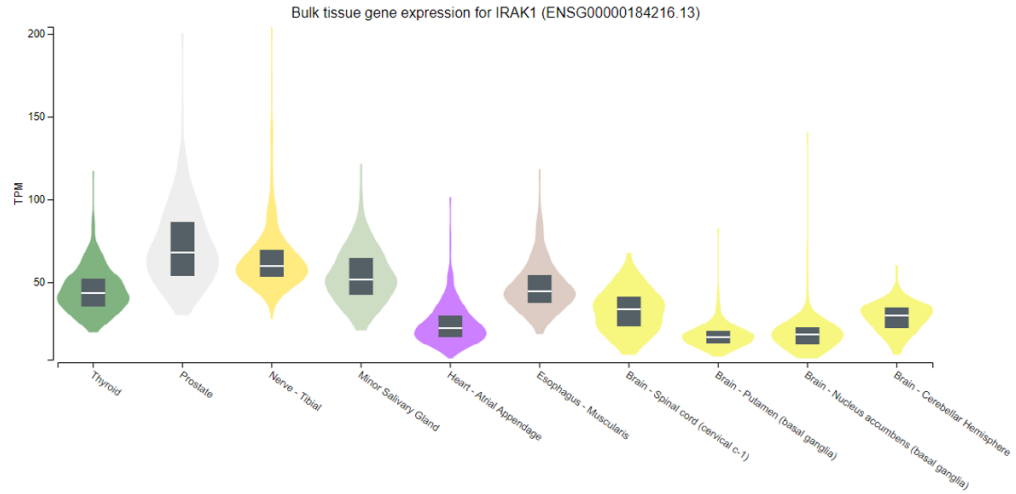

Figure 3. Expression of *IRAKI1* Gene

Expression of significant eQTLs for IRAK1.

Using GtexPortal, 755 eQTLs were identified in 17 networks. The dominant tissue was the thyroid with 194 variants (**Figure 4**). These results indicate that the thyroid is the tissue that most expresses the *IRAK1* gene and explains its relationship with the incidence of SLE.

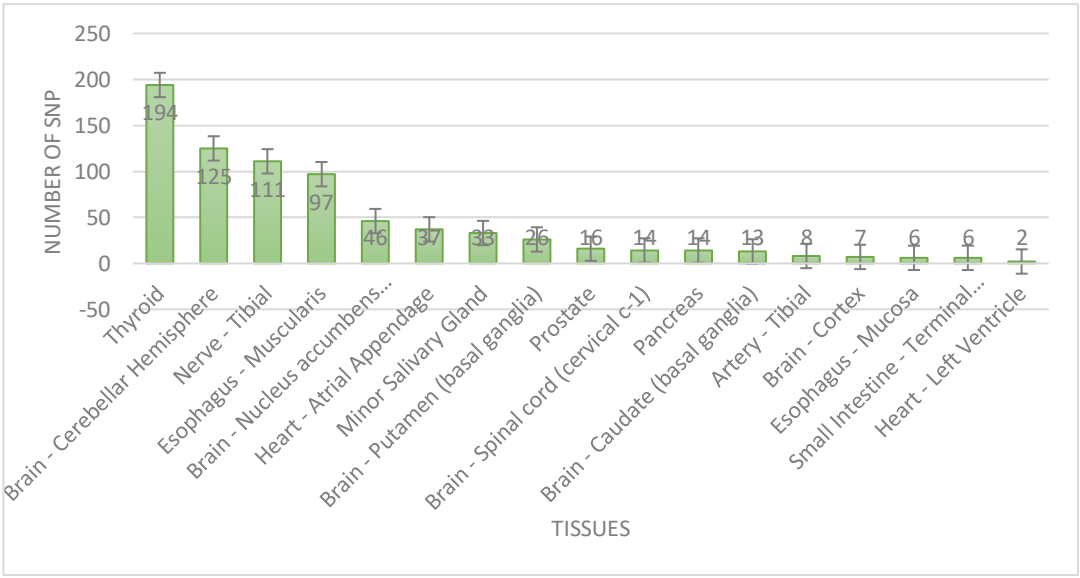

Figure 4. Gene expression in tissues

DISCUSSION

The response to treatment and disease prognosis can be predicted using SNP. By studying the impact of SNPs on proteins, a novel substance can be devised to mitigate the effects of mutations within a population. Many in silico studies have been conducted to study SNP polymorphisms, both damaging and neutral, for example, the TNF-α gene (Dabhi & Mistry, 2014).

This study used many tools to predict damaging SNP polymorphisms in the *IRAK1* gene. The first is SIFT to predict changes in amino acids that affect protein function (Ng & Henikoff, 2003), and continued analysis using the PROVEAN, SNAP, PhD-SNP, and Panther programs.

Based on the analysis of the four tools, three missense SNPs are hazardous because they can cause changes in amino acid sequences, resulting in changes in protein structure and function changes, namely rs1059702, rs1059703, and rs11465830. Other studies have also reported that the same thing where SNPs rs1059702 (Zhai, 2013) and rs1059703 (Doudar et al., 2019) are associated with disease. Studies of rs1059702 and rs1059703 have reported that both are associated with SLE (Li, 2015). rs1059702 is predicted to be the leading cause of SLE; its allele produces an S196F amino acid change in the *IRAK1* gene (Kaufman, 2013). A previous study reported that mice lacking *IRAK1* are susceptible to autoimmunity (Heiseke et al., 2015).

A study on the genotypes and alleles of SNP rs1059703 on the G allele reported a homozygous GG genotype with less frequency in patients with SLE than in healthy individuals. Meanwhile, heterozygous AG and homozygous AA genotypes are higher than those in healthy individuals (Doudar et al., 2019).

Based on the use of the I-Mutant tool, four missense SNPs (rs10127175, rs1059702, rs1059703, and rs11465830) were unstable, had reduced stability, and allowed changes in structure and function. The resulting proteins had a DDG value less than zero. This explains why the results of this study align with several previous studies that reported that missense SNPs from the *IRAK1* gene were one of the factors causing SLE incidence.

The interaction network between *IRAK1* and the other genes was identified using the GenMania program. The results showed that IRAK1 interacted significantly with *IRAK3*, *PELI2*, and *MYD88*. MYD88 belongs to the Toll-Like Receptor family, and interleukin-1 is a canonical adapter for downstream inflammatory signaling pathways (Deguine & Barton, 2014). *IRAK1* triggers innate immune responses to infections (Hatcher, 2020). Blood and immune tissues (peripheral blood, lymph nodes, thymus, and bone marrow) have the highest expression of *IRAK1* (Cao et al., 1996).

Inflammatory signaling downstream of the Toll-interleukin receptor (TIR) is influenced by *IRAK1* by regulating NF-κB regulatory factors and IFN regulatory factors (IRFs) (Rao et al., 2005). The interaction between IRAK1 and *MYD88* can mediate signaling downstream of TIRs that rapidly bind ligands to IL-1R or TLRs. *IRAK1* can bind TRAF6, releasing *IRAK1* homodimers from *MYD88* and ultimately activating downstream NF-κB (Jain et al., 2014).

LIMITATIONS OF THE STUDY

The in silico approach is used as the first step in the prediction process or as a screening test before moving on to more comprehensive methods. Because this is a prediction, it would be wise to run more tests to provide a more reliable validation of the results.

CONCLUSION

An in silico approach identified four missense SNPs (rs1059702, rs1059703, rs10127175, and rs11465830). These four SNPs have the opportunity to disrupt protein stability, which can change protein structure and function, and are susceptible to SLE. Bioinformatics analysis for predicting pathogenic properties from SNPs can save costs and time, but validation of the results in experimental tests is still needed.

ACKNOWLEDGMENT

Our appreciation goes to LPPM UAD for sponsoring this study.

REFERENCES

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics (SUPPL.76)*. doi: 10.1002/0471142905.hg0720s76

Cao, Z., Henzel, W. J., & Gao, X. (1996). IRAK: A Kinase Associated with the Interleukin-1 Receptor. *Science*, *271*(5252), 1128–1131. https://www.science.org/doi/10.1126/science.271.5252.1128

Capriotti, E., & Altman, R. B. (2011). Improving the prediction of disease-related variants using protein three-dimensional structure. *BMC Bioinformatics*, *12*(4), S3. doi: 10.1186/1471-2105-12-S4-S3.

Capriotti, E., Calabrese, R., & Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*, *22*(22), 2729–2734.

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE*, *7*(10). https://doi.org/10.1371/journal.pone.0046688

Collins, F. S., Brooks, L. D., & Chakravarti, A. (1999). Erratum: A DNA polymorphism discovery resource for research on human genetic variation (Genome Research (1998) 8 (1229-1231)). *Genome Research*, *9*(2), 210.

Dabhi, B., & Mistry, K. N. (2014). In silico analysis of single nucleotide polymorphism (SNP) in human TNF-α gene. *Meta Gene*, *2*, 586–595. https://doi.org/10.1016/j.mgene.2014.07.005

Deguine, J., & Barton, G. M. (2014). MyD88: A central player in innate immune signaling. *F1000Prime Reports*, *6*(November), 1–7. https://doi.org/10.12703/P6-97

Doudar, N. A., Abdelshafy, S. S., Rady, S. A. K., & Mokhtar, A. M. (2019). Systemic lupus erythematosus: Genetic variants in Xq28 region. *Reumatologia*, *57*(5), 264–270. https://doi.org/10.5114/reum.2019.89517

Emilsson, V. (2008). *Genetics of gene expression and its effect on disease*. doi: 10.1038/nature06758.

Goulielmos, G. N., Zervou, M. I., Vazgiourakis, V. M., Ghodke-Puranik, Y., Garyfallos, A., & Niewold, T. B. (2018). The genetics and molecular pathogenesis of systemic lupus erythematosus (SLE) in populations of different ancestry. *Gene*, *668*(March), 59–72.

Hatcher, J. M. (2020). Discovery of a Selective, Covalent IRAK1 Inhibitor with Antiproliferative Activity in MYD88 Mutated B-Cell Lymphoma. *ACS Medicinal Chemistry Letters*, *11*(11), 238–2243.

Hecht, M., Bromberg, Y., & Rost, B. (2015). Better prediction of functional effects for sequence variants From VarI-SIG 2014: Identification and annotation of genetic variants in the context of structure, function and disease. *BMC Genomics*, *16*(8), 1–12.

Heiseke, A. F., Jeuk, B. H., Markota, A., Straub, T., Lehr, H. A., Reindl, W., & Krug, A. B. (2015). IRAK1 Drives Intestinal Inflammation by Promoting the Generation of Effector Th Cells with Optimal Gut-Homing Capacity. *The Journal of Immunology*, *195*(12), 5787–5794.

Ilhan, I., & Tezel, G. (2013). *How to Select Tag SNPs in Genetic Association Studies ? The CLONTagger Method with Parameter Optimization*. *17*(7), 368–383.

Jain, A., Kaczanowska, S., & Davila, E. (2014). IL-1 receptor-associated kinase signaling and its role in inflammation, cancer progression, and therapy resistance. *Frontiers in Immunology*, *5*(NOV), 1–8. https://doi.org/10.3389/fimmu.2014.00553

Julià, A., López-Longo, F. J., Pérez Venegas, J. J., Bonàs-Guarch, S., Olivé, À., Andreu, J. L., Aguirre-Zamorano, M. Á., Vela, P., Nolla, J. M., de la Fuente, J. L. M., Zea, A., Pego-Reigosa, J. M., Freire, M., Díez, E., Rodríguez-Almaraz, E., Carreira, P., Blanco, R., Taboada, V. M., López-Lasanta, M., … Fernández-Nebro, A. (2018). Genome-wide association study meta-analysis identifies five new loci for systemic lupus erythematosus. *Arthritis Research and Therapy*, *20*(1), 1–10. https://doi.org/10.1186/s13075-018-1604-1

Kaufman, K. M. (2013). Fine mapping of Xq28: both MECP2 and IRAK1 contribute to risk for systemic lupus erythematosus in multiple ancestral groups. *Annals of the Rheumatic*

Diseases, *72*(3), 437–444.
Lander, E. S. (1996). The New Genomics: Global Views of Biology. *Science*, *274*(5287), 536–539.
Lee, J. E., Choi, J. H., Lee, J. H., & Lee, M. G. (2005). Gene SNPs and mutations in clinical genetic testing: Haplotype-based testing and analysis. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *573*(1–2), 195–204.
Li, C. (2015). Susceptibility of autoimmune diseases in three polymorphisms of infection-associated gene IRAK1. *The Journal of Infection in Developing Countries*, *9*(6), 614–623.
Marshak-rothstein, A. (2006). *Toll-like receptors in systemic autoimmune disease*. 823–835.
Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, *31*(13), 3812–3814. https://doi.org/10.1093/nar/gkg509
Patnala, R., Clements, J., & Batra, J. (2013). Candidate gene association studies: A comprehensive guide to useful in silico tools. *BMC Genetics*, *14*. https://doi.org/10.1186/1471-2156-14-39
Petukh, M., Kucukkal, T. G., & Alexov, E. (2019). On human disease-causing amino acid variants: statistical study of sequence and structural patterns. *Methods Molecular Biology*, *176*(5), 139–148.
Rao, N., Nguyen, S., Ngo, K., & Fung-Leung, W.-P. (2005). A novel splice variant of interleukin-1 receptor (IL-1R)-associated kinase 1 plays a negative regulatory role in Toll/IL-1R-induced inflammatory signaling. *Molecular and Cellular Biology*, *25*(15), 6521–6532.
Tanzilia, M. F., Tambunan, B. A., & Dewi, D. N. S. S. (2021). Tinjauan Pustaka: Patogenesis Dan Diagnosis Sistemik Lupus Eritematosus. *Syifa' MEDIKA: Jurnal Kedokteran Dan Kesehatan*, *11*(2), 139. https://doi.org/10.32502/sm.v11i2.2788
Thomas, Kejariwal, A., Guo, N., Mi, H., Campbell, M. J., Muruganujan, A., & Lazareva-Ulitsky, B. (2006). Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Research*, *34*(WEB. SERV. ISS.), 645–650. https://doi.org/10.1093/nar/gkl229
Thomas, R., McConnell, R., Whittacker, J., Kirkpatrick, P., Bradley, J., & Sandford, R. (1999). Identification of mutations in the repeated part of the autosomal dominant polycystic kidney disease type 1 gene, PKD1, by long-range PCR. *American Journal of Human Genetics*, *65*(1), 39–49. https://doi.org/10.1086/302460
Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., & Ng, P. C. (2016). SIFT missense predictions for genomes. *Nature Protocols*, *11*(1), 1–9. https://doi.org/10.1038/nprot.2015.123
Venkata Subbiah, H., Ramesh Babu, P., & Subbiah, U. (2020). In silico analysis of non-synonymous single nucleotide polymorphisms of human DEFB1 gene. *Egyptian Journal of Medical Human Genetics*, *21*(1). https://doi.org/10.1186/s43042-020-00110-3
Venkata Subbiah, H., Ramesh Babu, P., & Subbiah, U. (2022). Determination of deleterious single-nucleotide polymorphisms of human LYZ C gene: an in silico study. *Journal of Genetic Engineering and Biotechnology*, *20*(1). https://doi.org/10.1186/s43141-022-00383-8
Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C. T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G. D., & Morris, Q. (2010). The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research*, *38*(SUPPL. 2), 214–220. https://doi.org/10.1093/nar/gkq537
Zhai, Y. (2013). Association of interleukin-1 receptor-associated kinase (IRAK1) gene polymorphisms (rs3027898, rs1059702) with systemic lupus erythematosus in a Chinese Han population. *Inflammation Research : Official Journal of the European Histamine Research Society*, *62*(6), 555–560.
Zhernakova, A., Diemen, C. C., Van, & Wijmenga, C. (2009). *Detecting shared pathogenesis from the shared genetics of immune-related diseases*. *10*.