ISSN 2816-8089

RESEARCH ARTICLE

A Scalable Algorithm for Interpreting DNA Sequence and Predicting the Response of Killer T-Cells in Systemic Lupus Erythematosus Patients

Nwoye O Ephraim[1], Fidelis P Obinna[2], Nwosu O I[1], Balogun O Jessy[1], Raid Rafi Al-Nima[3], Wai Lok Woo[4*]

[1]Department of Biomedical Engineering, University of Lagos, Lagos, Nigeria
[2]Department of Biomedical Technology, Federal University of Technology, Akure, Nigeria
[3]Technical Engineering College, Northern Technical University, Iraq
[4]Department of Computer and Information Sciences, Northumbria University, UK

Abstract

The incidence and prevalence of Systemic Lupus Erythematosus (SLE) in North America are 23.2 and 241 per 100,000 people per year respectively, while the incidence in Africa is 0.3 per 100,000 people per year. This study aims to predict the autoimmune response of killer T-cells in patients suffering from SLE by searching for variations in genes regulating the activities of killer T-cells. An approximate matching algorithm applying the Boyer-Moore algorithm was used for the matching. The threshold on all single nucleotide polymorphisms (SNPs) was set to 10% of the nucleotide sequence length of the gene. For 50% of susceptibility genes with no match, the patient is susceptible. Sixteen (16) patients showed that they are all guaranteed to manifest autoimmune killer T-cells according to the study. The algorithm can predict the response of killer T-cells and improve the early detection and treatment of SLE patients.

Key Words: *Systemic lupus erythematosus; Boyer moore; Approximate matching; SNP; DNA; Killer T-cells*

**Corresponding Author*: *Wai Lok Woo, Department of Computer and Information Sciences, Northumbria University, UK; E-mail: wailok.woo@northumbria.ac.uk*

***Received Date**: January 09, 2022, **Accepted Date**: February 15, 2022, **Published Date**: February 28, 2022*

***Citation**: Nwoye O Ephraim, Fidelis P Obinna, Nwosu O I, Balogun O Jessy, Raid Rafi Al-Nima, Wai Lok Woo. A Scalable Algorithm for Interpreting DNA Sequence and Predicting the Response of Killer T-Cells in Systemic Lupus Erythematosus Patients. Int J Bioinfor Intell Comput. 2022;1(1):48-71.*

1. Introduction

Systemic Lupus Erythematosus (SLE) is a chronic inflammatory rheumatic disease characterized by antibody production and tissue destruction. According to the American College of Rheumatology revised criteria (1997) for the classification of SLE, a combination of the following clinical manifestations is observed in an SLE patient. These clinical manifestations include malar rash, erythematous raised patches with adherent keratotic scaling and follicular plugging, a skin rash from reaction to sunlight or drugs, Oral Ulcers, Nonerosive Arthritis, Pleuritis or Pericarditis, Renal Disorder (Persistent proteinuria, Cellular casts), Immunologic disorder (Anti-DNA or Anti-Sm), seizures and psychosis. At least 4 of these symptoms must manifest before SLE can be suspected. In an SLE patient the immune tolerance or self-recognition that should prevent the system from attacking itself fails.

Over the past 40 years, more than 100 genetic risk factors have been defined in systemic lupus erythematosus through a combination of case studies, linkage analyses of multiplex families and case-control analyses of a single gene [1]. The Human Leukocyte Antigen *(HLA-DR)* genes, which encodes the major histocompatibility complex (MHC) proteins, was the first described genetic link to SLE [2]. The *HLA* complex helps the immune system distinguish the body's proteins from proteins made by foreign invaders such as viruses and bacteria. Aside the *HLA* gene, other genes that have been identified to signal or activate T-cells in SLE patients include; interferon induced with helicase C domain 1(IFIH1), signal transducer and activator of transcription 4(STAT4), solute carrier family 15 member 4(SLC15A4), Fyn binding protein *(FYB)*, in the type I interferon(IFN) pathway and protein tyrosine phosphatase nonreceptor 22 *(PTPN22)*, cytotoxic T lymphocyte-associated antigen 4*(CTLA4)*, *(CD8)*, PR/SET domain 1*(PRDM1)*, interleukin 21*(IL21)* in T-cell signaling and FC fragment of IgGreceptor IIa/IIIa *(FCGR2A/FCGR3A)* in the immune complex clearance pathway [3].

2. SLE, the Disease, Classification, Manifestation, and Diagnosis

SLE, a complex autoimmune disease, causes defects of multiple immunologic components of both the innate immune system and the adaptive immune system including altered immune tolerance mechanism, hyperactivation of T and B cells, decreased ability to clear immune complexes and apoptotic cells, and failure of multiple regulatory networks [4]. SLE is known to depend on the interaction of genetic, environmental and hormonal factors, although its pathogenesis may not be fully understood. Usually, SLE is triggered by environmental factors like sunlight, drugs, and stress, and is more common amongst women than men in the ratio 15: 1 [5]. As a result of the varying clinical manifestations in SLE patients, 11 criteria of diagnosis of SLE was established by the American College of Rheumatology (ACR), in which the simultaneous presence of at least four criteria can confirm the SLE diagnosis [6].

2.1. Immune response in an SLE patient

For a patient having susceptibility genes, exposure to UV radiation from sunlight is one of the environmental factors linked to the onset of SLE. In the case of sunburn, the cell deoxyribonucleic acid (DNA) on the skin may be severely damaged causing programmed cell death (Apoptosis). Resultantly, the inside of the cell is exposed including part of the nucleus DNA, histones and other proteins. Due to susceptibility genes present in the individual, the immune system misrecognizes the exposed parts of the nucleus as nuclear antigens. Also, less effective clearance of apoptotic bodies after apoptosis is promoted by susceptibility genes.

B-cells bind with the nuclear antigens and start producing antibodies called Anti –Nuclear Antibodies. These antibodies bind with antigens to produce the antigen-antibody complexes which enter the blood and drift away to the vessel walls and different organs and tissues. At the organ or tissue, a local inflammatory reaction takes place which causes damage to activation of the complement system, and after a large enzyme cascade cause tissue cells to die. Other triggers of SLE are bacteria, drugs, and sex hormones. This is known as type III hypersensitivity.

For type II sensitivity, the Anti-Nuclear antibodies are produced to target other cells like red and white blood cells and molecules like phospholipids which marks them for phagocytosis and destruction leading to symptoms.

Alternatively, antigen-antibody complexes are internalized by plasma dendritic cells. The antigens are then delivered to Toll-like receptors to stimulate the production of Type I interferons (alpha, beta, omega) which when released binds to B cells causing the production of more antibodies to bind the nuclear antigens. Type I interferon can promote maturation of dendritic cells which can ingest additional nuclear antigens which can then be presented to CD4+ helper T-cells and CD8+ killer and transforming them to auto-reactive cells.

2.2. Pathogenesis of SLE

The pathogenesis of SLE is characterized by a complex interplay of genetic predisposition and environmental exposures and loss of tolerance and immune activation [7].

In general, there are three stages in which the pathogenesis is subdivided, and they include Genetic predisposition, environmental exposures and Loss of tolerance Immune activation.

In stage one, there is a long period of predisposition to autoimmunity with genetic susceptibility and environmental exposures each contributing to disease development. Environmental factors, such as ultraviolet light exposure, and exposure to infection, such as that of Epstein–Barr virus, have been suspected as inducers or enhancers of SLE [8].

Stage two develops when there is a loss of tolerance to self-antigens, and autoantibodies (autoAb) are generated. Abnormal B- and T-cell functions perpetuate autoantibody production by B cells and creates autoreactive T-cells [9].

The third and final stage is inadequate regulation of autoAb production with T- and B- cells hyperactivity. Abnormal clearance of immune complexes(autoAb) results in their deposition in tissues, activation of complement and defective cellular apoptosis that generates a pool of potential autoantigen [9], eventually leading to tissue damage and clinical manifestations of the disease [8].

2.3. Genetics of SLE

Genome-wide association studies (GWAS) have revealed many implicated loci, most of them shared with other autoimmune diseases. The majority of genetic variations, including Single Nucleotide Polymorphisms (SNPs), associated with SLE are within immune response–related genes and HLA (Human leukocyte antigen) gene variants [10]. Each Single nucleotide polymorphism (SNP) contributes a relatively small risk.

In the pathophysiology of SLE, the following is observed:

After apoptosis, apoptotic debris binds with autoantibodies and activates plasmacytoid dendritic cells. Dendritic cells activate autoreactive T- and B- cells and propagate inflammation. Tissue injury by cytotoxic T-cells and autoantibodies. The human leukocyte antigen is a gene complex encoding the major histocompatibility complex (MHC) proteins. Genes linked with susceptibility to SLE can be classified to HLA and Non-HLA genes.

2.4. Gene associations

Most disease-associated variants lie in the noncoding regions regulating gene expression through transcriptional/ posttranscriptional mechanisms or epigenetic modifications including SLE [11]. The majority of established SLE susceptibility genes encode products involved in innate and adaptive immune responses, particularly the three key immunological pathways relevant to the pathogenesis of SLE: (1) activation of toll-like receptor (TLR), type I interferon (IFN), and nuclear factor κB (NF-κB) signaling; (2) clearance of apoptotic cells and immune complexes (ICs); and (3) dysfunctions in lymphocyte signaling [11].

Based on region, the genes implicated for susceptibility to SLE can be grouped in human leukocyte antigen (HLA) genes, which lie within the major histocompatibility complex (MHC) and non-HLA genes which lie outside it.

2.5. HLA genes – MHC region

The major histocompatibility complex (MHC), located on the short arm of chromosome 6, is one of the most extensively studied regions in the human genome because variants at this locus have long been associated with most autoimmune diseases, many infectious and inflammatory diseases as well as transplant compatibility. In particular, human leukocyte antigen (HLA) class I and class II molecules are critical in mediating host defense responses through antigen presentation, and immune tolerance through self/non-self-recognition [12]. The classical MHC encompasses approximately 3.6 mega base-pairs (Mb) on 6p21.3 and is divided into three subregions: the telomeric class I, class III and the centromeric class II regions (see Figure 1). Of the 224 genes within this region, 57% (128 genes) are expressed, and 40% of these have a putative immunoregulatory function [12].

In the meta-analysis of genome-wide linkage studies of SLE, Fora Bosco et al. established a relationship between the MHC region and SLE in chromosome 6p21 [13]. Essentially, hereditary and acquired deficiencies of classical HLA class I and class II genes (for their role in antigen presentation to T cells), as well as complement C4 alleles, leads to a lupus-like syndrome [12].

Predominantly, European-derived populations have shown the most consistent HLA associations with SLE in the class II alleles, HLA-DR3 and HLA-DR2 and their respective haplotypes [14]. HLA class II molecules mediate host defense responses through antigen presentation and immune tolerance by self/non-self-recognition. Given their roles in T-cell dependent antibody responses, cumulative studies have reported the association of class II alleles with autoantibody production, especially HLA-DR3 with anti-Ro/La antibodies [15]. A haplotype is a combination of alleles that lie close together in the genome and are inherited together as a consequence of minimal genetic recombination.

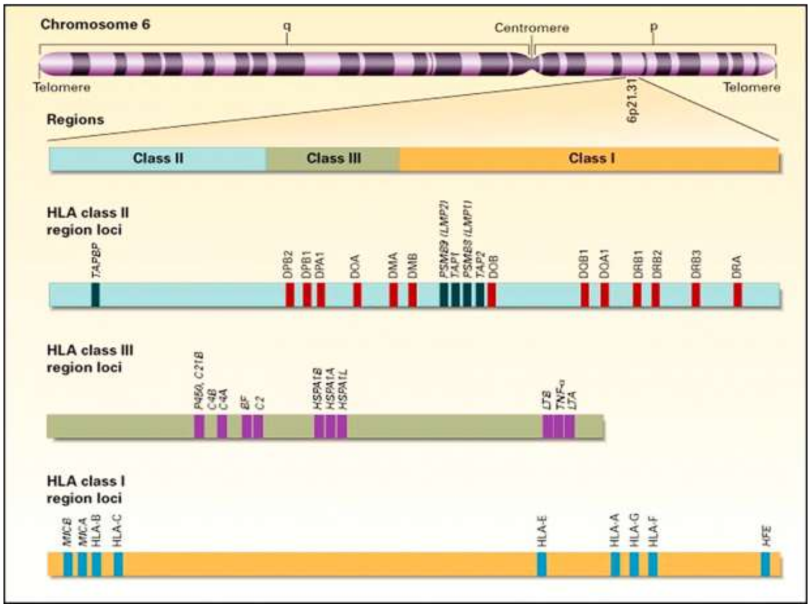

Figure 1: *Location and organization of the MHC on chromosome 6. The MHC is classically divided into three regions: class I, class II and class III. Each region contains several genes, but not all are shown. The classical class I and class II HLA genes encoding antigen presenting molecules are illustrated, as are three gene clusters in class III, from left to right: the complement cluster C4A, C4B, CFB, C2, the heat shock protein cluster, and the TNF cluster, LTB, TNF and LTA [16]. Copyright 2000 Massachusetts Medical Society. All rights reserved.*

The MHC class III region contains genes that encode complement component C4 (C4A and C4B), C2, and factor B *(Bf gene)*; some that encode cytokines such as tumor necrosis factor-alpha *(TNF-alpha, TNF gene)*; and some that encode lymphotoxins (LTA and LTB) (Namjou, Kelly, & Harley, 2007).

2.6. Non-HLA genes

These genes reside outside the major histocompatibility complex region. These genes can further be grouped based on the pathways linked with the pathogenesis of SLE [Table 1].

Table 1: *Pathways and description of non- HLA Genes.*

SNo	Pathways	Genes	Description
1	Type I IFN Pathway	i. IF1H1 ii. IRF- 5, IRF – 6, IRF- 7, TLR 7, 8 and 9 iii. STAT4 iv. TYK2	The Interferon-induced Helicase C domain 1 (IFIH1) gene is located at chromosome 2 (2q24 [17, 18]. Are transcription factors that regulate the expression of a wide range of genes [19, 20]. Expressed in T and B lymphocytes, monocytes, macrophages, natural killer cells, and dendritic cells. [21, 22]. Tyrosine Kinase 2 (TYK2) is located at chromosome 19 (19p13.2), a linkage locus for SLE. [23]

2	NFkB Pathway SLE risk genes in this pathway include; TNFAIP3, UBE2L3, TNIP1, IRAK1 [24] [25]	i.	TNFAIP3 (6q23) and TNIP1 (5q33)	critical regulators of NF-kB signaling pathway, modulating cell activation, cytokine signaling and apoptosis [26]. The role of TNFAIP3 in SLE of several SNPs has been identified [27, 28, 29, 30].
		ii.	IRAK1 - MECP2 (Xq28)	located on chromosome Xq28 [31]. There is evidence for altered methylation in SLE [32], as well as differential expression of potentially methylated genes [33].
3	B-cell and T-cell Signaling Pathway	i.	PTPN22	Research shows that the PTPN22 gene can limit T-cell receptor (TCR) signalling [11].
		ii.	CTLA-4	signalling regulate and maintain self-tolerance [34].
		iii.	PDCD1	Reported as associated to SLE from Taiwan and Poland populations [35][36].
		iv.	FYB	encodes an adaptor protein implied in positive regulation of T cells [37][38, 39].
		v.	TNFSF4	Found to be involved in proliferation and differentiation of T and B lymphocytes [40, 41].
		vi.	IL10	Possesses both immunosuppressive and immunostimulatory properties [42][43][44].
		vii.	RASGRP3	Responsible for the regulation a small membrane-bound GTPase implicated in immune B cell receptor signaling [45, 46].
		viii.	BANK1	Causes actions which results in calcium ion release from the stores of the endoplasmic reticulum [47, 48].
		ix.	CSK	The SLE-risk CSK variant is associated with high CSK expression, increased Lyn phosphorylation [49].
		x.	ETS1	V-etserythroblastosis virus E26 oncogene homolog 1 (avian) (ETS-1) is a member of the ETS family transcriptional factors [50]
		xi.	IKZF1	IKZF1 is located on chromosome 7 (7p12.2) [51].
4	Yamaguchi sarcoma viral (v-yes-1) related oncogene homolog (LYN)	i.	LYN	LYN is found on chromosome 8 (8q12.1) [47, 52]
		ii.	PRDM1 – ATG5	The protein binds specifically to the PRDI (positive regulatory domain I element) of the beta-IFN gene promoter [24, 53].
		iii.	BLK	It is located on chromosome 8 (8p23.1). [54].
		iv.	TREX1	TREX1 (3P21), DNA repair exonuclease 1 (TREX1, located on 3p21.3- p21.2) [24, 55].
5	Clearance and of Apoptotic Cells and Immune Complexes (ICs)	i.	FCGR2A/B, FCGR3A/B.	Different functional variants of FCGR2A, FCGR2B, and FCGR3A as risk factors for SLE [56].
		ii.	ITGAM	Alters immune complex clearance and deposition resulting in tissue damage [57].

	iii.	C1q	Deficiency has been described to be the most potent risk factor for developing SLE [58, 59].
	iv.	C4A	complex complement of C4 gene shows a strong disease risk in SLE [60, 61].

Early detection and diagnosis of SLE have been difficult due to its multiple symptoms which are also shared by other diseases. This research therefore aims to develop and evaluate a novel, faster and computer aided means of detecting the manifestation of SLE based on the DNA analysis for faster treatment response in patients.

3. Materials and Methods

According to Bengalio et al. (2010), the threshold for single nucleotide polymorphism (SNP) allowed for identifying significant variations or mutations in a gene is 10% of the sequence [62]. The autoimmune response of killer T cells is hence predicted to occur when more than 50% of genes acting in the adaptive immune pathway of SLE pathogenesis show more than 10% SNP variations.

3.1. The implementation of this algorithm requires the following research materials:

Laptop or computer with intel Pentium core dual; Sequence Read Archive Toolkit; Printers; Python Development Environment; Ubuntu/Linux Operating System; Human Reference DNA sequence; The DNA sequence from 16 Individuals; Google Cloud Platform and Statistical Package for The Social Sciences Software (SPSS).

3.2. Modified Boyer Moore algorithm for approximate pattern matching

3.2.1. Statement

In searching for all occurrences of a pattern string $p=p_1\ldots p_m$ in a text string $t=t_1\ldots t_n$ with at most k mismatches, $1 <k<m$.

For some position j, the state S_j is a bit number, each in $S_j[i]$, *for* $1 < i <m$, and contains the number of mismatches between $p_1\ldots p_i$ *and* $t_{j-m+1}\ldots t_j$.

Proposition: For some position j in t, let l be the largest index i, $1 <i<m-1$, such that $S_j[i] < k$, if such an index exists and 0 if otherwise. Let $d = m - l$. Then, the next position after j where P is likely to occur is $j_{next} = j+d$ *i.e.*, $Sj'[m] > k$, for every j' *such* $j < j' <j_{next}$, d is the next shift and 1<d<m-k

Then for each character a in ∑, a vector T_a of size m is constructed such that: for i, $1 < i < m$,

$$T_a\left[i\right] = \begin{cases} 0, if\ a = p_i \\ 1, otherwise \end{cases} \qquad [1]$$

The T array defined above contains information about the occurrence of a given character a, such that at a given position in the pattern p we have

$$S_{j+d}[i]=\left\{\begin{matrix} S_j[i-d]+\sum_{r=0}^{d-1} T_{tj+d-r}[i-r] if\ d<i\leq m \\ \sum_{r=0}^{i-1} T_{(tj+d-r)}[i-r] otherwise \end{matrix}\right\} \quad [2]$$

In order to obtain S_{j+d} as a sum of numbers in base 2^b let D denote the $|\sum|$ x m matrix such that, element D[a][m-r] for each a contains $\sum$ and 0 < r < m -1, is denoted by $D_{a,m-r}$ is defined as

$$D_{(a,m-r)}=\sum_{i=r+1}^{m} T_a[i-r]2^{(i-r)b} \quad [3]$$

$D_{a,m-r}$ denotes positions in $p_1 \ldots p_{m-r}$ containing character a. For a fixed r, $D_{a,m-r}$ is obtained by a left shift T_a. Therefore,

$$S_{(j+d)}=(S_j << bd)+\sum_{r=0}^{d-1} D_{t_{j+d-r,m-r}} \quad [4]$$

with initial values S_0 =0 and d = m.

3.3. Character skip

In a character skip, assuming j is the last position scanned in text t and d is the next shift. The substring is of text still to be scanned at the step of the search is then $t_{j+1} \ldots t_{j+d.}$

Let $S_{j+d,0}$= S_j<<bd be the obtained number, then, each of the d characters $t_{j+d.r+1}$, with $1 < r <d$, is processed.

Let $S_{j+d,r}$ be the partial state obtained after processing characters $t_{j+d} \ldots t_{j+d-r+1}$ of t. Then $S_{j+d,r}$ = $S_{j+d,r-1}$ + $Dt_{j+d-r+1,\ m-r+1}$ *and* S_{j+d}= $S_{j+d,d}$.

Hence for given indexes r, $1< r <d$, and i, $1 <i<m$, we have the following conditions:

a. If $S_{j+d,r}[i] > k$ the prefix of length I of P does not occur in $j + d$.

b. If $S_{j+d,r}[i] < k$, and no more comparisons have to be performed for the prefix of length I of P, then the prefix matches at position $j + d$.

The computation stops at the first partial state giving enough information for further processing. Let S_j' be this partial state. Differences between S_j and $S_j'$are located only in individual states exceeding $k+1$ [63].

3.4. Sample size determination

The formula for sample size is given by:

$$N=\frac{Z^2 \times SD[1-SD]}{ME^2} \quad [5]$$

where
Z is the Z-value or Z-score corresponds with the chosen confidence level.
N is the number of samples
SD is the standard deviation
ME is the margin of error, also known as the confidence interval (CI)

For this research, the following parameters are used:

Margin of Error or Confidence Interval =+/-5%; Confidence interval (CI)=90%; Corresponding Z value = 1.645

Hence the Sample Size determine from is approximately = 271; *i.e.,* 271 SLE data.

3.5. Statistical analysis

Statistical Analysis aims to determine the accuracy and predictive power of the hypothesis chosen to predict the response of killer T cells in SLE patients', *i.e.,* for more than 50% of susceptibility genes in the adaptive immunity pathway showing significant variations or polymorphisms. Since the proposed algorithm returns binary categorical variables FOUND (1) or NOT FOUND (0) when matches exist or no matches due to significant SNP variations, respectively, a multivariate binary logistic regression model is appropriate for testing the significance, accuracy and predictive power of the hypothesis.

A multivariate binary logistic regression model is one that has several independent variables and one categorically dependent variable. In this research, the independent variables are binary values found (1) and not found (0) returned for each of the susceptibility genes in the adaptive immunity pathway by the matching algorithm, while the dependent variable is whether or not the patient would experience the autoimmune response of killer T cells. The regression model is defined by the formula:

$$P(Y)=\frac{e^{b_0+b_1x_1+\ldots+b_nx_n}}{1+e^{b_{0+}b_1x_1+\ldots b_nx_n}} \quad [6]$$

The data collected after running the matching algorithm for each of the samples are analyzed using the Statistical Package for the Social Sciences (SPSS) software. The statistical importance, predictive power and accuracy are finally reported.

3.6. Proposed algorithm implementation

Proposed algorithm for predicting the response of Killer T-cells in the DNA sequence of an SLE patient is summarized in the following steps:
For each of the genes involved in signaling or activating killer T-cells, get the Nucleotide sequence for that gene from the Human Reference DNA sequence.
Run an approximate sequence matching algorithm on each of the DNA sequences from the SLE patients.

Set a threshold on all single nucleotide polymorphisms (SNPs) to 10% of the nucleotide sequence length of the gene. 10% SNP variation in the nucleotide sequence of genes is sufficient to induce harmful variation in the performance of the gene as identified by [62].

Find the percentage of the number of all genes with variations more significant than 10%.

The flow charts in Figure 2 and Figure 3 below show the algorithm used by the software and the overview of the software in operation, respectively.

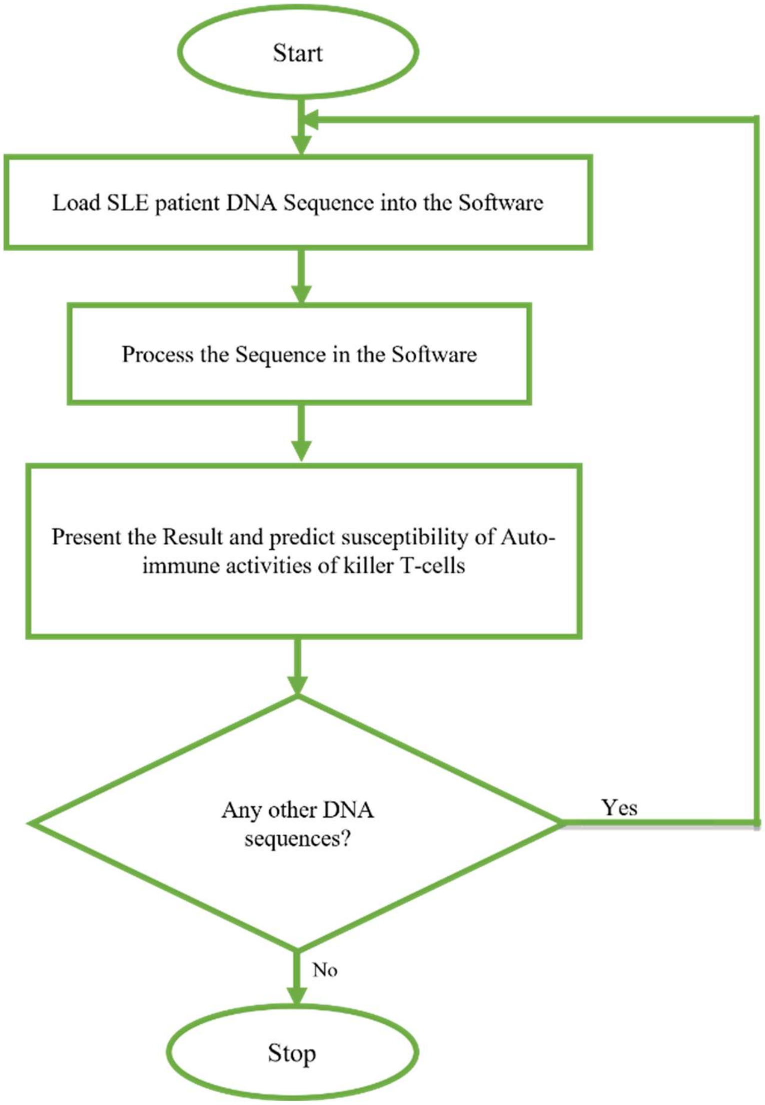
Start
Load SLE patient DNA Sequence into the Software
Process the Sequence in the Software
Present the Result and predict susceptibility of Auto-immune activities of killer T-cells
Any other DNA sequences?
Yes
No
Stop

Figure 2: *A Flow chart for the Overview of Software in Operation.*

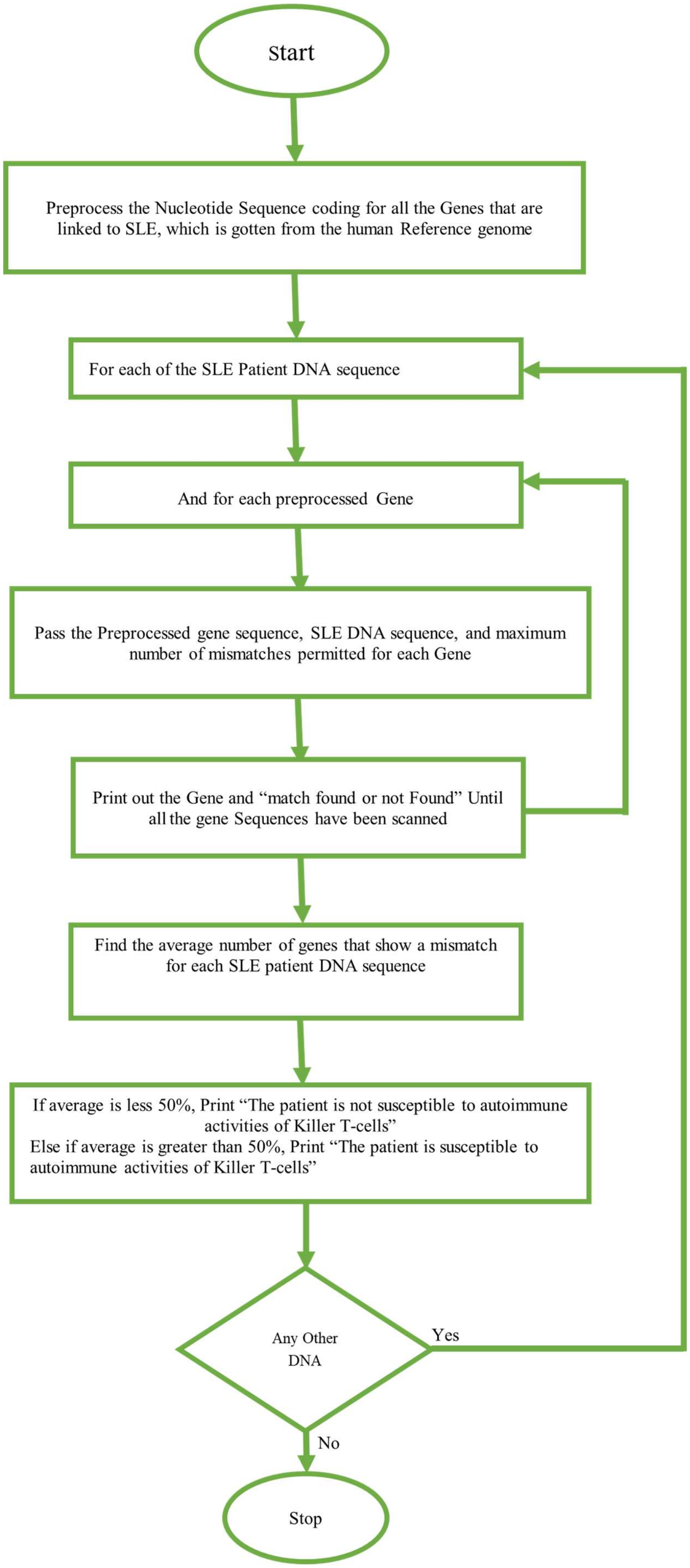
Start
Preprocess the Nucleotide Sequence coding for all the Genes that are linked to SLE, which is gotten from the human Reference genome
For each of the SLE Patient DNA sequence
And for each preprocessed Gene
Pass the Preprocessed gene sequence, SLE DNA sequence, and maximum number of mismatches permitted for each Gene
Print out the Gene and "match found or not Found" Until all the gene Sequences have been scanned
Find the average number of genes that show a mismatch for each SLE patient DNA sequence
If average is less 50%, Print "The patient is not susceptible to autoimmune activities of Killer T-cells"
Else if average is greater than 50%, Print "The patient is susceptible to autoimmune activities of Killer T-cells"
Any Other DNA
Yes
No
Stop

Figure 3: *Flow Chart showing the Algorithm Implementation.*

4. Results & Discussion

4.1. Analysis of genes in NFkB pathway

From the Literature Review, genes acting in the NFkB pathway include *tnfaip3*, *tnip1*, *irak1*, and *mecp2* genes. Binary Logistic Regression for the covariance of these genes shows statistical significance as shown in Table 2. Nagelkerke R Square test show that between 46.3% and 62.7% of the variability in the independent variable (the autoimmune response of killer T cells), is accounted for by dependent variables (the genes acting in the NFkB pathway). Furthermore, the closeness of the observed and expected in the contingency table for Hosmer and Lemeshow Test gives credence to the predictive power of the hypothesis with insignificant values of p>0.0. The chi-square test value is χ^2 (degree of freedom =4, number of samples =270) =167.766, p<0.001

Table 2a: *Model Summary for genes in NFkB Pathway.*

Step	Result = 0.00		Result = 1.00		Total
	Observed	Expected	Observed	Expected	
1	23	20.814	0	2.186	23
2	11	16.185	7	1.815	18
3	31	27.794	0	3.206	31
4	20	17.107	6	8.893	26
5	9	9.806	14	13.194	23
6	0	4.36	20	15.64	20
7	6	4.142	14	15.858	20
8	0	1.156	20	18.844	20
9	0	1.79	33	31.21	33
10	6	2.847	50	53.153	56

Table 2b: *Contingency table for genes in NFkB Pathway.*

Step	-2 log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	193.976	0.463	0.627

Table 2c: *Classification Table when independent variables are considered in NFkB Pathway.*

Observed			Predicted		
			Result		Percentage
			0.00	1.00	Correct
Step 1	Result	0.00	89	17	84.0
		1.00	20	144	87.8
	Overall Percentage				86.3

4.2. Analysis of genes in the apoptosis and disposal of cellular debris pathway

Genes acting in this pathway include *atg5*, *fcgr2b*, *irf5*, *trex1* genes. Binary Logistic Regression for the covariance of these genes shows statistical significance as shown in Table 3. Nagelkerke R Square test shows that between 46.2% and 62.6% of the variability in the independent variable (the autoimmune response of killer T cells is), accounted for or explained by dependent variables (the genes acting in the apoptotic pathway). Furthermore, the closeness of the observed and expected in the contingency table for Hosmer and Lemeshow Test gives credence to the predictive power of the hypothesis with insignificant values of p>0.001. The chi-square test value is χ^2 (degree of freedom =4, number of samples =270) =167.476, p<0.001.

Table 3a: *Model Summary for genes in the apoptotic pathway*

Step	-2 log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	194.266^a	0.462	0.626

Table 3b: *Classification table for genes in the apoptotic pathway*

Observed			Predicted		
			Result		Percentage
			0.00	1.00	Correct
Step 1	Result	0.00	82	24	77.4
		1.00	21	143	87.2
	Overall Percentage				83.3

Table 3c: *Contingency table for genes in the apoptotic pathway*

Step	Result = 0.00		Result = 1.00		Total
	Observed	Expected	Observed	Expected	
1	28	26.254	0	1.746	28
2	20	24.547	7	2.453	27
3	22	17.175	0	4.825	22
4	12	14.959	14	11.041	26
5	13	10.559	13	15.441	26
6	6	7.591	27	25.409	33
7	5	2.386	14	16.614	19
8	0	1.564	33	31.436	33
9	0	0.964	56	55.036	56

4.3. Analysis of genes in the phagocyte function and antigen presentation pathway

Genes acting in this pathway include *fcgr2b*, *fcgr3a/b*, *il10*, *itgam* genes. Binary Logistic Regression for the covariance of these genes shows statistical significance as shown in Table

4. Nagelkerke R Square test shows that between 57.6% and 78.0% of the variability in the independent variable (the autoimmune response of killer T cells), is accounted for or explained by dependent variables (the genes acting in the antigen presentation pathway). Furthermore, the closeness of the observed and expected in the contingency table for Hosmer and Lemeshow Test gives credence to the predictive power of the hypothesis with insignificant values of p>0.0. The chi-square test value is χ^2 (degree of freedom =4, number of samples =270) =209.614, p<0.001.

Table 4a: *Model Summary for genes in the antigen presentation pathway*

Step	-2 log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	130.102a	0.576	0.78

Table 4b: *Contingency table for genes in the antigen presentation pathway*

Step	Result = 0.00		Result = 1.00		Total
	Observed	Expected	Observed	Expected	
1	61	58.735	0	2.265	61
2	12	16.466	7	2.534	19
3	22	18.964	7	10.036	29
4	6	8.869	27	24.131	33
5	0	1.533	19	17.467	19
6	5	0.737	14	18.263	19
7	0	0.643	28	27.357	28
8	0	0.048	25	24.952	25
9	0	0.005	37	36.995	37

Table 4c: *Classification Table for genes in the antigen presentation pathway*

	Observed		Predicted		
			Result		Percentage
			0.00	1.00	Correct
Step 1	Result	0.00	95	11	89.6
		1.00	7	157	95.7
	Overall Percentage				93.3

4.4. Analysis of genes in T-cell function pathway

Genes acting in this pathway include *csk*, *ets1*, *HLA-dr2/dr3*, *ikzf1*, *ikzf3*, *il10*, *stat4*, *pdcd1*, *prdm1*, *ptpn22*, *tnfsf4*, *tyk2* genes. Binary Logistic Regression for the covariance of these genes shows statistical significance as shown in Table 5. Nagelkerke R Square test show that between 72.3% and 100% of the variability in the independent variable (the autoimmune response of killer T cells), is accounted for and explained by dependent variables (the genes acting in the T cell function pathway). Furthermore, the closeness of the observed and expected in the contingency table for Hosmer and Lemeshow Test gives credence to the predictive power of the hypothesis with insignificant values of p>0.0. The chi-square test value is χ^2 (degree of freedom =12, number of samples =270) =319.685, p<0.001.

Table 5a: *Model summary for genes in the T-cell function pathway*

Step	-2 log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	0.000a	0.723	1

Table 5b: *Contingency table for genes in the T-cell function pathway*

Step	Result = 0.00		Result = 1.00		Total
	Observed	Expected	Observed	Expected	
1	26	26.000	0	0.000	25
2	26	26.000	0	0.000	26
3	24	24.000	0	0.000	24
4	9	9.000	14	14.000	23
5	0	0.000	28	28.000	28
6	0	0.000	7	7.000	7
7	0	0.000	115	115.000	115

Table 5c: *Classification table for genes in the T-cell function pathway*

			Predicted		
	Observed		Result		Percentage
			0.00	1.00	Correct
Step 1	Result	0.00	85	0.00	100.0
		1.00	0.00	164	100.0
	Overall Percentage				100.0

4.5. Analysis of genes in B-cell function pathway

Genes acting in this pathway include *bank1*, *blk*, *csk*, *elf1*, *ets1*, *fcg2b*, *hla-dr2/dr3*, *ikzf1*, *il10*, *il21*, *lyn*, *prdm1*, *rasgrp3* genes. Binary Logistic Regression for the covariance of these genes shows statistical significance as shown in Table 6 below. Nagelkerke R Square test show that between 54.0% and 73.1% of the variability in the independent variable, (the autoimmune response of killer T cells) is accounted for or explained by dependent variables (the genes acting in b-cell function pathway). Furthermore, the closeness of the observed and expected in the contingency table for Hosmer and Lemeshow Test gives credence to the predictive power of the hypothesis with insignificant values of p>0.0. The chi-square test value is χ^2 (degree of freedom =15, number of samples =270) =319.685, p<0.001.

Table 6a: *Model Summary for genes in B-cell function pathway*

Step	-2 log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	152.129a	0.54	0.731

Table 6b: *Classification table for genes in B-cell function pathway*

Observed			Predicted		
			Result		Percentage
			0.00	1.00	Correct
Step 1	Result	0.00	85	0.00	100.0
		1.00	0.00	164	100.0
	Overall Percentage				100.0

Table 6c: *Contingency table for genes in B-cell function pathway*

Step	Result = 0.00		Result = 1.00		Total
	Observed	Expected	Observed	Expected	
1	27	27.000	0	0.000	27
2	26	26.000	0	0.000	26
3	27	27.000	0	0.000	27
4	5	5.000	21	21.000	26
5	0	0.000	28	28.000	28
6	0	0.000	19	19.000	19
7	0	0.000	96	96.000	96

4.6. Analysis of genes in signal transduction, cell cycle, growth, energy, metabolism, epigenetic modifications, DNA repair pathway

Genes acting in this pathway include *mecp2*, *ppp2ca*, *slc29a3* genes. Binary Logistic Regression for the covariance of these genes shows statistical significance as shown in Table 7. Nagelkerke R Square test show that between 48.2% and 65.3% of the variability in the independent variable (the autoimmune response of killer T cells) is accounted for or explained by dependent variables (the genes acting in the NFkB pathway). Furthermore, the closeness of the observed and expected in the contingency table for Hosmer and Lemeshow Test gives credence to the predictive power of the hypothesis with insignificant values of $p>0.01$. The chi-square test value is χ^2 (degree of freedom =3, number of samples =270) =177.612, $p<0.001$.

Table 7a: *Model summary for genes in Signal transduction, cell cycle pathway*

Step	-2 log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	184.130a	0.482	0.653

Table 7b: *Contingency table for genes in Signal transduction, cell cycle pathway*

Step	Result = 0.00		Result = 1.00		Total
	Observed	Expected	Observed	Expected	
1	24	23.504	0	0.496	24
2	42	43.498	14	12.502	56
3	25	22.38	14	16.62	39
4	10	12.228	14	11.772	24
5	0	1.618	18	16.382	18
6	5	1.77	20	23.23	25
7	0	0.888	31	30.112	31
8	0	0.114	53	52.886	53

Table 7c: *Classification table for genes in Signal transduction, cell cycle pathway*

Observed			Predicted		
			Result		Percentage
			0.00	1.00	Correct
Step 1	Result	0.00	101	5	95.3
		1.00	42	122	74.4
	Overall Percentage				82.6

4.7. Analysis of genes in the complement and clearance of immune complexes pathway

Genes acting in this pathway include *c1q*, *fcgr2a/2b*, *fcgr3b* genes. Binary Logistic Regression for the covariance of these genes shows statistical significance as shown in Table 8. Nagelkerke R Square test show that between 73.8% and 100% of the variability in the independent variable (the autoimmune response of killer T cells), is accountable for or explained by dependent variables (the genes acting in the immune complex pathway). Furthermore, the closeness of the observed and expected in the contingency table for Hosmer and Lemeshow Test gives credence to the predictive power of the hypothesis with insignificant values of p>0.05. The chi-square test value is χ^2 (degree of freedom =5, number of samples =270) =361.743, p<0.001.

The analysis gives credence to the hypothesis; hence variation (SNPs) in more 50% of genes would predispose a patient to autoimmune activities of killer T cells.

Table 8a: *Model summary for genes in immune complexes pathway*

Step	-2 log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	0.000a	0.738	1

Table 8b: *Contingency table for genes in immune complexes pathway*

Step	Result = 0.00		Result = 1.00		Total
	Observed	Expected	Observed	Expected	
1	50	50.000	0	0.000	50
2	29	29.000	0	0.000	29
3	27	27.000	0	0.000	27
4	0	0.000	28	28.000	28
5	0	0.000	19	19.000	19
6	0	0.000	21	21.000	21
7	0	0.000	96	96.000	96

4.8. Approximate matching algorithm

The approximate matching algorithm involves adding a Boyer-Moore approach to a shift-add method when searching all occurrences of a pattern string $P = P1 \dots Pm$ in a text string $T = T1$ …. Tn with at most k mismatches, $1 < k < m$ [63]. For some position j, the state vector S_j is a bit number (represented in base 2b), and S_j contains individual states of the search between each prefix of P and the corresponding substring of t. Individual state $S_j[i]$ contains the number of mismatches between $P1 \dots Pi$ *and* $Tj\text{-}m{+}1 \dots Tj$. Furthermore, P matches at j if and only if $S_j[m] < k + 1$ [63].

Software development and testing of the software were done in Ubuntu 16.04 LTS platform on a virtual machine on Google cloud with specifications Intel Xeon, 52 Gigabytes RAM and 500GB Hard disk size. The software was written in Python programming Language and run-on Ubuntu terminal interface.

Open-source DNA sequence data (16) located on the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) in the United States were used for this research. Data was taken from locals within the Nigeria tribes with run or accession numbers; SRR017528, SRR016407 SRR799760, SRR020178, SRR023623, SRR023628, SRR1596548, SRR1611133, SRR1611134, SRR1611136, SRR630584, SRR630583, SRR618674, SRR618673, SRR618672, SRR618671.

Table 9 below gives a summary of the results gotten from 16 patients. The results of the findings suggest all the autoimmune response of killer T – cells are likely to manifest in all the sixteen patients as Single Nucleotide Polymorphisms (SNPs) as seen to manifest for all the 40 genes, and this goes in-line with the stated hypothesis. The patients DNA have experienced a significant mutation as all the pathways that produce autoreactive killer T cells are activated.

Table 9: *Result of Analysis of study participants*

Run Number	Selection	Base Pairs	Date Published (DD/MM/YY)	SLE Susceptibility
SRR017528	Random	32.1Gb	30-06-2009	YES
SRR016407	Random	25.1Gb	10-07-2009	YES
SRR799760	Random	1.4Gb	25-03-2013	YES
SRR020178	Unspecified	25GB	31-03-2011	YES
SRR023623	Unspecified	8.4Gb	16-03-2010	YES
SRR023628	Unspecified	12Gb	16-03-2010	YES
SRR1596548	Unspecified	72.6Mb	13-10-2014	YES
SRR1611133	Size Fractionation	106.4M	22-07-2015	YES
SRR1611134	Unspecified	566.2Mb	13-10-2014	YES
SRR1611136	Unspecified	564.2Mb	14-10-2014	YES
SRR630584	Random	3.4Gb	21-12-2012	YES
SRR630583	Random	3.4Gb	21-12-2012	YES
SRR618671	Random	22.1Gb	21-12-2012	YES
SRR618673	Random	20.9Gb	21-12-2012	YES
SRR618674	Random	21Gb	21-12-2012	YES
SRR618672	Random	21.9Gb	21-12-2012	YES

However, it is observed that variations set at 0.1% of the length of the patient genome have no match to any of the genes. This unexpected occurrence could well result from structural variation in the genome like multiple copy number variations, deletions, insertions, and duplications account for this outcome.

5. Conclusion

The goal of this work was to create an algorithm that could analyze the DNA sequence of a patient with Systemic Lupus Erythematosus (SLE) and predict killer T-cell responses. The method used was to match the DNA coding sequence of a gene from the reference genome to the coding sequence of that gene in the patients' genome to discover a gene variant in patient DNA. The threshold for significant single nucleotide polymorphisms is 0.1 percent of the overall length of the gene-coding DNA sequence. The approximate sequence matching algorithm developed was used to conduct the matching. The findings from 16 patients reveal that they will all have autoimmune killer T-cells. Furthermore, the algorithm's predictive power and accuracy are both at 80%. Future direction for the present study will involve large number of data set and extensive clinical testing to evaluate the consistency of the model for predicting the response of killer T-cells.

References

1. Sestak AL, Nath SK, Sawalha AH, et al. Current status of lupus genetics. Arthritis Res Ther. 2007;9:1-9.

2. https://emedicine.medscape.com/article/1884084-overview

3. Deng Y, Tsao BP. Advances in lupus genetics and epigenetics. Curr Opin Rheumatol. 2014;26:482-92.

4. Firestein GS, Budd R, Gabriel SE, McInnes IB and O'Dell JR: Kelley's Textbook of Rheumatology E-Book. (11th edn), Elsevier Health Sciences. Netherlands, Europe. 2012.

5. https://www.freepatentsonline.com/6280941.html

6. Hochberg MC. Updating the american college of rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheumatol. 1997;40:1725-34.

7. Lloyd P, Doaty S, and B. Hahn. Aetiopathogenesis of systemic lupus erythematosus. Systemic lupus erythematosus. (1st edn), Oxford University Press. Oxford, UK.2016.

8. Wallace D and Hahn BH. Dubois' Lupus Erythematosus and Related Syndromes (9thedn), Elsevier Health Sciences. Neatherlands, Europe. 2018.

9. Krishnan S, Chowdhury B, Juang YT, et al. Overview of the pathogenesis of systemic lupus erythematosus. Systemic lupus erythematosus: acompanion to Rheumatology. Mosby Inc. 2007;pp.55-63.

10. Criswell LA. The genetic contribution to systemic lupus erythematosus. Bull NYU Hosp Jt Dis. 2008;66:176-83.

11. Deng Y and Tsao BP. Genes and genetics in human systemic lupus erythematosus. In: Tsokos GC (ed). Systemic lupus erythematosus. Academic Press, Massachusetts. 2016;pp. 69-76.

12. Fernando M and Vyse T. Major Histocompatibility Complex Class II.(5th edn), Academic Press. Massachusetts. USA. 2011.

13. Forabosco P, Gorman J, Cleveland C, et al. Meta-analysis of genome-wide linkage studies of systemic lupus erythematosus. Genes Immun. 2006;7:609-14.

14. B. P. Tsao. Update on human systemic lupus erythematosus genetics. Curr Opin Rheumatol 2004;16:513-21.

15. Morris D, Fernando M, Taylor K, et al. MHC associations with clinical and autoantibody manifestations in European SLE. Genes Immun. 2004;15:210-7.

16. Klein J, Sato A. The HLA system. N Engl J Med. 2000;343:702-9.

17. Chistiakov DA. Interferon induced with helicase C domain 1 (IFIH1) and virus-induced autoimmunity: a review. Viral Immunol. 2010;23:3-15.

18. Molineros JE, Maiti AK, Sun C, et al. Admixture mapping in lupus identifies multiple functional variants within IFIH1 associated with apoptosis, inflammation, and autoantibody production. PLoS Genet. 2013;9:1-19.

19. Møller-Larsen S, Nyegaard M, Haagerup A, et al. Association analysis identifies TLR7 and TLR8 as novel risk genes in asthma and related disorders. Thorax. 2008;63:1064-9.

20. Jensen MA, Niewold TB. Interferon regulatory factors: critical mediators of human lupus. Transl Res. 2015;165:283-95.

21. Azevedo Silva JD, Addobbati C, Sandrin-Garcia P, et al. Systemic lupus erythematosus: old and new susceptibility genes versus clinical manifestations. Curr Genomics. 2014;15:52-65.

22. Remmers EF, Plenge RM, Lee AT, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med. 2007;357:977-86.

23. Cen H, Leng RX, Wang W, et al. Association study of IFIH1 rs1990760 polymorphism with systemic lupus erythematosus in a Chinese population. Inflammation. 2013;36:444-8.

24. Han JW, Zheng HF, Cui Y, et al. Genome-wide association study in a chinese han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genet. 2009;41:1234-7.

25. Kaufman KM, Zhao J, Kelly JA, et al. Williams. Fine mapping of Xq28: both MECP2 and IRAK1 contribute to risk for systemic lupus erythematosus in multiple ancestral groups. Ann Rheum Dis. 2013;72:437-44.

26. Beyaert R, Heyninck K, Van Huffel S. A20 and A20-binding proteins as cellular inhibitors of nuclear factor-κB-dependent gene expression and apoptosis. Biochem Pharmacol. 2000;60:1143-51.

27. Graham RR, Cotsapas C, Davies L, et al. Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nat Genet. 2008;40:1059-61.

28. Musone SL, Taylor KE, Lu TT, et al. 4Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. Nat Genet. 2008;40:1062-4.

29. Yang W, Shen N, Ye DQ, et al. Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. PLoS Genet. 2010;6:1-11.

30. Verstrepen L, Carpentier I, Verhelst K, et al. ABINs: A20 binding inhibitors of NF-κB and apoptosis signaling. Biochem Pharmacol. 2009;78:105-14.

31. Jacob CO, Zhu J, Armstrong DL, et al. Identification of IRAK1 as a risk gene with critical role in the pathogenesis of systemic lupus erythematosus. Proc Nat Acad Sci. 2009;106:6256-61.

32. Webb R, Wren JD, Jeffries M, et al. Variants within MECP2, a key transcription regulator, are associated with increased susceptibility to lupus and differential gene expression in patients with systemic lupus erythematosus. Arthritis Rheum. 2009;60:1076-84.

33. Pan Y, Sawalha AH. Epigenetic regulation and the pathogenesis of systemic lupus erythematosus. Transl Res. 2009;153:4-10.

34. Chistiakov DA, Turakulov R. CTLA-4 and its role in autoimmune thyroid disease. J Mol Endocrinol. 2003;31:21-36.

35. Mostowska M, Wudarski M, Chwalińska-Sadowska H, et al. The programmed cell death 1 gene 7209 C> T polymorphism is associated with the risk of systemic lupus erythematosus in the polish population. Clin Exp Rheumatol. 2008;26:457-60.

36. Wang SC, Chen YJ, Ou TT,et al. Programmed death-1 gene polymorphisms in patients with systemic lupus erythematosus in Taiwan. J Clin Immunol. 2006;26:506-11.

37. Griffiths EK, Krawczyk C, Kong YY, et al. Positive regulation of T cell activation and integrin adhesion by the adapter Fyb/Slap. Science. 2001;293:2260-3.

38. Peterson EJ, Woods ML, Dmowski SA, et al. Coupling of the TCR to integrin activation by Slap-130/Fyb. Science. 2001;293:2263-5.

39. Sandrin-Garcia P, Junta CM, Fachin AL, et al. Shared and unique gene expression in systemic lupus erythematosus depending on disease activity. Ann N Y Acad Sci. 2009;1173:493-500.

40. Ito T, Wang YH, Duramad O, et al. TSLP-activated dendritic cells induce an inflammatory T helper type 2 cell response through OX40 ligand. J Exp Med. 2005;202:1213-23.

41. Linton PJ, Bautista B, Biederman E, et al. Costimulation via OX40L expressed by B cells is sufficient to determine the extent of primary CD4 cell expansion and Th2 cytokine secretion in vivo. J Exp Med. 2003;197:875-83.

42. Gateva V, Sandling JK, Hom G, et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nat Genet. 2009;41:1228-33.

43. Madan R, Demircik F, Surianarayanan S, et al. Nonredundant roles for B cell-derived IL-10 in immune counter-regulation. J Immunol. 2009;183:2312-20.

44. Sakurai D, Zhao J, Deng Y, et al. Preferential binding to Elk-1 by SLE-associated IL10 risk allele upregulates IL10 expression. PLoS Genet. 2013;9:1-12.

45. Stone JC. Regulation and function of the RasGRP family of ras activators in blood cells. Genes Cancer. 2011;2:320-34.

46. Wang C, Ahlford A, Jaervinen TM, et al. Genes identified in asian SLE GWASs are also associated with SLE in caucasian populations. Eur J Hum Genet. 2013;21:994-9.

47. Kozyrev SV, Abelson AK, Wojcik J, et al. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nat Genet. 2008;40:211-6.

48. Sanchez E, Rasmussen A, Riba L, et al. Impact of genetic ancestry and sociodemographic status on the clinical expression of systemic lupus erythematosus in American Indian–European populations. Arthritis Rheum. 2012;64:3687-94.

49. Manjarrez-Orduño N, Marasco E, Chung SA, et al. CSK regulatory polymorphism is associated with systemic lupus erythematosus and influences B-cell signaling and activation. Nat Genet. 2012;44:1227-30.

50. Moisan J, Grenningloh R, Bettelli E, et al. Ets-1 is a negative regulator of Th17 differentiation. J Exp Med. 2007;204:2825-35.

51. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013;45:1238-43.

52. Hom G, Graham RR, Modrek B, et al. Association of systemic lupus erythematosus with C8orf13–BLK and ITGAM–ITGAX. N Engl J Med. 2008;358:900-9.

53. Zhou XJ, Lu XL, Lv JC, et al. Genetic association of PRDM1-ATG5 intergenic region and autophagy with systemic lupus erythematosus in a Chinese population. Ann Rheum Dis. 2011;70:1330-7.

54. Guthridge JM, Lu R, Sun H, et al. Two functional lupus-associated BLK promoter variants control cell-type-and developmental-stage-specific transcription. Am J Hum Genet. 2014;94:586-98.

55. Crow YJ, Hayward BE, Parmar R, et al. Mutations in the gene encoding the 3′-5′ DNA exonuclease TREX1 cause aicardi-goutieres syndrome at the AGS1 locus. Nat Genet. 2006;38:917-20.

56. Nimmerjahn F, Ravetch JV. Fcγ receptors as regulators of immune responses. Nat Rev Immunol. 2008;8:34-47.

57. Fossati-Jimack L, Ling GS, Cortini A, et al. Phagocytosis is the main CR3-mediated function affected by the lupus-associated variant of CD11b in human myeloid cells. PloS One. 2013;8:1-11.

58. Bock M, Heijnen I, Trendelenburg M. Anti-C1q antibodies as a follow-up marker in SLE patients. PLoS One. 2015;10:1-11.

59. Schejbel L, Skattum L, Hagelberg S, et al. Molecular basis of hereditary C1q deficiency-revisited: identification of several novel disease-causing mutations. Genes Immun. 2011;12:26-34.

60. Yih Chen J, Ling Wu Y, Yin Mok M, et al. Effects of complement C4 gene copy number variations, size dichotomy, and C4A deficiency on genetic risk and clinical presentation of systemic lupus erythematosus in east asian populations. Arthritis Rheumatol. 2016;68:1442-53.

61. Pereira KMC, Faria AGA, Liphaus BL, et al. Low C4, C4A and C4B gene copy numbers are stronger risk factors for juvenile-onset than for adult-onset systemic lupus erythematosus. Rheumatology (Oxford). 2016;55:869-73.

62. Benaglio P, Rivolta C. Ultra high throughput sequencing in human DNA variation detection: a comparative study on the NDUFA3-PRPF31 region. PLoS One. 2010;5:1-10.

63. El-Mabrouk N, Crochemore M. Boyer-Moore strategy to efficient approximate string matching. Lecture Notes in Computer Science. 7th Annual Symposium. California USA. 1996.