

## Research Article

# Application of Genetic Algorithm-Based Support Vector Machine in Identification of Gene Expression Signatures for Psoriasis Classification: A Hybrid Model

Leili Tapak <sup>1,2</sup>, Saeid Afshar <sup>3,4</sup>, Mahlagha Afrasiabi <sup>5</sup>, Mohammad Kazem Ghasemi <sup>1</sup>, and Pedram Alirezaei <sup>6</sup>

<sup>1</sup>Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

<sup>2</sup>Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

<sup>3</sup>Research Center for Molecular Medicine, Hamadan University of Medical Sciences, Hamadan, Iran

<sup>4</sup>Department of Medical Biotechnology, School of Advanced Medical Sciences and Technologies, Hamadan University of Medical Sciences, Hamadan, Iran

<sup>5</sup>Department of Computer, Hamedan University of Technology, Hamedan, Iran

<sup>6</sup>Department of Dermatology, Psoriasis Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

Correspondence should be addressed to Saeid Afshar; safshar.h@gmail.com

Received 5 March 2021; Accepted 23 August 2021; Published 8 September 2021

Academic Editor: Khac-Minh Thai

Copyright © 2021 Leili Tapak et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Background.** Psoriasis is a chronic autoimmune disease impairing significantly the quality of life of the patient. The diagnosis of the disease is done via a visual inspection of the lesional skin by dermatologists. Classification of psoriasis using gene expression is an important issue for the early and effective treatment of the disease. Therefore, gene expression data and selection of suitable gene signatures are effective sources of information. **Methods.** We aimed to develop a hybrid classifier for the diagnosis of psoriasis based on two machine learning models of the genetic algorithm and support vector machine (SVM). The method also conducts gene signature selection. A publically available gene expression dataset was used to test the model. **Results.** A number of 181 probe sets were selected among the original 54,675 probes using the hybrid model with a prediction accuracy of 100% over the test set. A number of 10 hub genes were identified using the protein-protein interaction network. Nine out of 10 identified genes were found in significant modules. **Conclusions.** The results showed that the genetic algorithm improved the SVM classifier performance significantly implying the ability of the proposed model in terms of detecting relevant gene expression signatures as the best features.

## 1. Introduction

Psoriasis is a chronic autoimmune/inflammatory and hyperproliferative disease with primary manifestations on skin and joints [2]. Psoriasis has been reported to be associated with increased hyperlipidemia, hypertension, coronary artery disease (CAD), diabetes (type II), and obesity [3] as well as the increased risks of stroke and myocardial infarction [4, 5]. Moreover, psoriasis is associated with anxiety, isolation, and mental disorders (e.g., depression), and it reduces socialization for patients and prevents them from having close and intimate relationships [6–8].

Psoriasis can occur at any age, and studies have shown that the inception of the disease takes place between the ages of 20–30 and 50–60 [3]. Nevertheless, most psoriasis patients are <30 years old with a large proportion less than 10 years of age [9], and its prevalence varies according to the climatic/geographical situations such that the Caucasian population has a higher rate and the Asian population has a lower rate among others [10, 11]. There are several phenotypic manifestations including epidermal hyperplasia, angiogenesis, and changed keratinocyte differentiation. Moreover, patients' skin was infiltrated with neutrophils, dendritic cells, and T lymphocytes and chemokine as well as cytokine

[12–15]. Visual examination of cutaneous lesion biopsy is the main but inefficient route of diagnosis of psoriasis [14].

While the genetic base of psoriasis has been confirmed, its molecular structure still requires more investigations to be well-enough understood to be useful in the diagnosis of psoriasis [16, 17]. In this regard, quantitative polymerase chain reaction (qPCR) and microarray (high-throughput) techniques have been employed by large-scale studies to reveal the molecular basis of psoriasis and to explore the gene expression patterns in lesional samples compared with nonlesional samples in the disease [15, 18]. In a systematic screening of about 54,000 probe sets conducted by using the Affymetrix HG-U133 Plus 2 platform, a number of 179 unique differential genes (out of 223 probe sets) were identified in the uninvolved psoriatic cutaneous samples and the dysregulated genes were demonstrated to be modulated with the three transcription factors associated with lipid metabolism including sterol regulatory element-binding protein (SREBF), peroxisome proliferator activator receptor alpha (PPARA), and estrogen receptor 2 (ESR2) [16]. Moreover, by screening expressions of approximately 22,000 probe sets using the Affymetrix HG-U133A platform in a study [19], a number of 179 genes among 203 probe sets were detected where the expression of genes in the psoriatic lesion skins changed at least two-fold, and it was shown that there is an association between the Wnt pathway, regulating the stem cell proliferation, and the psoriasis development [15]. Nevertheless, its causative role has remained elusive. There have been detected many up/downregulated probe sets related to lesional samples of skin obtained from psoriasis patients with at least 3-fold changes in expression of some of them like multiple T cell genetic markers and two type I interferon-inducible genes [15, 19]. There are also studies that have compared the level of gene expression of psoriatic and healthy samples in human and mice [20, 21]. While it was observed that the lists of the top 5,000 fold change (up/downregulated probes) were in common with different phenotypes of psoriasis in mice, there is no consistency between different studies in terms of the lists of marker genes. Therefore, developing classification models and identifying target genes are necessary.

During the last decades, machine learning methods have received much attention as they have provided a computer-aided bed for constructing modern classifiers and they have been shown to provide promising results. It has been also shown that the combination of these methods with each other can enhance the prediction power of the classifier. These models can be also used for marker gene selection. Among machine learning methods, support vector machine (SVM) owing to the kernel method has shown promising performance in classification/regression in several medical problems and gene expression studies. SVM fits a hyperplane to separate groups and projects the input space into a higher space by using kernel functions [22]. In this regard, utilizing heuristic approaches like the genetic algorithm (GA) leads to determining a subset of inputs to fit a hyperplane that is most robust among others. GA is an evolutionary algorithm that has received much attention in optimization problems [23]. Mimicking the biology of

changes that occur in a DNA sequence like mutation and crossover, GA works by generating an initial set of solutions, which are then assessed through a fitness function. In this way, a subset consisting of the best solutions from the initial solutions is selected and operators like mutation are utilized to generate the consequent sets of solutions [23]. Studies have shown that an automated hybrid system for the diagnosis of diseases like cancer designed based on GA and SVM outperforms the SVM. In this hybrid approach, the GA is used to reduce the dimension of the feature space [24, 25].

To our knowledge, there is no study that proposes hybrid models for the diagnosis of psoriasis. This study proposed a hybrid approach for identifying the psoriasis-associated features (probe sets). The used method is the combination of a GA and SVM for feature selection.

## 2. Methods

*2.1. Data Source and Preprocessing.* A publicly available dataset of psoriasis whole blood transcriptome dataset (available in GEO repository: GSE55201 dataset generated using the Affymetrix U133 Plus (microarray) with platform ID GPL570) was used. This dataset consisted of expression data of 30 healthy controls and 44 psoriasis patients at baseline and 7 psoriasis patients after two weeks of treatment [26]. In this study, the differentially expressed genes (DEGs) between 30 healthy controls and 44 psoriasis patients at baseline samples were determined by using the limma package [27] under R software (version 4.0.3) [28] for subsequent analysis.

*2.2. Support Vector Machine.* The SVM is a machine learning technique that has been developed for classification and regression problems, and it encompasses parts of nonparametric statistics and machine learning. The fundamental idea is to map the covariates in the input space into a space with a higher dimension using some kernel functions, so that a linear regression handles the complex nonlinear regression of the primary input space. The hyperplane equation is  $\mathbf{w} \cdot \mathbf{x} + b = 0$ , and there is a need to minimize the norm of the coefficient vector  $\mathbf{w}$  and to maximize the margin  $1/\|\mathbf{w}\|$  between two classes [25].

This method utilizes the structural risk minimization principle to fit a hyperplane separating two groups optimally (Figure 1). Let us assume that there is a subject  $j$  with an input vector of  $\mathbf{x}_j \in \mathbb{R}^p$  with  $p$  components, which should be classified as psoriasis patient ( $y_j = -1$ ) or normal subject ( $y_j = +1$ ). The SVM problem is represented by

$$f(\mathbf{x}) = \mathbf{w} \cdot \phi(\mathbf{x}) + b, \quad (1)$$

where  $\mathbf{w}$  and  $b$  stand for the weight vector (regression coefficients) and the bias term, respectively. Then, by considering an  $\epsilon$ -insensitivity loss function for equation (1), the following optimization problem ((2) and (3)) can be considered a convex optimization problem:

$$\frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum \xi_j + C \sum \xi_j^*, \quad (2)$$

which is optimized given the following restrictions:

$$\begin{cases} w^T \phi(x_j) + b - y_j \leq \varepsilon + C\xi_j, \\ y_j - w^T \phi(x_j) - b \leq \varepsilon + C\xi_j^*, \\ j = 1, \dots, N, \end{cases} \quad (3)$$

where  $\xi_j^*$  and  $\xi_j$  are nonnegative slack variables (which penalize the training errors in the loss function) with the error tolerance of  $\varepsilon$ . Also,  $C > 0$  is the tradeoff parameter which shows the capacity (tuning parameter) and determines the empirical error's degree (lower values of  $C$  are related to a wider margin and a reduced risk of overfitting an SVM but larger in-sample classification errors). Optimizing of the problem [2] is conducted through minimization of the Lagrange function (4):

$$L(w, b, \xi; \alpha, \nu) = \frac{1}{2} w^T w + C \sum \xi_j - \sum \alpha_j \{ y_j (w^T \phi(x_j) + b) - 1 + \xi_j \} - \sum \nu_j \xi_j, \quad (4)$$

with the  $\alpha_j$  and  $\nu_j$  as the Lagrange multipliers. This convex optimization problem can be solved through nonlinear programming tools or via a convex quadratic programming problem in  $\alpha_j$ . In a nonlinear SVM setting, the score of a subject is computed by substituting the scalar product of the covariates with a kernel function (e.g., polynomial, Gaussian radial basis (GRBF), and exponential radial basis) [29, 30].

**2.3. Genetic Algorithm.** A genetic algorithm is an exploratory algorithm utilized to solve optimization problems. In a GA, a set of candidate solutions (individuals), which is called the initial population, evolved toward better solutions for an optimization problem. Each individual has a set of properties. Solutions are considered strings of 0/1s (in binary). The solutions are generated using natural evolution, such as inheritance, mutation, and selection. In a GA, chromosomes of a population encode candidate solutions [31].

This evolution is usually created by a random population and occurs in subsequent generations. In each generation, the fitness of each individual in the population is calculated, while several individuals are randomly chosen from the recent population. Then, it is modified to create a new better population. Then, this modification of population is repeated in the next iterations. Usually, the algorithm ends when the maximum number of generations is produced or a predefined level of fitness is obtained for the most recent generated population [23].

In the present study, because we intended to address parameter optimization and feature selection simultaneously, the chromosome (i.e., each individual in the population) was a combination of parameter genes ( $C$ ,  $\varepsilon$ , and other parameters in kernel functions such as  $\gamma$  in a kernel function of the form  $\phi(x_j, x_j') = \exp(-\gamma \|x_j, x_j'\|^2)$ ) and feature gene ( $f_1, \dots, f_G; f_n \in \{1, 2, \dots, G\}$  and  $G$  is the number

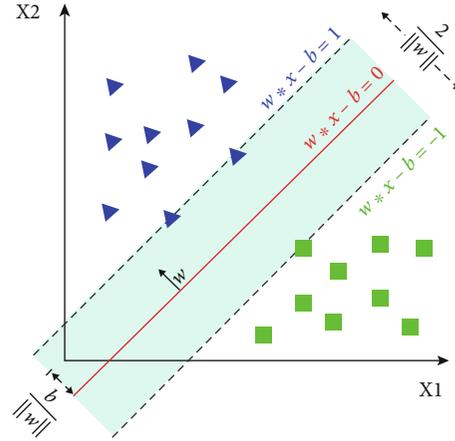


FIGURE 1: Linear separation of two classes with a support vector machine classifier. Samples on the margin are called the support vectors.

of candidate features for constructing the model) [32]. Here, a chromosome stands for an individual in GAs, and parameters included in a chromosome are used for SVM modeling [23].

Evaluation and comparison of each candidate were conducted and quantified through a fitness function. In the present study, the accuracy of the SVM classifier of 10-fold cross-validation (CV) was utilized as the fitness function, and greater fitness value was related to a better individual. Given training data  $((x_j, y_j) \in (\mathbb{R}^p, \{-1, +1\}))$  for  $N$  subjects, the objective function can be calculated by equation (4). The accuracy of 10-fold CV for the SVM was calculated by

$$\text{Accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{total sample}}, \quad (5)$$

which is the proportion of correct predictions among the total number of samples examined (true positive indicates the number of psoriasis patients who were predicted as patients truly, and true negative indicates the number of healthy cases who were predicted as healthy truly).

**2.4. Hybrid GA-SVM Model and Tuning Parameters.** This study utilized a method for enhancing the performance of the SVM. The enhancement is to select the best subset of features and to optimize the parameters of the model. GA was utilized to handle both aforementioned aspects in the SVM classification problem simultaneously for psoriasis diagnosis. In GA, the fitness function was the accuracy of the SVM classifier. The GA algorithm was used to optimize the objective function defined for the SVM classifier (equation (4)) and to find suitable features (feature selection) for the diagnosis of psoriasis and detecting a healthy/patient person. Figure 2 shows the block diagram of the hybrid GA-SVM model.

In the GA used, the number of populations was five. The solutions were represented in binary as strings of 0s and 1s. The value 0 indicated that the feature/attribute was not selected and value 1 indicated that the feature/attribute

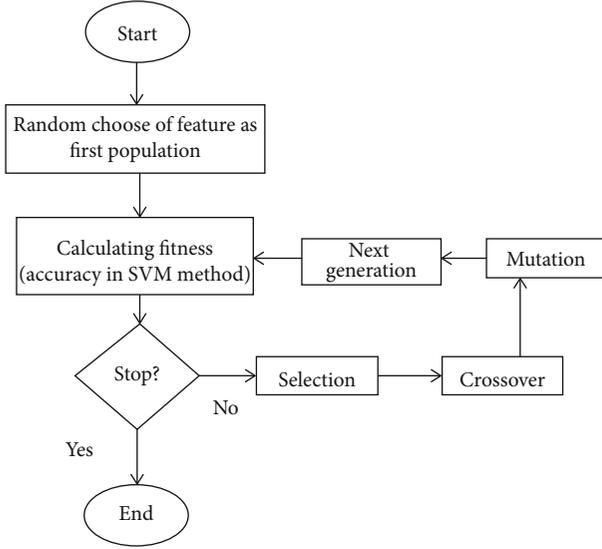


FIGURE 2: Overview of the hybrid GA-SVM model.

was selected. The roulette wheel model, single-point crossover operator with 0.8 of crossover rate, and mutation operator with 0.01 of the mutation rate were used for the selection of the appropriate chromosomes to produce the next generation.

In the roulette wheel selection, the probability of selecting an individual upbringing of the next generation is proportional to its fitness. A better fitness (here, greater prediction accuracy) is related to a higher probability of selecting an individual. Selecting a solution/individual can be considered spinning roulette with pockets for each individual with sizes depending on their probability ( $p_j = f_j / \sum_{j=1}^G f_j$ , where  $G$  is the size of the current generation).

To optimize the parameters of the SVM, they were encoded with binary chains on two fix search intervals of

$$\begin{aligned} C_{\max} < C < C_{\min}, \\ \gamma_{\max} < \gamma < \gamma_{\min}. \end{aligned} \quad (6)$$

Thus, a 32-bit encoding scheme of  $C$  (i.e.,  $C_{b1}, \dots, C_{b32}$ ) and  $\gamma$  where

$$\begin{aligned} C_b &= \sum_{i=1}^{32} C_{bi} 2^{i-1} = \frac{g_{\max}(C - C_{\min})}{C_{\max} - C_{\min}}, \\ \gamma_b &= \sum_{i=1}^{32} \gamma_{bi} 2^{i-1} = \frac{g_{\max}(\gamma - \gamma_{\min})}{(\gamma_{\max} - \gamma_{\min})}, \end{aligned} \quad (7)$$

with  $g_{\max} = 2^{32} - 1$  was considered [33].

**2.5. Evaluation Criteria.** We exerted the method on 74 samples (50 samples for training and 24 samples for testing).

TABLE 1: The performance criteria of the SVM and GA-SVM.

Number of used features	Method	Total accuracy (%)	AUC*
54,657	SVM	62.500	0.625
27,265	GA+SVM	79.167	0.792
445	SVM	87.500	0.878
181	GA+SVM	100.000	1.000

\* Area under the ROC curve.

In this study, the total accuracy and the area under the ROC curve (AUC) were used to evaluate the performance of the models.

$$\text{Accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{total sample}}. \quad (8)$$

Analyses were performed by using the MATLAB software programming [34].

**2.6. Protein-Protein Interaction (PPI) Network.** The protein-protein interaction (PPI) network was constructed using the STRING version 11.0 [35] with a confidence cutoff of 0.7 for selected probes with GA+SVM. The constructed PPI network was visualized and analyzed using Cytoscape version 3.6.0 [36]. The CytoHubba plugin under Cytoscape software was used to determine the hub genes. Modules of the constructed network were evaluated with the MCODE plugin under Cytoscape software with cutoff criteria of the number of nodes more than 5 and MCODE score more than 3 and default parameters (node score cutoff = 0.2,  $K$ -core = 2, degree cutoff = 2, and max depth = 100).

**2.7. Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway and Gene Ontology (GO) Enrichment Analysis.** The KEGG pathway and GO enrichment analysis including the biological process, molecular function, and cellular component were performed using the DAVID database (<https://david.ncifcrf.gov/>) for selected probes with GA+SVM.

### 3. Results

**3.1. Comparison of the Models.** Table 1 shows the performance of the SVM and GA-SVM. The first two lines of the table are related to the original 54,675 probe sets which led to a total accuracy and AUC of 62.500% and 0.625, respectively. By using features selected by GA (27,265 features), the total accuracy and AUC of the SVM increased to 79.167% and 0.792, respectively. Moreover, after an initial screening of the probe sets using the limma package based on the  $|FC| \geq 1$  and adjusted  $P$  value  $< 0.05$ , a number of 445 DEGs were selected for further evaluation. Then, the SVM and GA-SVM were trained using them. According to the results (Table 1), the total accuracy and AUC of the SVM using 445 DEGs were 87.500% and 0.878, respectively. After feature selection using GA, the number of 181 probe sets was selected and the total accuracy and AUC of the SVM (using 181 features) increased to 100 and 1, respectively. Also,

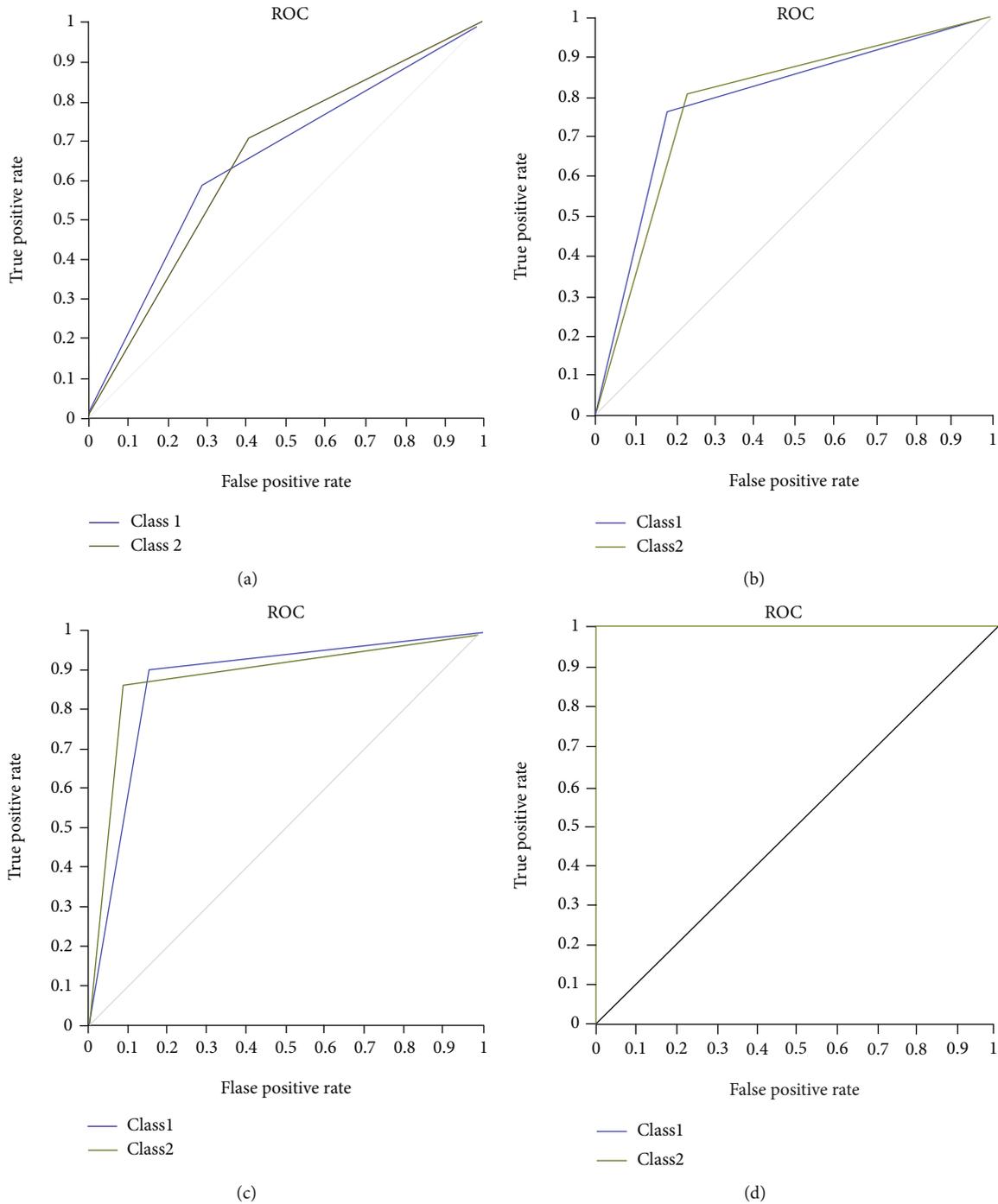


FIGURE 3: The ROC curves of the four scenarios of classification of psoriasis patients using (a) SVM with 54,657 features, (b) GA+SVM with 27,265 features, (c) SVM with 445 features, and (d) GA+SVM with 181 features.

ROC curves related to the four scenarios are provided in Figure 3.

3.2. *PPI Network Analysis.* The constructed PPI network for selected probes with GA+SVM, consisting of 100 nodes and 244 edges, is shown in Figure 4. The top 10 genes with a high degree including ribosomal protein S3 (RPS3), ribosomal protein S5 (RPS5), ribosomal protein S20 (RPS20), ribo-

somal protein S15a (RPS15A), ribosomal protein S3A (RPS3A), X-linked ribosomal protein S4 (RPS4X), ribosomal small subunit protein 7 (RPS7), ribosomal protein L13 (RPL13), ribosomal protein L35 (RPL35), and heat shock protein family A member 8 (HSPA8) were selected.

Evaluation of the PPI network indicated that a significant module with a score of 14.570 consisted of 15 nodes and 102 edges. As seen in Figure 4, among the top 10 genes,

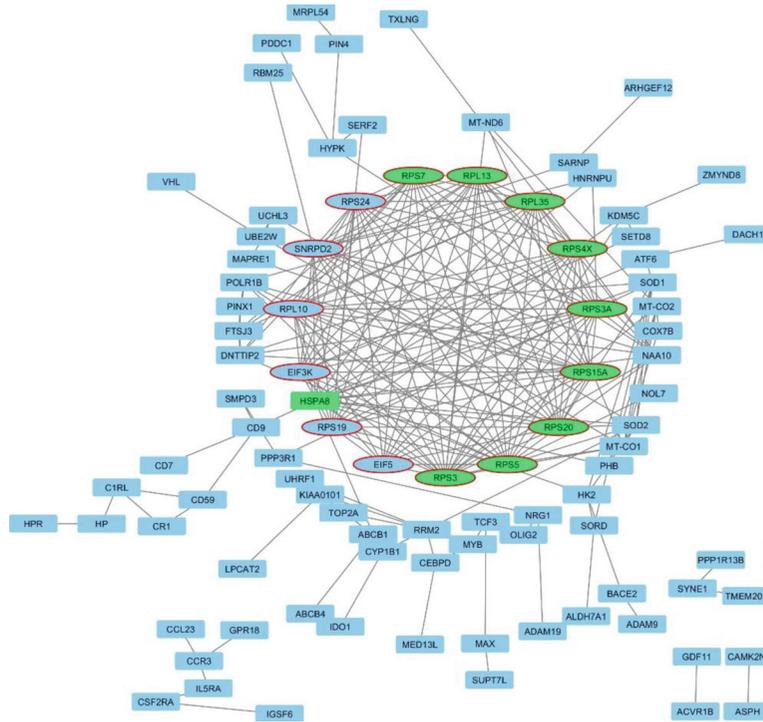


FIGURE 4: Protein-protein interaction network. The PPI network was constructed using STRING and visualized with Cytoscape. The selected top 10 genes with a high degree were shown in green. The nodes related to the significant module were shown in ellipse shapes.

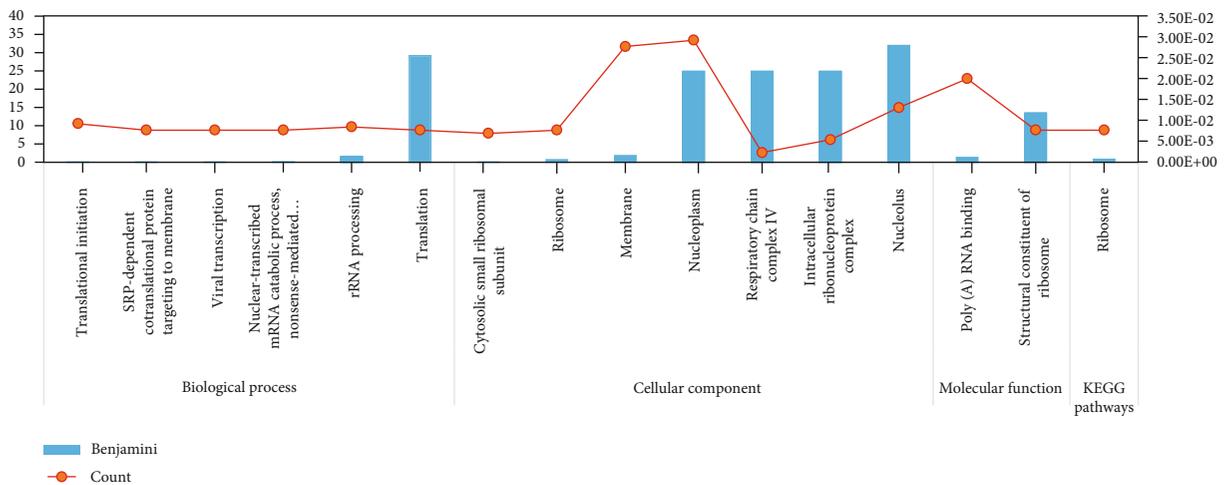


FIGURE 5: Gene Ontology and KEGG pathway enrichment analysis. The KEGG pathway and GO enrichment analysis for selected probes with GA+SVM were performed using the DAVID database.

9 hub genes including RPS3, RPS5, RPS20, RPS15A, RPS3A, RPS4X, RPS7, RPL13, and RPL35 were found in the significant module.

3.3. GO and KEGG Pathway Analysis. The KEGG pathway analysis (Figures 5–7) indicated that the ribosome pathway was significantly enriched in 181 selected genes. GO enrichment analysis showed that the structural constituent of the ribosome and poly(A) RNA binding were enriched in molecular function GO terms; cytosolic small ribosomal subunit, ribosome, membrane, nucleoplasm, respiratory chain complex IV, intracellular ribonucleoprotein complex,

nucleolus, and small ribosomal subunit were enriched in cellular compartments; and translational initiation, SRP-dependent cotranslational protein targeting to membrane, viral transcription, nuclear-transcribed mRNA catabolic process, nonsense-mediated decay, rRNA processing, and translation were enriched in biological process terms.

#### 4. Discussions

This study proposed a hybrid machine learning model to provide a precise psoriasis prediction model, using the gene expression profiles of human samples. The results of the

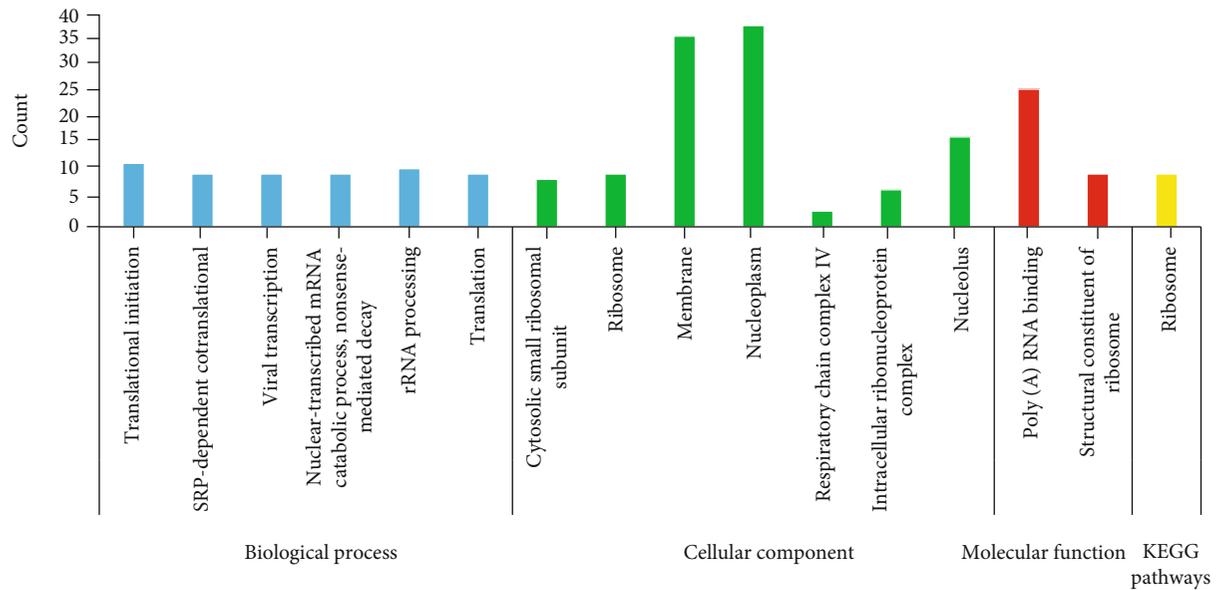


FIGURE 6: Gene Ontology and KEGG pathway enrichment analysis.

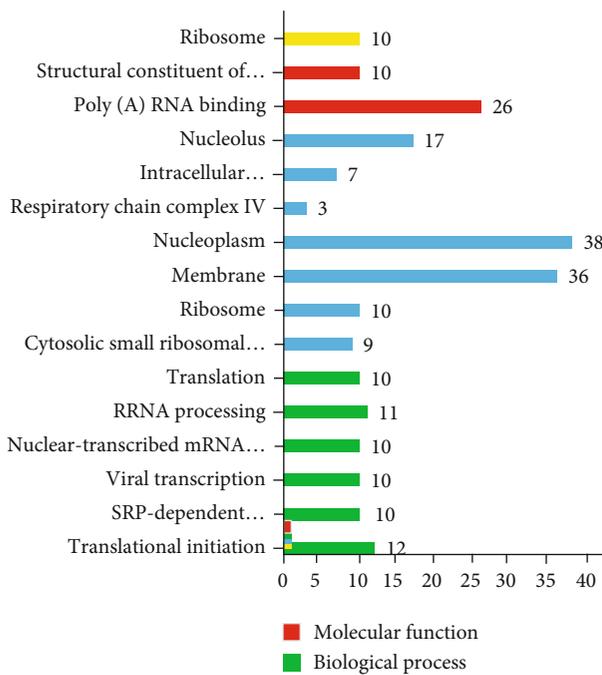


FIGURE 7: Molecular  $g$  = function and biological process.

present study showed the ability of the proposed model to discriminate psoriasis cases from normal controls. To evaluate the performance of the model, the predicted diagnosis (binary predicted response: psoriasis vs. control) returned by the proposed model during the validation stage (i.e., over test set) was compared against the true target value (i.e., known diagnosis of binary observed response: psoriasis vs. control). The best prediction model would return a high AUC and high prediction accuracy. The traditional SVM (without GA) was also trained and compared with the hybrid model. The finding of this study showed a

higher performance for the proposed hybrid prediction model and showed that GA has significantly improved the performance of the SVM classifier by achieving a total accuracy of 100%.

By comparing the results of the proposed GA+SVM of this study with those of Le et al. [37] who compared the performance of four classifiers of the random forest, naive Bayes,  $K$ -nearest neighborhood, and SVM for the classification of psoriasis and achieved a 98.3% total accuracy for the random forest classifier (using all 54,675 probe sets and a larger sample size of  $n = 180$  compared with the present study with  $n = 74$ ), a greater accuracy was obtained based on 181 probe sets using a much smaller sample size ( $n = 74$ ) indicating the need for noteworthy performance of the hybrid models. Moreover, feature selection based on the genetic algorithm reduces the dimension of the feature space, so the redundant, noisy, or irrelevant data are removed, the quality of the data and the accuracy of the resulting model improve due to searching from a population of points instead of a single point, and finally the selected feature set avoids wasting of resources in the next round of information collection or throughout utilization [38]. Moreover, in other studies, GA+SVM has been shown to outperform the traditional SVM in detecting other diseases or other outcomes [22, 24]. So it is suggested to use and evaluate the performance of other hybrid models. Apparently, the proposed technique illustrates competitive performance against the state-of-the-art models.

PPI network analysis identified a set of 9 hub genes (including RPS3, RPS5, RPS20, RPS15A, RPS3A, RPS4X, RPS7, RPL13, and RPL35) that is proposed to be associated with psoriasis or to be differentially expressed in psoriasis samples. The protein encoded by RPS3 has an important role in DNA repair through the cleavage of damaged DNA. Moreover, RPS3 enhances the inflammatory response through proteasomal degradation of  $I\kappa B\alpha$  [39, 40].

RPS5 encodes the nucleotide-binding protein which is activated proteolytically with PBS1. RPS5 regulates several biological processes such as proliferation and differentiation. Furthermore, RPS5 suppresses the inflammatory responses induced by lipopolysaccharide [41, 42]. The protein encoded by RPS20 plays a role in P53 activation through the Mdm2 binding and ribosomal biogenesis [43, 44]. RPS7 regulates the cell cycle and apoptosis via the MDM2-P53 interaction and through the regulation of MAPK and PI3K/AKT signaling pathways [45]. The protein encoded by PS4X is one of the components of the ribosomal small subunit complex. PS4X plays an essential role in biological processes such as proliferation and translation [46, 47]. The protein encoded by RPL13 plays an essential role in pre-rRNA processing and ribosome assembly [48]. RPL35 as an important component of the ribosomal large subunit plays an essential role in protein translation [49]. RPS15A has an essential role in mRNA binding to the small subunit of the ribosome and control of the cell cycle [50]. RPS15A activates the NF- $\kappa$ B signaling pathway by inducing I $\kappa$ B- $\alpha$  degradation [51]. RPS3A has an essential role in the regulation of translation, apoptosis, and differentiation. RPS3A also plays a role in NF- $\kappa$ B signaling pathway enhancement [52, 53]. In this study, KEGG pathway analysis indicated that the ribosome pathway was significantly enriched for selected genes. The structural constituent of the ribosome and poly(A) RNA binding were significantly enriched in molecular function GO terms. Translational initiation, SRP-dependent cotranslational protein targeting to membrane, nonsense-mediated decay, viral transcription, rRNA processing, translation, and nuclear-transcribed mRNA catabolic process were enriched in biological process GO terms.

A similar study showed that the ribosome pathway might be associated with inflammation [54]. Moreover, results of Li et al.'s study indicated that selected DEGs in psoriasis were enriched in the ribosomal pathway [55]. rRNA processing is required for 28S and 5.8S rRNA maturation and proper ribosome biogenesis and plays a role in innate immune signaling [56, 57]. Wu et al. in their study showed that in ankylosing spondylitis as an autoimmune disease, the structural constituent of the ribosome, SRP-dependent cotranslational protein targeting to membrane, nonsense-mediated decay, translation, viral transcription, nuclear-transcribed mRNA catabolic process, poly(A) RNA binding, and translational initiation term were enriched [58].

## 5. Conclusions

This study proposed a hybrid method of GA and SVM for the diagnosis of psoriasis. The proposed method was assessed using a real dataset and compared with the conventional SVM. The current study results revealed that the hybrid method outperformed the traditional SVM demonstrating the feasibility of identifying the best features using GA. Finally by considering the results of such hybrid method, assessing other heuristic approaches such as ant colony or particle swarm optimization is suggested for future studies.

## Data Availability

A publicly available dataset was analyzed in this study. This data can be found at <https://www.ncbi.nlm.nih.gov/geo/> (the NCBI Gene Expression Omnibus).

## Ethical Approval

This study was approved by the Ethics Committee of Hamadan University of Medical Sciences (approval code: IR.UM-SHA.REC.1398.637). All methods were performed in accordance with the relevant guidelines and regulations.

## Conflicts of Interest

The authors declare that there is no conflict of interests.

## Acknowledgments

The financial support for this study was provided by the Deputy of Research and Technology of Hamadan University of Medical Sciences (no. 9808286317).

## References

- [1] D. WHEELER and L. GRANDINETTI, "Psoriasis: evolving treatment for a complex disease," *Cleveland Clinic Journal of Medicine*, vol. 79, no. 6, p. 413, 2012.
- [2] T. Chu and T. Tsai, "Psoriasis and cardiovascular comorbidities with emphasis in Asia," *Giornale Italiano di Dermatologia e Venereologia*, vol. 147, no. 2, pp. 189–202, 2012.
- [3] A. Rendon and K. Schäkel, "Psoriasis pathogenesis and treatment," *International journal of molecular sciences*, vol. 20, no. 6, p. 1475, 2019.
- [4] J. M. Gelfand, E. D. Dommasch, D. B. Shin et al., "The risk of stroke in patients with psoriasis," *Journal of Investigative Dermatology*, vol. 129, no. 10, pp. 2411–2418, 2009.
- [5] L. Puig, "Cardiovascular risk and psoriasis: the role of biologic therapy," *Actas Dermo-Sifiliográficas (English Edition)*, vol. 103, no. 10, pp. 853–862, 2012.
- [6] G. Parra and E. Dauden, "Psoriasis and depression: the role of inflammation," *Actas Dermo-Sifiliográficas*, vol. 110, no. 1, pp. 12–19, 2019.
- [7] J. Koo, L. Marangell, M. Nakamura et al., "Depression and suicidality in psoriasis: review of the literature including the cytokine theory of depression," *Journal of the European Academy of Dermatology and Venereology*, vol. 31, no. 12, pp. 1999–2009, 2017.
- [8] N. Da Silva, C. von Stülpnagel, A. Langenbruch, A. Danckworth, M. Augustin, and R. Sommer, "Disease burden and patient needs and benefits in anogenital psoriasis: developmental specificities for person-centred healthcare of emerging adults and adults," *Journal of the European Academy of Dermatology and Venereology*, vol. 34, no. 5, pp. 1010–1018, 2020.
- [9] J. Krueger and A. Bowcock, "Psoriasis pathophysiology: current concepts of pathogenesis," *Annals of the rheumatic diseases*, vol. 64, supplement 2, pp. ii 30–iii6, 2005.
- [10] R. Langley, G. Krueger, and C. Griffiths, "Psoriasis: epidemiology, clinical features, and quality of life," *Annals of the rheumatic diseases*, vol. 64, supplement 2, pp. ii18–ii23, 2005.

- [11] N. P. Foundation, "About psoriasis," *NPF*, 2013.
- [12] F. Chamian, M. A. Lowes, S.-L. Lin et al., "Alefacept reduces infiltrating T cells, activated dendritic cells, and inflammatory genes in psoriasis vulgaris," *Proceedings of the National Academy of Sciences.*, vol. 102, no. 6, pp. 2075–2080, 2005.
- [13] F. Chamian and J. G. Krueger, "Psoriasis vulgaris: an interplay of T lymphocytes, dendritic cells, and inflammatory cytokines in pathogenesis," *Current opinion in Rheumatology.*, vol. 16, no. 4, pp. 331–337, 2004.
- [14] J. K. Kulski, W. Kenworthy, M. Bellgard et al., "Gene expression profiling of Japanese psoriatic skin reveals an increased activity in molecular stress and immune response signals," *Journal of molecular medicine.*, vol. 83, no. 12, pp. 964–975, 2005.
- [15] P. Guo, Y. Luo, G. Mai et al., "Gene expression profile based classification models of psoriasis," *Genomics*, vol. 103, no. 1, pp. 48–55, 2014.
- [16] J. E. Gudjonsson, J. Ding, X. Li et al., "Global gene expression analysis reveals evidence for decreased lipid biosynthesis and increased innate immunity in uninvolved psoriatic skin," *Journal of Investigative Dermatology.*, vol. 129, no. 12, pp. 2795–2804, 2009.
- [17] Y. Abo-zeid, N. S. M. Ismail, G. R. McLean, and N. M. Hamdy, "A molecular docking study repurposes FDA approved iron oxide nanoparticles to treat and control COVID-19 infection," *European Journal of Pharmaceutical Sciences.*, vol. 153, p. 105465, 2020.
- [18] A. M. Bowcock, W. Shannon, F. du et al., "Insights into psoriasis and other inflammatory diseases from large-scale gene expression studies," *Human molecular genetics.*, vol. 10, no. 17, pp. 1793–1805, 2001.
- [19] J. Reischl, S. Schwenke, J. M. Beekman, U. Mrowietz, S. Stürzebecher, and J. F. Heubach, "Increased expression of Wnt5a in psoriatic plaques," *Journal of Investigative Dermatology.*, vol. 127, no. 1, pp. 163–169, 2007.
- [20] W. R. Swindell, A. Johnston, S. Carbajal et al., "Genome-wide expression profiling of five mouse models identifies similarities and differences with human psoriasis," *PLoS One*, vol. 6, no. 4, article e18266, 2011.
- [21] for the Collaborative Association Study of Psoriasis, R. P. Nair, K. C. Duffin et al., "Genome-wide scan reveals association of psoriasis with IL-23 and NF- $\kappa$ B pathways," *Nature genetics.*, vol. 41, no. 2, pp. 199–204, 2009.
- [22] M. T. Tally and H. Amintoosi, "A hybrid method of genetic algorithm and support vector machine for intrusion detection," *International Journal of Electrical & Computer Engineering*, vol. 11, no. 1, pp. 2088–8708, 2021.
- [23] D. Zhang, J. Xiao, N. Zhou et al., "A genetic algorithm based support vector machine model for blood-brain barrier penetration prediction," *BioMed research international.*, vol. 2015, article 292683, pp. 1–13, 2015.
- [24] A. Alzubaidi, G. Cosma, D. Brown, and A. G. Pockley, "Breast cancer diagnosis using a hybrid genetic algorithm for feature selection based on mutual information," in *2016 International Conference on Interactive Technologies and Games (ITAG)*, Nottingham, UK, 2016.
- [25] S.-H. Min, J. Lee, and I. Han, "Hybrid genetic algorithms and support vector machines for bankruptcy prediction," *Expert systems with applications.*, vol. 31, no. 3, pp. 652–660, 2006.
- [26] C. Q. Wang, M. Suárez-Fariñas, K. E. Nogales et al., "IL-17 induces inflammation-associated gene products in blood monocytes, and treatment with ixekizumab reduces their expression in psoriasis patient blood," *The Journal of investigative dermatology.*, vol. 134, no. 12, pp. 2990–2993, 2014.
- [27] G. K. Smyth, "Limma: linear models for microarray data," in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, pp. 397–420, Springer, 2005.
- [28] R. C. Team, *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013, <http://www.R-project.org/>.
- [29] L. Auria and R. A. Moro, *Support vector machines (SVM) as a technique for solvency analysis*, 2008.
- [30] V. Jakkula, *Tutorial on support vector machine (SVM)*, School of EECS, Washington State University., 2006.
- [31] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Machine Learning*, vol. 3, no. 2/3, pp. 95–99, 1988.
- [32] C. R. Houck, J. Joines, and M. G. Kay, "A genetic algorithm for function optimization: a Matlab implementation," *Ncsu-ie tr.*, vol. 95, no. 9, pp. 1–10, 1995.
- [33] F. Temitayo, O. Stephen, and A. Abimbola, "Hybrid GA-SVM for efficient feature selection in e-mail classification," *Computer Engineering and Intelligent Systems.*, vol. 3, no. 3, pp. 17–28, 2012.
- [34] V. Matlab, *7.10.0 (R2010a)*, The MathWorks Inc, Natick, Massachusetts, 2010.
- [35] D. Szklarczyk, A. L. Gable, D. Lyon et al., "STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic acids research.*, vol. 47, no. D1, pp. D607–D613, 2019.
- [36] P. Shannon, A. Markiel, O. Ozier et al., "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [37] N. Q. K. Le, D. T. Do, N. T. K. Nguyen, T. N. K. Hung, and N. T. T. Trang, "Identification of gene expression signatures for psoriasis classification using machine learning techniques," *Medicine in Omics.*, vol. 1, p. 100001, 2020.
- [38] T. K. Mansoori, A. Suman, and S. K. Mishra, "Feature selection by genetic algorithm and SVM classification for cancer detection," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 9, 2014.
- [39] H. W. Yang, H. D. Kim, and J. Kim, "The DNA repair domain of human rpS3 protects against photoaging by removing cyclobutane pyrimidine dimers," *FEBS letters.*, vol. 593, no. 15, pp. 2060–2068, 2019.
- [40] Y. Li, B. Huang, H. Yang et al., "Latexin deficiency in mice up-regulates inflammation and aggravates colitis through HECTD1/Rps3/NF- $\kappa$ B pathway," *Scientific Reports*, vol. 10, no. 1, p. 9868, 2020.
- [41] D. Qi, B. J. DeYoung, and R. W. Innes, "Structure-function analysis of the coiled-coil and leucine-rich repeat domains of the RPS5 disease resistance protein," *Plant Physiology*, vol. 158, no. 4, pp. 1819–1832, 2012.
- [42] J. Xu, K.-Q. Wang, W.-H. Xu et al., "The matrine derivate MASM prolongs survival, attenuates inflammation, and reduces organ injury in murine established lethal sepsis," *The Journal of Infectious Diseases.*, vol. 214, no. 11, pp. 1762–1772, 2016.
- [43] T. T. Nieminen, M. F. O'Donohue, Y. Wu et al., "Germline Mutation of *\_RPS20\_*, Encoding a Ribosomal Protein, Causes Predisposition to Hereditary Nonpolyposis Colorectal

- Carcinoma Without DNA Mismatch Repair Deficiency,” *Gastroenterology*, vol. 147, no. 3, pp. 595–598.e5, 2014.
- [44] L. Daftuar, Y. Zhu, X. Jacq, and C. Prives, “Ribosomal proteins RPL37, RPS15 and RPS20 regulate the Mdm2-p53-MdmX network,” *PLoS one*, vol. 8, no. 7, article e68667, 2013.
- [45] W. Zhang, D. Tong, F. Liu et al., “RPS7 inhibits colorectal cancer growth via decreasing HIF-1 $\alpha$ -mediated glycolysis,” *Oncotarget*, vol. 7, no. 5, pp. 5800–5814, 2016.
- [46] S. P. Tsoufack, L. Meunier, L. Sanchez et al., “Low expression of the X-linked ribosomal protein S4 in human serous epithelial ovarian cancer is associated with a poor prognosis,” *BMC Cancer*, vol. 13, no. 1, 2013.
- [47] J. Kuang, Q. Y. Li, F. Fan et al., “Overexpression of the X-linked ribosomal protein S4 predicts poor prognosis in patients with intrahepatic cholangiocarcinoma,” *Oncology Letters*, vol. 14, no. 1, pp. 41–46, 2017.
- [48] C. le Caignec, B. Ory, F. Lamoureux et al., “\_RPL13\_ Variants Cause Spondyloepimetaphyseal Dysplasia with Severe Short Stature,” *American journal of human genetics.*, vol. 105, no. 5, pp. 1040–1047, 2019.
- [49] N. Jiang, L. Hu, C. Liu, X. Gao, and S. Zheng, “60S ribosomal protein L35 regulates  $\beta$ -casein translational elongation and secretion in bovine mammary epithelial cells,” *Archives of biochemistry and biophysics.*, vol. 583, pp. 130–139, 2015.
- [50] W. Feng, C. Liang, C. Wang, X. Yu, Q. Li, and H. Yang, “Knockdown of ribosomal protein S15A inhibits proliferation of breast cancer cells through induction of apoptosis in vitro,” *Cytotechnology*, vol. 70, no. 5, pp. 1315–1323, 2018.
- [51] C. Liu, X. He, X. Liu et al., “RPS15A promotes gastric cancer progression via activation of the Akt/IKK- $\beta$ /NF- $\kappa$ B signalling pathway,” *Journal of Cellular and Molecular Medicine*, vol. 23, no. 3, pp. 2207–2218, 2019.
- [52] K.-H. Lim, K.-H. Kim, S. I. Choi et al., “RPS3a over-expressed in HBV-associated hepatocellular carcinoma enhances the HBx-induced NF- $\kappa$ B signaling via its novel chaperoning function,” *PLoS one*, vol. 6, no. 8, article e22258, 2011.
- [53] C. Zhou, J. Weng, C. Liu et al., “High RPS3A expression correlates with low tumor immune cell infiltration and unfavorable prognosis in hepatocellular carcinoma patients,” *American Journal of Cancer Research*, vol. 10, no. 9, pp. 2768–2784, 2020.
- [54] S. J. Gan, B. Ye, S. X. Qian et al., “Immune- and ribosome-related genes were associated with systemic vasculitis,” *Scandinavian Journal of Immunology.*, vol. 81, no. 2, pp. 96–101, 2015.
- [55] Y. Li, P. Lin, S. Wang et al., “Quantitative analysis of differentially expressed proteins in psoriasis vulgaris using tandem mass tags and parallel reaction mass monitoring,” *Clinical Proteomics*, vol. 17, no. 1, 2020.
- [56] C. Bianco and I. Mohr, “Ribosome biogenesis restricts innate immune responses to virus infection and DNA,” *eLife*, vol. 8, 2019.
- [57] C. F. Spurlock, J. T. Tossberg, Y. Guo, S. Sriram, P. S. Crooke, and T. M. Aune, “Defective structural RNA processing in relapsing-remitting multiple sclerosis,” *Genome Biology.*, vol. 16, no. 1, p. 58, 2015.
- [58] B. Wu, J. Yu, Y. Liu et al., *Potential pathogenic genes and mechanism of ankylosing spondylitis: a study based on WGCNA and bioinformatics analysis*, Research Square, 2021.