



# Multiple Machine Learning Methods Reveal Key Biomarkers of Obstructive Sleep Apnea and Continuous Positive Airway Pressure Treatment

Jie Zhu<sup>1</sup>, Larry D. Sanford<sup>2</sup>, Rong Ren<sup>1</sup>, Ye Zhang<sup>1</sup> and Xiangdong Tang<sup>1\*</sup>

<sup>1</sup>Sleep Medicine Center, Department of Respiratory and Critical Care Medicine, Mental Health Center, West China Hospital, Sichuan University, Chengdu, China, <sup>2</sup>Sleep Research Laboratory, Center for Integrative Neuroscience and Inflammatory Diseases, Pathology and Anatomy, Eastern Virginia Medical School, Norfolk, VA, United States

## OPEN ACCESS

### Edited by:

Silong Peng,  
Institute of Automation (CAS), China

### Reviewed by:

Jingpeng Sun,  
Chinese Academy of Sciences (CAS),  
China

Shekh Md Mahmudul Islam,  
University of Dhaka, Bangladesh

### \*Correspondence:

Xiangdong Tang  
2372564613@qq.com

### Specialty section:

This article was submitted to  
Genetics of Common and Rare  
Diseases,

a section of the journal  
Frontiers in Genetics

**Received:** 24 April 2022

**Accepted:** 24 June 2022

**Published:** 13 July 2022

### Citation:

Zhu J, Sanford LD, Ren R, Zhang Y and  
Tang X (2022) Multiple Machine  
Learning Methods Reveal Key  
Biomarkers of Obstructive Sleep  
Apnea and Continuous Positive Airway  
Pressure Treatment.  
Front. Genet. 13:927545.  
doi: 10.3389/fgene.2022.927545

Obstructive sleep apnea (OSA) is a worldwide health issue that affects more than 400 million people. Given the limitations inherent in the current conventional diagnosis of OSA based on symptoms report, novel diagnostic approaches are required to complement existing techniques. Recent advances in gene sequencing technology have made it possible to identify a greater number of genes linked to OSA. We identified key genes in OSA and CPAP treatment by screening differentially expressed genes (DEGs) using the Gene Expression Omnibus (GEO) database and employing machine learning algorithms. None of these genes had previously been implicated in OSA. Moreover, a new diagnostic model of OSA was developed, and its diagnostic accuracy was verified in independent datasets. By performing Single Sample Gene Set Enrichment Analysis (ssGSEA) and Counting Relative Subsets of RNA Transcripts (CIBERSORT), we identified possible immunologic mechanisms, which led us to conclude that patients with high OSA risk tend to have elevated inflammation levels that can be brought down by CPAP treatment.

**Keywords:** obstructive sleep apnea, machine learning, continuous positive airway pressure, bioinformatic analysis, random forest (bagging) and machine learning, artificial neural network, SVM-support vector machine

## INTRODUCTION

The prevalence of obstructive sleep apnea (OSA) is estimated at one billion people worldwide, including over 400 million who have moderate-to-severe symptoms (Benjafield et al., 2019). The main characteristic of OSA is excessive sleepiness due to collapsed upper airways during sleep, resulting in oxygen desaturations, heart rate changes, neurological arousal, and therefore disturbed sleep (Young et al., 2004). In the absence of proper treatment, OSA contributes to a higher mortality rate from cardio- or cerebro-vascular events (Lévy et al., 2015). A variety of treatments have been developed to correct the narrowing of the upper airway in OSA patients. Continuous positive airway pressure (CPAP) treatment is the most extensively studied and proven therapy.

There has long been an indication that family history is a powerful risk factor for OSA. According to the Cleveland Family Study (Grilo et al., 2013), approximately one-third of the variance in the apnea-hypopnea index (AHI) can be explained by genetic factors shared across families. In addition, epidemiological evidence suggests a strong association between OSA susceptibility and genetic polymorphisms (Sun et al., 2015). Genes associated with cardiovascular consequences can be

hypermethylated by hypoxia (Stenvinkel et al., 2007; Watson et al., 2010). Forkhead Box P3 (FOXP3), an immune-related gene, activates regulatory T cells and prevents atherosclerosis by modulating lipoprotein metabolism (Klingenberg et al., 2013). An increase in methylation of FOXP3 promoter has been linked to systemic inflammation among children with OSA. A strong correlation between DNA methylation levels and total C reactive protein (CRP) levels has been observed in OSA, suggesting a possible underlying mechanism (Kim et al., 2012). In fact, extensive inflammation is believed to be a major contributing factor to OSA. Multiple inflammatory biomarkers such as interleukin-6 (IL-6), tumor necrosis factor (TNF), CRP, and von Willebrand factor (VWF) antigen have been observed independently and consistently associated with OSA (De Luca Canto et al., 2015; Nowakowski et al., 2018; Hirsch et al., 2019). Nevertheless, the specific mechanisms that regulate this broadly activated inflammatory background remain unclear. Due to the recent emergence of next-generation sequencing, high-throughput techniques have enabled examining expression profiles for thousands of genes at a time. This has enabled identifying marker genes related to a wide range of diseases and has facilitated effective disease diagnosis and treatment.

As a result of excellent capabilities in handling large and complex datasets, artificial intelligence (AI) systems, which use multiple machine learning methods, have gained widespread popularity in evaluating genetic profiles. The recursive feature elimination (RFE) approach has demonstrated its effectiveness in selecting informative variables for disease classification. In order to aid in the identification of the least useful features to remove from consideration, a Support Vector Machine (SVM) classification model can be used to assign weights to features. Subsequently, each time the RFE procedure is executed, the least important feature, that having the smallest weight, can be eliminated thereby reducing the number of parameters and potentially increasing accuracy (Ding and Wilkins, 2006). The random forest (RF) algorithm is a method of reducing dimensions based on the creation of thousands of decision trees (Zheng et al., 2021). A random assignment of variables into testing and training groups is made first, followed by 10,000 iterations until the lowest error rate is achieved, and the optimal variable number and the optimal number of trees are determined. In addition, an RF model with variable importance values can be created (Bi et al., 2020). An artificial neural network (ANN), a method for supervised learning, consists of an interconnected group of artificial neurons arranged in layers (Alizadeh Savareh et al., 2020). An ANN is developed by changing weights of connection during the training phase to improve network performance.

Our primary aim in this study was to develop a novel prediction tool for OSA risk and CPAP treatment utilizing these three machine learning algorithms (SVM-RFE, RF, and ANN). We utilized transcriptome microarray data obtained from a public database, Gene Expression Omnibus (GEO). To establish machine learning models and discover key biomarkers associated with OSA and its therapeutic response to CPAP treatment, we integrated five independent microarray datasets related to moderate-severe OSA patients and their CPAP treatment. The

results were also revalidated in another two separate datasets. Additionally, we assessed possible immunological mechanisms of OSA using multiple gene sets and enrichment analysis. As part of this research, we hoped to identify key genes that are implicated in OSA pathogenesis in addition to determining whether OSA is accompanied by dysfunctional innate immunity.

## MATERIALS AND METHODS

### Data Collecting and Downloading

Patients with OSA, controls, and patients who had undergone CPAP treatment were included in this study. Genome profiles were derived from the GEO (Gene Expression Omnibus) database, which provides array- and sequence-based data. As the study utilized a public database, no approval from an institutional review board was required. A total of five microarray datasets were obtained. GSE133601 included 15 patients with moderate to severe OSA and who adhered to CPAP therapy over 3 months. Peripheral blood mononuclear cells were collected before and after CPAP treatment. GSE75097 involved 42 treatment-naïve subjects and patients with moderate to severe OSA that had received at least 1 year of adequate CPAP treatment. Peripheral blood mononuclear cells were collected. GSE71356 contained eight controls whose whole blood was collected. GSE61463 consisted of 16 OSA patients and five controls, and peripheral blood mononuclear cell samples were analyzed. GSE49800 was comprised of 18 subjects with severe OSA who had undergone CPAP therapy. Transcriptional profiles of peripheral blood leukocytes were assessed. GSE38792 and GSE135917 were used as independent validation sets. They provided information on subcutaneous and visceral adipose tissue transcription of OSA patients (including those who took CPAP treatment) and of controls.

### Data Processing and Batch Effect Control

R software (version 3.6.2) was used for statistical analyses. If multiple microarray probes were mapped to a single gene, its mean expression level was used in the analysis. The gene expression values were log<sub>2</sub>-transformed before normalization. All genes and samples were checked to ensure that no missing values were contained in the dataset. Quantile normalization was used to standardize data. The SVA package was used to conduct batch effect processing to remove differences caused by different platforms and technologies. Principal component analysis (PCA) was performed to verify whether or not the batch effect was eliminated. Genetic comparisons were performed between OSA patients and controls or between treatment-naïve and CPAP therapy groups. Differential expression analysis was performed using the stringr and limma packages in R software. The fold change was derived from the average expression values. A log<sub>2</sub>FoldChang (log<sub>2</sub>FC) greater than 1.5 and *p*-value < 0.05 were set to identify the differentially expressed genes (DEGs). Continuous variables were compared between groups using equal variance. The significance threshold for the *p*-value was also set to <0.05.

## Gene Set Enrichment Analysis

GSEA analyses of RNA-seq profiles revealed DEGs-related signaling pathways in OSA patients and those who had undergone CPAP treatment. Screening of the enriched set was based on FDR (False Discovery Rate)  $< 0.25$  and a  $p < 0.05$  after 1,000 permutations. In GSEA, gene expression profiles from patient samples and controls were analyzed according to specific datasets (Subramanian et al., 2005). The GSEA website and the MsigDB database were used to obtain the c2. cp.kegg.v7.4. symbols.gmt dataset and c5. go.v7.4. symbols.gmt dataset for enrichment analyses, presenting Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analyses for biologic pathways and function annotations. Statistically significantly enriched gene sets were defined as those with a minimum number of samples per group of 5.

## Cell-Type Identification by Estimating Relative Subsets of RNA Transcript Analyses for Infiltrating Immune Cells

CIBERSORT is a computational technique for quantifying cell fractions based upon bulk tissue gene expression profiles, which can distinguish 22 human hematopoietic cell phenotypes (Newman et al., 2015). It is widely applied in various diseases to accurately estimate underlying immunocellular landscapes. The CIBERSORT gene matrix contains 547 genes and distinguishes 7 T cell subsets, naïve and memory B cells, plasma cells, NK cells, and myeloid subsets (Wang et al., 2020). Heatmap analysis and correlation analysis of multiple immune cells in OSA patients and controls, or in those receiving CPAP treatment, were conducted using the pheatmap and corrplot packages, respectively.

## Single Sample Gene Set Enrichment Analysis Algorithm

ssGSEA identifies gene sets by their common biological functions, spatial localization, and physiological significance, and then it calculates separate enrichment scores for each pairing of a sample and gene set. Gene sets consisting of 782 genes are used for predicting the abundance of 28 types of immune cells and functions in individual tissue samples. The immune cells include activated dendritic cells (aDCs), B cells, CD8<sup>+</sup> T cells, natural killer cells (NK cells), dendritic cells (DCs), neutrophils, macrophages, mast cells, plasmacytoid dendritic cells (pDCs), immature dendritic cells (iDCs), follicular helper T cells (Tfh), T helper cells, type-1 T helper cells (Th1), and type-2 T helper cells (Th2), tumor infiltrating lymphocytes (TILs), macrophages, regulatory T cells (Tregs). The immune-related functions consist of antigen presenting cell (APC) co-inhibition, APC co-stimulation, chemokine receptor (CCR), checkpoint, cytolytic activity, human leukocyte antigen (HLA), inflammation-promoting, MHC class I, para-inflammation, T cell co-inhibition, T cell co-stimulation, Type I IFN response, and Type II IFN response.

## Support Vector Machine Recursive Feature Elimination and Random Forest Algorithm

DEGs between OSA patients and controls or between treatment-naïve and CPAP therapy groups were treated as variables in machine learning procedures. In the first step, SVM-RFE screening was performed for candidate genes. The SVM-RFE model employs a backward selection approach by which variables can be identified based on their weights on the model. A first ranking criterion is calculated using the SVM weights, then the features with the smallest ranking criteria are eliminated. The process is then repeated until the highest accuracy of classification is achieved.

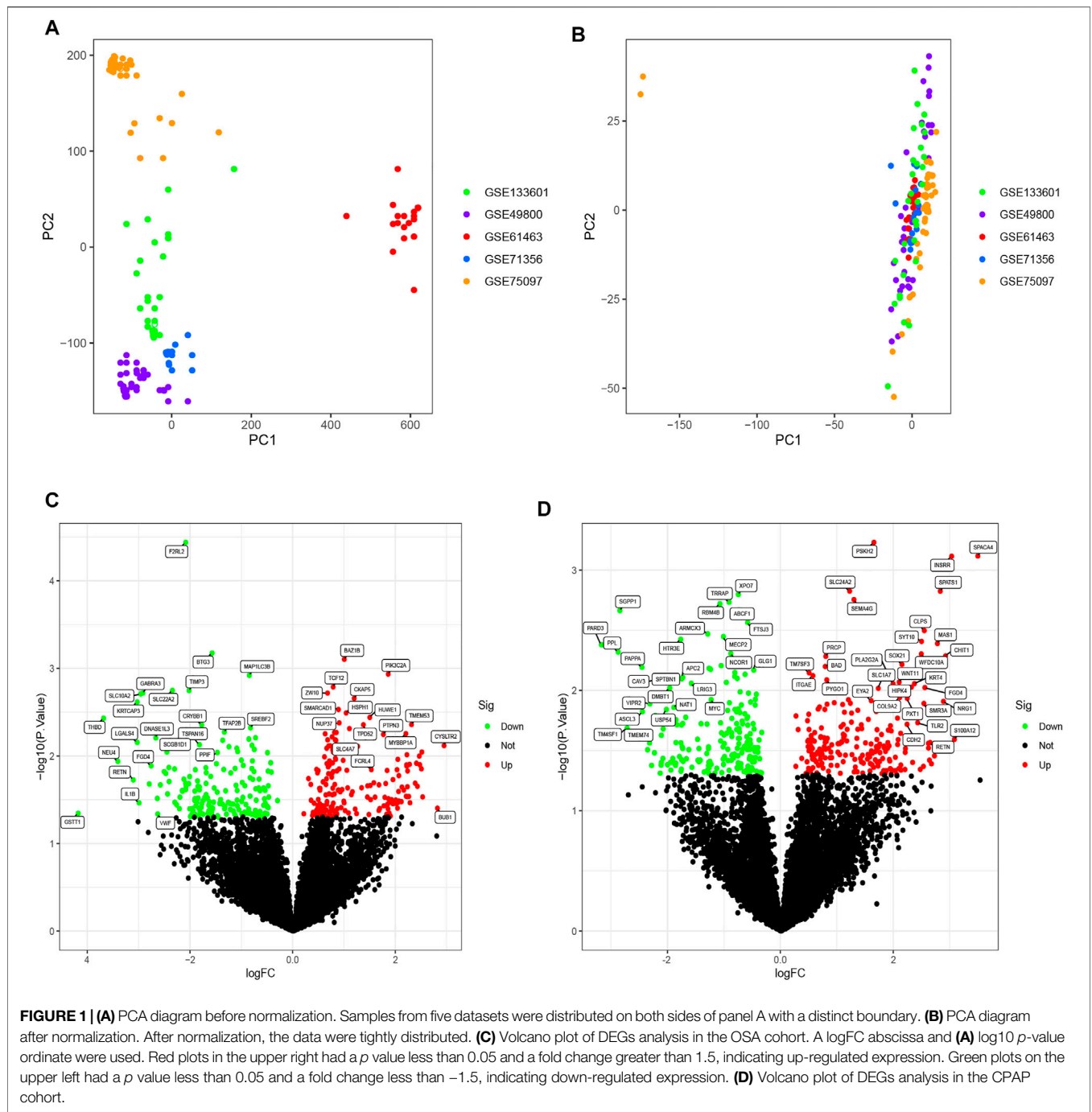
Using the e1071 package and the svmRFE function in R software, we eliminated the recursive features of DEGs. All genes were sorted by their SVM weights in the linear SVM model, and those with low weights were eliminated. For the purpose of avoiding overfitting, a 15-fold cross-validation approach was used to increase the number of estimates. Fifteen subsamples from the original sample were randomly distributed. The model was tested using one subsample while the other subsamples were used as training data. After the 15-fold cross-validation was completed, the loop function provided an estimation of generalization accuracy and error rate. The best list of variables was determined based on the highest accuracy rate and lowest error rate. The Root Mean Square Error (RMSE) is defined as the standard deviation of the prediction errors, which measures the difference between the observed value and the actual value. The RMSEs were calculated from the 15-fold CV to verify the results of SVM-RFE.

A second machine learning algorithm, RF, was then applied to the candidate genes obtained from the SVM-RFE algorithm. It utilized a large number of decision trees and combined the bootstrap aggregation method to select features at random (Deo, 2015). Training data was collected for each tree through repeated subsampling (bootstrapping) (Hanko et al., 2021). Bootstrap subsamples excluding approximately 33% of the data provide an out of bag (OOB) sample (Hanko et al., 2021). The error rate was minimized by minimizing the number of decision trees included in the model, preventing overfitting. Internal validation of the RF was estimated using the OOB sample. The decreasing accuracy method was used to obtain the dimensional importance value based on the RF model. The genes with the highest importance values were selected for inclusion as variables in establishing an ANN.

## Construction of Gene Signature by Artificial Neural Network

R software neuralnet and NeuralNetTools packages were used to build an ANN model of the candidate genes screened by SVM-RFE and RF (Beck, 2018). Typically, an ANN, another form of supervised learning algorithm, includes an interconnected set of artificial neurons in the form of intermittent layers.

There are typically three types of layers in neural networks: hidden, input, and output. Neurons are the basic components of computation, also known as nodes or units. A node receives



inputs from other nodes or from outside and produces output after completing calculations. Each connection between two nodes represents a weighted value ( $W$ ) for the signal passing through the connection. Each node applies a function  $f$  to the weighted sum of the inputs. The function  $f$  is nonlinear and is known as the activation function. In order to achieve the purpose of nonlinear representation of neurons and meet the data requirements of the actual environment, it introduces nonlinearity into the output of neurons. The function includes the inputs ( $x_1, x_2, \dots$ ), the weights ( $w_1,$

$w_2, \dots$ ) associated with the inputs, and input weights ( $b$ ): Output of neuron =  $f(w_1.x_1 + w_1.x_2 + \dots + b)$ . In each neuron, inputs are received at the previous layer, then output is sent to the next layer, and so forth until the output layer is reached. There is only one input layer and one output layer in a neural network, but there can be multiple hidden layers. Layers and neurons are not allocated according to a fixed rule. According to a broad consensus, one hidden layer can approximate any function that involves a continuous mapping from one finite space to another (Heaton, 2008).

There are very few cases where a second or more hidden layer improves performance, and most of the time, one hidden layer suffices. The number of neurons in hidden layers are decided by input and output dimensionality. In cases where the input-output relationship is fairly simple, ideally, the optimal dimensionality for a hidden layer should be in the middle of the input layer and output layer (Heaton, 2008; Beck, 2018). In our ANN model, one hidden layer and five neurons were established for the classification model of OSA and CPAP treatment based on the principles outlined above.

A min-max normalization method was used to preprocess the data before training the neural network. Classification scores were calculated by multiplying the weight scores by the expression levels of the important genes. In a 5-fold cross-validation method, a training set and a verification set were randomly selected from the dataset. The training set served as the basis for determining the weights of candidate genes, while the verification set served as the basis for assessing the efficiency of classification. The R software pROC package was employed to assess classification accuracy.

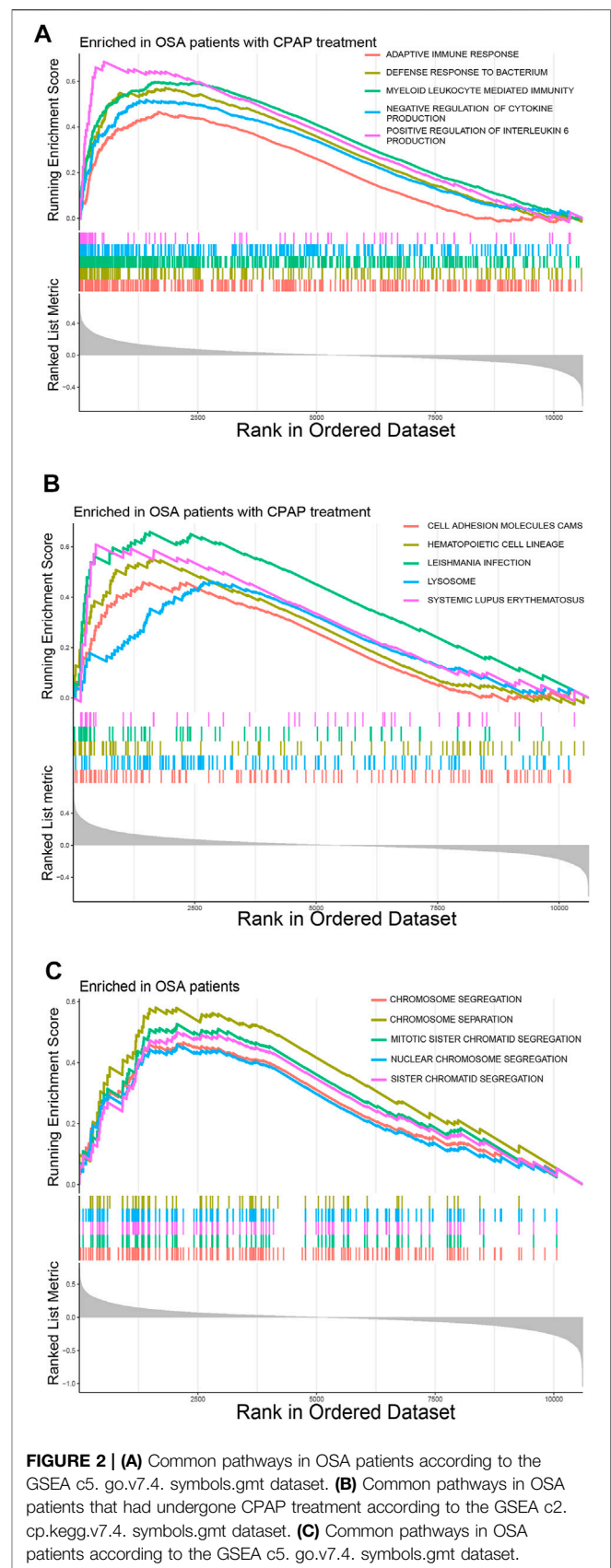
## RESULTS

### The Removal of Batch Effect Through Cross-Platform Normalization

The R software ComBat function was used to eliminate batch effects due to non-ignorable technical differences across experiments, platforms, or studies. A total of 10,613 genes were detected in datasets from five different microarray platforms. Unnormalized and normalized PCA plots are shown in **Figures 1A,B**, respectively. Scatter plots illustrate the top two principal components (PCs) of expressed values. Unnormalized data plots indicate that the samples were loosely clustered and have distinct boundaries. As the samples clustered more tightly after normalization, they were more similar across datasets.

### Differentially Expressed Genes Analysis

Differential expression analysis was performed on two cohorts, between OSA patients and controls and between treatment-naive and CPAP therapy groups. The OSA cohort consisted of 77 OSA patients and 19 controls. The CPAP cohort comprised 47 individuals undergoing CPAP therapy and 61 treatment-naive OSA patients. Further, using the R software limma package, DEGs were identified between OSA and controls or between CPAP and treatment-naive samples. The results of the DEGs are presented in a volcano plot (**Figures 1C,D**). Based on fold changes  $>1.5$  and significance thresholds of  $p < 0.05$ , 360 significant DEGs linked to OSA and 393 significant DEGs linked to CPAP were identified. A cross-comparison of two cohorts of DEGs identified 37 intersection genes: BAZ1B, MAP1LC3B, TCF12, HUWE1, TPD52, CRYBB1, PTPN3, BAD, CAND1, TXLNA, BHLHB9, GRPEL1, FGD4, REV3L, EXOSC10, SMAD4, TBX3, RETN, PPL, MGAT5, GLT1D1, SLC44A5, FAAH, FLT3, DOCK9, MGP, EPN3, TMEM121,



**FIGURE 2 | (A)** Common pathways in OSA patients according to the GSEA c5. go.v7.4. symbols.gmt dataset. **(B)** Common pathways in OSA patients that had undergone CPAP treatment according to the GSEA c2. cp.kegg.v7.4. symbols.gmt dataset. **(C)** Common pathways in OSA patients according to the GSEA c5. go.v7.4. symbols.gmt dataset.

ZNF214, CLEC10A, FKBP4, EYA2, MRO, TFF2, ABCF1, MOAP1, DNMT1.

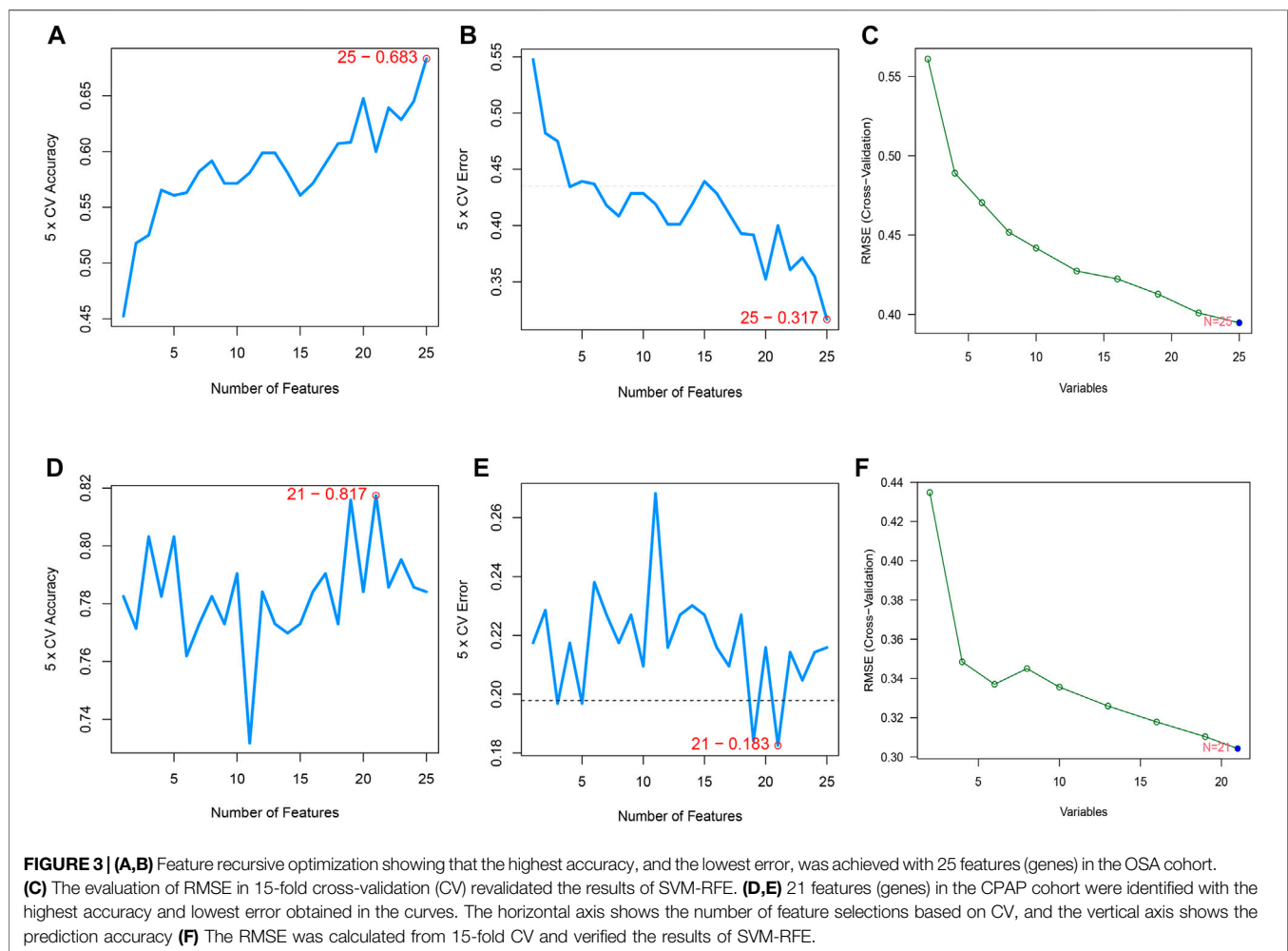
## GSEA of DEGs

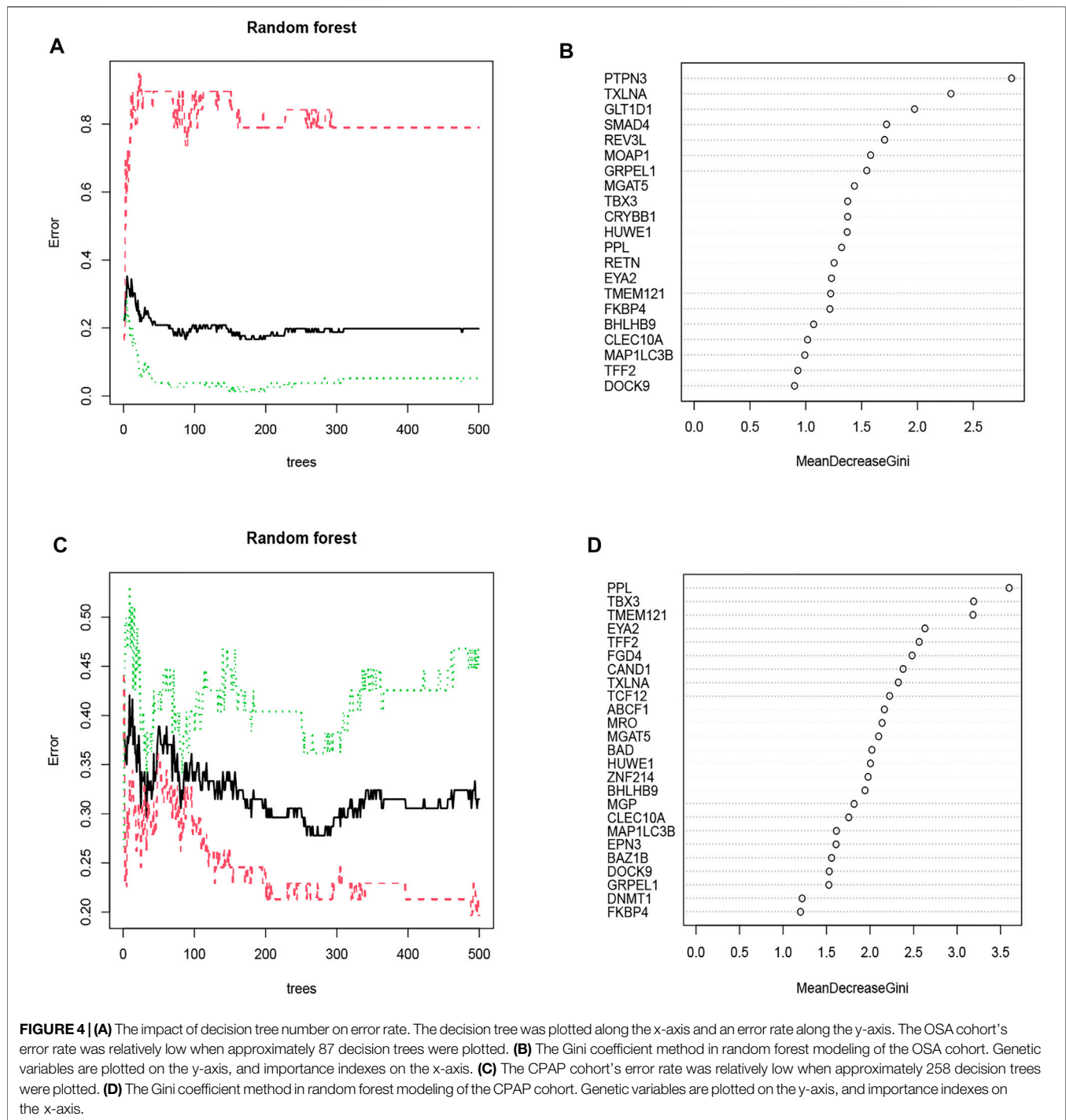
Some immune-related pathways were included in the GSEA results. DEGs of OSA patients who had undergone CPAP treatment were enriched in the adaptive immune response, defense response to bacterium, myeloid leukocyte mediated immunity, and negative regulation of cytokine production per the c5. go.v7.4. symbols.gmt dataset (**Figure 2A**). Several cellular structure-related pathways were also present in the results in OSA patients that had undergone CPAP treatment including cell adhesion molecules, hematopoietic cell lineages, and lysosomes per the c2. cp.kegg.v7.4. symbols.gmt dataset (**Figure 2B**). Some cell cycle-related activities were also observed in OSA patients, including chromosome segregation, mitotic sister chromatid segregation, nuclear chromosome segregation per the c5. go.v7.4. symbols.gmt dataset (**Figure 2C**).

## Selection of Candidate Genes and Construction of Predictive Signatures Using Multiple Machine Learning Algorithms Across Cohorts

The SVM-RFE algorithm, which searches for genes with the smallest classification error, and RF, which detects genes with the highest importance, resulted in the selection of candidate genes. When the accuracy of the SVM-RFE algorithm was highest, and the estimation error was the lowest, 25 genes were identified in the CPAP cohort (**Figures 3A,B**) and 21 genes were identified in the OSA cohort (**Figures 3D,E**). We then input these genes into the RF classifier. By evaluating the RMSE, the best models were also determined to have a better balance of prediction errors (**Figures 3C,F**).

Each possible number of variables was analyzed with a recurrent RF classification to determine the average error rate. The error rate was relatively small when the number of decision trees was approximately 87 in the OSA cohort and 258 in the CPAP cohort (**Figures 4A,C**). Next, an RF model was built, and the Gini coefficient method was used to calculate the dimensional



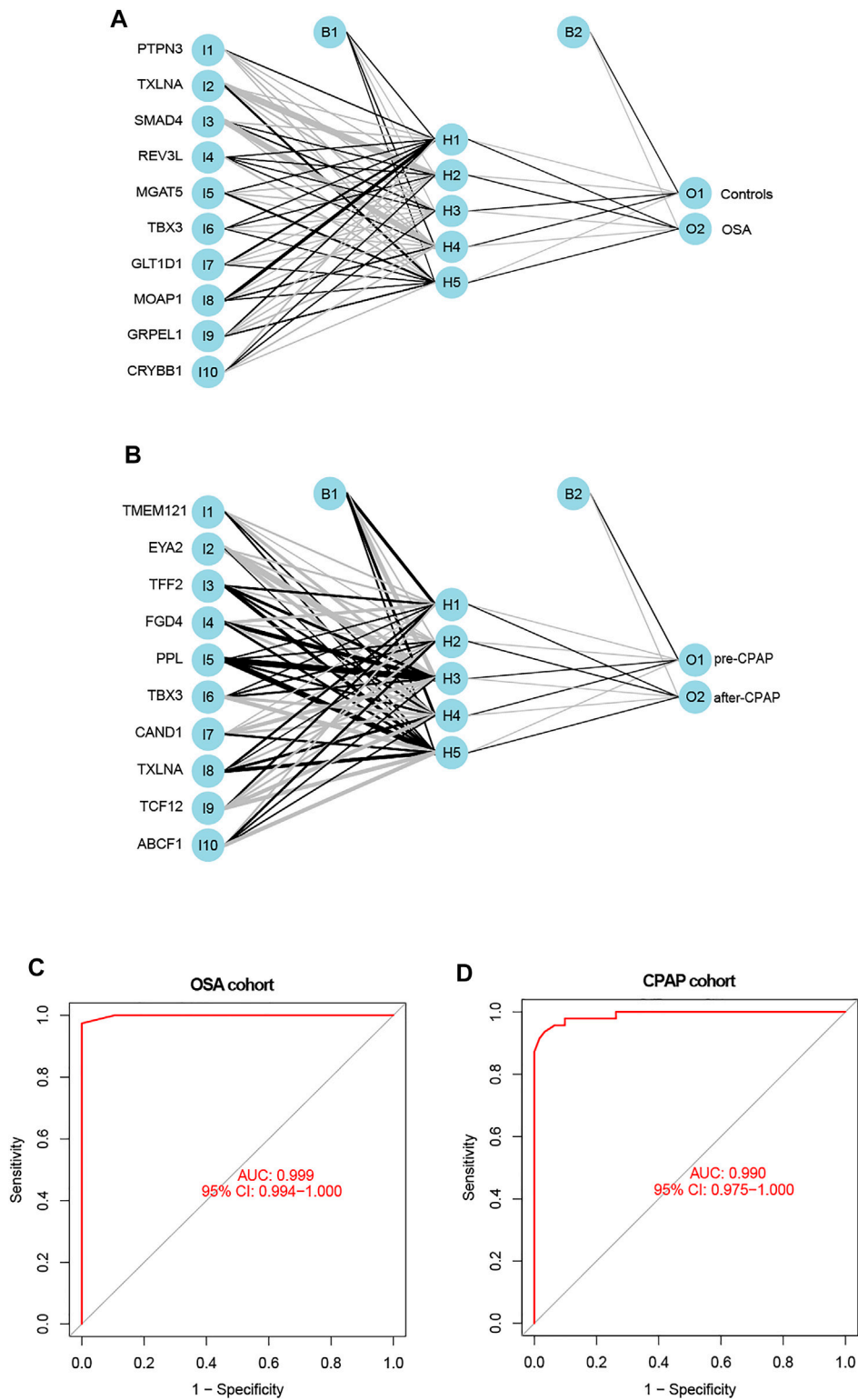


importance value. We chose the top ten genes with the greatest importance value as variables in each cohort's subsequent construction of an ANN. In the OSA cohort, the top 10 genes were: PTPN3, TXLNA, GLT1D1, SMAD4, REV3L, MOAP1, GRPEL1, MGAT5, TBX3, and CRYBB1 (Figure 4B). In the CPAP cohort, the top 10 genes were: PPL, TBX3, TMEM121, EYA2, TFF2, FGD4, CAND1, TXLNA, TCF12, and ABCF1 (Figure 4D).

## Creating a Model of an Artificial Neural Network

The following formula was constructed to calculate the classification score for the ANN model: neural score =  $\sum(\text{Gene Expression} \times \text{Neural Network Weight})$ .

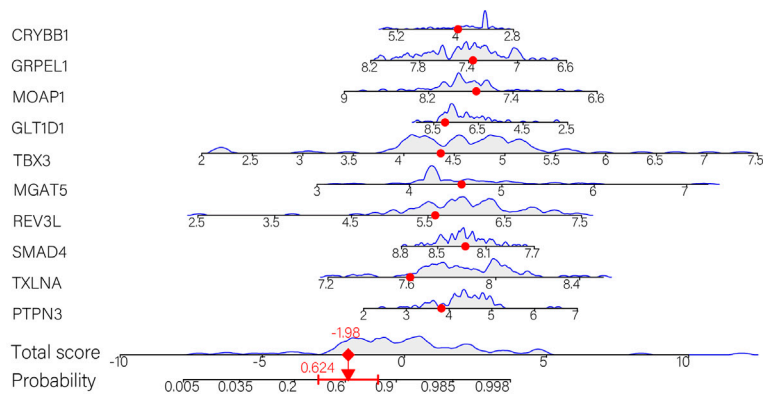
The weight predictions for the OSA cohort were 2.84 (PTPN3), -2.78 (TXLNA), -6.57 (SMAD4), 2.00 (REV3L), 1.41



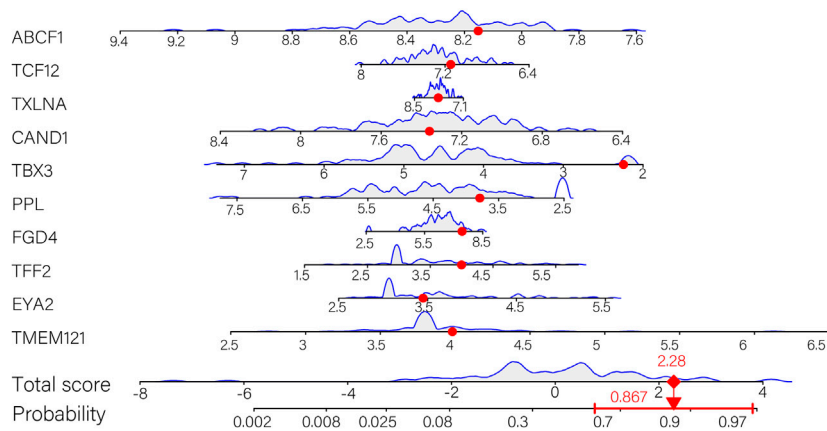
**FIGURE 5 | (A,B)** The constructions of an ANN for the OSA cohort and the CPAP cohort were comprised of one input layer, one hidden layer, and one output layer. **(C)** ROC curves of the ANN-based OSA diagnostic model. **(D)** ROC curves of the ANN-based CPAP treatment model.



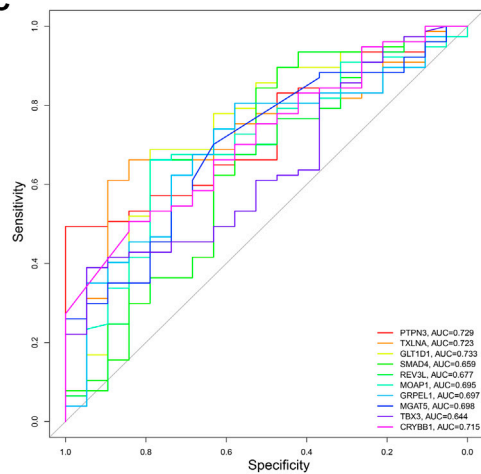
**A** Nomogram of OSA cohort



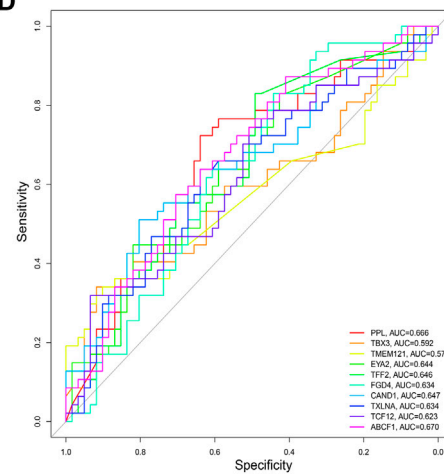
**B** Nomogram of CPAP cohort



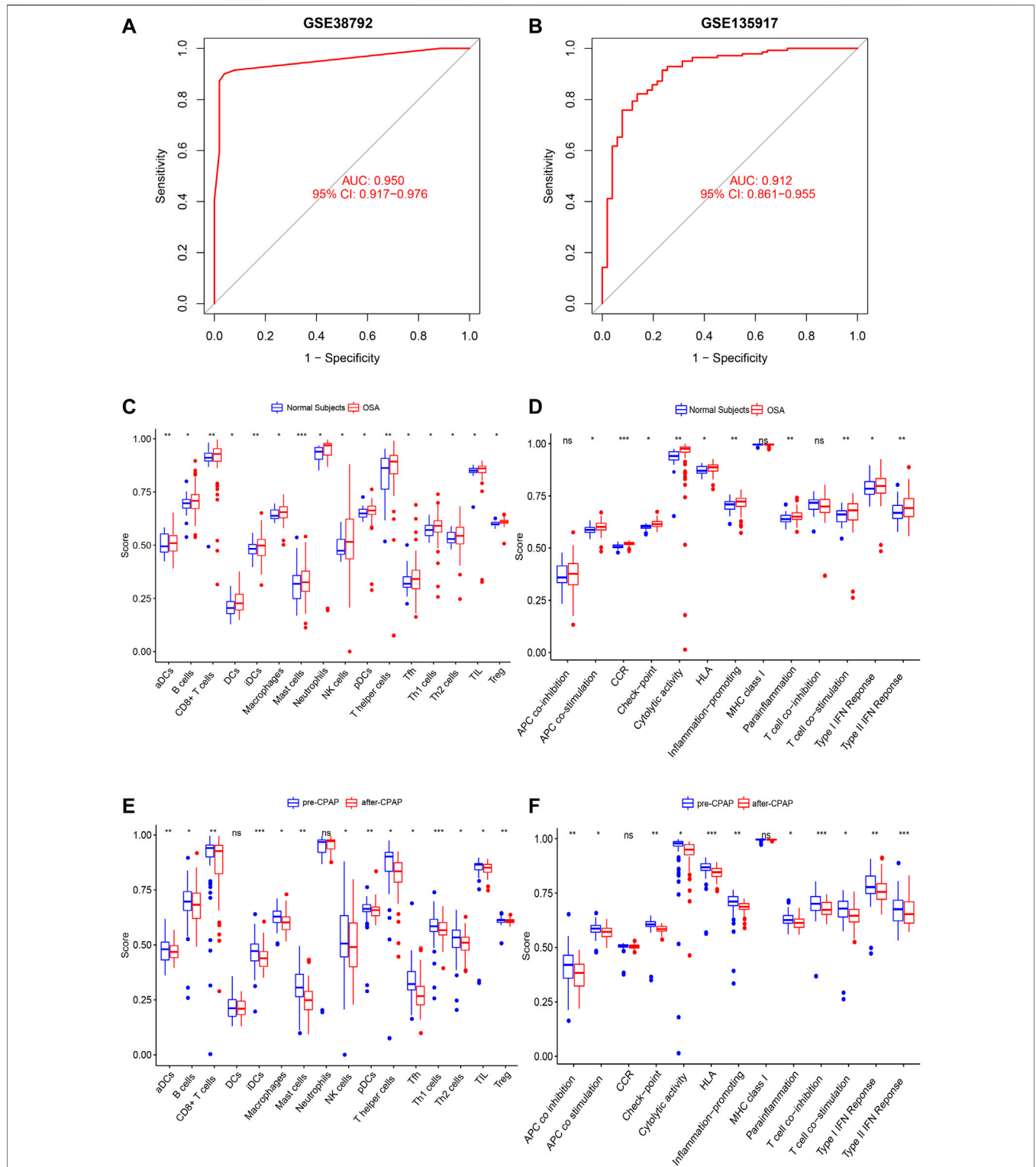
**C**



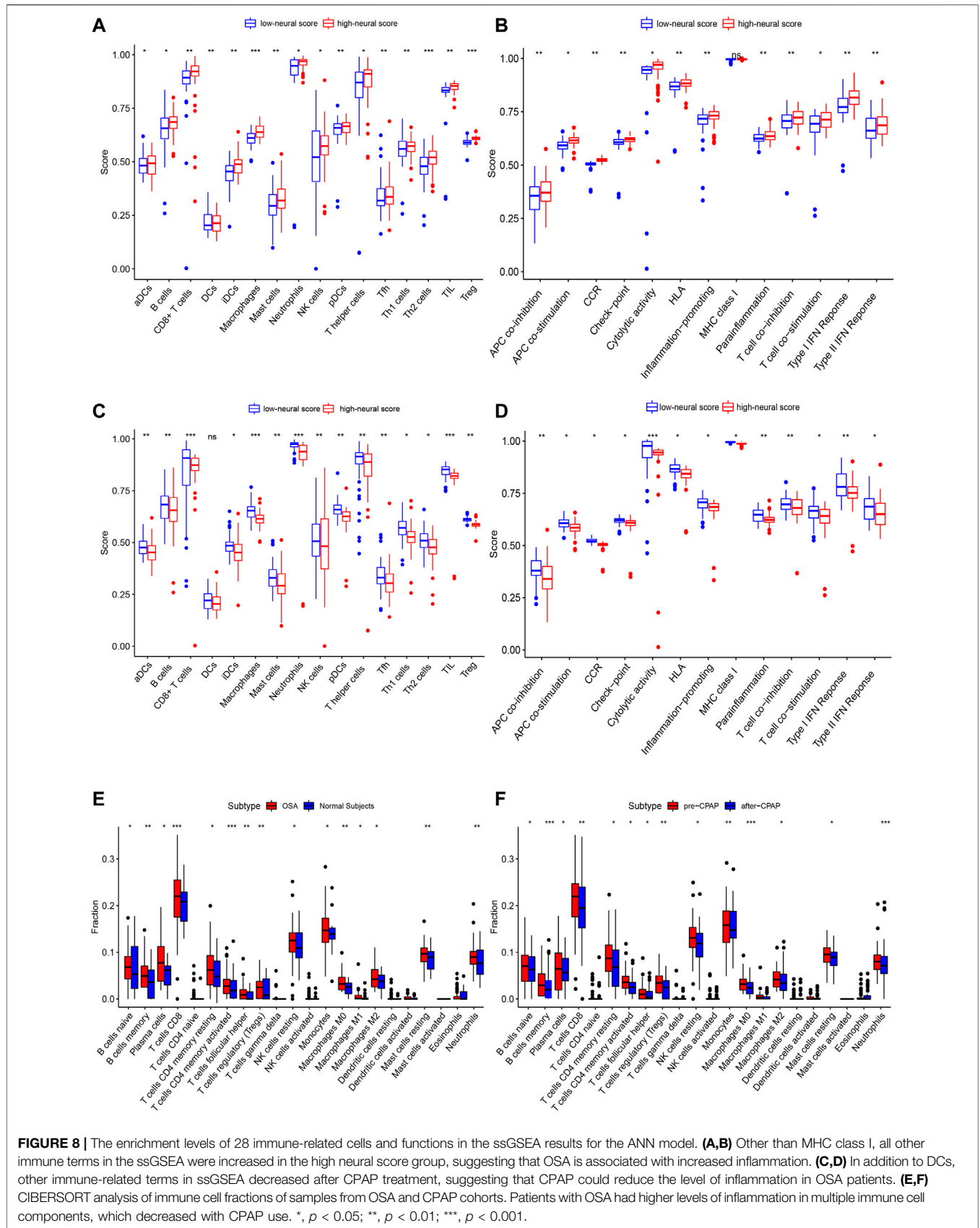
**D**



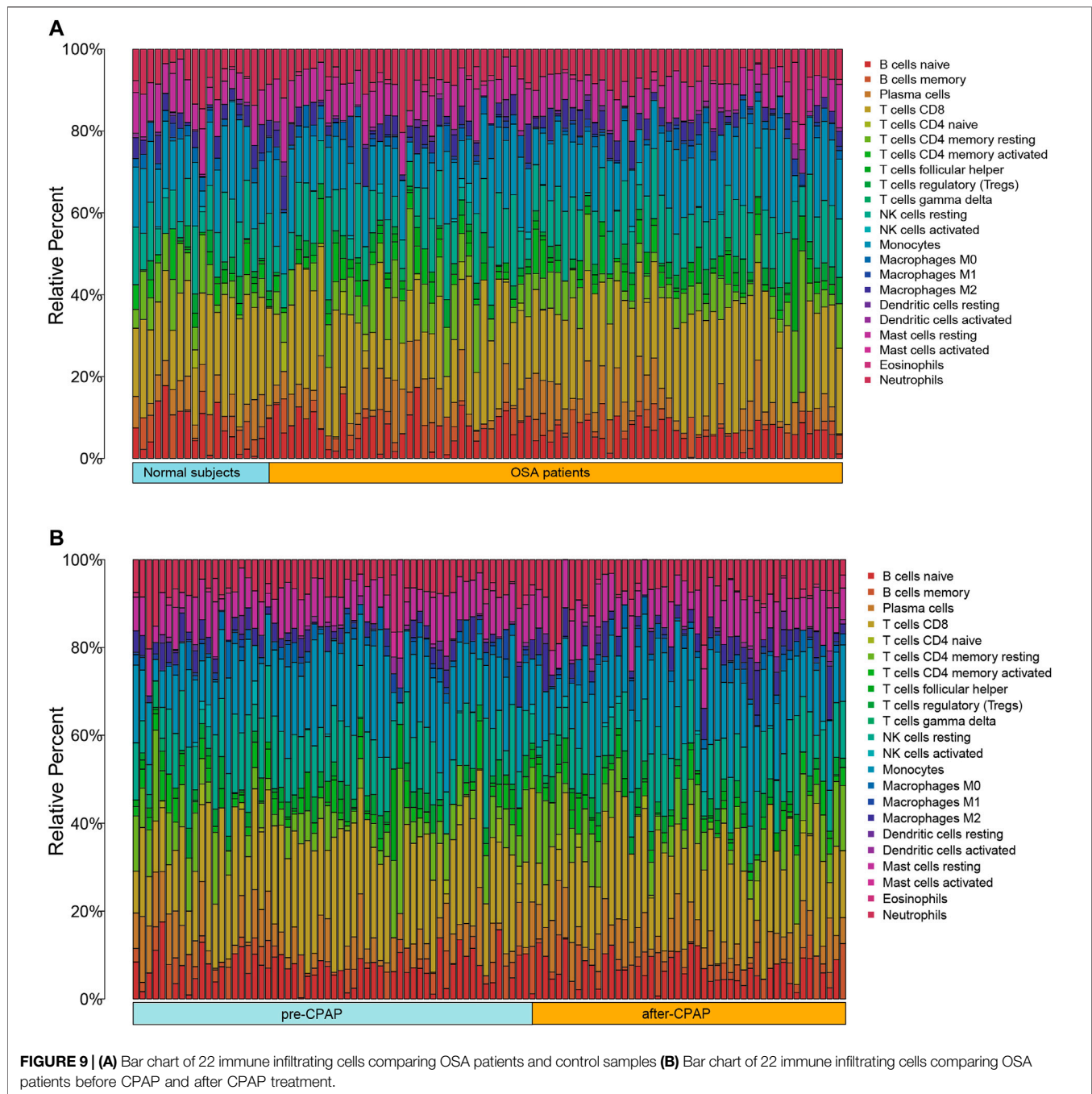
**FIGURE 6 | (A,B)** Nomograms with key genes were constructed for OSA risk prediction and CPAP therapeutic response. A point line is shown on the horizontal axis for each variable. An axis for total score was plotted, and a line for probability was drawn downward to determine the risk or response to treatment. **(C)** ROC curves of ANN model genes in the OSA cohort. **(D)** ROC curves of ANN model genes in the CPAP cohort.



**FIGURE 7 | (A,B)** Verification of the ROC curves by the ANN model for the GSE38792 and GSE135917 datasets. **(C,D)** The enrichment levels of 28 immune-related cells and functions in the ssGSEA results for the OSA cohort. Besides APC co-inhibition, MHC Class I, and T cell co-inhibition, other cell components, and functional pathways showed higher ssGSEA scores in OSA patients. **(E,F)** The enrichment levels of 28 immune-related cells and functions in the ssGSEA results for the CPAP cohort. Aside from DCs, Neutrophils, CCRs, and MHC class I, other immune cells and functions showed a downward trend in ssGSEA scores after CPAP treatment. \*,  $p < 0.05$ ; \*\*,  $p < 0.01$ ; \*\*\*,  $p < 0.001$ .

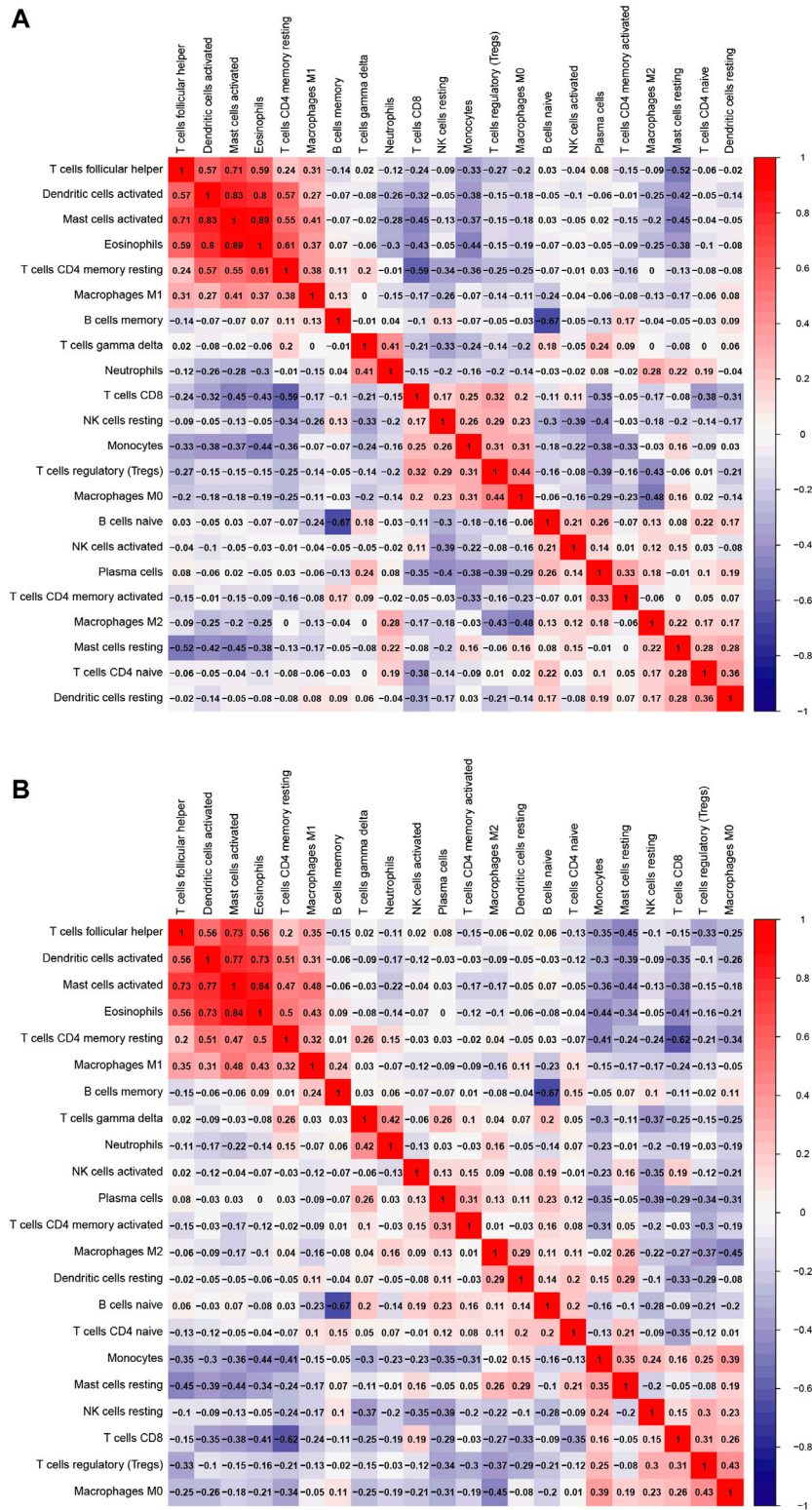


**FIGURE 8 |** The enrichment levels of 28 immune-related cells and functions in the ssGSEA results for the ANN model. **(A,B)** Other than MHC class I, all other immune terms in the ssGSEA were increased in the high neural score group, suggesting that OSA is associated with increased inflammation. **(C,D)** In addition to DCs, other immune-related terms in ssGSEA decreased after CPAP treatment, suggesting that CPAP could reduce the level of inflammation in OSA patients. **(E,F)** CIBERSORT analysis of immune cell fractions of samples from OSA and CPAP cohorts. Patients with OSA had higher levels of inflammation in multiple immune cell components, which decreased with CPAP use. \*,  $p < 0.05$ ; \*\*,  $p < 0.01$ ; \*\*\*,  $p < 0.001$ .



(MGAT5), and 1.87 (TBX3), 17.77 (GLT1D1), 42.69 (MOAP1), 8.24 (GRPEL1), -7.96 (CRYBB1) (**Figure 5A**). Based on nomograms, MGAT5, REV3L, TXLNA, and PTPN3 were positively correlated with OSA risk, while the remaining six genes were negatively correlated (**Figure 6A**). The weight predictions for the CPAP cohort were -5.44 (TMEM121), -6.53 (EYA2), 8.75 (TFF2), -15.52 (FGD4), 2.54 (PPL), 1.72 (TBX3), -0.57 (CAND1), 8.66 (TXLNA), 7.99 (TCF12), -5.25 (ABCF1) (**Figure 5B**). Based on nomograms, the CPAP response was positively correlated

with FGD4, TFF2, EYA2, and TMEM121, but was negatively correlated with other genes (**Figure 6B**). With the receiver operating characteristic (ROC) curve, the 5-time cross-validation illustrated the model classification performance. The areas under the curves (AUC) showed the hardness of the model (average AUC >0.99) (**Figures 5C,D**). The AUC for each gene was also assessed within each cohort (**Figures 6C,D**). The AUC of the neural network score was much better than that of other genes. Furthermore, the same ANN model also had excellent performance across two



**FIGURE 10 | (A)** Heatmap of 22 immune cells comparing OSA patients and control samples **(B)** Heatmap of 22 immune cells comparing OSA patients before and after the use of CPAP.

independent validation cohorts from GSE38792 and GSE135917 (Figures 7A,B).

## Immune Cell Infiltration and the Neural Score

We used ssGSEA to examine immune infiltration in the transcriptomes of both OSA and CPAP cohorts, including twenty-eight immune-related terms to assess the abundance of immune cells.

In the OSA cohort, ssGSEA scores in multiple terms were higher in OSA patients than in controls, including diverse immune cells (DCs, B cells, T cells, macrophages, mast cells, neutrophils, NK cells, etc.) and a variety of immune pathways (APC co-stimulation, CCR, Checkpoint, Cytolytic activity, HLA, MHC class I, T cell co-stimulation, etc.) (Figures 7C,D). Almost all the elevated immune parameters responded to CPAP treatment. A decrease in the ssGSEA scores in immune cells and pathways was observed after CPAP treatment (Figures 7E,F). Our ANN model retained the differences in immunity between OSA and controls and before and after treatment with CPAP. The OSA patients were clustered into two groups based on the average neural score. Figure 8 shows the ssGSEA scores for the high-neural score and low-neural score groups. The high neural score group was associated with a higher ssGSEA score, indicating that a higher risk of OSA was associated with elevated immune infiltration (Figures 8A,B). Similarly, OSA patients with CPAP treatment were divided based on their average neural score. A higher neural score indicated a better CPAP treatment response, accompanied by lower immune infiltration (Figures 8C,D).

As a result of CIBERSORT, the proportion of 22 immune cell types in mixed tissue samples from the OSA and CPAP cohorts was estimated (Figures 9A,B). There were clear positive correlations among T follicular helper cells, activated mast cells, eosinophils, activated dendritic cells, and resting memory CD4 T cells (Figures 10A,B). Additionally, these immune cells exhibited negative correlations with monocytes, CD8 T cells, and resting mast cells. Diverse immune cells had a higher CIBERSORT fraction in OSA patients than controls (Figure 8E), such as B cells (naïve and memory B cells), plasma cells, T cells (CD8, resting CD4 memory cells, activated CD4 memory cells, etc.), NK cells, macrophages (M0, M1, M2), mast cells, and neutrophils. Additionally, nearly all types of immune cells showed decreased levels of elevation after CPAP use (Figure 8F).

## DISCUSSION

OSA can be predicted using patient-reported symptoms such as sleepiness, snoring, and observed apnea (Holfinger et al., 2022). Questionnaires based on typical symptoms are widely used to assess OSA risk (Chung et al., 2008). However, such screenings tend to miss patients with mild symptoms and may include patients with conditions that cause similar symptoms. It is also believed that some physiological or pathological indicators, such as sex, age, BMI, smoking history, alcohol consumption, obesity, and hypertension, are

highly associated with OSA risk (Gaspar et al., 2017; Peppard and Hagen, 2018; Drager et al., 2019). Furthermore, there is increasing recognition of the predictive role of inflammatory factors in OSA risk (Shamsuzzaman et al., 2002; Yokoe et al., 2003). However, there are no biological indicators to assess the risk of OSA that are as widely used in clinical practice as patient-reported symptoms. The development of high-throughput techniques has opened up numerous possibilities for gene-level analysis, providing an entirely new perspective on the assessment and treatment of OSA (Kim et al., 2012; Strausz et al., 2021). Significant advances made in targeting driver genes for cancer diseases suggests that new OSA prediction tools based on genetic data could contribute to identifying OSA risk and improving outcomes. Compared to traditional statistical tools such as logistic regression models, machine learning algorithms are more efficient at detecting multilevel, nonlinear relationships between variables and outcomes (Holfinger et al., 2022).

Machine learning also is rapidly being applied to various clinical models of diseases. According to Holfinger et al. (Holfinger et al., 2022), machine learning derived prediction tools based on age, sex, race, and BMI provide better diagnosis for OSA than logistic regression when used in community-based samples. Using sleep parameters and endoscopic findings to develop machine learning models for predicting the success rate of sleep surgery also showed higher accuracy than subjective prediction by sleep surgeons (Kim et al., 2021). Moreover, machine learning tools are effective at assessing long-term cardiovascular risk in OSA patients (Li et al., 2022). Machine learning has even been used to help build a CPAP compliance-monitoring system to improve the management of OSA patients (Turino et al., 2021). Additionally, machine learning is used to recognize data from polysomnograms (PSGs) and other monitoring equipment automatically. The data from PSGs is in the form of multichannel signals, making them ideal for machine learning techniques. The work of Linda Zhang has demonstrated how machine learning can assist in automating sleep staging and apnea/hypopnea detection, as well as building models to predict comorbid outcomes (Zhang et al., 2019; Zhang, 2021). As for other contactless devices, such as a microwave Doppler radar sleep monitoring system, machine learning is also capable of identifying OSA events (Snigdha et al., 2020). However, the performance of machine learning prediction models based on genetic datasets has not been examined. Utilizing three machine learning methods (ANN, RF, and SVM-REF), we performed a comprehensive analysis of transcriptome data for OSA risk and CPAP treatment.

Public databases were used to obtain microarray data on patients with OSA and those who had undergone CPAP treatment. Comparing OSA data with controls, including before and after CPAP treatment, DEG intersections were calculated. The final risk prediction model was comprised of ten genes and demonstrated excellent ROC performance as verified in multiple independent data sets. Several of these ten genes are involved in the immune system. PTPN3 is described as an immune checkpoint molecule. Increased expression of PTPN3 may act as a negative feedback mechanism that regulates the overactivation of lymphocytes and may be related to the PD-1/PD-L1 axis (Fujimura et al., 2019). PTPN3 has been implicated in various tumor studies as an immune system regulator and as an immunotherapy target (Gao et al., 2014; Peng et al., 2020; Koga et al., 2021). TXLNA, formerly known as interleukin-14 (IL-14), has been

identified as a key factor in intracellular vesicle traffic, essential for cellular functions, such as neurotransmitter release, cell division, and cell motility (Nogami et al., 2003). It has been demonstrated that IL-14 promotes the proliferation of B cells and the expansion of memory B cells (Leca et al., 2008) and enhances the functions of memory B cells (Ford et al., 1995). SMAD4 has a critical role in activating TGF- $\beta$  signaling pathways, a major immune-suppressive signal, affecting cytotoxic T cells and regulating the recruitment of regulatory T cells (Nakamura et al., 2001; Chen et al., 2003). In response to proinflammatory cytokines, SMAD4 proteins inhibit IFN- $\gamma$  secretion by NK cells (Yu et al., 2006). MGAT5 encodes a glycosyltransferase called N-acetylglucosaminyltransferase V (GnT-V), required for T cell function (Daniels et al., 2002). GnT-V is tightly involved in regulating T cell activity and signaling. GnT-V deficient mice showed increased T cell receptor clustering, which led to a reduced threshold of T cell activation and increased TH1 differentiation (Demetriou et al., 2001; Morgan et al., 2004).

Several other genes are involved in important cell cycle and function processes. REV3L encodes the catalytic subunit of DNA polymerase zeta (Pol zeta), which belongs to the B family of DNA polymerases. A key function of this protein is to contribute to the tolerance of DNA damage by translesion synthesis (Prakash et al., 2005; Waters et al., 2009). By interacting with the BAX protein, MOAP1 serves as one of the key regulators of apoptosis, contributing to mitochondrial and death receptor-mediated apoptosis (Su et al., 2022). GRPEEL1 is a subtype of the GRPE protein homolog. It acts as a nucleotide exchange factor in mitochondria to influence nonnative protein folding (Ma et al., 2022).

The main characteristic of OSA is recurrent episodes of upper airway narrowing, which results in intermittent hypoxia (IH) and thus induces systemic inflammation. The activation of systemic inflammation and proinflammatory pathways are important mechanisms of OSA-derived chronic health conditions such as cardiovascular disease and cognitive impairment (Thompson et al., 2022). Widespread increases in inflammatory factors like TNF- $\alpha$ , interleukin 8(IL-8), Interleukin 6 (IL-6) (NF)-kB, and CRP levels have been observed in OSA patients and can be alleviated by CPAP treatment (Garvey et al., 2009; Kheirandish-Gozal and Gozal, 2019; Thompson et al., 2022). These proinflammatory factors are part of complex interactive networks generated by immune cells and vascular endothelial cells, adipose cells, and liver cells (Lee et al., 2007; Baessler et al., 2013). These factors were also highly variable among individuals, according to different studies (Kaditis et al., 2014; Gaines et al., 2016; Huang et al., 2016). It appears that genetic variances are responsible for similar heterogeneity (Riha et al., 2005; Popko et al., 2008; Kong et al., 2017). Given this background, finding important markers, particularly those closely related to OSA patients' inflammation level and responding to CPAP treatment, will improve accuracy and provide new perspectives in identifying and treating these patients.

Accordingly, our ssGESA and CIBERSORT results provided similar conclusions. Multiple identified immune cells, such as B cells, T cells, plasma cells, NK cells, macrophages, and neutrophils, and diverse immune pathways were elevated in OSA patients and decreased after CPAP treatment. The predictive model based on machine learning algorithms

maintained this characteristic, with individuals at high risk for OSA showing extensive activation of immune cells and pathways. Moreover, as mentioned previously, most genes in the models play an important role in the immune system. In addition, all ten genes were highly expressed in OSA patients, and their expression levels decreased after CPAP treatment. Furthermore, all patients in our study suffered from moderate to severe OSA. A majority of the studies reporting increased inflammation in OSA were based on moderate to severe cases (Murphy et al., 2017; Díaz-García et al., 2022; Popadic et al., 2022).

As the severity of OSA increased, so did the level of inflammation (Karamanli et al., 2017; Huang et al., 2021), which may have a more serious effect on gene transcription. These findings confirm the critical role that the immune background plays in developing and treating OSA and identifying important genes involved in OSA pathophysiology. There is little information regarding the relative roles played by these genes in OSA development. Further experiments *in vivo* and *in vitro* will be necessary to validate and examine the specific mechanisms involved.

While our results are promising, there also are several limitations to the current study. Machine learning based on genetic data poses challenges, especially regarding how to best apply results to clinical practice, given that genetic testing for OSA patients is not shared and costly. Second, AUC is a powerful tool for measuring model discrimination; however, its clinical utility is determined by the threshold met for sleep apnea treatment. A definite threshold has not been determined and remains subject to interpretation. Our models used a chosen cut-point to calculate predictive characteristics. Further population testing and model corrections and improvements are required before this type of analysis reaches its potential as a research and clinical tool for OSA.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. The datasets can be found online (<https://www.ncbi.nlm.nih.gov/>). The names of the repository and accession number(s) can be found in the article.

## AUTHOR CONTRIBUTIONS

JZ and XT generated the original concept. JZ performed the statistical analysis and wrote the first draft of the manuscript. YZ and RR helped writing part of the manuscript. LS and XT supervised the entire study. All authors had full access to all study data and analyses, participated in preparing this report, and approved of its final, submitted form.

## FUNDING

The present work was supported by the National Natural Science Foundation of China (Grant No. 82120108002).

## REFERENCES

- Alizadeh Savareh, B., Asadzadeh Aghdaie, H., Behmanesh, A., Bashiri, A., Sadeghi, A., Zali, M., et al. (2020). A Machine Learning Approach Identified a Diagnostic Model for Pancreatic Cancer through Using Circulating microRNA Signatures. *Pancreatology* 20 (6), 1195–1204. doi:10.1016/j.pan.2020.07.399
- Baessler, A., Nadeem, R., Harvey, M., Madbouly, E., Younus, A., Sajid, H., et al. (2013). Treatment for Sleep Apnea by Continuous Positive Airway Pressure Improves Levels of Inflammatory Markers - a Meta-Analysis. *J. Inflamm.* 10, 13. doi:10.1186/1476-9255-10-13
- Beck, M. W. (2018). NeuralNetTools: Visualization and Analysis Tools for Neural Networks. *J. Stat. Soft.* 85 (11), 1–20. doi:10.18637/jss.v085.i11
- Benjafield, A. V., Ayas, N. T., Eastwood, P. R., Heinzer, R., Ip, M. S. M., Morrell, M. J., et al. (2019). Estimation of the Global Prevalence and Burden of Obstructive Sleep Apnoea: a Literature-Based Analysis. *Lancet Respir. Med.* 7 (8), 687–698. doi:10.1016/S2213-2600(19)30198-5
- Bi, G., Chen, Z., Yang, X., Liang, J., Hu, Z., Bian, Y., et al. (2020). Identification and Validation of Tumor Environment Phenotypes in Lung Adenocarcinoma by Integrative Genome-Scale Analysis. *Cancer Immunol. Immunother.* 69 (7), 1293–1305. doi:10.1007/s00262-020-02546-3
- Chen, W., Jin, W., Hardegen, N., Lei, K. J., Li, L., Marinos, N., et al. (2003). Conversion of Peripheral CD4+CD25- Naive T Cells to CD4+CD25+ Regulatory T Cells by TGF- $\beta$  Induction of Transcription Factor Foxp3. *J. Exp. Med.* 198 (12), 1875–1886. doi:10.1084/jem.20030152
- Chung, F., Yegneswaran, B., Liao, P., Chung, S. A., Vairavanathan, S., Islam, S., et al. (2008). STOP Questionnaire. *Anesthesiology* 108 (5), 812–821. doi:10.1097/ALN.0b013e31816d83e4
- Daniels, M. A., Hogquist, K. A., and Jameson, S. C. (2002). Sweet 'n' Sour: the Impact of Differential Glycosylation on T Cell Responses. *Nat. Immunol.* 3 (10), 903–910. doi:10.1038/ni1002-903
- De Luca Canto, L., Pacheco-Pereira, C., Aydinov, S., Major, P. W., Flores-Mir, C., and Gozal, D. (2015). Biomarkers Associated with Obstructive Sleep Apnea: A Scoping Review. *Sleep. Med. Rev.* 23, 28–45. doi:10.1016/j.smrv.2014.11.004
- Demetriou, M., Granovsky, M., Quaggin, S., and Dennis, J. W. (2001). Negative Regulation of T-Cell Activation and Autoimmunity by Mgat5 N-Glycosylation. *Nature* 409 (6821), 733–739. doi:10.1038/35055582
- Deo, R. C. (2015). Machine Learning in Medicine. *Circulation* 132 (20), 1920–1930. doi:10.1161/CIRCULATIONAHA.115.001593
- Díaz-García, E., García-Tovar, S., Alfaro, E., Jaureguizar, A., Casitas, R., Sánchez-Sánchez, B., et al. (2022). Inflammasome Activation: A Keystone of Proinflammatory Response in Obstructive Sleep Apnea. *Am. J. Respir. Crit. Care Med.* 205 (11), 1337–1348. doi:10.1164/rccm.202106-1445OC
- Ding, Y., and Wilkins, D. (2006). Improving the Performance of SVM-RFE to Select Genes in Microarray Data. *BMC Bioinforma.* 7 (Suppl. 2), S12. doi:10.1186/1471-2105-7-S2-S12
- Drager, L. F., Santos, R. B., Silva, W. A., Parise, B. K., Giatti, S., Aiello, A. N., et al. (2019). OSA, Short Sleep Duration, and Their Interactions with Sleepiness and Cardiometabolic Risk Factors in Adults. *Chest* 155 (6), 1190–1198. doi:10.1016/j.chest.2018.12.003
- Ford, R., Tamayo, A., Martin, B., Niu, K., Claypool, K., Cabanillas, F., et al. (1995). Identification of B-Cell Growth Factors (Interleukin-14; High Molecular Weight-B-Cell Growth Factors) in Effusion Fluids from Patients with Aggressive B-Cell Lymphomas. *Blood* 86 (1), 283–293. doi:10.1182/blood.v86.1.283.bloodjournal861283
- Fujimura, A., Nakayama, K., Imaizumi, A., Kawamoto, M., Oyama, Y., Ichimiya, S., et al. (2019). PTPN3 Expressed in Activated T Lymphocytes Is a Candidate for a Non-antibody-type Immune Checkpoint Inhibitor. *Cancer Immunol. Immunother.* 68 (10), 1649–1660. doi:10.1007/s00262-019-02403-y
- Gaines, J., Vgontzas, A. N., Fernandez-Mendoza, J., Calhoun, S. L., He, F., Liao, D., et al. (2016). Inflammation Mediates the Association between Visceral Adiposity and Obstructive Sleep Apnea in Adolescents. *Am. J. Physiology-Endocrinology Metabolism* 311 (5), E851–E858. doi:10.1152/ajpendo.0024910.1152/ajpendo.00249.2016
- Gao, Q., Zhao, Y. J., Wang, X. Y., Guo, W. J., Gao, S., Wei, L., et al. (2014). Activating Mutations in PTPN3 Promote Cholangiocarcinoma Cell Proliferation and Migration and Are Associated with Tumor Recurrence in Patients. *Gastroenterology* 146 (5), 1397–1407. doi:10.1053/j.gastro.2014.01.062
- Garvey, J. F., Taylor, C. T., and McNicholas, W. T. (2009). Cardiovascular Disease in Obstructive Sleep Apnoea Syndrome: the Role of Intermittent Hypoxia and Inflammation. *Eur. Respir. J.* 33 (5), 1195–1205. doi:10.1183/09031936.00111208
- Gaspar, L. S., Álvaro, A. R., Moita, J., and Cavadas, C. (2017). Obstructive Sleep Apnea and Hallmarks of Aging. *Trends Mol. Med.* 23 (8), 675–692. doi:10.1016/j.molmed.2017.06.006
- Grilo, A., Ruiz-Granados, E. S., Moreno-Rey, C., Rivera, J. M., Ruiz, A., Real, L. M., et al. (2013). Genetic Analysis of Candidate SNPs for Metabolic Syndrome in Obstructive Sleep Apnea (OSA). *Gene* 521 (1), 150–154. doi:10.1016/j.gene.2013.03.024
- Hanko, M., Grendár, M., Snopko, P., Opšćenák, R., Šutovský, J., Benčo, M., et al. (2021). Random Forest-Based Prediction of Outcome and Mortality in Patients with Traumatic Brain Injury Undergoing Primary Decompressive Craniectomy. *World Neurosurg.* 148, e450–e458. doi:10.1016/j.wneu.2021.01.002
- Heaton, J. (2008). *Introduction to Neural Networks with Java*. Heaton Research, Inc.
- Hirsch, D., Evans, C. A., Wong, M., Machaalani, R., and Waters, K. A. (2019). Biochemical Markers of Cardiac Dysfunction in Children with Obstructive Sleep Apnoea (OSA). *Sleep. Breath.* 23 (1), 95–101. doi:10.1007/s11325-018-1666-y
- Holfinger, S. J., Lyons, M. M., Keenan, B. T., Mazzotti, D. R., Mindel, J., Maislin, G., et al. (2022). Diagnostic Performance of Machine Learning-Derived OSA Prediction Tools in Large Clinical and Community-Based Samples. *Chest* 161 (3), 807–817. doi:10.1016/j.chest.2021.10.023
- Huang, T., Goodman, M., Li, X., Sands, S. A., Li, J., Stampfer, M. J., et al. (2021). C-Reactive Protein and Risk of OSA in Four US Cohorts. *Chest* 159 (6), 2439–2448. doi:10.1016/j.chest.2021.01.060
- Huang, Y. S., Guilleminault, C., Hwang, F. M., Cheng, C., Lin, C. H., Li, H. Y., et al. (2016). Inflammatory Cytokines in Pediatric Obstructive Sleep Apnea. *Med. Baltim.* 95 (41), e4944. doi:10.1097/MD.0000000000004944
- Kaditis, A. G., Gozal, D., Khalyfa, A., Kheirandish-Gozal, L., Capdevila, O. S., Gourgoulis, K., et al. (2014). Variants in C-Reactive Protein and IL-6 Genes and Susceptibility to Obstructive Sleep Apnea in Children: a Candidate-Gene Association Study in European American and Southeast European Populations. *Sleep. Med.* 15 (2), 228–235. doi:10.1016/j.sleep.2013.08.795
- Karamanli, H., Kizilirmak, D., Akgedik, R., and Bilgi, M. (2017). Serum Levels of Magnesium and Their Relationship with CRP in Patients with OSA. *Sleep. Breath.* 21 (2), 549–556. doi:10.1007/s11325-016-1402-4
- Kheirandish-Gozal, L., and Gozal, D. (2019). Obstructive Sleep Apnea and Inflammation: Proof of Concept Based on Two Illustrative Cytokines. *Ijms* 20 (3), 459. doi:10.3390/ijms20030459
- Kim, J., Bhattacharjee, R., Khalyfa, A., Kheirandish-Gozal, L., Capdevila, O. S., Wang, Y., et al. (2012). DNA Methylation in Inflammatory Genes Among Children with Obstructive Sleep Apnea. *Am. J. Respir. Crit. Care Med.* 185 (3), 330–338. doi:10.1164/rccm.201106-1026OC
- Kim, J. Y., Kong, H. J., Kim, S. H., Lee, S., Kang, S. H., Han, S. C., et al. (2021). Machine Learning-Based Preoperative Datamining Can Predict the Therapeutic Outcome of Sleep Surgery in OSA Subjects. *Sci. Rep.* 11 (1), 14911. doi:10.1038/s41598-021-94454-4
- Klingenberg, R., Gerdes, N., Badeau, R. M., Gisterå, A., Strodthoff, D., Ketelhuth, D. F., et al. (2013). Depletion of FOXP3+ Regulatory T Cells Promotes Hypercholesterolemia and Atherosclerosis. *J. Clin. Invest.* 123 (3), 1323–1334. doi:10.1172/JCI63891
- Koga, S., Onishi, H., Masuda, S., Fujimura, A., Ichimiya, S., Nakayama, K., et al. (2021). PTPN3 Is a Potential Target for a New Cancer Immunotherapy that Has a Dual Effect of T Cell Activation and Direct Cancer Inhibition in Lung Neuroendocrine Tumor. *Transl. Oncol.* 14 (9), 101152. doi:10.1016/j.tranon.2021.101152
- Kong, D., Qin, Z., Wang, W., and Kang, J. (2017). Effect of Obstructive Sleep Apnea on Carotid Artery Intima Media Thickness Related to Inflammation. *Cim* 40 (1), 25–E33. doi:10.25011/cim.v40i1.28051
- Leca, N., Laftavi, M., Shen, L., Matteson, K., Ambrus, J., Jr., and Pankewycz, O. (2008). Regulation of Human Interleukin 14 Transcription *In Vitro* and *In Vivo* after Renal Transplantation. *Transplantation* 86 (2), 336–341. doi:10.1097/TP.0b013e31817c6380
- Lee, W. Y., Allison, M. A., Kim, D. J., Song, C. H., and Barrett-Connor, E. (2007). Association of Interleukin-6 and C-Reactive Protein with Subclinical Carotid Atherosclerosis (The Rancho Bernardo Study). *Am. J. Cardiol.* 99 (1), 99–102. doi:10.1016/j.amjcard.2006.07.070
- Lévy, P., Kohler, M., McNicholas, W. T., Barbé, F., McEvoy, R. D., Somers, V. K., et al. (2015). Obstructive Sleep Apnoea Syndrome. *Nat. Rev. Dis. Prim.* 1, 15015. doi:10.1038/nrdp.2015.15



- Li, A., Roveda, J. M., Powers, L. S., and Quan, S. F. (2022). Obstructive Sleep Apnea Predicts 10-year Cardiovascular Disease-Related Mortality in the Sleep Heart Health Study: a Machine Learning Approach. *J. Clin. Sleep Med.* 18 (2), 497–504. doi:10.5664/jcsm.9630
- Ma, C., Gao, B., Wang, Z., You, W., Yu, Z., Shen, H., et al. (2022). GrpEL1 Regulates Mitochondrial Unfolded Protein Response after Experimental Subarachnoid Hemorrhage *In Vivo* and *In Vitro*. *Brain Res. Bull.* 181, 97–108. doi:10.1016/j.brainresbull.2022.01.014
- Morgan, R., Gao, G., Pawling, J., Dennis, J. W., Demetriou, M., and Li, B. (2004). N-acetylglucosaminyltransferase V (Mgat5)-Mediated N-Glycosylation Negatively Regulates Th1 Cytokine Production by T Cells. *J. Immunol.* 173 (12), 7200–7208. doi:10.4049/jimmunol.173.12.7200
- Murphy, A. M., Thomas, A., Crinoin, S. J., Kent, B. D., Tambuwala, M. M., Fabre, A., et al. (2017). Intermittent Hypoxia in Obstructive Sleep Apnoea Mediates Insulin Resistance through Adipose Tissue Inflammation. *Eur. Respir. J.* 49 (4), 1601731. doi:10.1183/13993003.01731-2016
- Nakamura, K., Kitani, A., and Strober, W. (2001). Cell Contact-dependent Immunosuppression by Cd4+Cd25+Regulatory T Cells Is Mediated by Cell Surface-Bound Transforming Growth Factor  $\beta$ . *J. Exp. Med.* 194 (5), 629–644. doi:10.1084/jem.194.5.629
- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337
- Nogami, S., Satoh, S., Nakano, M., Shimizu, H., Fukushima, H., Maruyama, A., et al. (2003). Taxilin; a Novel Syntaxin-binding Protein that Is Involved in Ca<sup>2+</sup>-dependent Exocytosis in Neuroendocrine Cells. *Genes Cells* 8 (1), 17–28. doi:10.1046/j.1365-2443.2003.00612.x
- Nowakowski, S., Matthews, K. A., von Känel, R., Hall, M. H., and Thurston, R. C. (2018). Sleep Characteristics and Inflammatory Biomarkers Among Midlife Women. *Sleep* 41 (5). doi:10.1093/sleep/zsy049
- Peng, X. S., Yang, J. P., Qiang, Y. Y., Sun, R., Cao, Y., Zheng, L. S., et al. (2020). PTPN3 Inhibits the Growth and Metastasis of Clear Cell Renal Cell Carcinoma via Inhibition of PI3K/AKT Signaling. *Mol. Cancer Res.* 18 (6), 903–912. doi:10.1158/1541-7786.MCR-19-1142
- Peppard, P. E., and Hagen, E. W. (2018). The Last 25 Years of Obstructive Sleep Apnea Epidemiology—And the Next 25? *Am. J. Respir. Crit. Care Med.* 197 (3), 310–312. doi:10.1164/rccm.201708-1614PP
- Popadic, V., Brajkovic, M., Klasnja, S., Milic, N., Rajovic, N., Lisulov, D. P., et al. (2022). Correlation of Dyslipidemia and Inflammation with Obstructive Sleep Apnea Severity. *Front. Pharmacol.* 13, 897279. doi:10.3389/fphar.2022.897279
- Popko, K., Gorska, E., Potapinska, O., Wasik, M., Stoklosa, A., Plywaczewski, R., et al. (2008). Frequency of Distribution of Inflammatory Cytokines IL-1, IL-6 and TNF-Alpha Gene Polymorphism in Patients with Obstructive Sleep Apnea. *J. Physiol. Pharmacol.* 59 (Suppl. 6), 607–614.
- Prakash, S., Johnson, R. E., and Prakash, L. (2005). Eukaryotic Translesion Synthesis DNA Polymerases: Specificity of Structure and Function. *Annu. Rev. Biochem.* 74, 317–353. doi:10.1146/annurev.biochem.74.082803.133250
- Riha, R. L., Brander, P., Vennelle, M., McArdle, N., Kerr, S. M., Anderson, N. H., et al. (2005). Tumour Necrosis Factor- $\alpha$  (-308) Gene Polymorphism in Obstructive Sleep Apnoea-Hypopnoea Syndrome. *Eur. Respir. J.* 26 (4), 673–678. doi:10.1183/09031936.05.00130804
- Shamsuzzaman, A. S., Winnicki, M., Lanfranchi, P., Wolk, R., Kara, T., Accurso, V., et al. (2002). Elevated C-Reactive Protein in Patients with Obstructive Sleep Apnea. *Circulation* 105 (21), 2462–2464. doi:10.1161/01.cir.0000018948.95175.03
- Snigdha, F., Islam, S. M. M., Boric-Lubecke, O., and Lubecke, V. (2020). “Obstructive Sleep Apnea (OSA) Events Classification by Effective Radar Cross Section (ERCS) Method Using Microwave Doppler Radar and Machine Learning Classifier,” in 2020 IEEE MTT-S International Microwave Biomedical Conference (IMBioC), (IEEE), 1–3. doi:10.1109/IMBioC47321.2020.9385028
- Stenvinkel, P., Karimi, M., Johansson, S., Axelsson, J., Suliman, M., Lindholm, B., et al. (2007). Impact of Inflammation on Epigenetic DNA Methylation? a Novel Risk Factor for Cardiovascular Disease? *J. Intern. Med.* 261 (5), 488–499. doi:10.1111/j.1365-2796.2007.01777.x
- Strausz, S., Ruotsalainen, S., Ollila, H. M., Karjalainen, J., Kiiskinen, T., Reeve, M., et al. (2021). Genetic Analysis of Obstructive Sleep Apnoea Discovers a Strong Association with Cardiometabolic Health. *Eur. Respir. J.* 57 (5), 2003091. doi:10.1183/13993003.03091-2020
- Su, Y., Wang, W., and Meng, X. (2022). Revealing the Roles of MOAP1 in Diseases: A Review. *Cells* 11 (5), 889. doi:10.3390/cells11050889
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102
- Sun, J., Hu, J., Tu, C., Zhong, A., and Xu, H. (2015). Obstructive Sleep Apnea Susceptibility Genes in Chinese Population: A Field Synopsis and Meta-Analysis of Genetic Association Studies. *PLoS One* 10 (8), e0135942. doi:10.1371/journal.pone.0135942
- Thompson, C., Legault, J., Moullec, G., Martineau-Dussault, M. E., Baltzan, M., Cross, N., et al. (2022). Association between Risk of Obstructive Sleep Apnea, Inflammation and Cognition after 45 Years Old in the Canadian Longitudinal Study on Aging. *Sleep. Med.* 91, 21–30. doi:10.1016/j.sleep.2022.02.006
- Turino, C., Benítez, I. D., Rafael-Palou, X., Mayoral, A., Lopera, A., Pascual, L., et al. (2021). Management and Treatment of Patients with Obstructive Sleep Apnea Using an Intelligent Monitoring System Based on Machine Learning Aiming to Improve Continuous Positive Airway Pressure Treatment Compliance: Randomized Controlled Trial. *J. Med. Internet Res.* 23 (10), e24072. doi:10.2196/24072
- Wang, L., Yang, Z., and Cao, Y. (2020). Regulatory T Cell and Activated Natural Killer Cell Infiltration in Hepatocellular Carcinoma: Immune Cell Profiling Using the CIBERSORT. *Ann. Transl. Med.* 8 (22), 1483. doi:10.21037/atm-20-5830
- Waters, L. S., Minesinger, B. K., Wiltrout, M. E., D’Souza, S., Woodruff, R. V., and Walker, G. C. (2009). Eukaryotic Translesion Polymerases and Their Roles and Regulation in DNA Damage Tolerance. *Microbiol. Mol. Biol. Rev.* 73 (1), 134–154. doi:10.1128/MMBR.00034-08
- Watson, J. A., Watson, C. J., McCann, A., and Baugh, J. (2010). Epigenetics: The Epicenter of the Hypoxic Response. *Epigenetics* 5 (4), 293–296. doi:10.4161/epi.5.4.11684
- Yokoe, T., Minoguchi, K., Matsuo, H., Oda, N., Minoguchi, H., Yoshino, G., et al. (2003). Elevated Levels of C-Reactive Protein and Interleukin-6 in Patients with Obstructive Sleep Apnea Syndrome Are Decreased by Nasal Continuous Positive Airway Pressure. *Circulation* 107 (8), 1129–1134. doi:10.1161/01.cir.0000052627.99976.18
- Young, T., Skatrud, J., and Peppard, P. E. (2004). Risk Factors for Obstructive Sleep Apnea in Adults. *JAMA* 291 (16), 2013–2016. Epub 2004/04/29. PubMed PMID: 15113821. doi:10.1001/jama.291.16.2013
- Yu, J., Wei, M., Becknell, B., Trotta, R., Liu, S., Boyd, Z., et al. (2006). Pro- and Antiinflammatory Cytokine Signaling: Reciprocal Antagonism Regulates Interferon-Gamma Production by Human Natural Killer Cells. *Immunity* 24 (5), 575–590. Epub 2006/05/23. PubMed PMID: 16713975. doi:10.1016/j.immuni.2006.03.016
- Zhang, L. (2021). “Classification and Characterization of Sleep Apnea Using Machine Learning Methods on Sleep Studies,” (Nashville (US): Vanderbilt University). Dissertation or Thesis.
- Zhang, L., Fabbri, D., Upender, R., and Kent, D. (2019). Automated Sleep Stage Scoring of the Sleep Heart Health Study Using Deep Neural Networks. *Sleep* 42 (11), zsz159. Epub 2019/07/11. PubMed PMID: 31289828; PubMed Central PMCID: PMC6802563. doi:10.1093/sleep/zsz159
- Zheng, Y., Bi, G., Bian, Y., Li, M., Huang, Y., Zhao, M., et al. (2021). The Tumor Environment Immune Phenotype of LUSC by Genome-wide Analysis. *Int. Immunopharmacol.* 96, 107564. Epub 2021/04/05. PubMed PMID: 33813369. doi:10.1016/j.intimp.2021.107564

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhu, Sanford, Ren, Zhang and Tang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.