

RESEARCH

Open Access



Construction of a molecular diagnostic system for neurogenic rosacea by combining transcriptome sequencing and machine learning

Rui Mao^{1,2,3} and Ji Li^{1,2,3*}

Abstract

Patients with neurogenic rosacea (NR) frequently demonstrate pronounced neurological manifestations, often unresponsive to conventional therapeutic approaches. A molecular-level understanding and diagnosis of this patient cohort could significantly guide clinical interventions. In this study, we amalgamated our sequencing data ($n=46$) with a publicly accessible database ($n=38$) to perform an unsupervised cluster analysis of the integrated dataset. The eighty-four rosacea patients were partitioned into two distinct clusters. Neurovascular biomarkers were found to be elevated in cluster 1 compared to cluster 2. Pathways in cluster 1 were predominantly involved in neurotransmitter synthesis, transmission, and functionality, whereas cluster 2 pathways were centered on inflammation-related processes. Differential gene expression analysis and WGCNA were employed to delineate the characteristic gene sets of the two clusters. Subsequently, a diagnostic model was constructed from the identified gene sets using linear regression methodologies. The model's C index, comprising genes PNPLA3, CUX2, PLIN2, and HMGCR, achieved a remarkable value of 0.9683, with an area under the curve (AUC) for the training cohort's nomogram of 0.9376. Clinical characteristics from our dataset ($n=46$) were assessed by three seasoned dermatologists, forming the NR validation cohort (NR, $n=18$; non-neurogenic rosacea, $n=28$). Upon application of our model to NR diagnosis, the model's AUC value reached 0.9023. Finally, potential therapeutic candidates for both patient groups were predicted via the Connectivity Map. In summation, this study unveiled two clusters with unique molecular phenotypes within rosacea, leading to the development of a precise diagnostic model instrumental in NR diagnosis.

Keywords Neurogenic rosacea, Diagnostic system, Transcriptome

Introduction

Rosacea is a prevalent chronic inflammatory cutaneous disorder of the face. Epidemiological surveys report its prevalence to range between 0 and 22% [1, 2]. Commonly characterized by facial erythema reminiscent of a "drunken" appearance and accompanied by sensations of burning and tingling, rosacea significantly affects patients' appearance, quality of life, and psychological well-being [3, 4]. A distinct subset of rosacea patients exhibits distinct clinical features, characterized by pronounced facial redness, burning, tingling, and sensory

*Correspondence:

Ji Li

lji_xy@csu.edu.cn

¹ Department of Dermatology, Xiangya Hospital, Central South University, Changsha, China

² Hunan Key Laboratory of Aging Biology, Xiangya Hospital, Central South University, Changsha, China

³ National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

disturbances extending beyond mere inflammation. In 2011, Scharschmidt designated this subgroup as neurogenic rosacea (NR), [5] based on sensations rather than cutaneous manifestations. NR patients commonly present with neurological or neuropsychiatric conditions such as complex regional pain syndrome, idiopathic tremor, depression, and obsessive–compulsive disorder [5–7]. Conventional rosacea treatments like metronidazole, [8] isotretinoin, [9] and oral tetracycline, [10] are largely ineffective for NR, while neurological treatments, including oral gabapentin, pregabalin, and tricyclic antidepressants, [5, 7] may yield superior therapeutic outcomes. Despite these distinctions, NR lacks international recognition as a separate diagnostic entity, and its diagnosis remains largely subjective, based on clinical judgment rather than objective molecular markers. This lack of standardized diagnostic criteria can lead to mismanagement and worsening of NR symptoms, highlighting the urgent need for a precise, molecularly based diagnostic model for NR, especially in primary care dermatological practice.

Although the pathogenesis of rosacea remains elusive, factors including innate immune system imbalance, [11] gut microbiota, [12] physical triggers like ultraviolet radiation, [13] skin barrier dysfunction, [14] and aberrant neurovascular signaling [15] are implicated in rosacea's onset and progression. The role of neurovascular homeostasis in rosacea has gained recognition, particularly regarding the hypersensitivity of neural responses in NR patients and prevalent neurological complications [5]. This highlights the significance of neurovascular dysfunction in NR, warranting exploration of the molecular mediators involved. Two principal molecular categories are implicated in mediating neural functions in rosacea: ion channel-associated proteins and neuropeptides. Ion channel-associated proteins, such as the Mas-related G protein-coupled receptors (Mrgpr) family and transient receptor potential (TRP) channels, play crucial roles [16]. For instance, the activation of neurogenic TRP channels can stimulate the skin's vascular systems, leading to flushing—a characteristic symptom of rosacea. Additionally, specific G-protein-coupled receptors within the Mrgpr family are primarily involved in cutaneous neurogenic inflammation, promoting interactions between mast cells, sensory nerves, and skin cells [17–19]. Neuropeptides, including substance P and calcitonin gene-related peptide (CGRP), are also key players in this complex interaction. These molecular mechanisms may be pivotal in the development of NR, highlighting potential targets for therapeutic intervention.

Currently, numerous machine learning studies are being conducted for the diagnosis of rosacea utilizing facial photos or dermoscopy. For instance, Ge et al.

implemented a machine learning approach based on dermoscopy results, where the accuracy of their Gradient Boosting Machine (GBM) algorithm for classifying skin diseases was significantly superior to that of less experienced physicians [20]. Similarly, Binol et al. investigated the efficacy of machine learning methods to automatically identify erythematous acne lesions in facial images [21]. Additionally, Aggarwal et al. developed a diagnostic model for rosacea that employed machine learning techniques on facial photos, achieving an Area Under the Curve (AUC) of 0.89 [22]. However, biomarker-based diagnostics, owing to its objectivity and accuracy, is employed across a range of diseases [23–25]. Construction of disease-specific biomarkers and auxiliary diagnostic models have become the prevailing diagnostic methodology. In this context, our study assembled 43 biomarkers, encompassing mRNAs encoding neuropeptides like substance P, PACAP, VIP, and CRGP, as well as mRNAs for ion channel-related proteins like TRP channels and the Mrgpr family. Through clustering, we differentiated two patient groups and constructed an auxiliary diagnostic model to facilitate NR diagnosis.

Materials and methods

Clinical diagnosis

Three seasoned dermatologists from the Xiangya Hospital's dermatology department, well-versed in diagnosing and treating rosacea, collaborated with three highly experienced psychiatric clinicians from the same hospital to categorize 46 rosacea patients in our cohort. The diagnostic criteria for NR encompassed: First, rosacea could be diagnosed according to the 2017 edition of the International Diagnostic guidelines [26] for rosacea and the patient's facial symptoms and other clinical data. Second, there were no obvious papules and pustules in the patient's facial photos. Third, the patient had obvious facial redness, burning, tingling, and sensory disorders beyond blushing or inflammation. Fourth, patients might have had neurological or neuropsychiatric disorders, including complex regional pain syndrome, idiopathic tremor, depression, and obsessive–compulsive disorder [27]. Fifth, topical or oral metronidazole, isotretinoin, and other first-line treatment drugs were ineffective, but the symptoms significantly improved after using gabapentin or tricyclic antidepressants [5, 6]. Finally, demodex acne, contact dermatitis, and other differential diagnoses were excluded. Participants had to meet the above six diagnostic criteria at the same time to be diagnosed as NR. We used a questionnaire survey and physician evaluation to evaluate the clinical phenotype of patients with rosacea. Specifically, we used the Global flush severity score (GFSS) scale to evaluate paroxysmal flashes [28]. The researcher erythema rating scale (CEA) [28] and

the patient erythema self-rating scale (PSA) [28] were used to evaluate persistent erythema. The evaluation of papules and pustules was carried out by the combination of “Statistics of the number of papules and pustules” and “researcher Global Evaluation (IGA) scale.” The psychological burden of patients with rosacea was mainly assessed by the Chinese version of the Rosacea Quality of Life (RosaQoL) scale, the Dermatology Life Quality Index (DLQI) [28], and the Penn State Worry Questionnaire (PSWQ) [28]. All evaluation forms are available in Table S1.

The clinical data (Table S2) of 46 patients with rosacea are provided as supplementary data.

Data acquisition and integration

Our previous research involved collecting skin biopsy samples from the central facial area of patients diagnosed with rosacea, aged between 20 and 60, and healthy females undergoing cosmetic procedures, herein referred to as the Healthy Skin (HS) group. These samples were obtained from the Dermatology Department at Xiangya Hospital, Central South University, between June 1, 2014, and April 4, 2020. The study included 46 clinically and pathologically confirmed rosacea patients and 19 age-matched HS participants. Upon collection, all samples were immediately preserved at -80°C until further analysis. Each sample and its corresponding clinical data were collected following informed consent from all participants. The study's protocol was rigorously reviewed and approved by the Ethics Review Board of Xiangya Hospital, Central South University, ensuring all methods adhered to relevant guidelines and regulations. Sequencing data from the rosacea patients were archived in the Genome Sequence Archive under accession number HRA000379 (<http://bigd.big.ac.cn/gsa-human/>). Additionally, the GSE65914 datasets, which include microarray data for 38 rosacea and 20 normal samples, were obtained from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/gds/>). These datasets underwent a \log_2 transformation and quantile normalization. To integrate the GSE65914 dataset with our local dataset and mitigate batch effects, we employed the 'combat' function from the R package 'sva'.

Unsupervised clustering

Utilizing the expression patterns of the 43 identified biomarkers (Table 1), we embarked on an unsupervised cluster analysis for the 84 rosacea patients via the ConsensusClusterPlus software package [29]. Detailed information about these 43 biomarkers is shown in Table 1. Using agglomerative km clustering with a Pearson correlation distance of 1 and resampling 80% of the samples for 10 repetitions, the number and stability of clusters

were determined according to the consensus clustering algorithm [30]. To enhance the stability of classification, this process was repeated 1,000 times.

Gene differential expression analysis

The linear model for microarray analysis (Limma) [31] is a differential expression screening method based on the generalised linear model. Here, we used the R software package limma (version 3.40.6) for differential analysis to obtain differential genes between different comparison and control groups. We used the criterion that the absolute value of the fold change (FC) is greater than 1.5 and the adjusted- P value is less than 0.05 to screen for significant differentially expressed genes.

Principal component analysis (PCA)

We conducted PCA using the "FactoMineR" package in R. Initially, the dataset was standardized to ensure each variable contributed equally to the analysis. We then computed the eigenvalues and explained variances to assess the significance of each principal component. The contribution of each variable to the principal components was also evaluated. The results of the PCA were visualized using the "factoextra" package.

Analysis of immune cell infiltration

Based on the merged data matrix, we used the ssGSEA method of the GSVA package [32], MCPcounter [33], and Xcell (<https://xcell.ucsf.edu/>) [34] to calculate the abundance of immune cells and excluded the samples with $P > 0.05$. Finally, the Mann–Whitney U test was used to analyse the differences in immune cell subtypes among different groups.

Gene enrichment analysis and gene set enrichment analysis

We used the clusterProfiler [35] package for GO-BP and KEGG enrichment analysis, and the cutoff values of the P value and adjusted- P value were 0.05. There were fewer than three mRNAs covered in the enrichment path. The clusterProfiler [35] package was used for gene set enrichment analysis (GSEA). The input files were gene names, and their values of \log_2 (fold change) were between the high- and low-risk groups.

Weighted gene co-expression network analysis (WGCNA)

WGCNA was initiated by calculating the Median Absolute Deviation (MAD) for each gene from the gene expression profile. We discarded the top 50% of genes with the lowest MAD values and removed outlier genes and samples utilizing the "good Sample Genes" method from the WGCNA package in R [36]. Subsequently, we constructed a scale-free co-expression network using

Table 1 Detailed information of 43 neurogenic rosacea related genes

SYMBOL	ENSEMBL	ENTREZID	GENENAME
ADCYAP1	ENSG00000141433	116	adenylate cyclase activating polypeptide 1
ADCYAP1R1	ENSG00000078549	117	ADCYAP receptor type I
ADM	ENSG00000148926	133	adrenomedullin
ADM2	ENSG00000128165	79,924	adrenomedullin 2
ADRA1A	ENSG00000120907	148	adrenoceptor alpha 1A
ADRA1B	ENSG00000170214	147	adrenoceptor alpha 1B
ADRA1D	ENSG00000171873	146	adrenoceptor alpha 1D
ADRA2A	ENSG00000150594	150	adrenoceptor alpha 2A
ADRA2B	ENSG00000274286	151	adrenoceptor alpha 2B
ADRA2C	ENSG00000184160	152	adrenoceptor alpha 2C
ADRB1	ENSG00000043591	153	adrenoceptor beta 1
ADRB2	ENSG00000169252	154	adrenoceptor beta 2
ADRB3	ENSG00000188778	155	adrenoceptor beta 3
CALCA	ENSG00000110680	796	calcitonin related polypeptide alpha
CALCR	ENSG00000004948	799	calcitonin receptor
CALCRL	ENSG00000064989	10,203	calcitonin receptor like receptor
DRD1	ENSG00000184845	1812	dopamine receptor D1
F2RL1	ENSG00000164251	2150	F2R like trypsin receptor 1
F2RL3	ENSG00000127533	9002	F2R like thrombin or trypsin receptor 3
HRH1	ENSG00000196639	3269	histamine receptor H1
HRH2	ENSG00000113749	3274	histamine receptor H2
HRH3	ENSG00000101180	11,255	histamine receptor H3
HRH4	ENSG00000134489	59,340	histamine receptor H4
HTR2A	ENSG00000102468	3356	5-hydroxytryptamine receptor 2A
HTR3A	ENSG00000166736	3359	5-hydroxytryptamine receptor 3A
MMEL1	ENSG00000277131	79,258	membrane metalloendopeptidase like 1
MRGPRX1	ENSG00000170255	259,249	MAS related GPR family member X1
MRGPRX2	ENSG00000183695	117,194	MAS related GPR family member X2
MRGPRX4	ENSG00000179817	117,196	MAS related GPR family member X4
NGFR	ENSG00000064300	4804	nerve growth factor receptor
NPFFR2	ENSG00000056291	10,886	neuropeptide FF receptor 2
NPY	ENSG00000122585	4852	neuropeptide Y
TAC1	ENSG00000006128	6863	tachykinin precursor 1
TACR1	ENSG00000115353	6869	tachykinin receptor 1
TRPA1	ENSG00000104321	8989	transient receptor potential cation channel subfamily A member 1
TRPV2	ENSG00000187688	51,393	transient receptor potential cation channel subfamily V member 2
TRPV3	ENSG00000167723	162,514	transient receptor potential cation channel subfamily V member 3
TRPV4	ENSG00000111199	59,341	transient receptor potential cation channel subfamily V member 4
TRPV5	ENSG00000127412	56,302	transient receptor potential cation channel subfamily V member 5
TRPV6	ENSG00000276971	55,503	transient receptor potential cation channel subfamily V member 6
VIP	ENSG00000146469	7432	vasoactive intestinal peptide
VIPR1	ENSG00000114812	7433	vasoactive intestinal peptide receptor 1
VIPR2	ENSG00000106018	7434	vasoactive intestinal peptide receptor 2

WGCNA. Specifically, first, Pearson's correlation matrices and average linkage method were performed for all pair-wise genes. Then, a weighted adjacency matrix was constructed using the power function $A_{mn} = |C_{mn}|^\beta$

(C_{mn} = Pearson's correlation between Gene m and Gene n ; A_{mn} = adjacency between Gene m and Gene n). Here, β is a soft-thresholding parameter that emphasises strong correlations between genes and

penalises weak correlations. After choosing the power of 4, we transformed the adjacency into a topological overlap matrix (TOM), which could measure the network connectivity of a gene defined as the sum of its adjacency with all other genes for the network gene ratio, and the corresponding dissimilarity (1-TOM) was calculated. To classify genes with similar expression profiles into gene modules, average linkage hierarchical clustering was conducted according to the TOM-based dissimilarity measure, with a minimum size (Gene group) of 30 for the genes' dendrogram. We set the sensitivity to 3. To further analyse the module, we calculated the dissimilarity of module eigen genes, chose a cut line for the module dendrogram and merged some modules. In addition, we merged the modules with a distance of less than 0.25 and finally obtained three co-expression modules. Note that the grey module is considered a collection of genes that cannot be assigned to any module.

Correlations between module membership and sample traits were assessed, and a correlation heatmap was generated. Modules were prioritized based on their correlation with traits, the significance of module eigengenes, and p-values. The module exhibiting the highest mean correlation with the trait across all gene expressions within the module was considered most significant. Identification of hub genes related to the trait involved calculating the gene's internal connectivity and module membership. Internal connectivity assessed the gene's role within its module, while module membership determined its module affiliation.

Protein–protein interaction analysis

We used the Metascape (<https://metascape.org/gp/index.html#/main/step1>) website [37] for protein–protein (PPI) analysis of the modular hub gene, and the PPI analysis of the website was based on the Molecular Complex Detection (MCODE [38]) tool. Finally, we used Cytoscape (version 3.8.2) to visualise the PPI network.

Drug prediction

The Connectivity Map (CMap) is an expansive resource comprising over 3 million perturbational profiles that are accessible to the global scientific community. The platform clue.io (<https://clue.io>) provides a suite of applications designed to analyze these data effectively. Utilizing the QUERY CMap module available on the CLUE website, we uploaded the hub gene identified in the WGCNA module and configured the query parameters to target "Gene expression (L1000)."

Random forest analysis

We used the expression matrix of 696 differential genes and 84 rosacea samples as input files and standardized

the data. After setting the random number seed, we analyzed the data with the "randomForest" R software package. When building 500 trees, the error rate estimated by OOB was 5.3%. The experimental and control group's classification error rates were 0.02 and 0.06, respectively. Then, we used the "caret" package to split the training and test sets and the "Boruta" package to select and identify the key classification variables. A total of 51 important variables, 40 potentially important variables (tentative variable, no statistical difference between the importance score and the best shadow variable score), and 605 unimportant variables were identified. Next, we used the "dplyr" package to define a function to extract the important values corresponding to each variable and used the "ImageGP" package to draw important variable results. Finally, we used the "Caret" package cross-validation to select parameters and fit the model; specifically, we defined a function to generate some columns of mtry (mtry=19) for testing (a series of values not greater than the total number of variables), select data related to key feature variables, and select data related to key feature variables. We rigorously assessed the performance of the random forest model using the training dataset. The model demonstrated high effectiveness with an Accuracy and Kappa value of 1. Additionally, during the blind evaluation phase with an independent test set, the model achieved an AUC value of 1, indicating perfect classification performance.

Construction of the diagnostic model

The least absolute shrinkage and selection operator (LASSO) method, which is suitable for a reduction in high-dimensional data [39, 40], was used to select the best predictive characteristics. Features with nonzero coefficients in the LASSO regression model were selected [41]. Based on the standard lambda.1se (lambda value=4), we identified 10 molecules for the construction of the model. All of the potential predictors selected by lasso analysis were applied to develop a predicting model for the diagnosis of NR (R packages "glmnet" and "rms") [42].

Calibration curves were plotted to assess the calibration of the diagnosis nomogram (R package "rms"). We employed the bootstrap method to assess both the stability and performance of a given model. To quantify the discrimination performance of the diagnosis nomogram, Harrell's C-index was measured (R package "rms") [43]. Decision curve analysis was conducted to determine the clinical usefulness of the diagnosis nomogram by quantifying the net benefits at different threshold probabilities in the cohort [44]. The net benefit was calculated by subtracting the proportion of all of the patients with false-positive results from the proportion of patients with

true-positive results and by weighing the relative harm of forgoing interventions compared with the negative consequences of an unnecessary intervention (R packages “rms” and “rmda”). Receiver operating characteristic (ROC) curve analysis was employed to compare predictions concerning the sensitivity and specificity (R packages “pROC” and “rms”).

Results

Data processing

We integrated GSE65914 with our own sequencing cohort after removing the batch effect through the *sva* package as a training cohort. As delineated in Figure S1A and S1B's bar chart, post-elimination of the batch effect, the data distribution between the two datasets demonstrated a trend towards uniformity with the median lying on the same axis. Analysis through UPSET [45] revealed the intersection of the two cohorts yielding a total of 18,384 common protein-coding genes (Figure S1C). The density maps (Figure S1D and S1E) reveal significant discrepancies in the sample distribution across each dataset prior to batch effect removal, indicative of a batch effect. Subsequent to this removal, data distribution within each dataset appeared congruent, with analogous mean and variance. Furthermore, the PCA diagrams (Figure S1F and S1G) illustrated that prior to batch effect removal, the dataset samples were homogeneously clustered, suggestive of a batch effect. However, post-removal, samples from each dataset were both clustered and interspersed, a strong indication of effective batch effect elimination. The final product was a data matrix comprised of 123 samples and 18,384 protein-coding genes.

Cluster Characteristics and Inflammatory Profiles

Cluster analysis was carried out utilizing using ConsensusClusterPlus (Fig. 1A). The optimal number of clusters was discerned via the empirical cumulative distribution function (CDF) plot (Fig. 1B and C) and the evaluation of average intra-group consistency (Fig. 1D and E). The analysis revealed that at $K=2$, intra-group consistency was at its zenith, thereby resulting in the most efficacious clustering. Consequently, 84 rosacea patients were categorized into two clusters: Cluster 1 ($n=46$) and Cluster 2 ($n=38$). The expression differences of 43 established neurovascular loop biomarkers in rosacea between clusters 1 and 2 were then assessed. A total of 17 biomarkers were discovered to be highly expressed in Cluster 1 (Fig. 1F). These included nerve growth factor receptor (NGFR), tachykinin precursor 1 (TAC1), adrenomedullin (ADM and ADM2), adrenoceptor (ADRA1A, ADRB1, and ADRA2A), calcitonin (CALCA and CALCRL), histamine (HRH1, HRH2, and HTR3A), and TRPV family genes (TRPA1, TRPV2, TRPV3, and TRPV6). TAC1 encodes

four products of the tachykinin family: substance P and neurokinin A, and neuropeptide K and neuropeptide γ . These hormones are thought to be neurotransmitters that interact with neuroreceptors and smooth muscle cells. It is well known that they induce behavioral responses and act as vasodilators and secretagogues [46]. CGRP induces vasodilation. It dilates various blood vessels, including coronary arteries, brain arteries, and systemic vascular system. The anchor-like receptor family (TRPA1) is a member of the ion channel transient receptor potential (TRP) superfamily. Receptor-activated non-selective cationic channels are involved in pain detection and may also be involved in cold perception [47] and itching [48].

Initially, PCA was performed on 84 rosacea patients using 43 biomarkers (Fig. 2A), followed by analysis on a complete cohort (Fig. 2B) that included the aforementioned patients and 39 normal samples. The results showed that these 43 biomarkers were able to distinguish between cluster 1, cluster 2, and normal people in the two cohorts. According to the predicted results of Xcell (<https://xcell.ucsf.edu/>) [34], the abundance of nerve-related biomarkers in cluster 1 was significantly higher than in cluster 2 (Fig. 2C). Subsequently, ssGSEA [32] was employed to assess differences in immune cell abundance between the clusters. The results indicated that the immuno-inflammatory cell abundance in Cluster 1 was significantly reduced compared to Cluster 2 (Fig. 2E). Specifically, most immune cells, including B cells, CD4-positive T cells, CD8-positive T cells, macrophages, mast cells, natural killer cells, and neutrophils were considerably less abundant in cluster1 than in cluster2 (Fig. 2D and F). Consistency in immune cell infiltration abundance was observed in the MCP (Figure S2) and Xcell (Figure S3) analyses, with Cluster 1 significantly lower than Cluster 2.

The pathways in cluster 1 are mostly concentrated in the synthesis, transmission, and function of neurotransmitters

In an endeavor to delineate the disparities in functionality and pathway engagement between clusters 1 and 2, Gene Set Enrichment Analysis (GSEA) was executed for both groups. Utilizing the KEGG database, the GSEA findings revealed that the functional pathways within cluster 1 predominantly encompass the synaptic vesicle cycle, cholesterol metabolism, valine, leucine and isoleucine degradation, in addition to the biosynthesis of the amino acid pathway (Figure S4A). Moreover, the GSEA analysis grounded in the GO database ascertained that the pathways of synaptic signaling, cellular amino acid metabolic process, and cholesterol biosynthesis are predominantly augmented in cluster 1 (Figure S4B). Conversely, the GSEA-KEGG and GSEA-GO results of cluster 2 unveiled that the functional pathways within this cluster are

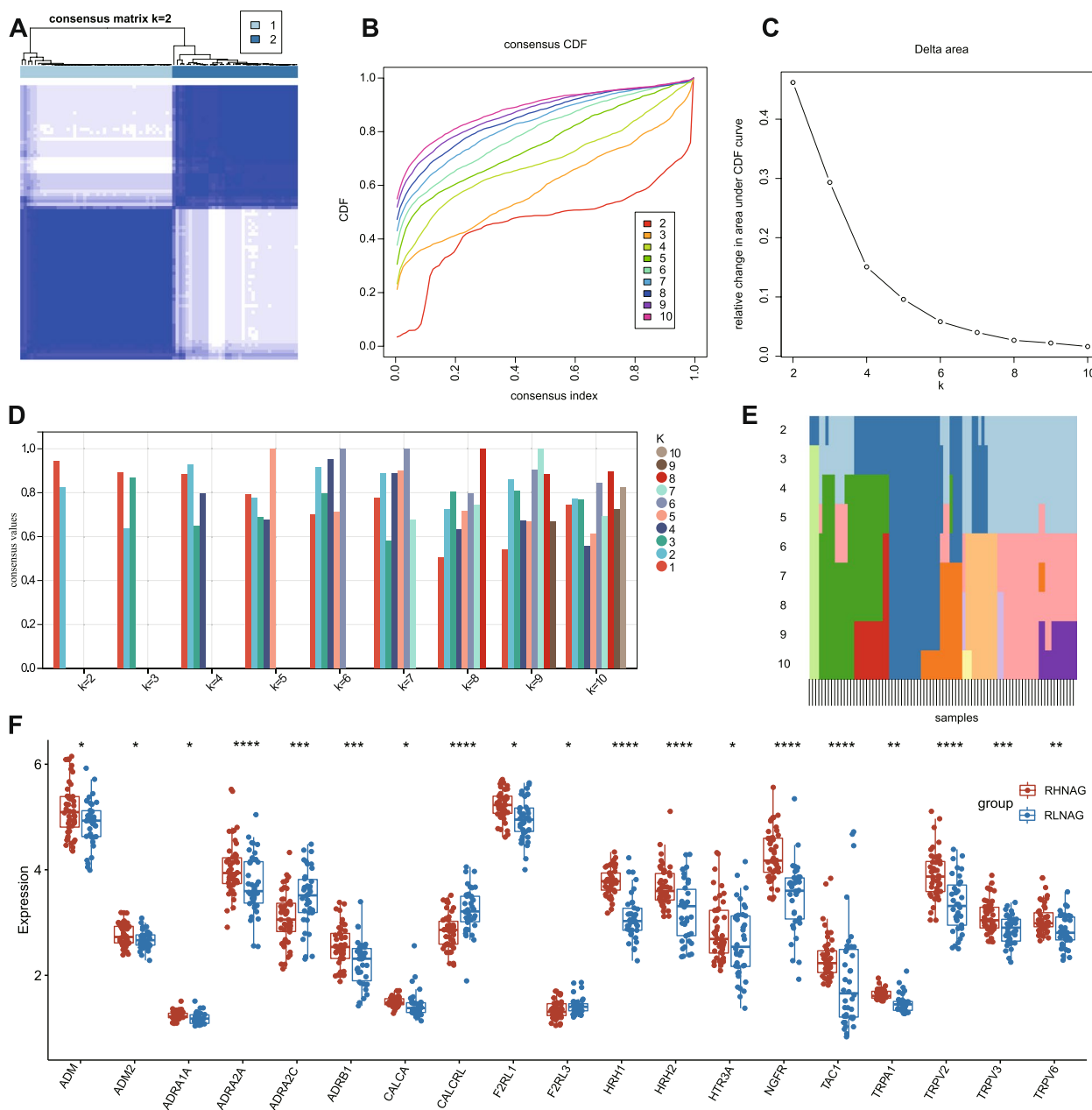


Fig. 1 Identification of neurogenic and non-neurogenic rosacea by unsupervised clustering. **A** Clustering heatmap; panel **(B)** shows the cumulative distribution function with different values of K, which is used to determine the approximate maximum value of CDF when the value of K is taken. At this time, the result of cluster analysis is the most reliable, usually taking the value of K with a small slope of CDF. Panel **(C)** shows the relative change of the area under the CDF curve between K and K minus 1; panels **(D)** and **(E)** show the consistent histogram and heat map of the group, respectively; panel **(F)** shows a histogram of NR-related biomarkers differentially expressed between RHNAG and RLNAG patients. *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; ****: $P < 0.0001$

primarily associated with immune- and inflammation-related processes, including but not limited to the IL-17 signaling pathway, NF- κ B signaling pathway, TNF signaling pathway, positive regulation of inflammatory response to an antigenic stimulus, regulation of CD8-positive, α - β T cell activation, and Th1 and Th2

cell differentiation (Figure S4C and S4D). The foregoing results underscore that the pathways in cluster 1 are overwhelmingly concentrated on the synthesis, transmission, and functionalization of neurotransmitters, whereas the functionality of cluster 2 is largely oriented toward inflammation and immune responsiveness.

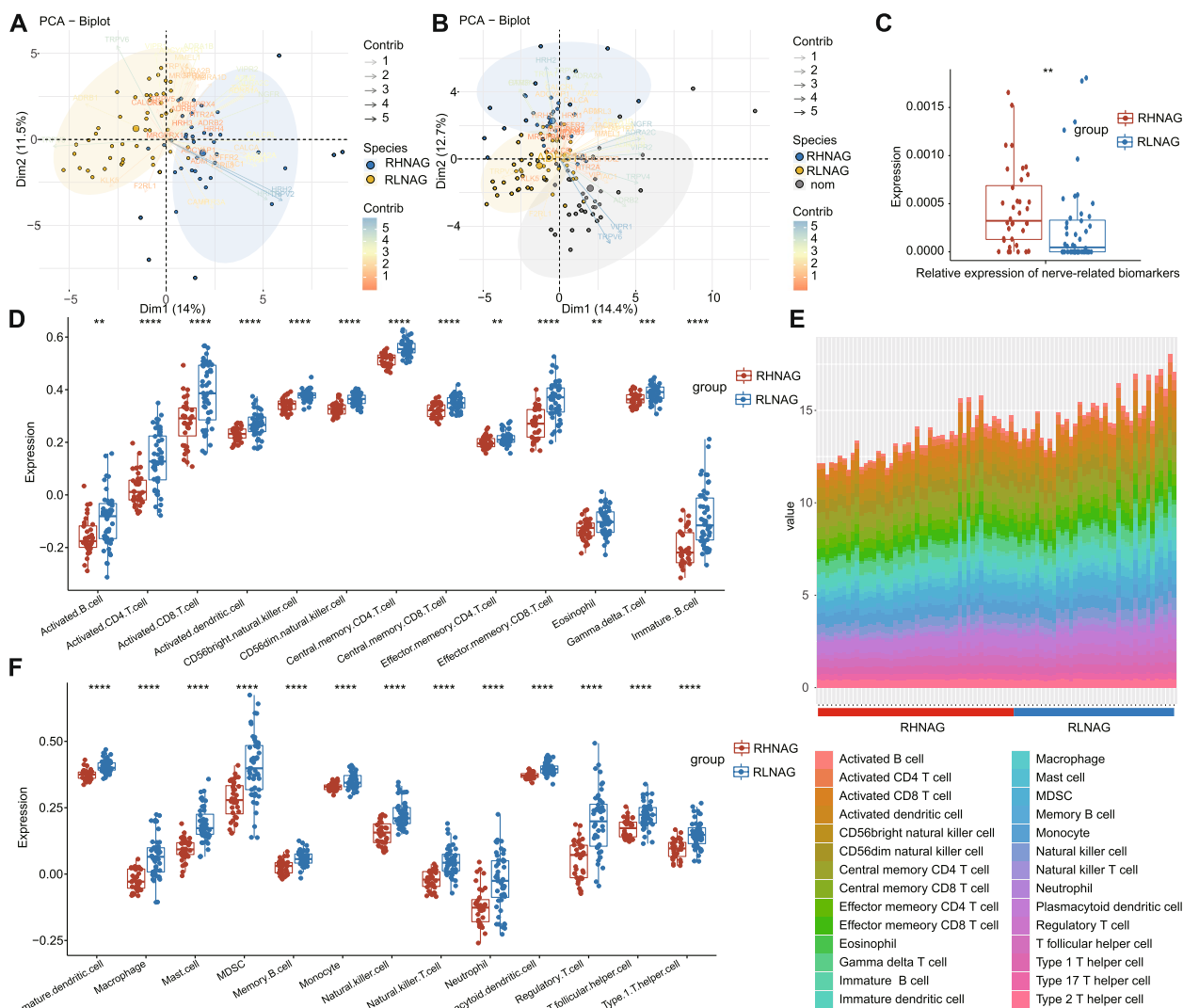


Fig. 2 The abundance of inflammatory cell infiltration in RLNAG was significantly higher than that in RHNAG. According to the expression of 43 NR biomarkers, PCA was performed in the cohort without normal samples (A) and the cohort containing 39 normal samples (B). We compared the abundance of nerve-related biomarkers expressed between the RHNAG and RLNAG, according to the predicted results of Xcell (C). According to the predicted results of ssGSEA, the differences in 26 immunocytes infiltration abundance between the RHNAG group and the RLNAG group were compared (D and F). E Columnar accumulation map of 26 types of immune cells' infiltration abundance predicted by ssGSEA in 84 patients with rosacea. *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; ****: $P < 0.0001$

Subsequently, we categorized cluster 1 as the rosacea group demonstrating a conspicuous neuromolecular phenotype (RHNAG) and cluster 2 as the rosacea group lacking a neuromolecular phenotype (RLNAG). As delineated in Table S2, the onset symptoms of RHNAG patients predominantly manifested as paroxysmal flushing ($n=16$; 66.7%), while a minority exhibited persistent erythema ($n=1$; 4.2%) or papular pustules ($n=6$; 25.0%). Conversely, the RLNAG patients chiefly presented with papular pustules ($n=15$; 68.2%), with a smaller proportion experiencing paroxysmal flushing ($n=7$; 31.8%). Furthermore,

an evaluation of the Global Flush Severity Score (GFSS) scale revealed significantly elevated scores in RHNAG patients relative to their RLNAG counterparts. Assessments employing the CEA, PSA, and IGA scales intimated that erythema was more pronounced in RHNAG patients, whereas papular pustules were markedly less prevalent compared to RLNAG patients. Noteworthy is the fact that according to the Dermatology Life Quality Index (DLQI), RHNAG patients suffered from a diminished quality of life compared to RLNAG patients, and the Penn State Worry Questionnaire (PSWQ) indicated exacerbated levels of

anxiety and depression within the RHNAG group. Collectively, these findings elucidate that the phenotype of RHNAG, characterized by elevated expression of neurotransmitters and receptor-associated markers, aggravated mental health conditions, and comparatively subdued inflammation, bears significant resemblance to that of NR in a clinical context.

The function of the hub gene in the blue module of WGCNA is mainly concentrated in the synthesis and function of neurotransmitters

In the quest to isolate the gene modules most pertinent to RHNAG or RLNAG, we scrutinized the genetic variances between these patient cohorts. Our examination unveiled a total of 1071 differentially expressed genes, comprising 411 genes that were upregulated in the RHNAG group and 660 genes that were upregulated in the RLNAG group (Figure S5A and S5B). Pursuant to this, we employed Weighted Gene Co-expression Network Analysis (WGCNA) on 3379 genes, all of which exhibited a variance exceeding 0.2 across the samples, and subsequently segmented these genes into four distinct modules (Figure S6A–S6D). Correlational analysis between the modules and clinical attributes revealed that the blue module exhibited positive correlations with the RHNAG group and negative correlations with immune cell abundance, including T cells, cytotoxic lymphocytes, and natural killer cells (Figure S6E). In contrast, the turquoise module manifested a pronounced positive correlation with the RLNAG group and an abundance of immune cell infiltration. The eigengene distance heatmap showcased the maximum distance between the blue and turquoise modules, indicative of a substantial divergence in gene expression patterns (Figure S6F). Furthermore, the scatter plot correlating Gene Significance (GS) with Module Membership (MM) confirmed that the genes most intimately associated with the RHNAG group were also pivotal within the blue module (Figure S6G).

To further explore the function of hub genes in the blue and turquoise modules, we carried out PPI analysis and functional enrichment analysis of hub genes in the two modules. The analysis results of MCODE software show that the hub genes of the blue module were mainly focused on the synthesis and function of neurotransmitter materials, such as cholesterol synthesis, amino acid synthesis, and steroid metabolism (Figure S7A). The results of KEGG and GO-BP enrichment analysis of the hub genes of the blue module also indicate that the functions of these genes are mainly focused on the pathways related to the synthesis and function of neurotransmitters such as amino acid metabolism, cholesterol metabolism, and steroid metabolism (Figures S7B and S7C). On the contrary, the functions of the core genes

of the turquoise module are mainly concentrated in immune and inflammatory pathways, such as activation of the immune response, T cell receptor signaling pathway, JAK–STAT signaling pathway, NF–kappa B signaling pathway, and positive regulation of T cell activation (Figures S8A–S8C).

Construction of the diagnostic model of RHNAG and RLNAG

Initially, a random forest analysis was executed on a set of 696 differentially expressed genes contrasting RHNAG and RLNAG, from which 51 pivotal genes were identified as key differentiators between the phenotypes of RHNAG and RLNAG (Fig. 3A). Subsequently, a scatterplot delineating the top 20 genes based on importance scores was generated (Fig. 3B). From this pool, 16 molecular entities boasting importance scores surpassing 50 were earmarked for ROC analysis. Out of these, 10 molecules demonstrating AUC values exceeding 0.9 were further shortlisted for lasso regression scrutiny (Fig. 3C). The subsequent lasso regression pinpointed four genes—HMGCR, CUX2, PLIN2, and PNPLA3—from the aforementioned 10 for model construction, with a lambda value set at 4 (Figs. 3D and E). Utilizing the nomogram approach, a linear diagnostic prediction model for RHNAG was formulated based on these four genes (Fig. 3G). According to the algorithm in the package, the scores corresponding to the expression of the four molecules are obtained from the following linear formulas, respectively.

$$(\text{score of PNPLA3} = -20.154(\exp \text{ of PNPLA3}) + 147.73;$$

$$\text{score of CUX2} = 19.133(\exp \text{ of CUX2}) - 19.4;$$

$$\text{score of PLIN2} = -21.473(\exp \text{ of PLIN2}) + 181.47;$$

$$\text{score of HMGCR} = -10.7(\exp \text{ of HMGCR}) + 120.85).$$

The formula for calculating the RHNAG diagnostic probability corresponding to the total score of the four molecules is as follows:

$$\text{probability} = 1.06777 * \ln(\text{totalscore}) - 19.4$$

For example, the expression of HMGCR, CUX2, PLIN2, and PNPLA3 in patient numbered GSM1611085 was 8.86, 4.4, 7.29, and 6.78, corresponding respectively to 73, 36, 74, and 91 points according to the formula. Therefore, the patient's overall score was 274, corresponding to the probability of 0.984 according to the formula that they had RHNAG. The final result shows

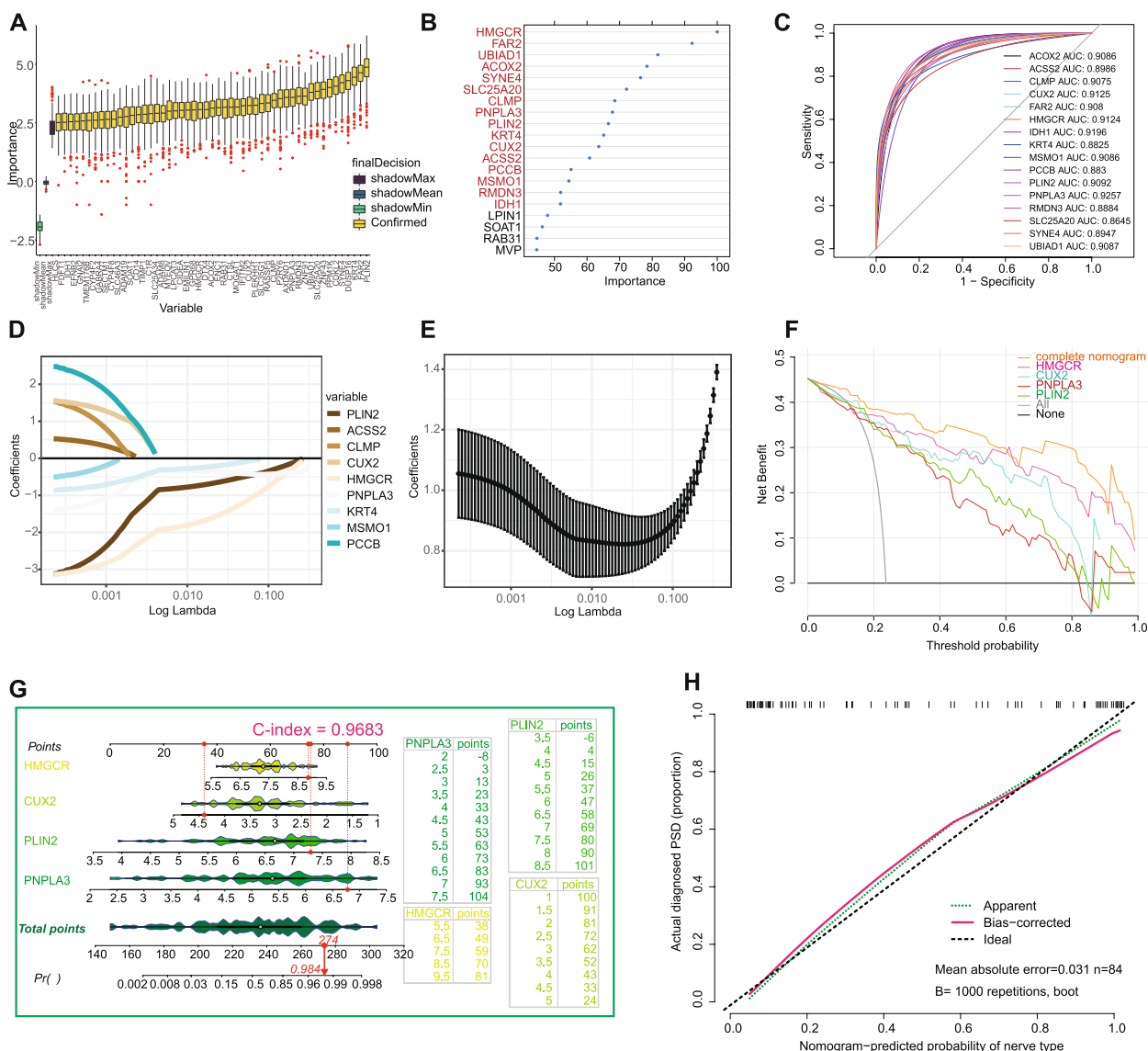


Fig. 3 Construction of the diagnostic model of RHNAG. **A** Histograms of standardized importance scores of 54 biomarkers that are very important for diagnostic typing obtained by random forest analysis; **B** The histogram of the biomarkers of the top 20 importance scores was obtained by random forest analysis. The red ones are markers with an importance score greater than 50; **C**, ROC analysis results of 16 biomarkers. **D** and **E** The result chart of LASSO regression analysis of 10 biomarkers with AUC value greater than 0.9. The red dotted line in the middle indicates that the lambda value of our selection is 4, so we have four genes to model; **F** Analysis of clinical decision curve of the whole model and single molecular model. The y-axis measures the net benefit. The black line represents the non-remission risk nomogram. The thin solid line represents the assumption that all patients are in RHNAG. The thick solid line represents the assumption that all patients are in RLNAG. The decision curve shows that if the threshold probability of a patient and a doctor is > 2% and < 100%, respectively, using this RHNAG nomogram in the current study to diagnose RHNAG adds more benefit than the intervention-all scheme or the intervention-none scheme; **G** Diagnostic model based on the expression of four biomarkers; on the right is the score corresponding to the expression of a specific biomarker; **H** Calibration curves of the RHNAG diagnostic nomogram in the primary cohorts

that the patient was indeed an RHNAG patient. Harrell’s C-index of this model was 0.9683. The results of BOOTSTRAT indicated a robust C-index of 0.908 (95% CI [0.901, 0.915]), suggesting high predictive accuracy and reliability of our model. The calibration curve analysis results show that the model’s predicted value was close

to the ideal (Fig. 3H). The above results indicate that the diagnostic model can predict whether a patient has the RHNAG type of rosacea. Finally, the results of DCA show that the benefits order of each of the four genes as a model to predict the clinical rate of return of RHNAG was HMGCGR > CUX2 > PLIN2 > PNPLA3. The clinical

benefit rate of the predictive model of the four genes was higher than that of each gene alone (Fig. 3F).

The model has high sensitivity and specificity in the diagnosis of NR

In line with the diagnostic categorization by three experienced dermatologists from Xiangya Hospital, 18 of the 46 rosacea patients were diagnosed with Neurogenic Rosacea (NR) while the remaining 28 were classified as Non-Neurogenic Rosacea (NNR) patients (Table S3). Visual documentation, comprising images of five NR and five NNR patients, revealed a notable absence of papules and pustules on the facial region of NR patients, replaced instead by pervasive flushing (Fig. 4A). Conversely, NNR patients exhibited pronounced papules and pustules either locally or encompassing the entire visible region (Fig. 4B). Finally, the total scores of 84 patients with rosacea were calculated according to the diagnostic model, and the ROC analysis was carried out according to the total scores. The results show that the AUC for the nomogram was 0.9376 (95% confidence interval (CI): 0.8898–0.9855; $P < 0.001$), the best cutoff value was 156, and the sensitivity and specificity were 0.974 and 0.861, respectively, in the total cohort of 84 patients with rosacea (Fig. 4C). Remarkably, an isolated ROC analysis on the cohort of 46 rosacea patients (NR: $n = 18$; NNR: $n = 28$), as assessed by the trio of dermatologists, showcased an AUC of 0.9023 for NR diagnosis (95% CI: 0.826–0.9925; $P < 0.001$). The prime cutoff value was determined to be 169, with a corresponding sensitivity and specificity of 0.926 and 0.870, respectively (Fig. 4D). Cumulatively, these findings underscore the robustness and elevated precision of our diagnostic model in clinical applications.

Prediction of Therapeutic Agents

Leveraging the cardinal genes identified within the blue and turquoise modules, potential therapeutic agents were prognosticated utilizing the Connectivity Map function on the clue.io platform. Agents with predictive scores beneath -0.6 were earmarked as potential treatment options. Our analysis indicated that the prospective therapeutic agents for RHNAG, as gleaned from the blue module, predominantly belonged to classes such as dopamine receptor antagonists, norepinephrine inhibitors, tricyclic antidepressants, opioid receptor antagonists, and sigma receptor antagonists (Refer to Figure S9A). Contrarily, the turquoise module's hub gene projections for RLNAG treatments were primarily aligned with mTOR inhibitors, JAK inhibitors, protein kinase inhibitors, MEK inhibitors, and STAT inhibitors (Figure S9B).

Discussion

In this study, leveraging neurovascular biomarkers, we stratified rosacea patients into two distinct groups: RHNAG and RLNAG. Intriguingly, the RLNAG group demonstrated a markedly heightened abundance of immune cells compared to the RHNAG cohort. Comprehensive GSEA, KEGG, and GO-BP enrichment analyses revealed that genes specific to RHNAG are predominantly enriched in pathways involving synaptic neurotransmitters, phospholipids, unsaturated fatty acids, amino acids, and cholesterol metabolism. Conversely, RLNAG-specific genes predominantly aligned with immune and inflammatory pathways. This molecular distinction was mirrored clinically; RHNAG patients typically presented absent of facial papules and pustules, but with pronounced flushing, burning, and tingling sensations. Additionally, a significant portion of the RHNAG group manifested neurotic symptoms including anxiety, depression, and neuroticism, and standard treatments often proved ineffectual. In stark contrast, the RLNAG patients, as observed in the clinical validation cohort, predominantly exhibited facial papules and pustules with a scant occurrence of neurological symptoms such as sensory aberrations and depression. Notably, their symptoms were ameliorated by conventional treatments. A salient observation from the clinical validation cohort, as assessed by a trio of seasoned dermatologists, was the unanimous categorization of all NR patients into the RHNAG group, while the majority of NNR patients predominantly clustered within the RLNAG cohort. Thus, we postulate that RHNAG patients share striking clinical parallels with NR patients, primarily presenting with neurological symptoms. In juxtaposition, RLNAG patients align closely with NNR patients, typified by pronounced inflammation and the presence of facial papules and pustules. This concordance underscores why the RHNAG diagnostic model boasts such elevated sensitivity and specificity in NR diagnosis, reinforcing its potential as a diagnostic surrogate for NR.

Cholesterol stands as the primary lipid collaborator with sphingolipids within membrane microdomains. Within the nervous system, cholesterol predominantly constitutes the primary lipid component of myelin (28%) [49]. Importantly, pivotal synaptic transmission processes, such as endocytosis, exocytosis, and the lateral diffusion of neurotransmitter receptors within the membrane, are profoundly modulated by cholesterol concentrations [50]. The functionality of neurotransmitter receptors is modulated by lipid domains [51], cholesterol [52], and sphingolipids [53]. Cholesterol's modulatory effects on membrane receptor function are largely attributed to either its direct receptor interaction or its overarching influence on the biophysical attributes

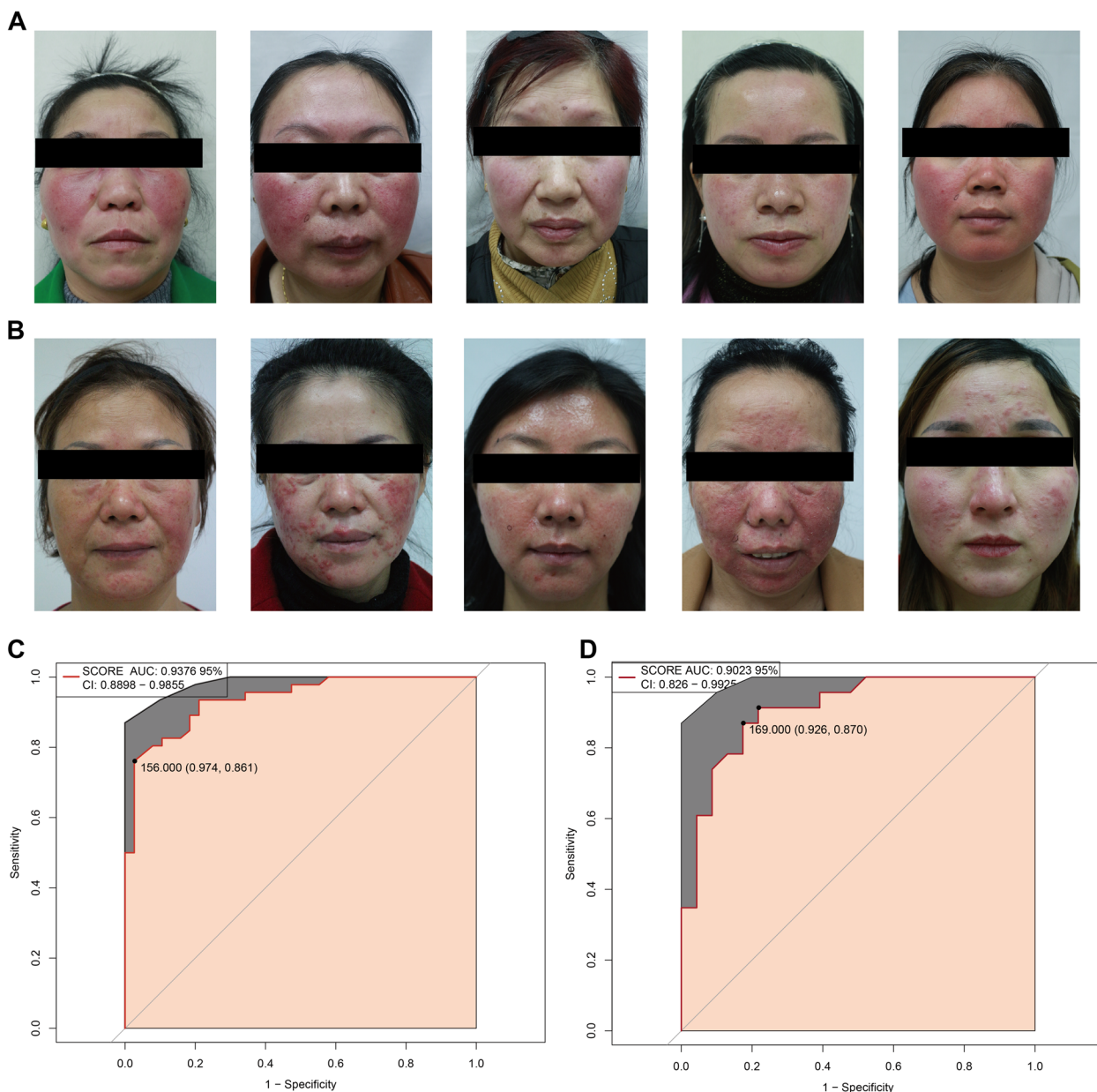


Fig. 4 The model has high sensitivity and specificity in the diagnosis of NR. **A** Typical facial photos of five patients with neurogenic rosacea in RHNA; **B** Typical facial photos of five patients with non-neurogenic rosacea in RLNA; **C** We performed ROC analysis on the total score of the model in a cohort of 84 patients with rosacea. **D** We performed ROC analysis on the total score of the model in a clinical validation cohort of 46 patients with rosacea (NR = 18, NNR = 28)

of the lipid bilayer of membranes [52, 53]. In the nervous milieu, lipids emerge as the predominant organic compounds. Glycerol phospholipids, sphingolipids, and cholesterol primarily constitute the neural lipid repertoire within both central and peripheral domains [49, 54]. Sphingolipids, glycerolipids, and their derivatives (namely glycosphingolipids, GSLs) are intricately associated with neurogenesis, synaptic transmission, and the synthesis,

functionality, and transport of neurotransmitter receptors [55]. Within the architecture of our NR diagnostic model, 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR) occupies a crucial position. HMGCR facilitates the conversion of (3S)-hydroxy-3-methylglutaryl-CoA (HMG-CoA) to mevalonic acid, representing the pivotal regulatory step in cholesterol synthesis and other isoprenoid formations, thereby orchestrating cellular

cholesterol equilibrium [56, 57]. Consequently, we postulate HMGCR's involvement in the synthesis of neurotransmitters and their receptor engagement, potentially mediated through cholesterol synthesis regulation, which may underscore its role in NR's pathogenesis. Further, the protein encoded by perilipin 2 (PLIN2) is aligned with the perilipin family, known for coating intracellular lipid storage vesicles, suggesting its potential as a lipid accumulation marker across diverse cellular environments and pathologies [58]. Another protein of interest, encoded by the patatin-like phospholipase domain containing 3 (PNPLA3), functions as a triacylglycerol lipase, mediating triacylglycerol hydrolysis in adipocytes and intricately involved in glycerol phospholipid biosynthesis [59]. In our investigations, these phospholipid biosynthesis-centric genes exhibited a positive correlation with the composite score of the NR diagnostic model, suggesting their potential to amplify the synthesis, transport, and functional attributes of neurotransmitters and their receptors, possibly through phospholipid and sphingomyelin synthesis, thereby elucidating their role in NR's etiology. The Cut-like homeobox 2 (CUX2) molecule, uniquely characterized by its negative correlation with the model score, encodes a protein housing three CUT domains and a homeodomain, both serving as DNA-binding moieties. This transcription factor critically governs neuronal proliferation and differentiation within the cerebral landscape, specifically modulating dendrite evolution, branching, dendritic spine genesis, and synaptogenesis within cortical strata II-III while demonstrating sequence-specific DNA binding. Markedly, CUX2 epitomizes a differentiation hallmark of glutamatergic pathways [60]. Suzuki et al. [61] established that a deficit in CUX2 significantly augments glutamatergic synaptic transmission within the hippocampal domain. Our findings resonate with this observation, underscoring that reduced CUX2 expression correlates with an elevated predisposition towards RHNAG diagnosis in rosacea-afflicted patients. This suggests that rosacea patients, due to the attenuated expression of CUX2, might experience amplified glutamate-mediated synaptic transmission, potentially triggering neurological manifestations that evolve into NR.

Our investigation represents a pioneering endeavor in the formulation of an auxiliary diagnostic model for NR predicated on biomarker profiling. Within a clinical milieu, it becomes feasible to ascertain the expression levels of these quartet molecules in peripheral blood. This allows for computation of the susceptibility to NR using the provided model equation. Subsequently, the derived prognostication can inform the prescription of appropriate therapeutic agents, an advancement that holds profound implications for the clinical diagnostic

and therapeutic paradigms of NR. To elucidate, RHNAG-afflicted individuals might benefit from the administration of tricyclic antidepressants such as nortriptyline and protriptyline (traditionally prescribed for depressive disorders), sigma receptor antagonists like rimcazole (prevalently utilized in schizophrenia management), and dopamine receptor antagonists, including trifluoperazine and fluphenazine (typically prescribed for schizophrenia). Conversely, for those diagnosed with RLNAG, therapeutic options might encompass JAK inhibitors like Ruxolitinib (predominantly used for primary myelodysplasia and post-thrombocytopenic myelodysplasia, with recent trials for autoimmune pathologies such as psoriasis) and mTOR inhibitors, e.g., everolimus (commonly incorporated in immunosuppressive regimens post-transplantation). Abnormal neurovascular regulation and imbalance of the inflammatory immune system are two intertwined pathogenetic factors of rosacea. Depending on the different clinical phenotypes of rosacea, the emphasis on these two mechanisms may differ. An intriguing query emerges: Why does NR predominantly exhibit neurological symptoms while manifesting attenuated inflammation? Studies by Kronfol et al. and Rothermundt et al. revealed that individuals without melancholic tendencies or depression often display proinflammatory states. Conversely, those with melancholic characteristics or diagnosed depression typically demonstrate diminished proinflammatory cytokine production [62–64]. Significantly, anti-inflammatory cytokines such as transforming growth factor (TGF)- β and IL-10 often present at elevated levels in major depression (MD) [65–67]. Melancholic and non-melancholic patients show different immune patterns. It's plausible that neurosystemic aberrations in RHNAG patients could be inhibiting systemic inflammation. On the other hand, RLNAG rosacea patients exhibit inflammatory levels that align with their clinical manifestations.

Nonetheless, our study is not devoid of limitations. A primary limitation is the paucity of valid samples, amounting to only 84, used in constructing the model. While we integrated all publicly available rosacea transcriptome data with our sequencing cohort, current constraints preclude further expansion of the sequencing sample size. Should additional rosacea data emerge in public repositories, we are poised to reassess our model. Another shortcoming stems from our model's reliance solely on mRNA expression levels within lesions. In a clinical setting, biomarkers sourced from blood or urine might be more feasible and palatable to patients. It's also noteworthy that our study observed five RHNAG patients manifesting pronounced papules and pustules, which are uncharacteristic for NR. This highlights the inherent challenge in perfectly aligning the NR phenotype through molecular modeling.

Nevertheless, our RHNAG diagnostic model retains commendable accuracy and sensitivity in diagnosing NR. Future endeavors encompassing more extensive sequencing of both NR and NNR patients might pave the way for a refined and more accurate NR diagnostic model.

In general, our pioneering approach combined machine learning with linear regression to devise a molecular diagnostic model for NR. This model, characterized by its high sensitivity and specificity, heralds a promising advancement in the diagnosis and treatment of NR. This could be particularly transformative in regions or among clinicians who might have previously overlooked NR.

Abbreviations

GEO	Gene Expression Omnibus
PCA	Principal Component Analysis
PLIN2	Perilipin 2
PNPLA3	Patatin Like Phospholipase Domain Containing 3
CUX2	Cut Like Homeobox 2
HMGCR	3-Hydroxy-3-Methylglutaryl-CoA Reductase

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-024-02008-0>.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.

Acknowledgements

The authors would like to thank the National Centre for Biotechnology Information staff. We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

Authors' contributions

Rui Mao: Conceptualization, Methodology, Software, Investigation, Visualization, Writing an original draft, Investigation. Ji Li: Conceptualization, Writing - original draft, Funding acquisition, Supervision.

Funding

This work was supported by the National Natural Science Funds for Distinguished Young Scholars (No. 82225039), the National Natural Science Foundation of China (No.82273557 and 8217344).

Availability of data and materials

Sequencing data from rosacea patients have been deposited in the genome sequence archive under accession number HRA000379 (<http://bigd.big.ac.cn/gsa-human/>). Clinical mugshots of 46 rosacea patients can be obtained by contacting the corresponding author.

Declarations

Ethics approval and consent to participate

The experiments conducted in this study were approved by the Clinical Medical Ethics Committee of Xiangya Hospital, Central South University (approval number: 201404361). Written informed consent was obtained from all participants prior to their participation in the study.

Consent for publication

All the authors in this study agreed to the publication of the manuscript. Informed consent for the publication of identifying images has been obtained from all subjects. It is important to note that no minors were included in the study cohort.

Competing interests

The authors declare no competing interests.

Received: 22 December 2023 Accepted: 9 September 2024

Published online: 13 September 2024

References

- Rainer BM, Fischer AH, da Luz Felipe Silva D, Kang S, Chien AL. Rosacea is associated with chronic systemic diseases in a skin severity-dependent manner: results of a case-control study. *J Am Acad Dermatol*. 2015;73(4):604–8.
- Thiboutot D, Anderson R, Cook-Bolden F, Draelos Z, Gallo RL, Granstein RD, et al. Standard management options for rosacea: The 2019 update by the National Rosacea Society Expert Committee. *J Am Acad Dermatol*. 2020;82(6):1501–10.
- Egeberg A, Fowler JF Jr, Gislason GH, Thyssen JP. Nationwide Assessment of Cause-Specific Mortality in Patients with Rosacea: A Cohort Study in Denmark. *Am J Clin Dermatol*. 2016;17(6):673–9.
- Vera N, Patel NU, Seminario-Vidal L. Rosacea Comorbidities. *Dermatol Clin*. 2018;36(2):115–22.
- Scharschmidt TC, Yost JM, Truong SV, Steinhoff M, Wang KC, Berger TG. Neurogenic rosacea: a distinct clinical subtype requiring a modified approach to treatment. *Arch Dermatol*. 2011;147(1):123–6.
- Parkins GJ, Maan A, Dawn G. Neurogenic rosacea: an uncommon and poorly recognized entity. *Clin Exp Dermatol*. 2015;40(8):930–1.
- Kim HO, Kang SY, Kim KE, Cho SY, Kim KH, Kim IH. Neurogenic rosacea in Korea. *J Dermatol*. 2021;48(1):49–55.
- Miyachi Y. Potential antioxidant mechanism of action for metronidazole: implications for rosacea management. *Adv Ther*. 2001;18(6):237–43.
- van Zuuren EJ, Fedorowicz Z. Low-Dose Isotretinoin: An Option for Difficult-to-Treat Papulopustular Rosacea. *J Invest Dermatol*. 2016;136(6):1081–3.
- Del Rosso JQ, Webster GF, Jackson M, Rendon M, Rich P, Torok H, et al. Two randomized phase III clinical trials evaluating anti-inflammatory dose doxycycline (40-mg doxycycline, USP capsules) administered once daily for treatment of rosacea. *J Am Acad Dermatol*. 2007;56(5):791–802.
- Yamasaki K, Di Nardo A, Bardan A, Murakami M, Ohtake T, Coda A, et al. Increased serine protease activity and cathelicidin promotes skin inflammation in rosacea. *Nat Med*. 2007;13(8):975–80.
- Lacey N, Delaney S, Kavanagh K, Powell FC. Mite-related bacterial antigens stimulate inflammatory cells in rosacea. *Br J Dermatol*. 2007;157(3):474–81.
- Yamasaki K, Kanada K, Macleod DT, Borkowski AW, Morizane S, Nakatsuji T, et al. TLR2 expression is increased in rosacea and stimulates enhanced serine protease production by keratinocytes. *J Invest Dermatol*. 2011;131(3):688–97.
- Dirschka T, Tronnier H, Fölster-Holst R. Epithelial barrier function and atopic diathesis in rosacea and perioral dermatitis. *Br J Dermatol*. 2004;150(6):1136–41.
- Choi JE, Di Nardo A. Skin neurogenic inflammation. *Semin Immunopathol*. 2018;40(3):249–59.
- Buddenkotte J, Steinhoff M. Recent advances in understanding and managing rosacea. *F1000Res*. 2018;7.
- Del Rosso JQ. Management of facial erythema of rosacea: what is the role of topical α -adrenergic receptor agonist therapy. *J Am Acad Dermatol*. 2013;69(6 Suppl 1):S44–56.
- Metzler-Wilson K, Toma K, Sammons DL, Mann S, Jurovcik AJ, Demidova O, et al. Augmented supraorbital skin sympathetic nerve activity responses to symptom trigger events in rosacea patients. *J Neurophysiol*. 2015;114(3):1530–7.

19. Steinhoff M, Schmelz M, Schaubert J. Facial Erythema of Rosacea - Aetiology, Different Pathophysiologies and Treatment Options. *Acta Derm Venereol.* 2016;96(5):579–86.
20. Ge L, Li Y, Wu Y, Fan Z, Song Z. Differential Diagnosis of Rosacea Using Machine Learning and Dermoscopy. *Clin Cosmet Investig Dermatol.* 2022;15:1465–73.
21. Binol H, Plotner A, Sopkovich J, Kaffenberger B, Niazi M, Gurcan MN. Ros-NET: A deep convolutional neural network for automatic identification of rosacea lesions. *Skin Res Technol.* 2020;26(3):413–21.
22. Aggarwal S. Data augmentation in dermatology image recognition using machine learning. *Skin Res Technol.* 2019;25(6):815–20.
23. Gillessen S, Armstrong A, Attard G, et al. Management of patients with advanced prostate cancer: report from the advanced prostate cancer consensus conference 2021 [J]. *Eur Urol.* 2022;82(1):115–41.
24. Tayob N, Kanwal F, Alsarraj A, et al. The performance of AFP, AFP-3, DCP as biomarkers for detection of hepatocellular carcinoma (HCC): a phase 3 biomarker study in the United States [J]. *Clin Gastroenterol Hepatol.* 2023;21(2):415–23. e4.
25. Xie Q, Xue W. IgE-Mediated food allergy: Current diagnostic modalities and novel biomarkers with robust potential [J]. *Crit Rev Food Sci Nutr.* 2023;63(29):10148–72.
26. Gallo RL, Granstein RD, Kang S, Mannis M, Steinhoff M, Tan J, et al. Standard classification and pathophysiology of rosacea: The 2017 update by the National Rosacea Society Expert Committee. *J Am Acad Dermatol.* 2018;78(1):148–55.
27. Schram AM, James WD. Neurogenic rosacea treated with endoscopic thoracic sympathectomy. *Arch Dermatol.* 2012;148(2):270–1.
28. Norquist JM, Watson DJ, Yu Q, Paolini JF, McQuarrie K, Santanello NC. Validation of a questionnaire to assess niacin-induced cutaneous flushing. *Curr Med Res Opin.* 2007;23(7):1549–60.
29. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics.* 2010;26(12):1572–3.
30. McLachlan GJ, Bean RW, Ng SK. Clustering. *Methods Mol Biol.* 2017;1526:345–62.
31. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
32. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;14:7.
33. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016;17(1):218.
34. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017;18(1):220.
35. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–7.
36. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
37. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* 2019;10(1):1523.
38. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003;4:2.
39. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med.* 2007;26(30):5512–28.
40. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33(1):1–22.
41. Kidd AC, McGettrick M, Tsim S, Halligan DL, Bylesjo M, Blyth KG. Survival prediction in mesothelioma using a scalable Lasso regression model: instructions for use and initial performance using clinical predictors. *BMJ Open Respir Res.* 2018;5(1):e000240.
42. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol.* 2016;34(18):2157–64.
43. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Crit Care Med.* 2007;35(9):2052–6.
44. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak.* 2008;8:53.
45. Lex A, Gehlenborg N, Strobelt H, Vuilleumot R, Pfister H. UpSet: Visualization of Intersecting Sets. *IEEE Trans Vis Comput Graph.* 2014;20(12):1983–92.
46. González-Santana A, Marrero-Hernández S, Dorta I, Hernández M, Pinto FM, Báez D, et al. Altered expression of the tachykinins substance P/neurokinin A/hemokinin-1 and their preferred neurokinin 1/neurokinin 2 receptors in uterine leiomyomata. *Fertil Steril.* 2016;106(6):1521–9.
47. Moparthi L, Survery S, Kreir M, Simonsen C, Kjellboom P, Högestätt ED, et al. Human TRPA1 is intrinsically cold- and chemosensitive with and without its N-terminal ankyrin repeat domain. *Proc Natl Acad Sci U S A.* 2014;111(47):16901–6.
48. Paulsen CE, Armache JP, Gao Y, Cheng Y, Julius D. Structure of the TRPA1 ion channel suggests regulatory mechanisms. *Nature.* 2015;520(7548):511–7.
49. Curatolo W. Glycolipid function. *Biochim Biophys Acta.* 1987;906(2):137–60.
50. Paulazo MA, Sodero AO. SIRT-1 Activity Sustains Cholesterol Synthesis in the Brain. *Neuroscience.* 2021;476:116–24.
51. Criado M, Eibl H, Barrantes FJ. Effects of lipids on acetylcholine receptor. Essential need of cholesterol for maintenance of agonist-induced state transitions in lipid vesicles. *Biochemistry.* 1982;21(15):3622–9.
52. Burger K, Gimpl G, Fahrenholz F. Regulation of receptor function by cholesterol. *Cell Mol Life Sci.* 2000;57(11):1577–92.
53. Baier CJ, Barrantes FJ. Sphingolipids are necessary for nicotinic acetylcholine receptor export in the early secretory pathway. *J Neurochem.* 2007;101(4):1072–84.
54. Fantini J, Garmy N, Mahfoud R, Yahi N. Lipid rafts: structure, function and role in HIV, Alzheimer's and prion diseases. *Expert Rev Mol Med.* 2002;4(27):1–22.
55. Sonnino S, Chigorno V. Ganglioside molecular species containing C18- and C20-sphingosine in mammalian nervous tissues and neuronal cell cultures. *Biochim Biophys Acta.* 2000;1469(2):63–77.
56. Luskey KL, Stevens B. Human 3-hydroxy-3-methylglutaryl coenzyme A reductase. Conserved domains responsible for catalytic activity and sterol-regulated degradation. *J Biol Chem.* 1985;260(18):10271–7.
57. Cuccioloni M, Mozzicafreddo M, Spina M, Tran CN, Falconi M, Eleuteri AM, et al. Epigallocatechin-3-gallate potently inhibits the in vitro activity of hydroxy-3-methyl-glutaryl-CoA reductase. *J Lipid Res.* 2011;52(5):897–907.
58. Kaushik S, Cuervo AM. Degradation of lipid droplet-associated proteins by chaperone-mediated autophagy facilitates lipolysis. *Nat Cell Biol.* 2015;17(6):759–70.
59. Romeo S, Kozlittina J, Xing C, Pertsemliadis A, Cox D, Pennacchio LA, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet.* 2008;40(12):1461–5.
60. Teissier A, Griveau A, Vigier L, Piolot T, Borello U, Pierani A. A novel transient glutamatergic population migrating from the pallial-subpallial boundary contributes to neocortical development. *J Neurosci.* 2010;30(31):10563–74.
61. Suzuki T, Tatsukawa T, Sudo G, Delandre C, Pai YJ, Miyamoto H, et al. CUX2 deficiency causes facilitation of excitatory synaptic transmission onto hippocampus and increased seizure susceptibility to kainate. *Sci Rep.* 2022;12(1):6505.
62. Rothermundt M, Arolt V, Fenker J, Gutbrodt H, Peters M, Kirchner H. Different immune patterns in melancholic and non-melancholic major depression. *Eur Arch Psychiatry Clin Neurosci.* 2001;251(2):90–7.
63. Rothermundt M, Arolt V, Peters M, Gutbrodt H, Fenker J, Kersting A, et al. Inflammatory markers in major depression and melancholia. *J Affect Disord.* 2001;63(1–3):93–102.
64. Kronfol Z. Immune dysregulation in major depression: a critical review of existing evidence. *Int J Neuropsychopharmacol.* 2002;5(4):333–43.
65. Howren MB, Lamkin DM, Suls J. Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis. *Psychosom Med.* 2009;71(2):171–86.

66. Dowlati Y, Herrmann N, Swardfager W, Liu H, Sham L, Reim EK, et al. A meta-analysis of cytokines in major depression. *Biol Psychiatry*. 2010;67(5):446–57.
67. Köhler CA, Freitas TH, Maes M, de Andrade NQ, Liu CS, Fernandes BS, et al. Peripheral cytokine and chemokine alterations in depression: a meta-analysis of 82 studies. *Acta Psychiatr Scand*. 2017;135(5):373–87.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.