

基于可解释深度学习神经模型PASNet的系统性红斑狼疮疾病特征基因挖掘与分析

张瑀欣¹, 周旭丹¹, 杨清琳², 孙向荣¹, 郭爽¹, 侯燕燕¹, 罗金¹, 陈长龙¹, 马国辉¹, 舒伟^{1*}

¹桂林医学院智能医学与生物技术学院, 广西 桂林

²广西国际商务职业技术学院数学艺术学院, 广西 南宁

收稿日期: 2025年3月24日; 录用日期: 2025年4月19日; 发布日期: 2025年4月27日

摘要

目的: 通过可解释深度学习神经模型(pathway-associated sparse deep neural network, PASNet)结合生物信息学方法寻找系统性红斑狼疮疾病的特征基因。方法: 从NCBI的GEO数据库收集SLE外周血转录组数据, 使用R语言筛选差异基因用于模型训练。基于PASNet模型构建用于SLE疾病研究的信号通路相关深度学习神经网络模型, 获得SLE疾病相关的关键基因和关键通路。LASSO回归和随机森林方法进一步筛选疾病特征基因。验证SLE疾病特征基因在外部数据集上的表达情况并验证特征基因的临床诊断效果。结果: 构建的神经网络模型在深度学习训练后, 测试集上验证的预测效果较好, ROC曲线下面积(area under the curve, AUC)为0.82。模型获得47个疾病相关的关键基因, 进一步机器学习算法分析及GSE61635数据验证筛选出ADHFE1、BCL11B、BLVRA和BOLAL四种SLE疾病特征基因。SLE单细胞数据验证结果显示, BCL11B和BLVRA基因与免疫细胞有不同程度的相关性, 用其构建的临床诊断模型能很好地区分疾病。结论: 通过深度学习结合生物信息学方法筛选的BCL11B和BLVRA基因可作为SLE的潜在生物标志物进一步研究。

关键词

深度学习, 生物信息学, 转录组, 系统性红斑狼疮

Mining and Analysis of Systemic Lupus Erythematosus Disease Characteristic Genes Based on Interpretable Deep Learning Neural Model PASNet Combined with Bioinformatics

*通讯作者。

文章引用: 张瑀欣, 周旭丹, 杨清琳, 孙向荣, 郭爽, 侯燕燕, 罗金, 陈长龙, 马国辉, 舒伟. 基于可解释深度学习神经模型 PASNet 的系统性红斑狼疮疾病特征基因挖掘与分析[J]. 临床医学进展, 2025, 15(4): 2834-2846.

DOI: 10.12677/acm.2025.1541247

Yuxin Zhang¹, Xudan Zhou¹, Qinglin Yang², Xiangrong Sun¹, Shuang Guo¹, Yanyan Hou¹, Jin Luo¹, Changlong Chen¹, Guohui Ma¹, Wei Shu^{1*}

¹School of Intelligent Medicine and Biotechnology, Guilin Medical University, Guilin Guangxi

²Department of Mathematics and Art, Guangxi International Business Vocational College, Nanning Guangxi

Received: Mar. 24th, 2025; accepted: Apr. 19th, 2025; published: Apr. 27th, 2025

Abstract

Objective: To identify characteristic genes of SLE using an interpretable deep learning neural model (PASNet) combined with bioinformatics methods. **Methods:** SLE peripheral blood transcriptome data were collected from the GEO database of NCBI, and differentially expressed genes were screened using R language for model training. A pathway-associated deep learning neural network model based on PASNet was constructed for SLE disease research to identify key genes and pathways related to SLE. LASSO regression and random forest methods were further employed to screen disease characteristic genes. The expression of these characteristic genes in external datasets was validated, and their clinical diagnostic efficacy was evaluated. **Results:** The constructed neural network model demonstrated good predictive performance on the test set after deep learning training, with an area under the ROC curve (AUC) of 0.82. The model identified 47 disease-related key genes, and machine learning algorithms analyze and GSE61635 data validation further screened out four SLE characteristic genes: ADHFE1, BCL11B, BLVRA, and BOLAL. SLE single-cell data validation results showed that BCL11B and BLVRA genes were correlated with immune cells to varying degrees, and the clinical diagnostic model constructed using these genes effectively distinguished the disease. **Conclusion:** The BCL11B and BLVRA genes screened through deep learning combined with bioinformatics methods can serve as potential biomarkers for SLE and warrant further research.

Keywords

Deep Learning, Bioinformatics, Transcriptome, Systemic Lupus Erythematosus

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

系统性红斑狼疮(Systemic lupus erythematosus, SLE)是一种复杂的自身免疫性疾病,其发病率在中青年女性群体中尤为显著,可导致多器官和系统受到炎症损害[1]。该疾病以免疫系统功能失调为特征,导致机体产生针对自身组织的异常免疫反应,进而引发多器官和系统的慢性炎症性损害[2]。研究表明,T细胞、B细胞、自然杀伤细胞(NK细胞)以及单核巨噬细胞等免疫细胞亚群在SLE患者体内均存在显著的功能异常。这些细胞不仅丧失了正常的免疫耐受机制,还会对自身抗原产生过度的免疫应答,通过释放促炎性细胞因子、形成免疫复合物等途径,导致全身性炎症反应和多器官损伤[3]。SLE的临床表现具有高度异质性,主要包括特征性皮肤病变、关节炎、肾脏损害(如狼疮性肾炎)、心肺功能异常以及血液系统受累等。其病因学机制涉及遗传易感性、环境触发因素、性激素水平以及免疫调节异常等多重因素的相互作用。流行病学数据显示,全球范围内SLE的患病人数已超过340万,且发病率呈逐年上升趋势,对公共卫生系统造成了沉重负担[4]。

SLE 的发病机制至今尚未完全阐明，寻找新的治疗策略和诊断方法成为研究的关键[5]。近年来，随着全基因组关联研究(Genome-wide association studies, GWAS)的不断发展，研究者们已经识别出多个与 SLE 相关的基因位点[6]-[8]。这些基因的变异可能导致免疫系统功能异常，进而引发自身免疫反应。一些特定的基因变异，如 HLA-DRB1，已被证实与 SLE 的风险显著增加有关。HLA-DRB1 是目前已知的最强的遗传风险因素之一，携带该基因变异的个体患病风险明显高于普通人群[9] [10]。目前除了已知的基因变异外，仍有大量潜在的致病基因尚待发掘。

近年来，随着人工智能技术的快速发展，其在医学领域的应用为疾病研究提供了新的方法论支持。机器学习与深度学习算法凭借其强大的数据处理能力和特征提取能力，在疾病诊断、预后预测和药物研发等方面展现出显著优势[11]。海量的基因组学、转录组学和表观遗传学数据为揭示疾病机制提供了丰富的信息资源。通过深度学习与生物信息学分析方法的有机结合，研究者能够从多维度、多层次解析疾病的分子特征，识别关键生物标志物，构建精准的疾病预测模型，为疾病的早期诊断和个体化治疗提供理论依据和技术支持[12]。这种基于人工智能的研究范式不仅提高了疾病研究的效率，也为深入理解疾病的复杂发病机制开辟了新的研究途径。

本研究利用深度学习结合生物信息学方法，基于 PASNet 构建可解释的深度学习神经网络模型从 GEO 数据集中找出 SLE 疾病的关键相关基因和通路，然后采用机器学习方法进一步筛选出 SLE 疾病的特征基因，并验证特征基因与免疫细胞之间相关性，构建基因预测模型评估其临床诊断效果，以期寻找 SLE 疾病相关的诊断生物标志物。

2. 材料与方法

2.1. 数据集下载及预处理

从 GEO 数据库以“SLE”和“Homo sapiens”为关键词载 SLE 转录组数据集。详细信息见表 1。使用 R 包 sva 对 GSE126307 [13]和 GSE81622 [14]数据集进行去批次处理合并，消除不同数据集之间的差异，用于后续分析。GSE61635 和 GSE135779 [15]作为外部数据集对特征基因进行后续验证。

Table 1. Basic information of the datasets in this study
表 1. 本研究所用数据主要信息

数据集 ID	测序平台	数据类型
GSE126307	GPL13369	微阵列数据
GSE81622	GPL10558	微阵列数据
GSE61635	GPL570	微阵列数据
GSE135779	GPL20301	高通量测序数据

2.2. SLE 疾病相关差异表达基因筛选

利用 R 语言 Limma 包对 SLE 样本和正常样本筛选差异基因(differentially expressed genes, DGEs)，筛选标准为 $P < 0.05$ ， $|\log_2FC| > 0.25$ ，用来消除差异不显著的基因，以作为后续深度学习模型的输入。通过 R 包“ggplot2”和“pheatmap”分别绘制火山图和热图。在后续试验的分析中，将数据集两组样本随机选取相同数量(SLE 组：34，正常人组：34)，以消除神经网络模型在运行过程中出现偏侧化问题。

2.3. 构建可解释性的通路相关深度学习神经网络模型

以 PASNet [16] 模型为基础, 构建一个与系统性红斑狼疮路径密切相关的深度学习网络模型。该模型包括四个不同的层次: 输入层(基因层)、路径层、隐藏层和输出层。经过数据预处理和分析后的数据集的 80% 数据集作为模型的训练数据集, 其余 20% 数据集作为模型的测试数据集。从 Reactome 数据库 [17] 中收集数据集中差异表达基因所参与的所有通路, 筛选出对应基因数大于 10 个的通路为 SLE 通路层的节点。建立基因 - 通路之间的二进制值矩阵, 有效地立基因层和通路层之间的关系。隐藏层代表了更复杂的生物途径和过程, 从而增强了网络设计的可解释性。采用 AdamW 作为优化器, 通过引入权重衰减机制, 有效解决了传统 Adam 优化器在某些场景下可能导致的模型性能波动问题。为进一步增强模型的鲁棒性和解释性, 我们在训练过程中引入了 dropout 技术。该技术通过在每次训练迭代中随机丢弃部分神经元, 降低了神经网络的复杂度, 有效防止了过拟合现象的发生。同时, 还采用了 L2 正则化方法, 通过在损失函数中加入权重的平方范数惩罚项, 约束模型的复杂度。这种双重正则化策略不仅有效抑制了模型在训练数据上的过拟合倾向, 显著提升了模型在未见数据上的泛化性能, 为模型的稳定性和可靠性提供了有力保障。

2.4. SLE 疾病特征基因的发现和验证

采取 LASSO 回归算法和随机森林算法对获取的 PASNet 模型发现的 TOP 50 基因和 TOP 10 通路基因进行进一步分析。分析过程均由 R 语言实现。使用 R 包 “glmnet”, 引入 L1 正则化系数拟合 LASSO 模型, 将基因数据通过 10 倍交叉验证确定 λ 的最佳值以获得特征基因。利用随机森林(Random Forest, RF) 分类器对已整合 SLE 相关基因进行重要性排序以获得特征基因, ntree 阈值设置为 1000。将 RF 和 LASSO 算法得到的基因取交集作为特征基因。使用外部数据集 GSE61635 和 GSE135779 验证特征基因表达水平。取表达趋势一致基因, 绘制受试者工作特征曲线(Receiver Operating Characteristic, ROC) [18] 评价对诊断 SLE 的准确性。

2.5. 特征基因在单细胞数据上表达分析

从 GSE135779 中获得 5 例正常样本和 7 例成人 SLE 样本 PBMC 表达矩阵。使用 R 包 Seurat 处理数据, 数据质量控制。以线粒体相关基因 <10% 为标准筛选细胞。对每个样本的前 2000 高变异基因进行归一化后分析。使用 ScaleData 功能对筛选出的前 2000 高变异基因进行缩放, 并使用 RunPCA 功能降低 PCA 维度。选择 $\text{dim} = 20$, 并使用 FindNeighbors 和 FindClusters 函数将细胞聚类为 20 个细胞群(分辨率 = 0.8)。然后使用 RunUMAP 功能进行可视化。

2.6. 临床风险模型构建

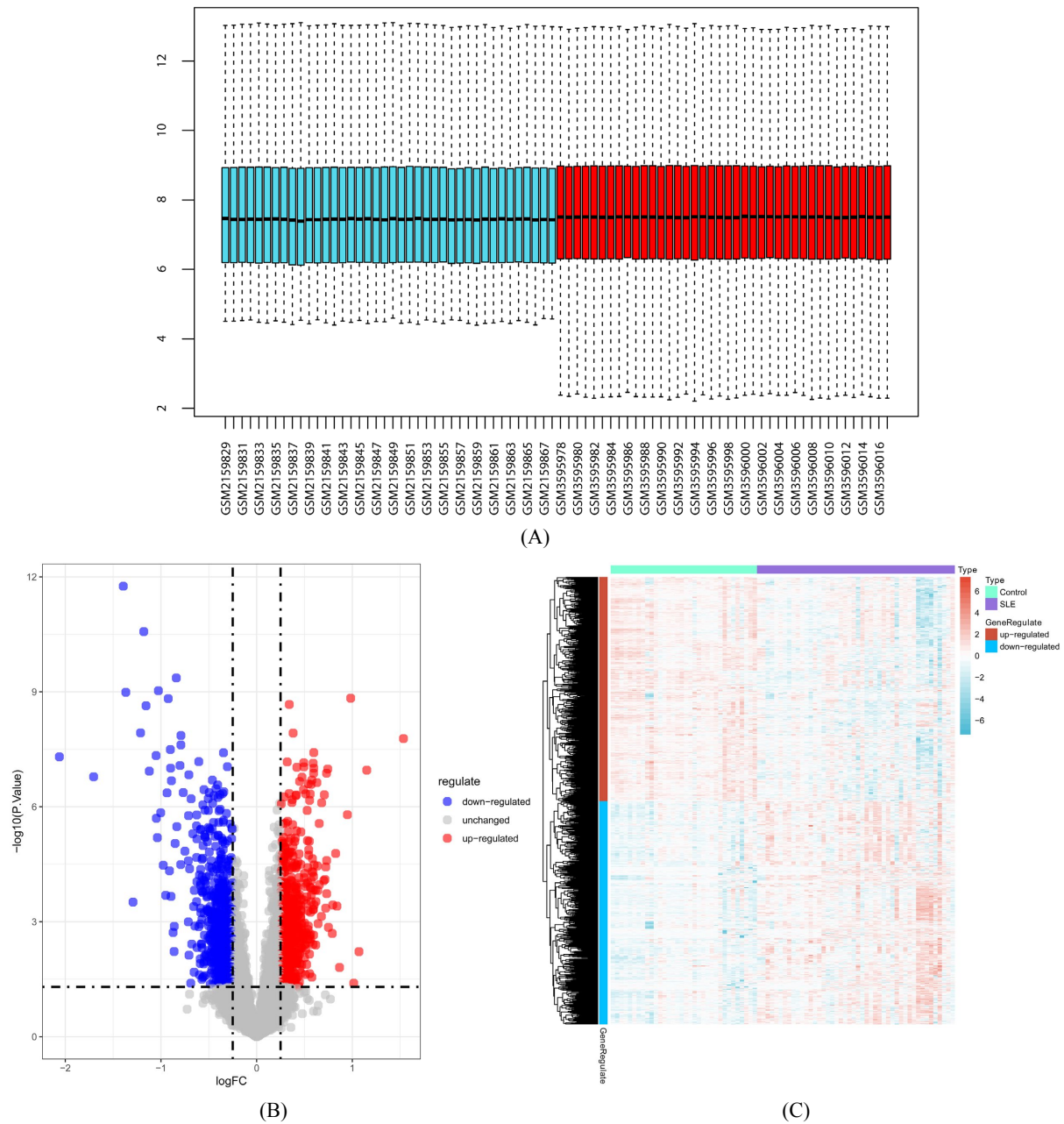
采用 Logistic 回归模型拟合的关键基因后, 构建 Nomogram 模型。Nomogram 模型是一种图形工具, 用于预测特定事件的概率, 结合多个变量的信息。其构建方法包括使用 R 中的 “rms” 包, 计算 ROC 曲线下面积(Area under the curve, AUC), 生成校准曲线评估模型诊断准确性。

3. 结果

3.1. SLE 数据集预处理和差异基因筛选

GSE81622、GSE126307 批次校正合并后共获得 80 个样本(control 组: 34; SLE 组: 46), 11,356 个基因表达数据(图 1(A))。使用 R 包 “Limma” 从 80 个样本中选出 1399 个差异基因, 其中包括上调基因 701, 下调基因 698。用 “ggplot2” 包和 “pheatmap” 包绘制火山图和热图(图 1(B), 图 1(C))。

GSE81622+GSE126307



Note: (A) The expression levels of the dataset after batch effect removal; (B) The volcano plot of differentially expressed genes in SLE; (C) The heatmap of differentially expressed genes in SLE.

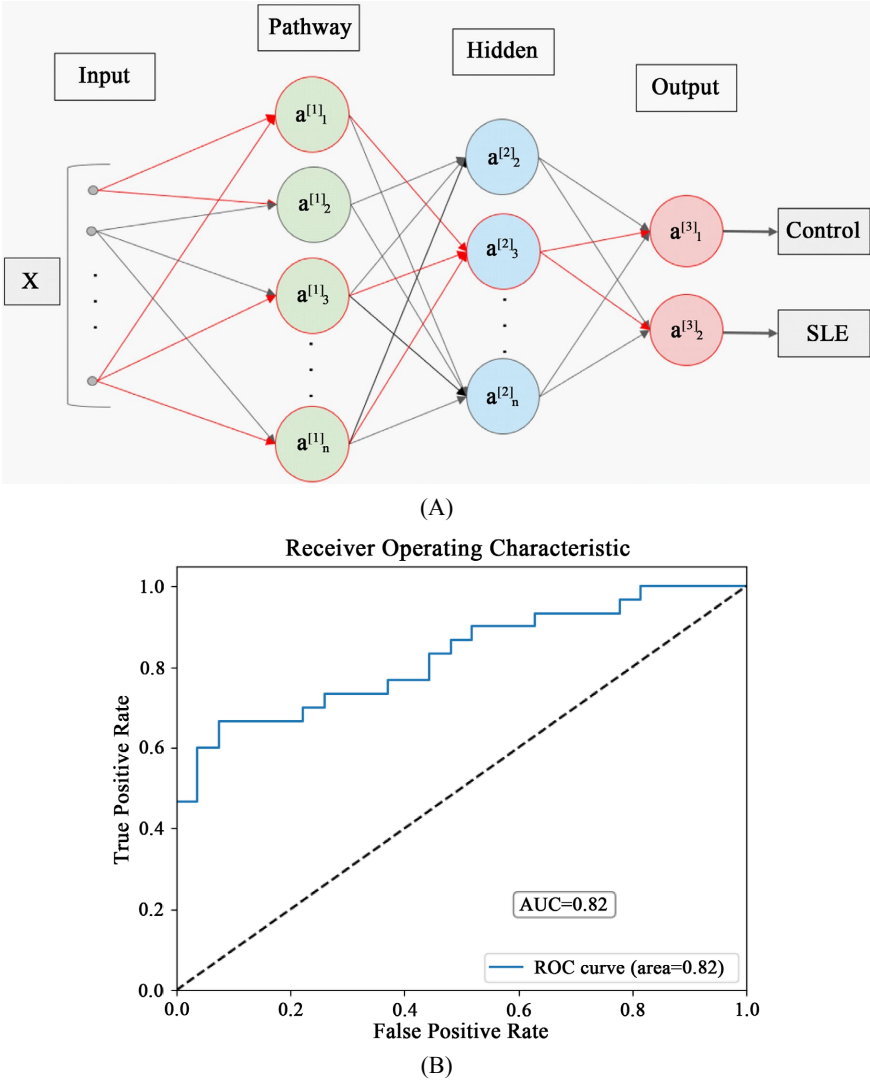
Figure 1. Data preprocessing and SLE related differentially expressed genes identification
图 1. 数据预处理和 SLE 差异表达基因的筛选

3.2. 可解释性深度学习模型构建和性能验证

为解决深度神经网络的“黑匣子”特性，结合基因 - 通路关系的生物学知识，以确定与 SLE 相关的基因和通路。基于 PASNet 模型的可解释方法将生物学知识整合到神经网络中，样本的基因表达数据作为输入，以样本状态作为标签对患者进行预测，模型框架见图 2(A)。模型训练完成后，提取模型的通路

层权重特征,通过提前加入的生物学先验,解释权重并用于后续的关键基因和通路的筛选。本研究将模型的学习率(η)设置为 0.0004, L2 正则化因子(λ)设置为 0.0007,中间层的 dropout 为 0.7 和 0.6, Epochs 上限为 3500。

为了评估模型对 SLE 和正常样本的预测性能,我们对测试数据集进行了十折交叉验证。这种评估方法通过将数据集分成十个子集,在每次迭代中使用九个子集进行训练,剩余一个子集进行测试,以进行全面的测试,以确保我们分析的结果稳健可靠。在测试集上,模型的 AUC 为 0.82 (图 2(B))。



Note: (A) Framework diagram of the model; (B) The area under the curve (AUC) value obtained from the receiver operating characteristic (ROC) analysis of the SLE test dataset is 0.82.

Figure 2. The framework and model performance of deep learning model PASNet used to discover SLE related genes
图 2. 用于研究 SLE 疾病相关基因的可解释深度学习神经网络模型 PASNet 框架及其性能

3.3. PASNet 模型发现的 SLE 疾病相关基因进一步筛选和外部数据集验证

通过运行神经网络模型,对模型得到权重参数进行分析,从而确定能够区分 SLE 和正常人的关键基因和关键通路途径。将 TOP50 基因和 TOP10 通路两部分基因进行整合,共获得 47 个基因。LASSO 回归

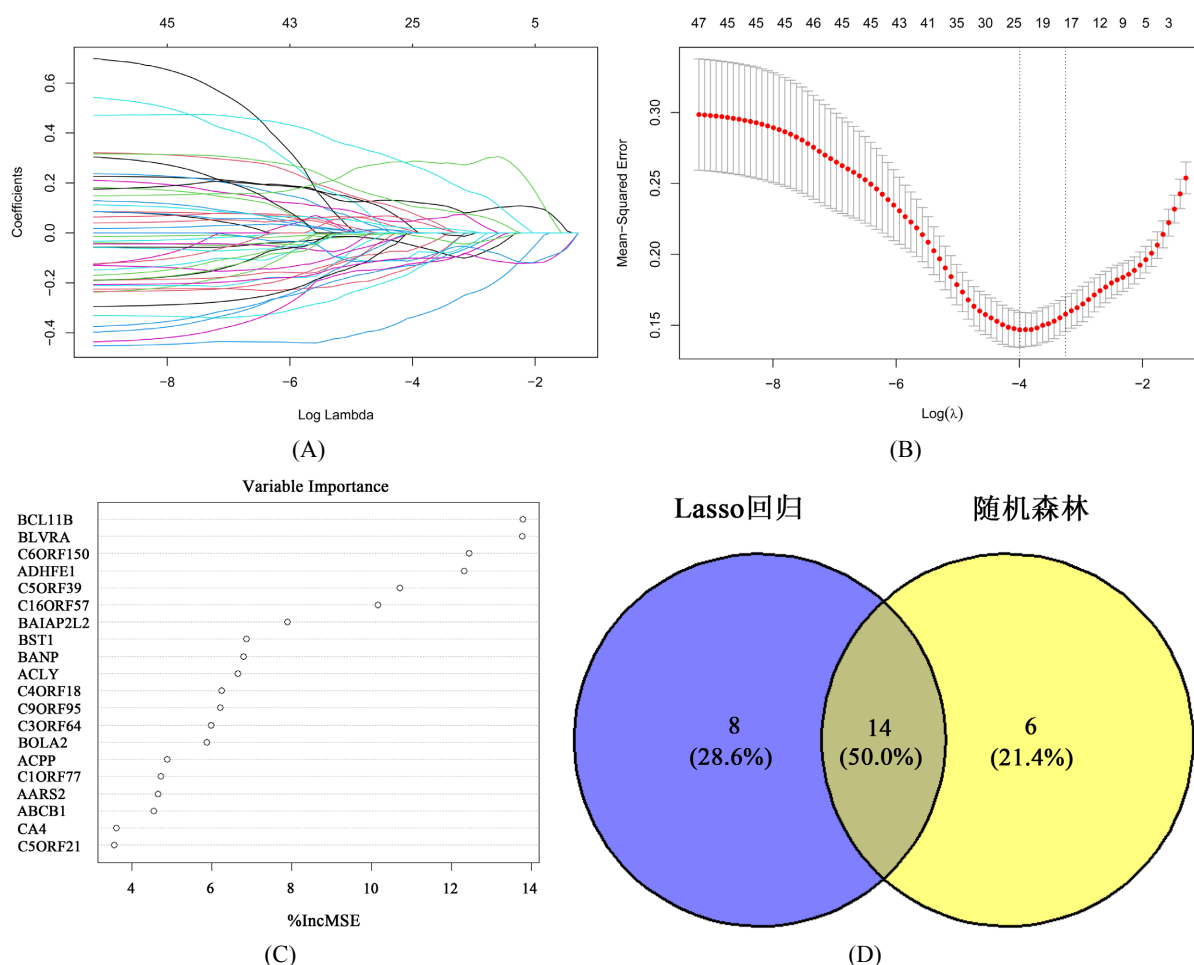
算法筛选出 22 个基因(图 3(A), 图 3(B)), 最佳 λ 值为 0.01844573。随机森林算法筛选前 20 基因, 随机森林决策树个数 $n_{tree} = 1000$ (图 3(C))。这两种算法结果取交集, 获得 14 个 SLE 相关基因(图 3(D))。以 GSE61635 作为外部数据集验证 14 个 SLE 相关基因的表达水平, 其中 ADHFE1、BCL11B、BLVRA、BOLAL 这几个特征基因在训练集和外部数据集中的表达趋势一致且具有显著差异(图 3(E), 图 3(F))。

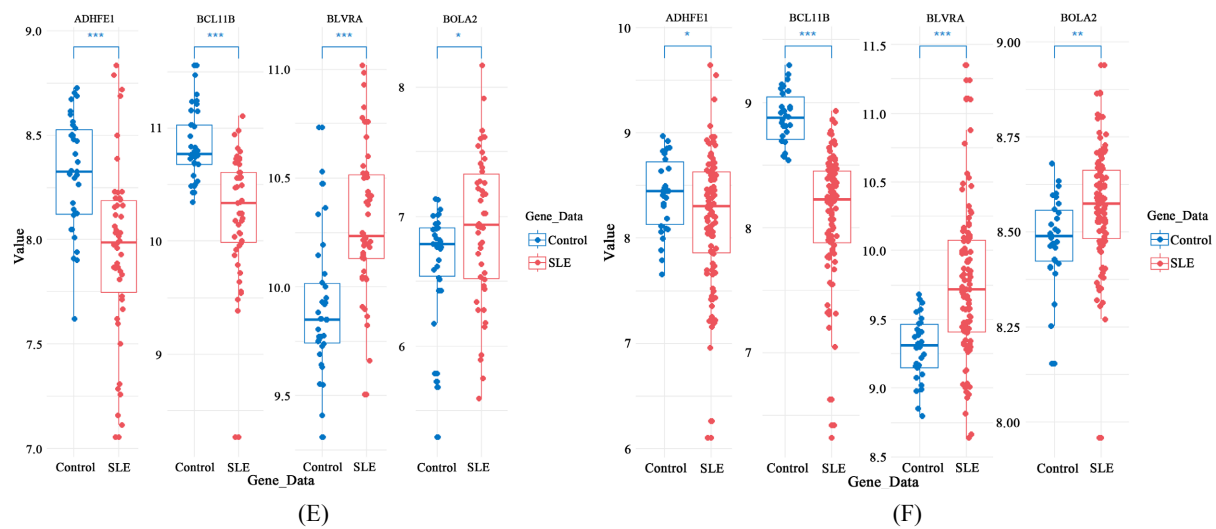
3.4. 特征基因的单细胞数据表达分析

为进一步验证 ADHFE1、BCL11B、BLVRA、BOLAL 基因在单细胞水平上的表达情况, 我们对收集的 SLE 单细胞 PBMC 测序数据进行分析。将 5 例正常样本和 7 例 SLE 样本进行聚类分簇, 共获得 20 个细胞簇。根据免疫细胞的特异性基因将 20 个细胞簇注释成 B 细胞、T 细胞、单核细胞、巨噬细胞和自然杀伤细胞 5 种细胞类型(图 4(A), 图 4(B))。BLVRA 在 SLE 样本中表达明显高于正常样本的表达, BCL11B 表达低于正常样本(图 4(C), 图 4(D)), 与 GSE61635 数据集中表达趋势一致。

3.5. 构建 SLE 疾病特征基因与临床诊断的列线图

为了探究最后获得的特征基因 BCL11B 和 BLVRA 是否能用于 SLE 疾病的诊断, 我们使用“rms”包构建列线图(图 5(A))。该临床模型在训练集上 AUC 值为 0.9035, 测试集上 AUC 值为 0.9139, 显示了其良好的预测能力(图 5(C), 图 5(D))。此外, 绘制校准曲线估算模型的诊断效果, 结果表明观察到与理想状态差距很小, 模型诊断效果好(图 5(B))。这些发现为 SLE 临床诊断提供了新线索。

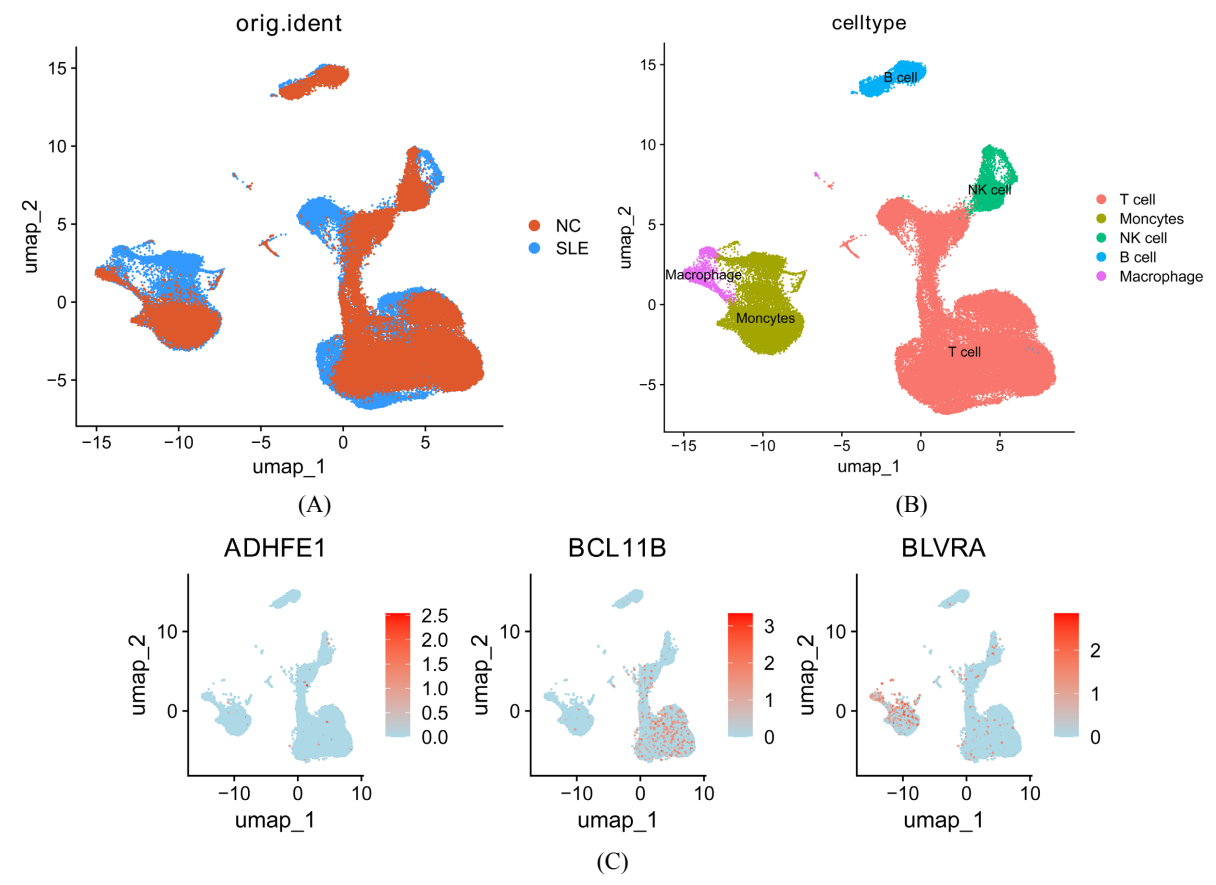


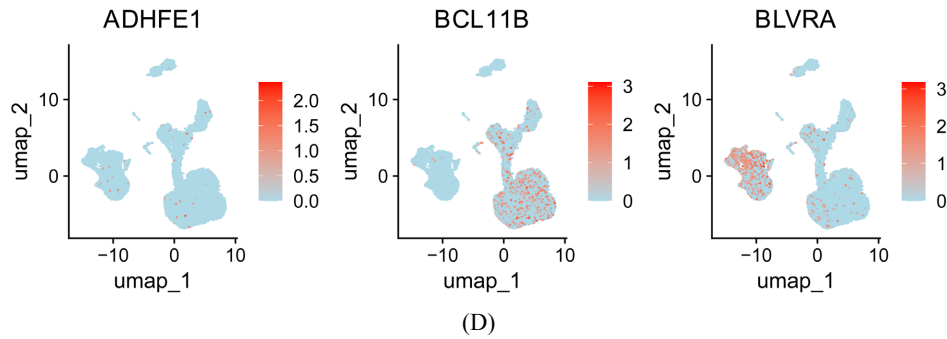


Note: (A) LASSO regression coefficient path plot. The vertical axis represents the regression coefficient values; the lower horizontal axis represents the Log Lambda values, and the upper horizontal axis indicates the number of genes with non-zero regression coefficients in the model as Log Lambda increases; (B) LASSO regression cross-validation plot. The vertical axis represents the likelihood deviation, where smaller values indicate better model fit; (C) Random forest gene importance plot; (D) Venn diagram showing the intersection of genes identified by the LASSO algorithm and the random forest algorithm; (E) Differential gene expression levels in the training set; (F) Differential gene expression levels in the GSE61635 dataset.

Figure 3. LASSO regression and random forest algorithm screening of SLE disease characteristic genes and validation of expression level in external datasets

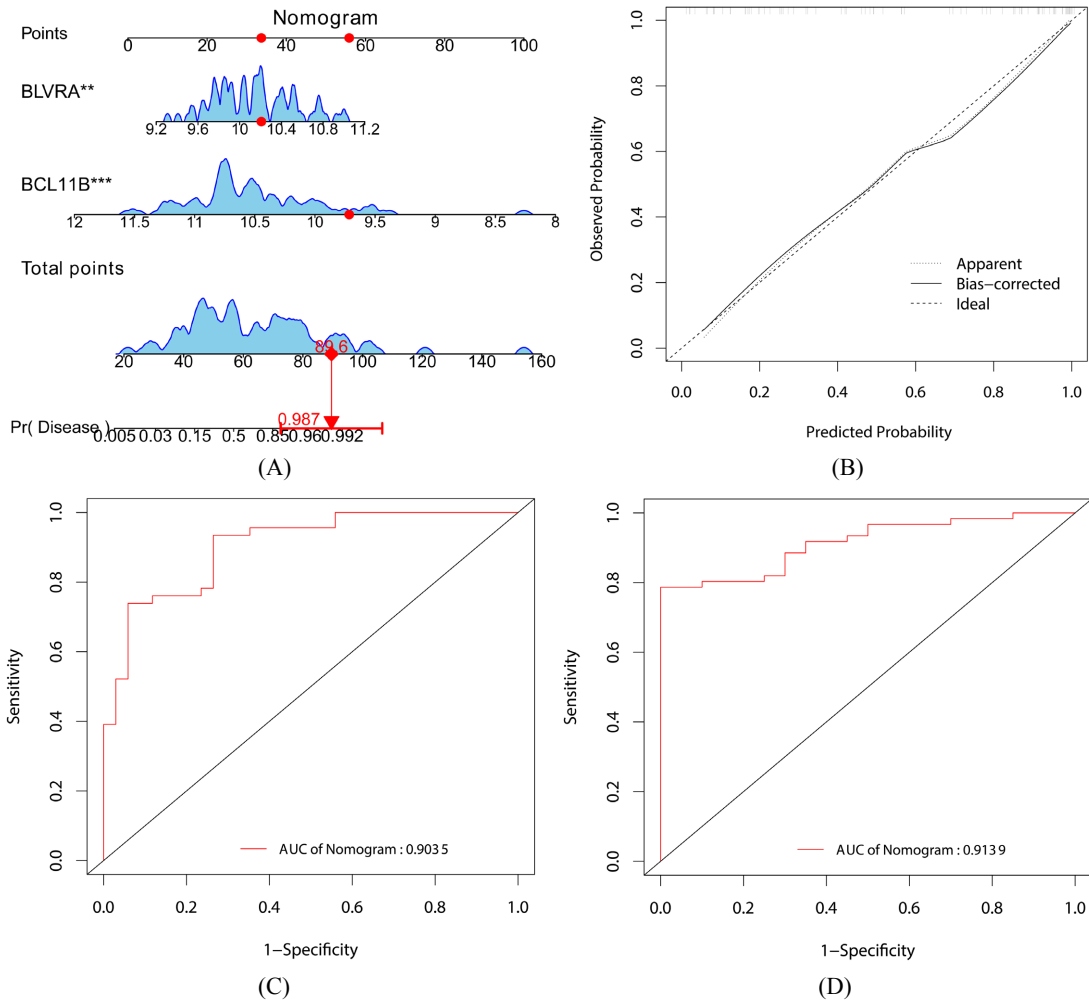
图 3. SLE 疾病特征基因的 LASSO 回归和随机森林算法筛选及其外部数据集表达水平验证





Note: (A) (B) UMAP analysis of single-cell data. Different groups and cells are marked with distinct colors; (C) Expression distribution of feature genes in normal samples; (D) Expression distribution of feature genes in SLE samples.

Figure 4. Expression analysis of characteristic genes in SLE PBMC single cells
图 4. 特征基因在 SLE PBMC 单细胞中的表达分析



Note: (A) Clinical prediction model diagram; (B) Calibration curve. The curve demonstrates that the accuracy of the prediction model closely aligns with the ideal scenario; (C) ROC curve and AUC value of the clinical prediction model in the training set. (D) Receiver operating characteristic (ROC) curve and area under the curve (AUC) of the clinical prediction model in the test set GSE61635.

Figure 5. Nomogram of SLE disease characteristic genes and clinical diagnosis
图 5. SLE 疾病特征基因与临床诊断的列线图

4. 讨论

系统性红斑狼疮(SLE)作为一种具有复杂发病过程和显著临床异质性的自身免疫性疾病,其确切病因至今尚不清楚。为了深入揭示 SLE 的致病分子机制,通过分析 SLE 患者的基因表达数据,挖掘出 SLE 发病过程中的关键分子信号和生物通路,本研究通过一种具有解释性深度学习网络模型结合生物信息学方法对 SLE 基因表达数据进行分析。该神经网络模型与以往传统的生物信息学和深度学习方法相比,其模型框架具有一定优势。首先,模型中引入一层细胞的信号转导通路,使得模型最后的预测结果具有一定的可解释性。其次,可以处理稀疏标记,能够在复杂的相互作用中提取真正的隐藏信号[12]。最后,通过增加样本量和维度,可以有效利用深度学习模型的力量来捕捉特征之间的非线性相互作用,从而提高预测准确性。

深度学习网络模型 PASNet 最初被设计用于分析和预测原发性脑癌患者的预后,与其他分类预测模型进行比较时,PASNet 模型预测准确性的 AUC 值达到了显著的 0.6622,位居所有相比较模型之首。所以本研究采用该模型分析了阿尔茨海默病帕金森病的关键基因,对这些疾病能进行精准的分类预测,其 AUC 值分别高达 0.82 和 0.73。所预测的生物通路和关键基因较多也与已知知识相符[19][20]。

本研究在 SLE 的 DGEs 中识别出 BLVRA 和 BCL11B 这 2 个关键基因,它们可能参与 SLE 发生和发展,同时在单细胞数据中发现关键基因都与免疫细胞存在一定的关系。并由此构建 SLE 临床预后模型列线图,ROC 曲线和校正曲线证实关键基因构建的诊断模型具有较好的预测准确性,AUC 值在 0.9 以上。

BLVRA 基因编码的胆绿素还原酶 A (Biliverdin reductase A)是一种关键的酶,在血红素代谢途径中起着至关重要的作用。属于胆绿素还原酶家族,是种非冗余 NADH 依赖性胆绿素还原酶,该家族的成员在 NADPH 或 NADH 存在的情况下催化胆绿素转化为胆红素来调节细胞氧化还原循环,参与了细胞内的氧化还原平衡调节[21]。胆红素作为一种强效的内源性抗氧化剂,能够清除自由基,保护细胞免受氧化应激的损伤[22]-[24]。研究表明,胆红素不仅具有抗氧化作用,还具有显著的抗炎和免疫调节功能,这些特性使其在预防和治疗多种疾病中具有潜在的应用价值,特别是在自身免疫性疾病、心血管疾病和神经退行性疾病等领域。

在免疫系统中,单核细胞和巨噬细胞是 BLVRA 基因表达的主要细胞类型。这些细胞在先天免疫和适应性免疫反应中发挥着核心作用。巨噬细胞通过吞噬病原体、呈递抗原和分泌细胞因子来调节免疫反应。研究发现,胆绿素还原酶 A 在巨噬细胞外表面的表达能够通过诱导抗炎细胞因子白细胞介素-10(IL-10)的合成,发挥抗炎和保护作用[25][26]。IL-10 是一种重要的抗炎细胞因子,能够抑制促炎因子的产生,调节免疫反应的平衡,从而减轻炎症反应和组织损伤[27]。

BCL11B 基因编码的锌指蛋白样转录因子是一种关键的调控分子,尤其在免疫系统中发挥着重要作用。作为 FOXP3 的辅助因子[28],BCL11B 在调节性 T 细胞(Treg 细胞)的功能维持和免疫耐受中扮演着核心角色[29]。FOXP3 是调节性 T 细胞的主要转录因子,而 BCL11B 通过与 FOXP3 的相互作用,能够增强其转录活性,从而促进 Treg 细胞的分化和功能[30]。Treg 细胞在抑制过度免疫反应和维持免疫稳态中至关重要,其功能异常与多种自身免疫性疾病的发生密切相关。在 B 细胞和 T 细胞的分化与增殖过程中,BCL11B 基因也发挥着重要的调节作用[31][32]。研究表明,BCL11B 通过调控一系列下游靶基因的表达,影响 B 细胞和 T 细胞的发育和功能。例如,在 T 细胞发育过程中,BCL11B 参与调控胸腺细胞的分化,确保 T 细胞受体(TCR)信号通路的正常传递[33]。

BCL11B 基因的功能缺失或突变会导致严重的免疫系统异常。在 BCL11B 基因缺陷的小鼠模型中,研究人员观察到显著的自身免疫表型,包括多器官炎症和调节性 T 细胞功能的严重紊乱[34]。这些小鼠表现出免疫耐受的破坏,导致免疫系统对自身抗原的攻击,进而引发自身免疫性疾病。这一发现进一步

证实 BCL11B 在维持免疫稳态中的关键作用。在人类疾病中, BCL11B 基因的罕见变异与某些原发性免疫缺陷病和自身免疫性疾病的发生相关。例如, 在儿童系统性红斑狼疮患者中, 研究发现 BCL11B 基因存在罕见性变异, 这些变异可能导致基因功能的改变, 从而增加儿童对系统性红斑狼疮的易感性[35]。

5. 结论

本研究基于可解释深度学习模型, 结合生物信息学方法对 SLE 的转录组数据进行系统性分析, 筛选出 BCL11B 和 BLVRA 作为 SLE 的关键疾病特征基因。研究表明, 这两个基因在 SLE 的发生和发展中具有重要的调控作用, 且与 SLE 的疾病诊断密切相关。然而, 由于本研究主要依赖于计算生物学分析, 缺乏分子生物学实验验证, BCL11B 和 BLVRA 在 SLE 中的具体分子机制尚未完全阐明。因此, 未来需要通过分子生物学实验进一步验证这两个基因的功能及其作用机制。

基金项目

国家自然科学基金项目(32060157); 国家级大学生创新创业训练计划(202210601048)。

参考文献

- [1] Accapezzato, D., Caccavale, R., Paroli, M.P., Gioia, C., Nguyen, B.L., Spadea, L., *et al.* (2023) Advances in the Pathogenesis and Treatment of Systemic Lupus Erythematosus. *International Journal of Molecular Sciences*, **24**, Article 6578. <https://doi.org/10.3390/ijms24076578>
- [2] Tsokos, G.C. (2020) Autoimmunity and Organ Damage in Systemic Lupus Erythematosus. *Nature Immunology*, **21**, 605-614. <https://doi.org/10.1038/s41590-020-0677-6>
- [3] Herrada, A.A., Escobedo, N., Iruretagoyena, M., Valenzuela, R.A., Burgos, P.I., Cuitino, L., *et al.* (2019) Innate Immune Cells' Contribution to Systemic Lupus Erythematosus. *Frontiers in Immunology*, **10**, Article 772. <https://doi.org/10.3389/fimmu.2019.00772>
- [4] Siegel, C.H. and Sammaritano, L.R. (2024) Systemic Lupus Erythematosus: A Review. *JAMA*, **331**, 1480-1491. <https://doi.org/10.1001/jama.2024.2315>
- [5] Parodis, I. and Sjöwall, C. (2024) Immune Mechanisms and Biomarkers in Systemic Lupus Erythematosus. *International Journal of Molecular Sciences*, **25**, Article 9965. <https://doi.org/10.3390/ijms25189965>
- [6] Yin, X., Kim, K., Suetsugu, H., Bang, S., Wen, L., Koido, M., *et al.* (2021) Meta-Analysis of 208370 East Asians Identifies 113 Susceptibility Loci for Systemic Lupus Erythematosus. *Annals of the Rheumatic Diseases*, **80**, 632-640. <https://doi.org/10.1136/annrheumdis-2020-219209>
- [7] Fazel-Najafabadi, M., Looger, L., Reddy-Rallabandi, H., *et al.* (2023) A Multilayered Post-GWAS Analysis Pipeline Defines Functional Variants and Target Genes for Systemic Lupus Erythematosus (SLE). medRxiv: The Preprint Server for Health Sciences. <https://doi.org/10.1101/2023.04.07.23288295>
- [8] Ghodke-Puranik, Y., Olferiev, M. and Crow, M.K. (2024) Systemic Lupus Erythematosus Genetics: Insights into Pathogenesis and Implications for Therapy. *Nature Reviews Rheumatology*, **20**, 635-648. <https://doi.org/10.1038/s41584-024-01152-2>
- [9] Wang, T., Wang, H., Qiu, L., Wu, L., Ling, H., Xue, Y., *et al.* (2022) Association of HLA-DR1, HLA-DR13, and HLA-DR16 Polymorphisms with Systemic Lupus Erythematosus: A Meta-Analysis. *Journal of Immunology Research*, **2022**, Article 8140982. <https://doi.org/10.1155/2022/8140982>
- [10] Selvaraja, M., Chin, V.K., Abdullah, M., Arip, M. and Amin-Nordin, S. (2021) HLA-DRB1*04 as a Risk Allele to Systemic Lupus Erythematosus and Lupus Nephritis in the Malay Population of Malaysia. *Frontiers in Medicine*, **7**, Article 598665. <https://doi.org/10.3389/fmed.2020.598665>
- [11] Rauschert, S., Raubenheimer, K., Melton, P.E. and Huang, R.C. (2020) Machine Learning and Clinical Epigenetics: A Review of Challenges for Diagnosis and Classification. *Clinical Epigenetics*, **12**, Article No. 51. <https://doi.org/10.1186/s13148-020-00842-4>
- [12] Karim, M.R., Islam, T., Shajalal, M., Beyan, O., Lange, C., Cochez, M., *et al.* (2023) Explainable AI for Bioinformatics: Methods, Tools and Applications. *Briefings in Bioinformatics*, **24**, bbad236. <https://doi.org/10.1093/bib/bbad236>
- [13] Jiang, S.H., Athanasopoulos, V., Ellyard, J.I., Chuah, A., Cappello, J., Cook, A., *et al.* (2019) Functional Rare and Low Frequency Variants in *BLK* and *BANK1* Contribute to Human Lupus. *Nature Communications*, **10**, Article No. 2201. <https://doi.org/10.1038/s41467-019-10242-9>

- [14] Zhu, H., Mi, W., Luo, H., Chen, T., Liu, S., Raman, I., *et al.* (2016) Whole-Genome Transcription and DNA Methylation Analysis of Peripheral Blood Mononuclear Cells Identified Aberrant Gene Regulation Pathways in Systemic Lupus Erythematosus. *Arthritis Research & Therapy*, **18**, Article No. 162. <https://doi.org/10.1186/s13075-016-1050-x>
- [15] Nehar-Belaid, D., Hong, S., Marches, R., Chen, G., Bolisetty, M., Baisch, J., *et al.* (2020) Mapping Systemic Lupus Erythematosus Heterogeneity at the Single-Cell Level. *Nature Immunology*, **21**, 1094-1106. <https://doi.org/10.1038/s41590-020-0743-0>
- [16] Hao, J., Kim, Y., Kim, T. and Kang, M. (2018) PASNet: Pathway-Associated Sparse Deep Neural Network for Prognosis Prediction from High-Throughput Data. *BMC Bioinformatics*, **19**, Article No. 510. <https://doi.org/10.1186/s12859-018-2500-z>
- [17] Fabregat, A., Korninger, F., Viteri, G., Sidiropoulos, K., Marin-Garcia, P., Ping, P., *et al.* (2018) Reactome Graph Database: Efficient Access to Complex Pathway Data. *PLOS Computational Biology*, **14**, e1005968. <https://doi.org/10.1371/journal.pcbi.1005968>
- [18] Mandrekar, J.N. (2010) Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology*, **5**, 1315-1316. <https://doi.org/10.1097/jto.0b013e3181ec173d>
- [19] Sun, X., Shu, W., Zhang, Y., Huang, X., Liu, J., Liu, Y., *et al.* (2023) Identification of Alzheimer's Disease Associated Genes through Explicable Deep Learning and Bioinformatic. 2023 *IEEE 4th International Conference on Pattern Recognition and Machine Learning (PRML)*, Urumqi, 4-6 August 2023, 320-327. <https://doi.org/10.1109/prml59573.2023.10348276>
- [20] Zhang, Y., Sun, X., Zhang, P., Zhou, X., Huang, X., Zhang, M., *et al.* (2024) Identification of Parkinson's Disease Associated Genes through Explicable Deep Learning and Bioinformatic. In: Huang, D.S., Premaratne, P. and Yuan, C., Eds., *Communications in Computer and Information Science*, Springer, 136-146. https://doi.org/10.1007/978-981-97-0903-8_14
- [21] Kim, S.Y., Kang, H.T., Choi, H.R. and Park, S.C. (2011) Biliverdin Reductase A in the Prevention of Cellular Senescence against Oxidative Stress. *Experimental and Molecular Medicine*, **43**, 15-23. <https://doi.org/10.3858/emm.2011.43.1.002>
- [22] Vitek, L., Hinds, T.D., Stec, D.E. and Tiribelli, C. (2023) The Physiology of Bilirubin: Health and Disease Equilibrium. *Trends in Molecular Medicine*, **29**, 315-328. <https://doi.org/10.1016/j.molmed.2023.01.007>
- [23] Canesin, G., Hejazi, S.M., Swanson, K.D. and Wegiel, B. (2020) Heme-Derived Metabolic Signals Dictate Immune Responses. *Frontiers in Immunology*, **11**, Article 66. <https://doi.org/10.3389/fimmu.2020.00066>
- [24] Adin, C.A. (2021) Bilirubin as a Therapeutic Molecule: Challenges and Opportunities. *Antioxidants*, **10**, Article 1536. <https://doi.org/10.3390/antiox10101536>
- [25] Hu, Z., Pei, G., Wang, P., Yang, J., Zhu, F., Guo, Y., *et al.* (2015) Biliverdin Reductase a (BVRA) Mediates Macrophage Expression of Interleukin-10 in Injured Kidney. *International Journal of Molecular Sciences*, **16**, 22621-22635. <https://doi.org/10.3390/ijms160922621>
- [26] Gibbs, P.E.M., Miralem, T. and Maines, M.D. (2015) Biliverdin Reductase: A Target for Cancer Therapy? *Frontiers in Pharmacology*, **6**, Article 119. <https://doi.org/10.3389/fphar.2015.00119>
- [27] Lobo-Silva, D., Carriche, G.M., Castro, A.G., Roque, S. and Saraiva, M. (2016) Balancing the Immune Response in the Brain: IL-10 and Its Regulation. *Journal of Neuroinflammation*, **13**, Article No. 297. <https://doi.org/10.1186/s12974-016-0763-8>
- [28] Hasan, S.N., Sharma, A., Ghosh, S., Hong, S., Roy-Chowdhuri, S., Im, S.-H., *et al.* (2019) Bcl11B Prevents Catastrophic Autoimmunity by Controlling Multiple Aspects of a Regulatory T Cell Gene Expression Program. *Science Advances*, **5**, eaaw0706. <https://doi.org/10.1126/sciadv.aaw0706>
- [29] Wing, J.B., Tanaka, A. and Sakaguchi, S. (2019) Human FOXP3⁺ Regulatory T Cell Heterogeneity and Function in Autoimmunity and Cancer. *Immunity*, **50**, 302-316. <https://doi.org/10.1016/j.immuni.2019.01.020>
- [30] Ono, M. (2020) Control of Regulatory T-Cell Differentiation and Function by T-Cell Receptor Signalling and Foxp3 Transcription Factor Complexes. *Immunology*, **160**, 24-37. <https://doi.org/10.1111/imm.13178>
- [31] García-Aznar, J.M., Alonso Alvarez, S. and Bernal del Castillo, T. (2024) Pivotal Role of BCL11B in the Immune, Hematopoietic and Nervous Systems: A Review of the Bcl11B-Associated Phenotypes from the Genetic Perspective. *Genes & Immunity*, **25**, 232-241. <https://doi.org/10.1038/s41435-024-00263-w>
- [32] Naito, T., Tanaka, H., Naoue, Y. and Taniuchi, I. (2011) Transcriptional Control of T-Cell Development. *International Immunology*, **23**, 661-668. <https://doi.org/10.1093/intimm/dxr078>
- [33] Drashansky, T.T., Helm, E.Y., Curkovic, N., Cooper, J., Cheng, P., Chen, X., *et al.* (2021) BCL11B Is Positioned Upstream of PLZF and Ror γ t to Control Thymic Development of Mucosal-Associated Invariant T Cells and MAIT17 Program. *iScience*, **24**, Article102307. <https://doi.org/10.1016/j.isci.2021.102307>
- [34] Drashansky, T.T., Helm, E., Huo, Z., Curkovic, N., Kumar, P., Luo, X., *et al.* (2019) Bcl11B Prevents Fatal Autoimmunity

by Promoting T_{reg} Cell Program and Constraining Innate Lineages in T_{reg} Cells. *Science Advances*, **5**, eaaw0480. <https://doi.org/10.1126/sciadv.aaw0480>

- [35] Wu, C., Fan, W., Yang, H., Chu, P., Liao, P., Chen, L., *et al.* (2023) Contribution of Genetic Variants Associated with Primary Immunodeficiencies to Childhood-Onset Systemic Lupus Erythematosus. *Journal of Allergy and Clinical Immunology*, **151**, 1123-1131. <https://doi.org/10.1016/j.jaci.2022.12.807>