

NFKBIA Centered Monocyte States Define a Single Cell Validated Oxidative Stress Signature for Sepsis

Kun Liu

The Fifth Hospital Of Xiamen

Qun Ye

First Affiliated Hospital of Xiamen University

kunhai Xiao

31315894@qq.com

The Fifth Hospital Of Xiamen

Research Article

Keywords: sepsis, oxidative stress, transcriptomics, machine learning, monocytes, single-cell RNA-seq, biomarker

Posted Date: September 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10964165/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

NFKBIA Centered Monocyte States Define a Single Cell Validated Oxidative Stress Signature for Sepsis

Kun Liu¹, Qun Ye², Kunhai Xiao^{1*}

¹ The Fifth Hospital of Xiamen, Xiamen, China

² The First Affiliated Hospital of Xiamen University, Xiamen, China

*Corresponding author: Kunhai Xiao, m18670377420@163.com

Running title: Oxidative stress signature for sepsis

Abstract

Early diagnosis of sepsis remains difficult because clinical features, microbiology, and conventional inflammatory biomarkers are heterogeneous. We reanalyzed public blood transcriptomic datasets to test whether oxidative stress-related expression patterns could provide an interpretable host-response signal for sepsis. Oxidative stress-related genes were profiled in a discovery cohort, summarized with an oxidative stress score, and evaluated through comparative diagnostic modeling and external validation. MAPK1, NFKBIA, and RAF1 formed a compact three-gene signature and were incorporated into an interpretable nomogram. In the GSE6535 training cohort, the three-gene nomogram achieved good discrimination (AUC, 0.860), while performance in the independent GSE54514 cohort was acceptable but moderate (AUC, 0.708), supporting cautious cross-dataset transportability. Immune and pathway analyses associated the signature with neutrophil-related inflammatory activity, translational regulation, and cellular stress-response programs. Complementary single-cell RNA-seq analysis localized oxidative stress activity and MAPK1/NFKBIA/RAF1 expression to discrete monocyte states, with coordinated NF-kB/NRF2-related activity, inflammatory ligand signals, trajectory patterns, and NFKBIA-centered network features supporting biological plausibility. These findings suggest that the MAPK1/NFKBIA/RAF1 signature may provide an interpretable auxiliary host-response signal for sepsis diagnosis, while prospective clinical and experimental validation remains required.

Keywords: sepsis; oxidative stress; transcriptomics; machine learning; monocytes; single-cell RNA-seq; biomarker.

Abbreviations: AUC, area under the receiver operating characteristic curve; AUC_{cell}, area under the recovery curve for single-cell gene-set activity; CI, confidence interval; DCA, decision curve analysis; GEO, Gene Expression Omnibus; GLM, generalized linear model; NHC, non-acutely ill healthy control; OS, oxidative stress; PBMC, peripheral blood mononuclear cell; PCA, principal component analysis; RF, random forest; scRNA-seq, single-cell RNA sequencing; ssGSEA, single-sample gene set enrichment analysis; SVM, support vector machine.

1. Introduction

Sepsis remains a leading cause of morbidity and mortality worldwide and continues to impose a substantial health care burden despite advances in antimicrobial therapy and critical care. Contemporary international guidance defines sepsis as life-threatening organ dysfunction caused by a dysregulated host response to infection, underscoring that host biology rather than pathogen burden alone contributes to clinical deterioration and outcome.[1] In practice, however, timely and accurate diagnosis remains difficult, particularly early in the disease course or in patients with competing infectious and sterile inflammatory phenotypes.[2]

Current diagnostic workflows integrate clinical assessment with routine laboratory parameters, imaging, and microbiological testing, yet each component has limitations. Cultures may be delayed or negative, and conventional inflammatory biomarkers provide imperfect discrimination between infectious and noninfectious systemic inflammation in critically ill populations.[2,3] These constraints motivate host-response-based diagnostics that capture the systemic biological perturbations of sepsis at the molecular level.[4]

Peripheral blood transcriptomics and host-response profiling have emerged as promising strategies to quantify sepsis biology. Recent work has moved beyond isolated gene lists toward consensus transcriptomic frameworks, clinically translated host-response assays, and biomarker-development pipelines that emphasize robustness, clinical heterogeneity, and

translational utility.[5-9] Such progress supports the broader premise that whole-blood gene-expression patterns can complement conventional assessment while remaining biologically interpretable.

Oxidative stress is widely recognized as a central component of sepsis pathophysiology. Excess production of reactive oxygen species, mitochondrial dysfunction, and impaired antioxidant defenses contribute to endothelial injury, microcirculatory dysfunction, metabolic collapse, and multi-organ failure.[10,11] Oxidative stress is also tightly interwoven with immune activation: neutrophils and monocytes generate reactive oxygen species as part of antimicrobial defense, yet dysregulated oxidative bursts can amplify inflammatory signaling and propagate tissue damage.[10,12] These features suggest that oxidative stress-related transcriptional programs may capture an integrative axis of sepsis biology spanning inflammation, innate effector activity, and cellular stress adaptation.

Concurrently, sepsis is increasingly understood as an immunologically heterogeneous syndrome. Many patients exhibit simultaneous innate immune activation, adaptive immune dysfunction, and context-dependent immune suppression rather than a single uniform inflammatory state.[13-15] This heterogeneity complicates biomarker development because a single inflammatory marker may not reflect the evolving balance between injurious hyperinflammation and clinically consequential immunosuppression.[2,13-15] Host-response signatures that encode multiple biological dimensions may therefore be better suited to address real-world diagnostic ambiguity.

Despite the biological plausibility of oxidative stress as a central sepsis axis, the diagnostic utility of oxidative stress-anchored gene signatures and their links to immune landscape alterations and downstream pathways remain incompletely characterized. In particular, it is unclear whether oxidative stress-related transcriptional features can be distilled into compact, interpretable diagnostic models with external validity, and which immune-cell states carry these signals in whole blood.

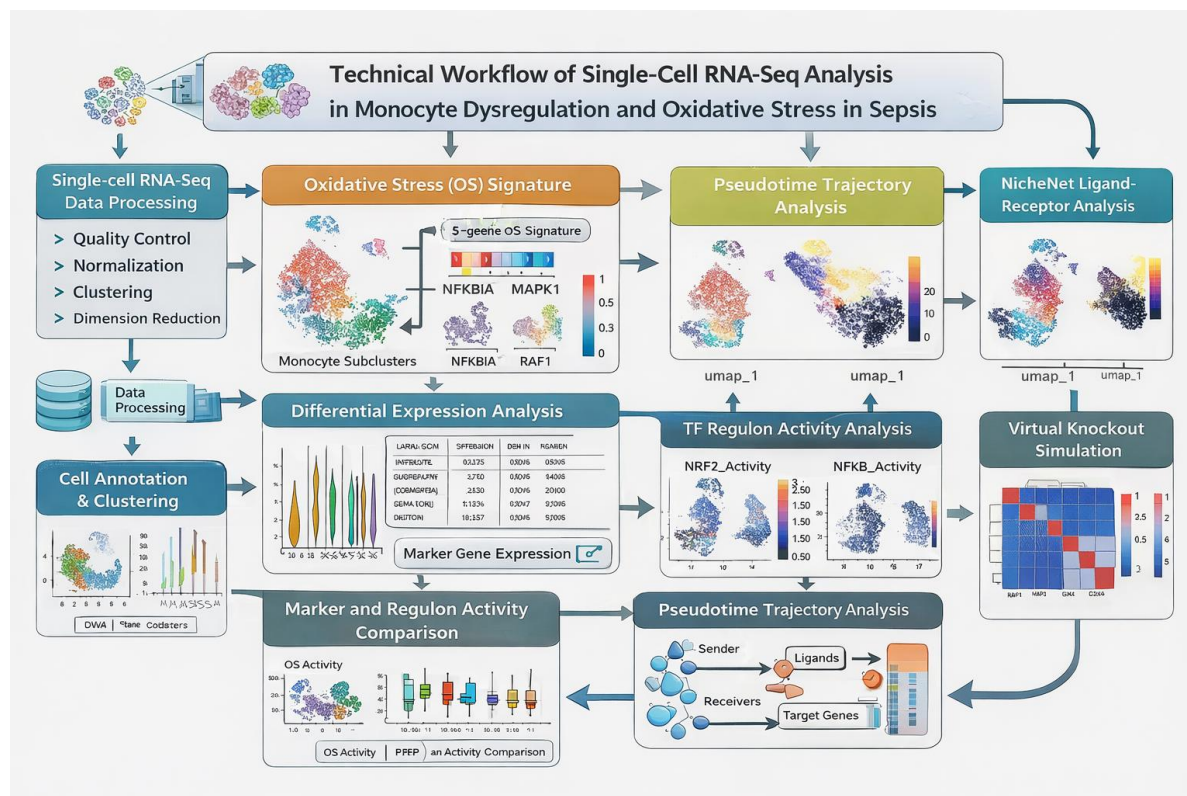
Accordingly, we hypothesized that oxidative stress-related gene expression patterns in whole blood reflect a biologically meaningful host-response dimension in sepsis and can be leveraged to develop diagnostically informative and mechanistically interpretable models.

Using publicly available bulk transcriptomic datasets and complementary single-cell analyses, we aimed to (1) characterize the expression patterns of oxidative stress-associated genes in sepsis, (2) construct an oxidative stress score using principal component analysis, (3) identify robust diagnostic biomarkers through systematic machine learning-based feature selection and model comparison, (4) evaluate diagnostic performance with external validation and clinically interpretable tools such as nomograms and decision curve analysis, and (5) localize the bulk-derived signature to specific immune-cell states and infer its potential regulatory context at single-cell resolution.

By anchoring diagnostic modeling to oxidative stress biology, linking predictive features to immune and pathway-level interpretation, and examining their monocyte-centered cellular context, this study seeks to advance interpretable host-response diagnostics and to provide mechanistic insights into sepsis-associated immune dysregulation and stress responses.

2. Methods

The overall analytical workflow of the study, including bulk transcriptomic modeling and monocyte-centered single-cell contextualization, is summarized in Scheme 1.



Scheme 1. Technical workflow of the study, highlighting the monocyte-focused single-cell supplementary analysis used to contextualize the bulk-derived oxidative stress signature.

2.1 Public bulk transcriptome datasets and preprocessing

Public bulk transcriptome datasets were obtained from the Gene Expression Omnibus (GEO) database and accessed on May 21, 2026. GSE6535 was used as the discovery cohort. This dataset was generated on the GPL4274 platform (NHICU Human 19K v1.0) and included 72 critically ill patients admitted to the intensive care unit, consisting of 55 microbiologically confirmed sepsis cases and 17 non-sepsis ICU controls. GSE54514 was used as an external whole-blood validation cohort. This dataset was generated on the GPL6947 platform (Illumina HumanHT-12 V3.0 expression beadchip) and contained longitudinal PAXgene whole-blood profiles from sepsis patients and healthy controls; to avoid repeated-measure inflation, only day-1 samples were retained for diagnostic validation, including 35 sepsis patients (26 survivors and 9 non-survivors) and 18 healthy controls. GSE46955 was used as an external monocyte validation cohort. This dataset was generated on the GPL6104 platform (Illumina humanRef-8 v2.0 expression beadchip) and contained blood monocytes from gram-negative

sepsis patients, recovery samples, and healthy donors, with or without ex vivo LPS stimulation. For validation against the basal disease state, only unstimulated basal samples were retained, including 8 sepsis-basal samples and 6 healthy-basal samples; recovery and LPS-stimulated samples were excluded from the primary validation analysis.

Samples were included if they were derived from human peripheral blood, whole blood, neutrophils, or purified monocytes and had an unambiguous sepsis or control phenotype. Samples were excluded if phenotype annotation was unclear, if the sample represented recovery rather than the acute disease state, if the sample was experimentally stimulated ex vivo and not relevant to the basal diagnostic comparison, or if the expression profile lacked a matched GEO sample annotation. For microarray datasets, probe identifiers were mapped to official gene symbols according to the corresponding GPL platform annotation files. Probes without valid gene symbols were removed. When multiple probes mapped to the same gene, the probe with the highest mean expression across all retained samples was selected to represent that gene. Genes with missing values in more than 50% of samples were removed; remaining missing values, if present, were imputed using k-nearest-neighbor imputation. Genes with consistently low abundance, defined as expression values below the 25th percentile in more than 80% of samples within a dataset, were filtered before downstream analysis. Expression values were log2 transformed when required and normalized within each dataset. External validation datasets were not merged with the discovery cohort; instead, the final gene features identified in the discovery cohort were extracted and evaluated independently in each validation dataset.

2.2 Identification of oxidative stress phenotypes

We curated 46 oxidative stress-related genes from the Molecular Signatures Database (MSigDB), focusing on genes implicated in redox regulation, inflammatory signaling, and cellular stress responses. After platform mapping in GSE6535, 41 genes were retained for downstream analysis. Expression differences between the sepsis and control groups were tested with the limma framework. Genes with a nominal $P < 0.05$ were regarded as differentially expressed for exploratory signature construction. Pearson correlation coefficients were calculated to evaluate co-expression patterns among oxidative stress-associated differentially

expressed genes, and gene-level distributions were displayed using boxplots.

2.3 Construction and sensitivity analysis of the oxidative stress score

To summarize the overall oxidative-stress transcriptional pattern at the sample level, principal component analysis (PCA) was performed using the expression matrix of oxidative stress-related differentially expressed genes after Z-score standardization. The oxidative stress score (OS score) was defined as the combined value of PC1 and PC2. PC1 alone was not used because the first two principal components jointly captured the major structured variance of the oxidative stress-related transcriptional program, whereas a single component may represent only part of a multidimensional immune-stress signal. To examine robustness, sensitivity analyses were performed using PC1 alone and an ssGSEA-based oxidative stress score derived from the same oxidative stress gene set. The major between-group trend remained consistent across the PC1-only, PC1 + PC2, and ssGSEA-based scoring strategies, supporting the stability of the OS score framework.

2.4 Machine-learning model construction and evaluation

Four machine-learning algorithms were compared for diagnostic marker selection and model construction: generalized linear model (GLM), support vector machine (SVM), random forest (RF), and extreme gradient boosting (XGBoost). Model training was performed in the discovery cohort, and independent GEO datasets were used for external validation. Within the discovery cohort, repeated ten-fold cross-validation was applied for internal model comparison, and the random seed was fixed at 1234. Candidate predictors consisted of oxidative stress-related differentially expressed genes retained after preprocessing and feature filtering. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, accuracy, and residual distribution. Ninety-five percent confidence intervals for AUC values were calculated using DeLong's method. No synthetic oversampling was applied; to reduce the influence of group-size imbalance, inverse class-frequency weights were used during model training where supported by the algorithm. The final diagnostic model was selected by jointly considering AUC, cross-validation stability, residual distribution, and external validation performance.

For GLM, binomial logistic regression was fitted with a logit link using predictors selected from the oxidative stress-related feature set. For SVM, a radial basis function kernel was used; cost C was searched over 0.01, 0.1, 1, 10, and 100, and gamma was searched over 0.001, 0.01, 0.1, and 1, with the optimal parameters selected by cross-validation AUC. For RF, 500 trees were fitted; mtry was searched from 1 to floor(sqrt(p)) and p/3 when applicable, and minimum node size was searched over 1, 5, and 10. For XGBoost, nrounds was searched over 50, 100, and 200; max_depth over 2-6; eta over 0.01, 0.05, 0.10, and 0.30; subsample and colsample_bytree over 0.70, 0.90, and 1.00; min_child_weight over 1, 3, and 5; and gamma over 0 and 1. The optimal XGBoost model was selected by cross-validation AUC.

2.5 Construction and validation of the diagnostic nomogram

An interpretable nomogram was built using the expression levels of MAPK1, NFKBIA, and RAF1 to estimate individual sepsis risk. The nomogram was developed with the rms package, and discrimination was quantified by ROC curves generated via pROC. Calibration was examined by comparing predicted probabilities with observed outcomes, and decision curve analysis (DCA) was conducted to evaluate net clinical benefit across a range of threshold probabilities. External validation was performed in GSE54514 without retraining the model, thereby testing whether the discovery-cohort signature retained transportability in an independent whole-blood dataset.

2.6 Immune microenvironment analysis

Immune-cell-associated signatures were inferred using single-sample gene set enrichment analysis (ssGSEA) to estimate the relative enrichment of immune cell types in sepsis and control samples. Group differences were visualized with violin plots and compared using non-parametric tests unless otherwise specified. Correlations between MAPK1, NFKBIA, RAF1 expression and immune signatures were evaluated using Pearson correlation analysis. For stratified comparisons, samples were split into high- and low-expression groups based on the median expression level of each key gene.

2.7 Functional enrichment and pathway analysis of key genes

To probe pathway-level implications of the three diagnostic genes, single-gene GSEA was

performed. Pathways with adjusted $P < 0.05$ were considered significant. Translation-initiation-related genes were then compared between sepsis and control samples using limma, and correlations between the three key genes and pathway-associated genes were assessed using Pearson analysis.

2.8 Single-cell RNA-seq data processing and annotation

Single-cell RNA-seq data were obtained from GEO accession GSE290679 and accessed on May 21, 2026. The dataset consisted of cryothawed peripheral blood mononuclear cells (PBMCs) profiled using 10x Genomics droplet-based single-cell RNA sequencing. The included samples represented three groups: 9 non-acutely ill healthy controls (NHC), 19 critically ill ICU patients without sepsis (CI-NS), and 19 critically ill ICU patients with sepsis (CI-Sep), yielding a reported 644,147-cell map of human critical illness with and without sepsis. Raw feature-barcode H5 matrices and the GEO series matrix file were downloaded and processed in Seurat. Cells with fewer than 200 detected genes, more than 5,000 detected genes, or mitochondrial transcript proportions greater than 15% were excluded. Genes detected in fewer than three cells were removed. Potential doublets were detected using DoubletFinder, with the expected doublet rate estimated according to the recovered cell number for each sample; predicted doublets were removed before clustering and downstream analyses. To minimize sample-level and batch-associated effects, normalization and scaling were performed after quality control, and integration/batch correction was performed using Harmony with sample identity and batch information as covariates. Principal component analysis, nearest-neighbor graph construction, UMAP visualization, and graph-based clustering were then performed.

Cell type annotation was based on canonical marker genes and cluster-specific marker expression. Major immune lineages were assigned using established markers, including CD3D/CD3E for T cells, CD4 and CD8A for CD4 and CD8 T-cell subsets, NKG7/GNLY for NK cells, MS4A1/CD79A for B cells, LYZ/CD14/FCGR3A for monocytes, FCER1A/CST3/HLA-DRA for dendritic cells and antigen-presenting cells, and PPBP/PF4 for platelets. Monocyte-enriched clusters were further extracted and reclustered for subpopulation-level analysis. Monocyte subcluster annotation was supported by differential marker genes,

oxidative-stress activity, inflammatory-regulon activity, and the distribution of key genes identified from the bulk analysis.

2.9 Single-cell downstream analysis parameters

AUCell analysis was used to quantify oxidative-stress signature activity at single-cell resolution. The oxidative-stress gene set was derived from the oxidative stress-related gene collection used in the bulk analysis and was cross-checked against MSigDB oxidative-stress-related terms. Genes not detected in the single-cell dataset were removed before scoring. AUCell scores were calculated using ranking-based gene-set enrichment within individual cells; cells with higher AUC values were interpreted as having stronger oxidative-stress transcriptional activity. The expression of MAPK1, NFKBIA, and RAF1 was mapped onto the monocyte UMAP space and compared with oxidative-stress activity to assess whether the bulk-derived signature localized to specific monocyte states.

NicheNet analysis was performed to identify upstream ligands that could be associated with the oxidative-stress and inflammatory transcriptional program in monocytes. Receiver cells were defined as monocyte subclusters with high oxidative-stress activity, whereas sender cells included the remaining annotated immune cell populations in the PBMC microenvironment. Candidate ligands were retained if they were expressed in at least 10% of sender cells and if their cognate receptors were detected in at least 10% of receiver monocytes. Ligands were ranked using ligand-target regulatory potential and Pearson correlation between predicted target activity and the observed monocyte signature. The final prioritized ligands were visualized with ligand activity scores, and top ligand-target pairs were retained for mechanistic interpretation.

Trajectory inference was performed within monocyte subclusters to reconstruct a pathological evolution axis. The root state was assigned to the monocyte state enriched in NHC samples and characterized by the lowest oxidative-stress activity, thereby representing the least disease-activated starting state. Pseudotime ordering was then used to examine progressive changes in oxidative-stress activity, NFKBIA expression, and monocyte activation state. Genes and scores varying along pseudotime were interpreted as candidate features of sepsis-

associated monocyte reprogramming. Transcription-factor regulon activity was evaluated using AUCell-based scoring of curated inflammatory and oxidative-stress regulon gene sets. NRF2-related and NF- κ B-related regulon activities were calculated for each monocyte and compared across monocyte subclusters and disease groups.

Virtual knockout and network perturbation analysis was used as an exploratory strategy to assess the regulatory impact of key genes linking the bulk and single-cell analyses. A gene-gene regulatory similarity network was constructed using oxidative-stress-related genes and diagnostic model genes detected in monocytes. Edge weights were calculated from pairwise Spearman correlations, and edges were retained when the absolute correlation coefficient exceeded 0.30 with false-discovery-rate-adjusted $P < 0.05$. Virtual knockout was simulated by removing the target node and recalculating network connectivity, edge stability, and downstream connectivity changes. Genes showing the largest reduction in connectivity or altered network stability after perturbation were considered putative downstream effectors of the targeted regulatory node. Because this analysis is computational and hypothesis-generating, the results were interpreted as evidence of regulatory network dependency rather than direct experimental gene knockout.

2.10 Statistical analysis

Unless otherwise specified, continuous variables were compared using the Wilcoxon rank-sum test for two-group comparisons or the Kruskal-Wallis test for comparisons involving more than two groups. Multiple-testing correction was performed using the Benjamini-Hochberg method. Adjusted $P < 0.05$ was considered statistically significant. For single-cell analyses, statistical comparisons were performed at the sample level whenever cell-proportion differences were evaluated, whereas cell-level visualizations were used to display within-cell transcriptional heterogeneity. All analyses were conducted in R (version 4.3.3) unless otherwise noted.

3. Results

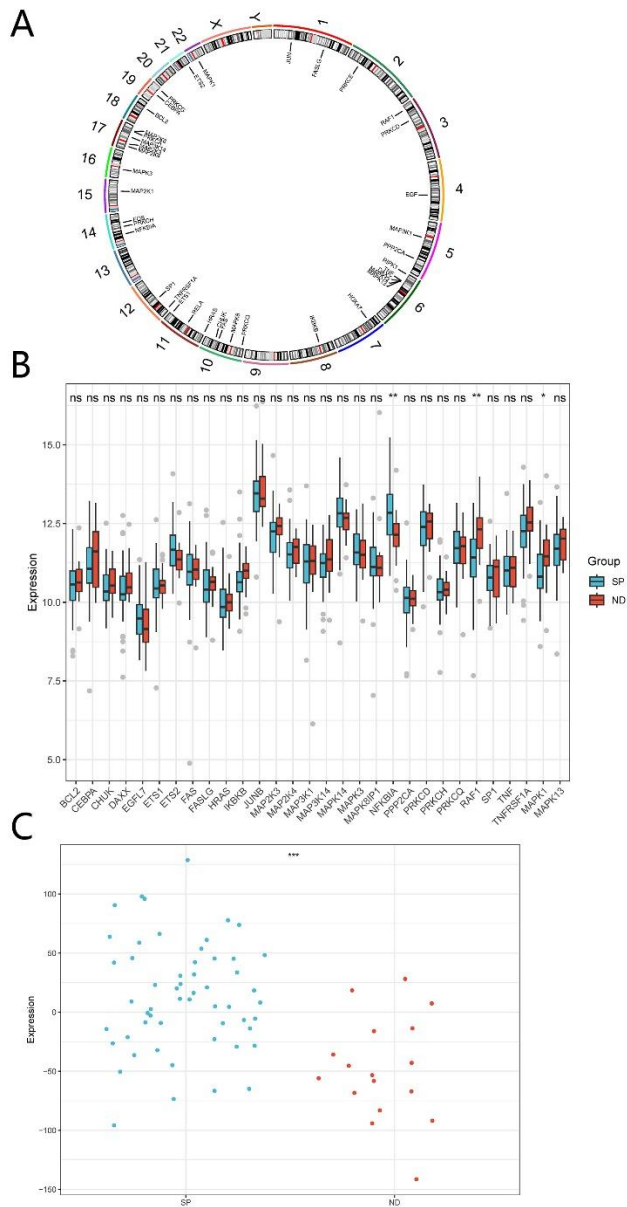
3.1 Identification of Dysregulated Oxidative Stress Phenotypes in Sepsis

We first identified a dysregulated oxidative stress-related transcriptional pattern in the GSE6535 discovery cohort. Forty-one oxidative stress-related genes retained after platform mapping were distributed across multiple chromosomes, indicating that the candidate feature set was not confined to a single genomic region (Figure 1A).

Differential expression analysis showed broad but selective disruption of oxidative stress-related genes between sepsis patients and non-sepsis ICU controls. Several regulators, including MAPK14, NFKBIA, and RAF1, differed significantly between groups, suggesting that oxidative stress-related host-response programs are associated with sepsis biology (Figure 1B).

The PCA-derived OS score was higher in the sepsis group than in the non-sepsis control group ($P < 0.001$), indicating that the combined oxidative stress-related expression pattern separated the two clinical states in the discovery cohort (Figure 1C). Sensitivity analyses using PC1 alone and ssGSEA-based scoring yielded consistent between-group trends, supporting the robustness of the OS score framework.

Figure 1. Landscape of oxidative stress (OS) phenotypes and construction of the OS scoring system in the GSE6535 discovery cohort. (A) Circos plot visualizing the chromosomal positions of the 41 OS-related genes retained after platform mapping. (B) Boxplots showing differential expression of OS-related genes between sepsis patients (SP) and non-sepsis ICU controls (ND); boxes indicate interquartile range, center lines indicate medians, whiskers indicate 1.5 interquartile range, and group comparisons were performed using the Wilcoxon rank-sum test unless otherwise specified. (C) Comparison of OS scores between SP and ND groups; *** $P < 0.001$.



3.2 Screening of Diagnostic Markers via Machine Learning Algorithms

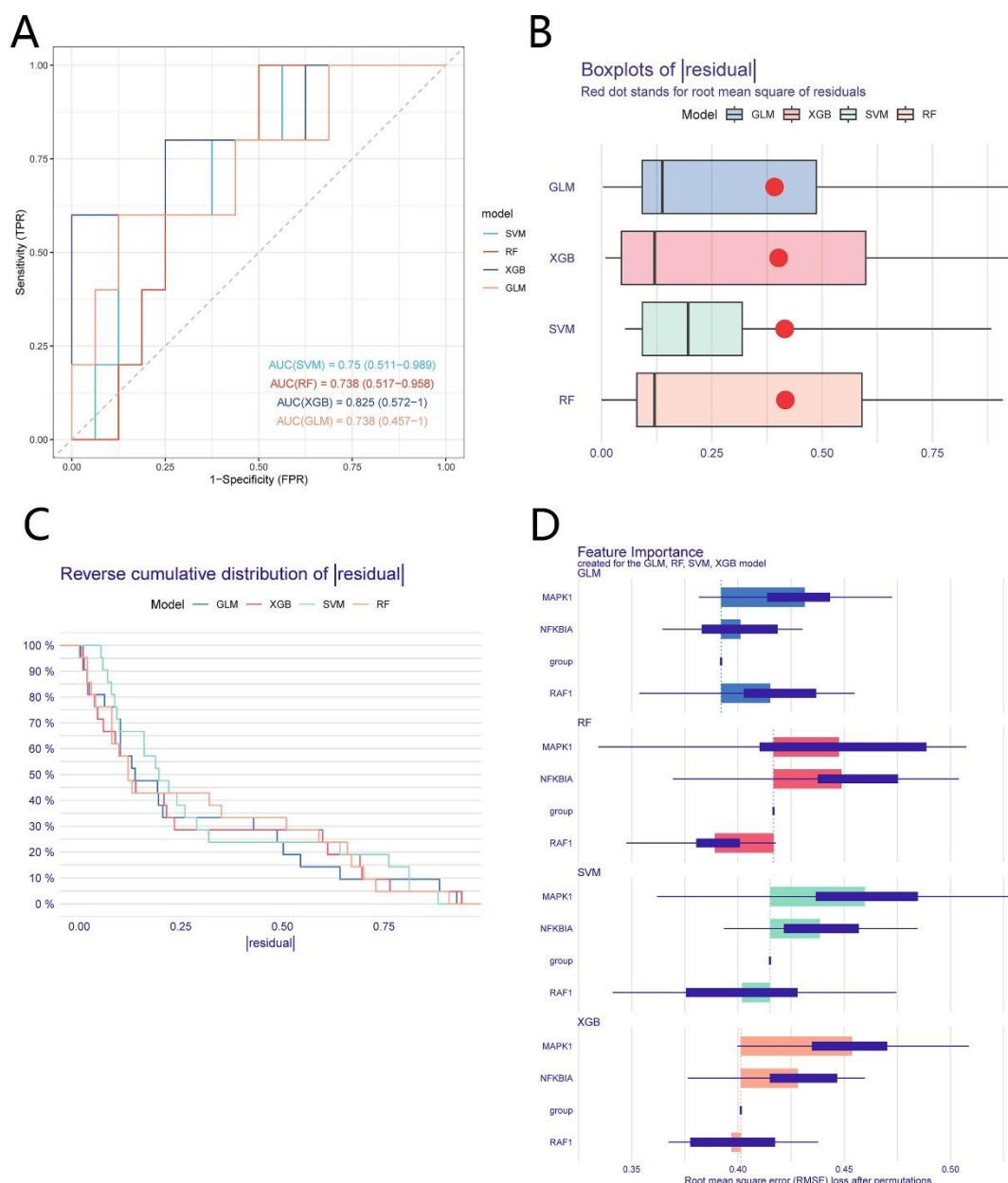
Among the four classifiers, XGBoost showed the highest discrimination and selected MAPK1, NFKBIA, and RAF1 as the final signature. The XGBoost model achieved an AUC of 0.825 (95% CI, 0.572-1.000), compared with 0.750 for SVM, 0.738 for RF, and 0.738 for GLM (Figure 2A).

Residual analysis provided a complementary stability check. XGBoost and RF showed lower residual dispersion than GLM and SVM, and the reverse cumulative residual curve of

XGBoost was closest to the coordinate axis, consistent with smaller prediction error among the tested classifiers (Figure 2B,C).

Feature-importance analysis of the final XGBoost model ranked MAPK1, NFKBIA, and RAF1 as the leading contributors to classification (Figure 2D). These three genes were therefore retained as a compact diagnostic signature for subsequent interpretation rather than as experimentally proven causal mediators.

Figure 2. Machine-learning-based selection of diagnostic markers in the GSE6535 discovery cohort. (A) ROC curves for SVM, RF, XGBoost, and GLM classifiers; XGBoost achieved the highest AUC (0.825; 95% CI, 0.572-1.000). (B) Residual distributions across models shown as boxplots; red markers indicate RMSE. (C) Reverse cumulative residual plots comparing prediction error profiles. (D) Variable-importance results from the final XGBoost model, identifying MAPK1, NFKBIA, and RAF1 as the three leading contributors to diagnostic classification.



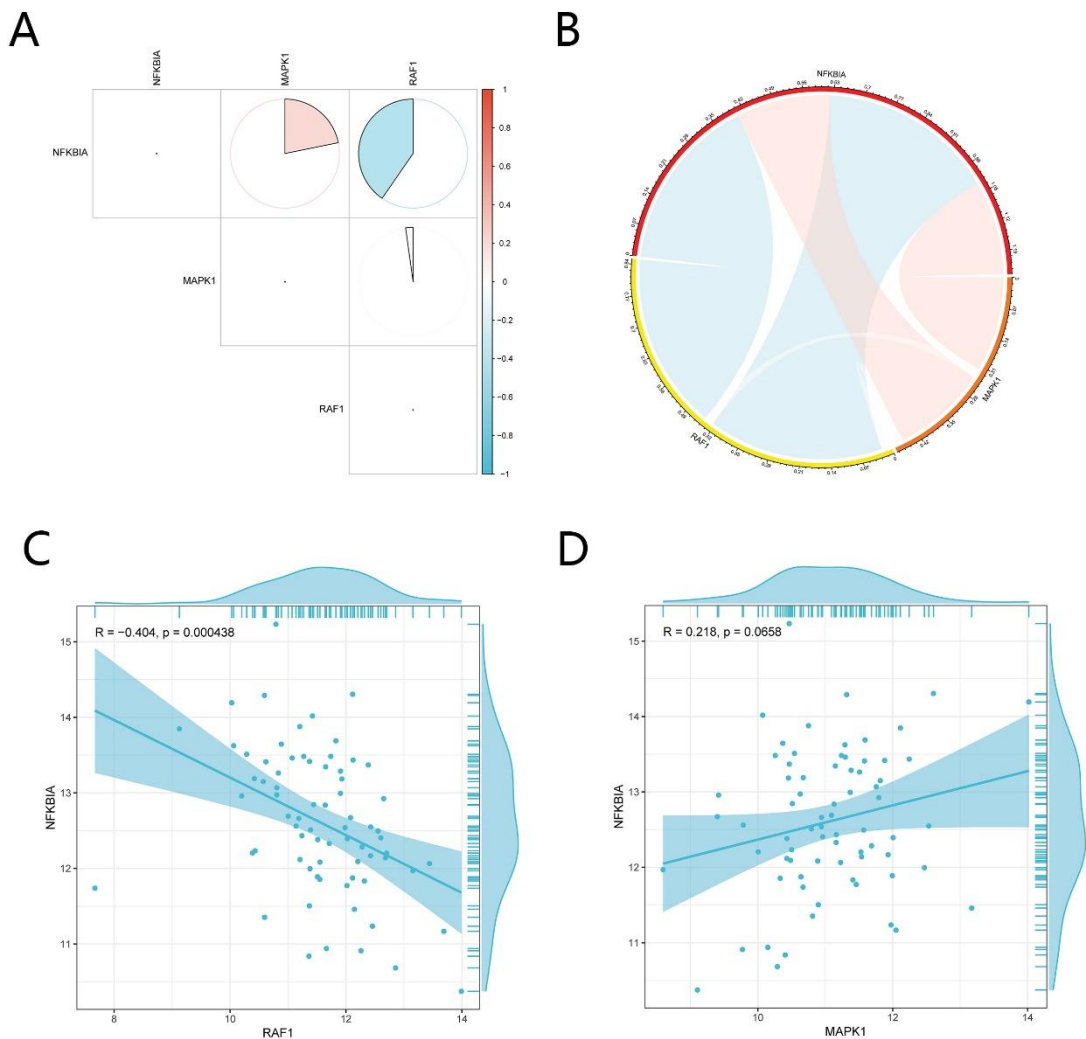
3.3 Key Diagnostic Biomarkers Correlation Analysis

After defining MAPK1, NFKBIA, and RAF1 as the core diagnostic signature, we examined whether their expression patterns showed coordinated relationships in sepsis. A correlation heatmap and chord plot summarized the pairwise expression structure among the three genes (Figure 3A,B).

Pearson correlation analysis showed a significant inverse correlation between NFKBIA and RAF1 ($R = -0.404$, $P < 0.001$; Figure 3C), whereas NFKBIA and MAPK1 showed a modest positive trend that did not reach conventional statistical significance ($R = 0.218$, $P = 0.0658$;

Figure 3D). These associations suggest coordinated expression behavior but should not be interpreted as direct regulatory proof.

Figure 3. Correlation patterns among the selected diagnostic genes in GSE6535. (A) Heatmap summarizing pairwise expression correlations among MAPK1, NFKBIA, and RAF1. (B) Chord diagram depicting the relative strength and direction of associations between the three biomarkers. (C) Scatter plot illustrating the inverse relationship between NFKBIA and RAF1 (Pearson $R = -0.404$, $P < 0.001$). (D) Scatter plot showing the modest positive association between NFKBIA and MAPK1 (Pearson $R = 0.218$, $P = 0.0658$).



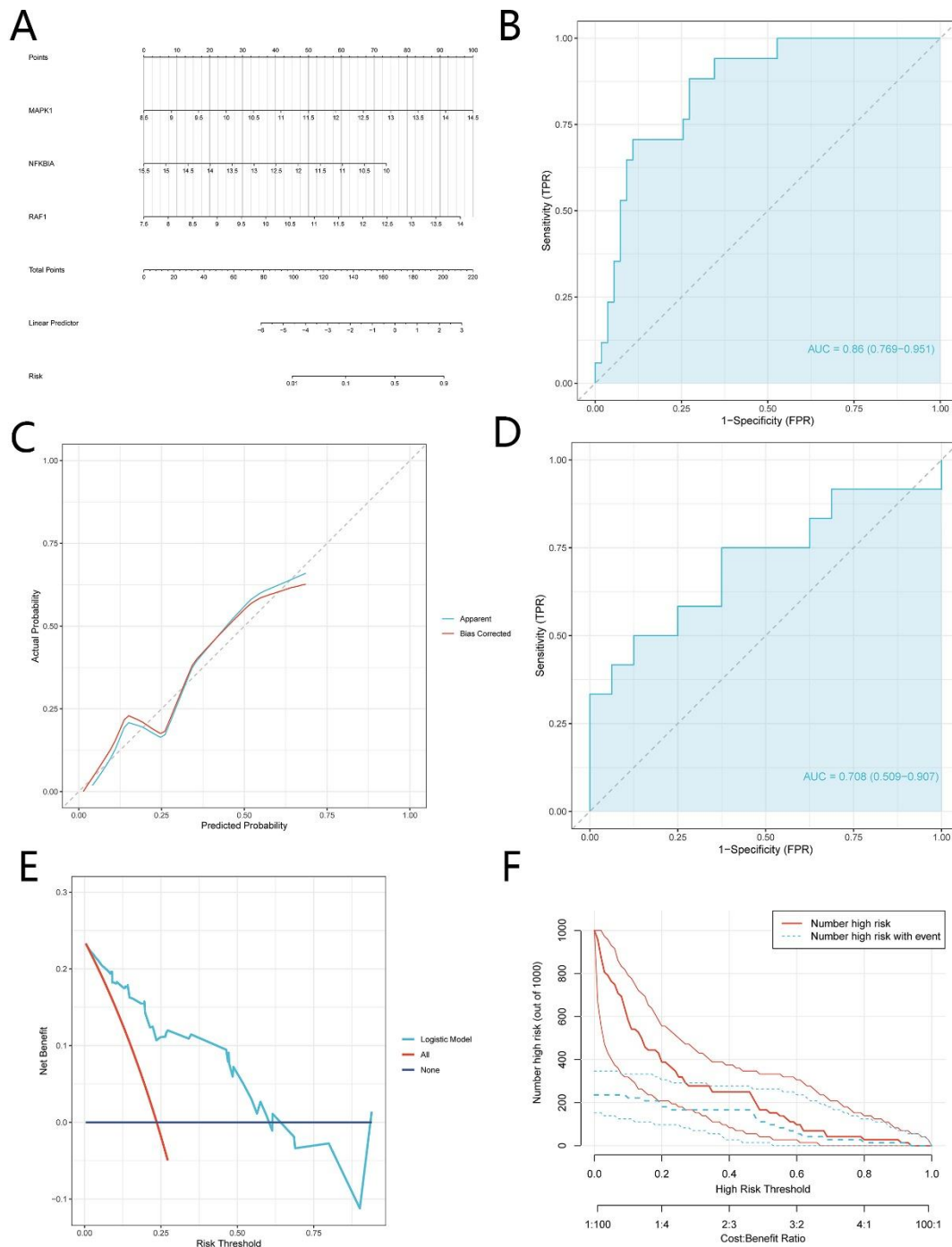
3.4 Construction and Evaluation of the Diagnostic Nomogram

To provide an interpretable risk-estimation tool, we translated the MAPK1/NFKBIA/RAF1 signature into a nomogram (Figure 4A). In the GSE6535 training cohort, the nomogram showed good discrimination on ROC analysis (AUC, 0.860; 95% CI, 0.769-0.951) and acceptable calibration, with the bias-corrected curve tracking the ideal reference line (Figure 4B,C).

In the independent GSE54514 validation cohort, discrimination decreased to an AUC of 0.708 (95% CI, 0.509-0.907; Figure 4D). This result supports external transportability at an acceptable but moderate level and should be interpreted with caution because the confidence interval was wide and the validation cohort differed clinically and technically from the discovery cohort.

Decision curve analysis and clinical impact analysis suggested potential net benefit across selected threshold probabilities (Figure 4E,F). These results support further evaluation of the nomogram as an auxiliary host-response risk stratification tool rather than as a replacement for clinical judgment, SOFA-based assessment, blood culture, or established biomarkers such as procalcitonin.

Figure 4. Development and validation of the three-gene diagnostic nomogram. (A) Nomogram integrating MAPK1, NFKBIA, and RAF1 for individualized sepsis risk estimation. (B) ROC curve in the GSE6535 training cohort (AUC, 0.860; 95% CI, 0.769-0.951). (C) Calibration plot comparing predicted probabilities with observed outcomes; the bias-corrected curve is shown against the ideal diagonal reference. (D) ROC curve in the external GSE54514 validation cohort (AUC, 0.708; 95% CI, 0.509-0.907), indicating acceptable but moderate discrimination. (E) Decision curve analysis comparing net benefit of the nomogram with treat-all and treat-none strategies across threshold probabilities. (F) Clinical impact curve illustrating the predicted number of high-risk individuals and corresponding true positives across thresholds.



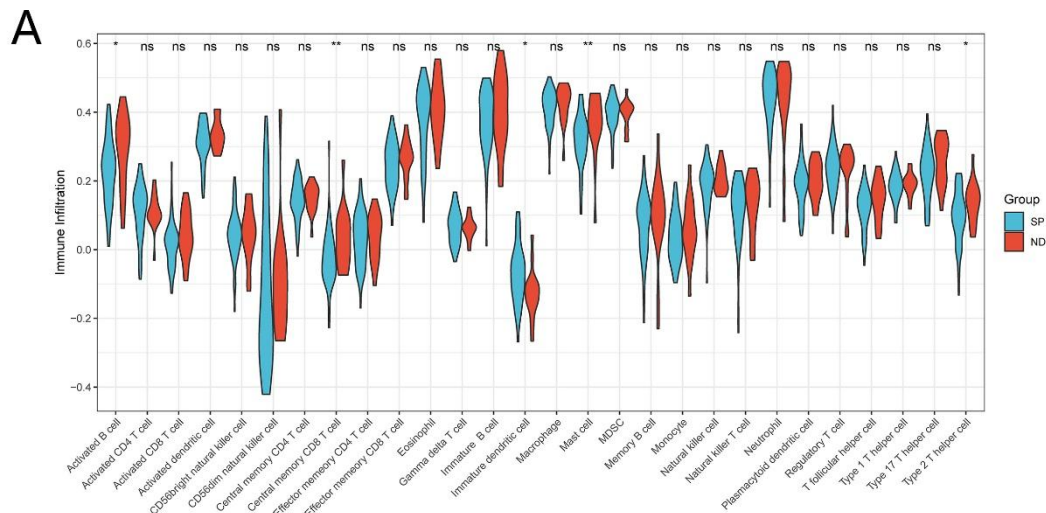
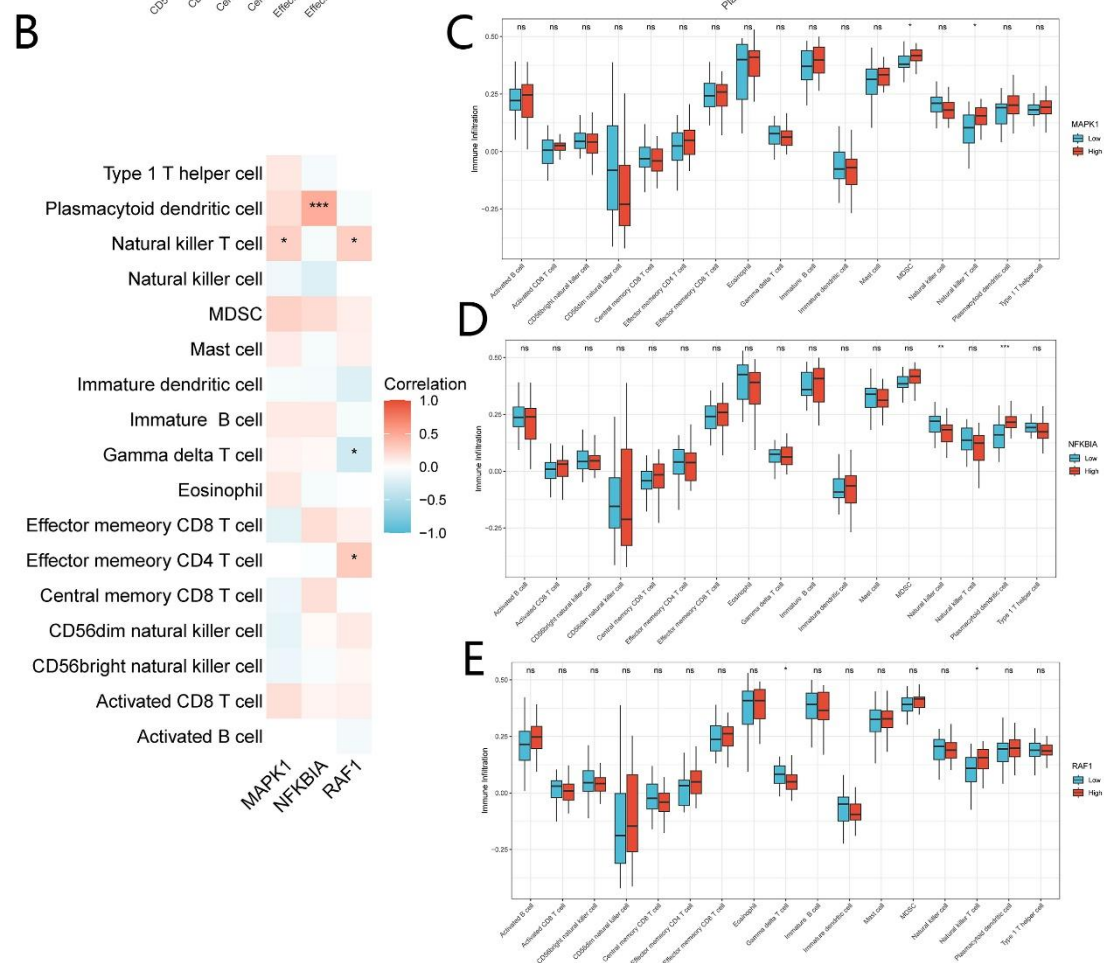
3.5 Immune Infiltration Landscape and Correlation with Key Genes

Immune dysregulation was evident at the inferred cell-signature level. ssGSEA analysis of 28 immune-cell signatures showed higher neutrophil-associated signal and reduced activated B-cell signal in sepsis compared with controls (Figure 5A).

We then asked whether the three signature genes tracked with changes in the inferred immune landscape. The correlation heatmap suggested gene-specific association patterns between transcript levels and immune-cell signatures (Figure 5B). In particular, NFKBIA showed a clear positive association with plasmacytoid dendritic cells ($P < 0.001$), whereas RAF1 was negatively associated with $\gamma\delta$ T cells ($P < 0.05$).

For a more direct comparison, samples were divided into high- and low-expression groups using the median expression of each gene. Higher MAPK1 expression was accompanied by increased MDSC and natural killer T-cell signatures (Figure 5C). In the same framework, the high-NFKBIA group showed greater enrichment of natural killer cells and plasmacytoid dendritic cells (Figure 5D). Higher RAF1 expression was, in turn, linked to elevated immature B-cell and natural killer T-cell signals (Figure 5E). Overall, the three OS-related genes were consistently associated with shifts in immune-cell signature patterns in sepsis.

Figure 5. ssGSEA-inferred immune-cell signature landscape and its association with MAPK1, NFKBIA, and RAF1 expression. (A) Violin plots comparing immune-cell enrichment scores between sepsis and control samples; group comparisons were performed using the Wilcoxon rank-sum test, with P-value markers shown in the plot. (B) Heatmap of Pearson correlations between gene expression and immune-cell signatures, with red indicating positive and blue indicating negative correlations. (C-E) Boxplots comparing immune-cell enrichment after stratification by median expression of MAPK1 (C), NFKBIA (D), and RAF1 (E); boxes indicate interquartile range, center lines indicate medians, and whiskers indicate 1.5 interquartile range.



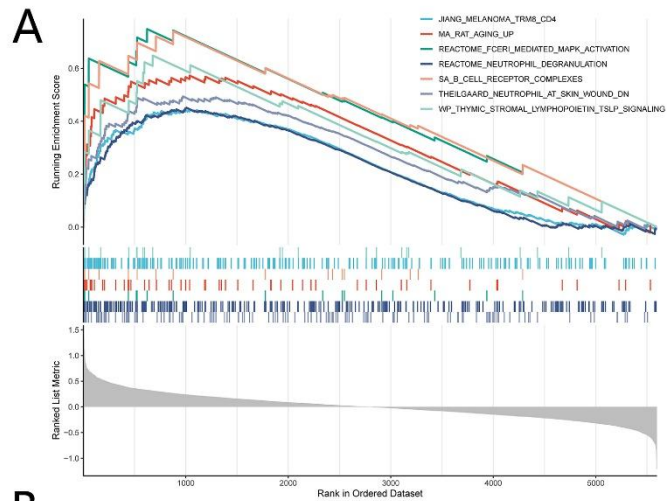
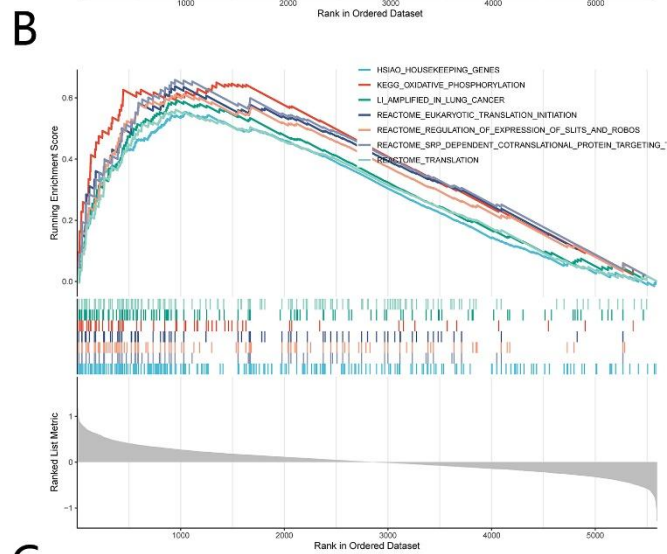
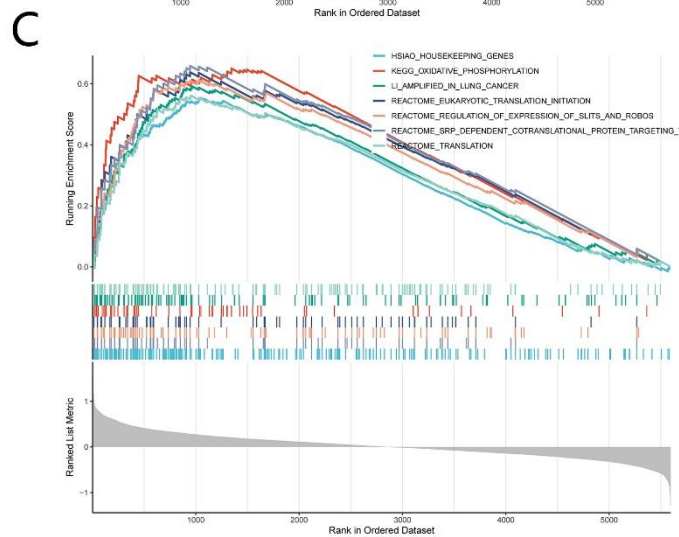
3.6 Functional Enrichment Analysis of Key OS Genes

Single-gene GSEA was used to explore pathways associated with the three signature genes.

MAPK1 was enriched in pathways related to signal transduction and immune response, including FCERI-mediated MAPK activation and neutrophil degranulation (Figure 6A). This pattern is consistent with, but does not prove, a link between MAPK1 expression and neutrophil-associated inflammatory activity in sepsis.

In contrast, NFKBIA and RAF1 displayed enrichment patterns associated with metabolic reprogramming and protein synthesis. NFKBIA was significantly correlated with "Eukaryotic Translation Initiation" and "Oxidative Phosphorylation" pathways, indicating its involvement in regulating cellular energy metabolism and translational control under oxidative stress conditions (Figure 6B). Interestingly, RAF1 exhibited a similar functional profile, showing strong enrichment in the "Eukaryotic Translation Initiation" pathway (Figure 6C). These results imply that while MAPK1 predominantly regulates immune signaling, NFKBIA and RAF1 may function synergistically to modulate protein synthesis and metabolic adaptation in response to sepsis-induced stress.

Figure 6. Functional enrichment analysis of key OS genes by single-gene GSEA. (A) MAPK1-associated enrichment results, including FCERI-mediated MAPK activation and neutrophil degranulation. (B) NFKBIA-associated enrichment results, including eukaryotic translation initiation and oxidative phosphorylation. (C) RAF1-associated enrichment results, including eukaryotic translation initiation. Enrichment results are interpreted as pathway associations rather than causal mechanisms.



439

440

441

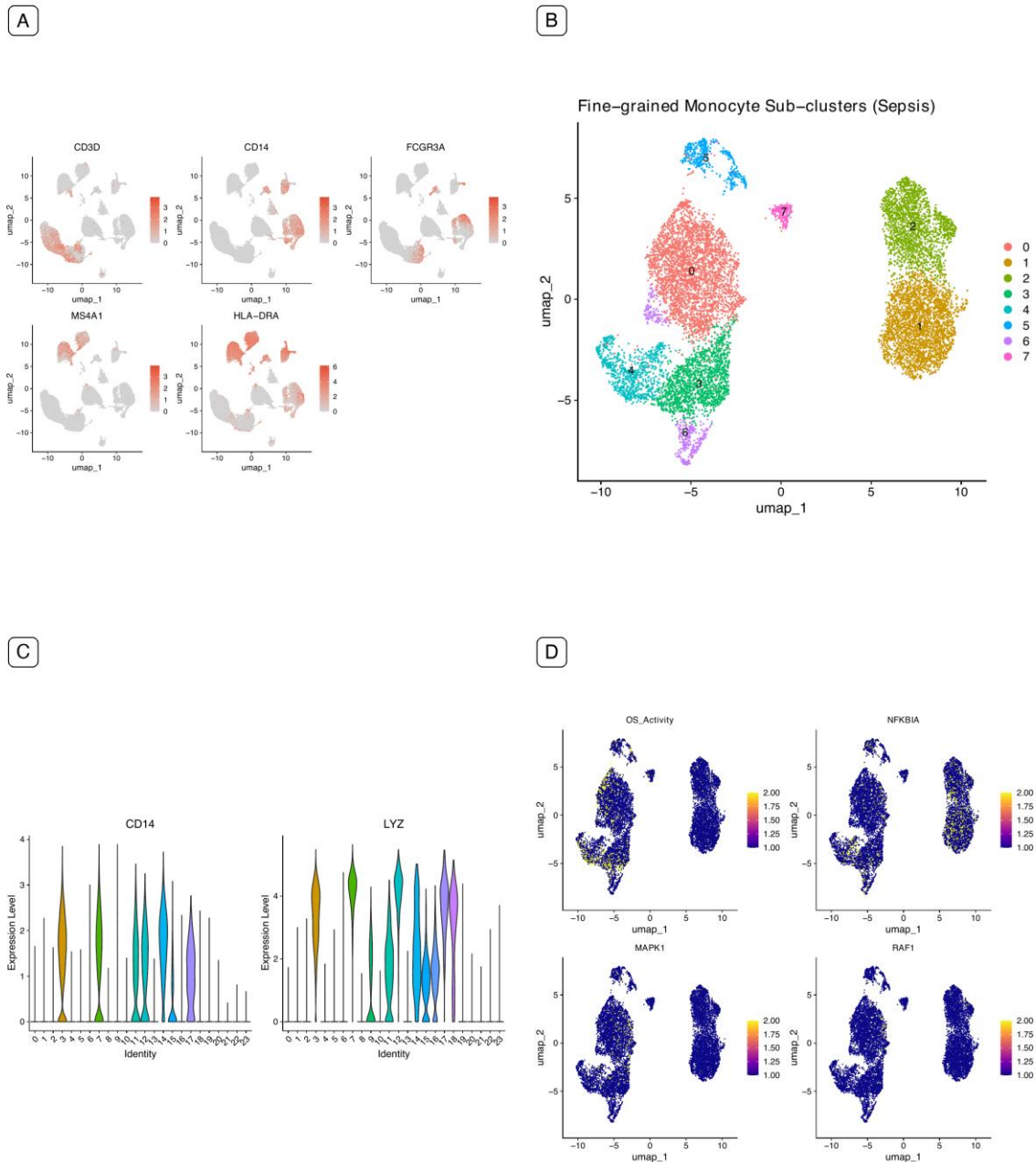
442 **3.7 Monocyte-Centered Single-Cell Mapping Localized the Bulk-Derived OS Signature**

to Discrete States

Manual inspection of canonical lineage markers confirmed the major immune compartments captured in the single-cell dataset and supported a monocyte-centered downstream strategy. CD3D, MS4A1, and HLA-DRA demarcated the expected T-cell, B-cell, and antigen-presenting cell compartments, whereas CD14 and FCGR3A were predominantly enriched in monocyte-lineage cells (Figure 7A). Focused reclustering of this compartment revealed eight transcriptionally distinct monocyte subclusters occupying discrete but partially continuous transcriptional spaces, consistent with the coexistence of relatively stable states and transitional populations in sepsis (Figure 7B). Independent verification using CD14 and LYZ confirmed that the selected clusters represented bona fide monocytes and minimized concern about contamination from adjacent immune populations (Figure 7C).

AUCell-based projection showed that oxidative stress activity was enriched in defined monocyte states rather than diffusely distributed across the entire compartment. NFKBIA, MAPK1, and RAF1 displayed overlapping localization patterns in oxidative stress-high regions (Figure 7D), indicating that the bulk-derived diagnostic signal retained cell-state context at single-cell resolution. These findings suggest that the three-gene signature may partly reflect pathologically activated monocyte states in sepsis.

Figure 7. Single-cell mapping of the oxidative stress signature in monocytes from GSE290679. (A) Canonical lineage markers used to annotate major PBMC immune compartments. (B) UMAP of reclustered monocytes showing eight transcriptionally distinct subclusters; colors denote monocyte subcluster identity. (C) Violin plots of CD14 and LYZ supporting monocyte identity of the selected subclusters. (D) AUCell/feature mapping of oxidative stress activity and MAPK1, NFKBIA, and RAF1 across the monocyte manifold; warmer colors indicate higher score or expression.



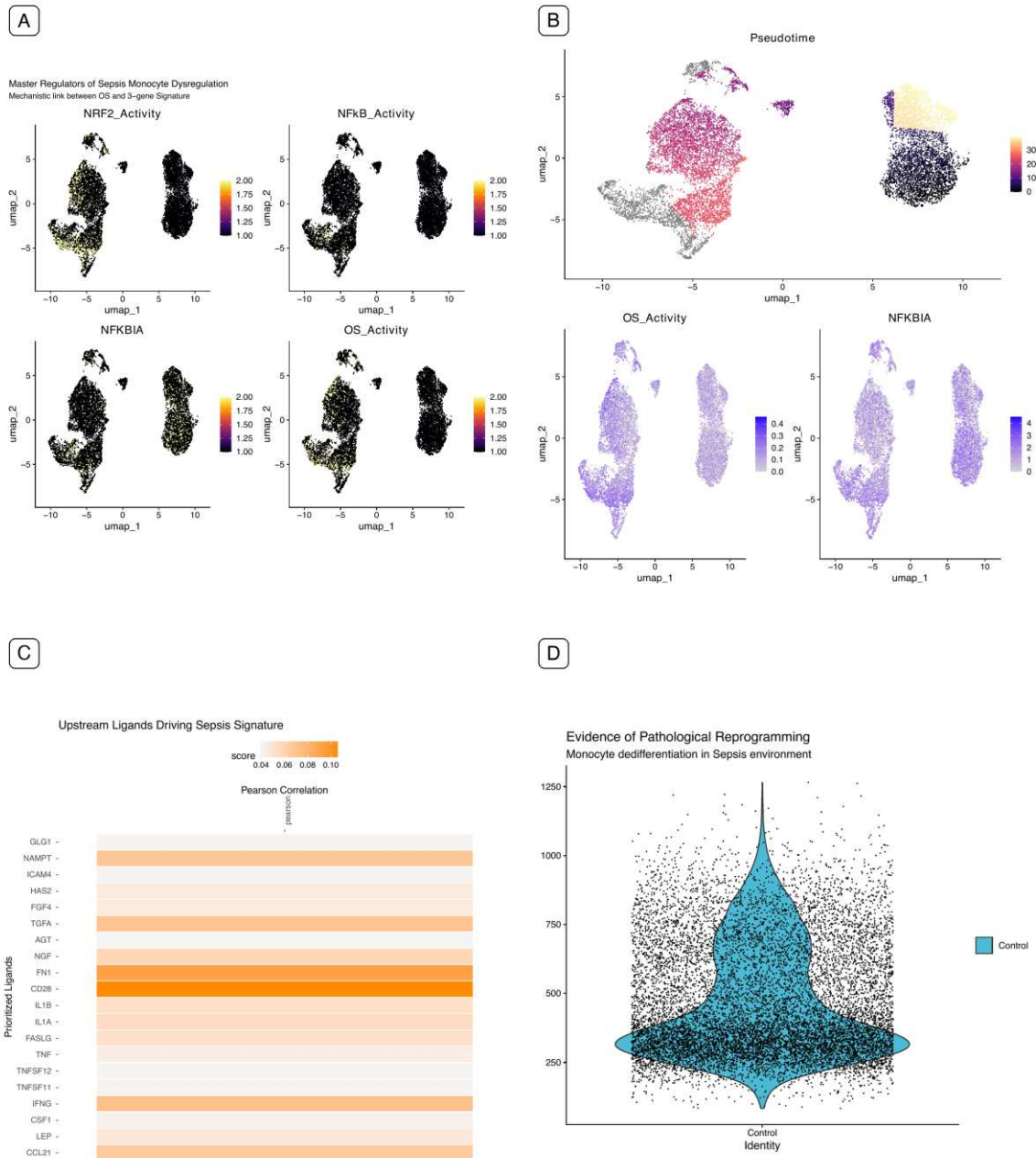
3.8 Regulatory Trajectory and Cell-Cell Signaling Analyses Support Pathological Monocyte Reprogramming

Because oxidative stress may be coupled to inflammatory transcriptional control, we examined regulon and trajectory patterns across monocyte states. NRF2-related activity, NF-kB activity, NFKBIA expression, and oxidative stress scores showed coordinated but non-identical spatial distributions (Figure 8A), consistent with concurrent redox adaptation and

inflammatory signaling. Pseudotime analysis revealed an ordered monocyte evolution axis along which oxidative stress activity and NFKBIA varied dynamically (Figure 8B).

Ligand prioritization, reprogramming scoring, and network perturbation analyses provided additional hypothesis-generating context. NicheNet prioritized candidate inflammatory ligands including IL1B, TNF, IFNG, TNFSF11, TNFSF12, CCL21, and CSF1 (Figure 8C). A reprogramming-related score showed heterogeneous distribution across septic monocytes (Figure 8D). The NFKBIA-centered virtual knockout framework identified a connected redox-inflammatory module involving SOD1, NQO1, RAF1, MAPK1, HMOX1, and GPX4 (Figure 9), supporting NFKBIA as a putative network node linking oxidative defense and inflammatory signaling.

Figure 8. Regulatory and intercellular mechanisms associated with oxidative stress-related monocyte remodeling in GSE290679. (A) Spatial mapping of NRF2-related activity, NF-kB activity, NFKBIA expression, and oxidative stress activity across monocyte states; color intensity indicates relative activity or expression. (B) Pseudotime analysis showing an ordered monocyte evolution axis with dynamic changes in oxidative stress activity and NFKBIA. (C) NicheNet-based prioritization of candidate upstream ligands expressed by sender cell populations and linked to receiver monocyte states. (D) Distribution of a reprogramming-related score indicating heterogeneous pathological remodeling within the monocyte compartment.



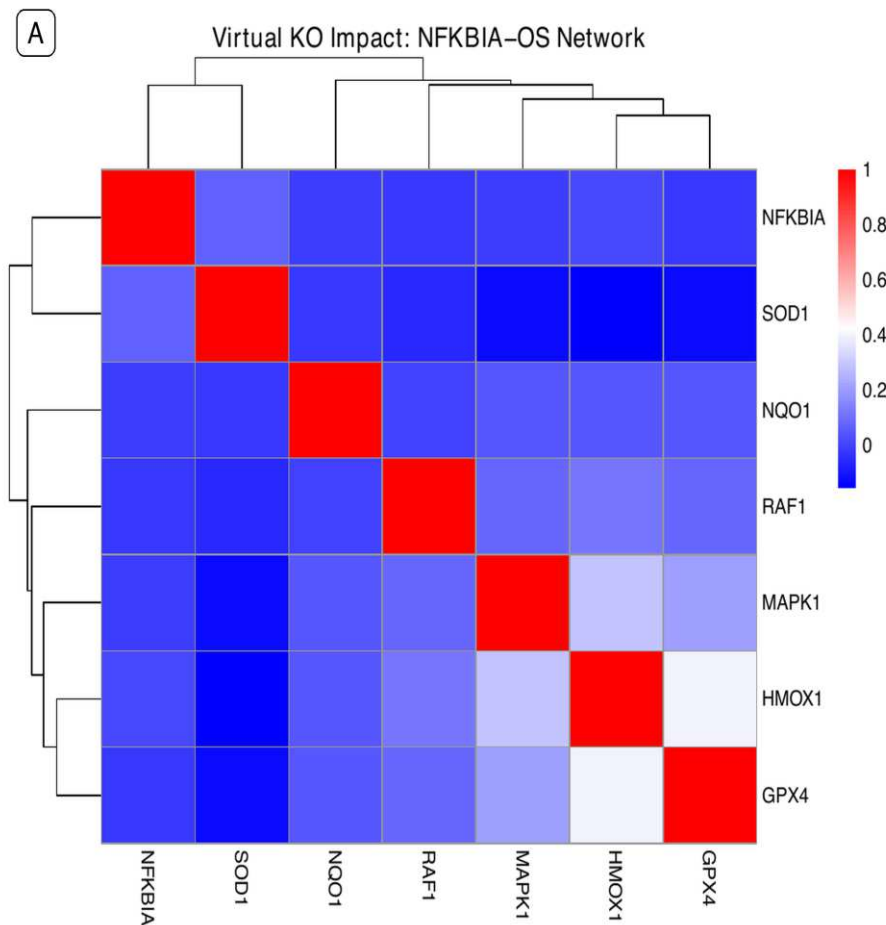


Figure 9. NFKBIA-centered network perturbation analysis in monocytes. Virtual knockout/stability analysis of the oxidative stress-related regulatory module identifies coordinated interactions among NFKBIA, SOD1, NQO1, RAF1, MAPK1, HMOX1, and GPX4. Edge retention was based on absolute Spearman correlation coefficient > 0.30 with false-discovery-rate-adjusted $P < 0.05$. The result is interpreted as computational evidence of putative network dependency rather than direct experimental knockout.

4. Discussion

In this study, we combined bulk blood transcriptomics with complementary single-cell interrogation to derive and biologically contextualize an oxidative stress-anchored diagnostic signature for sepsis. At the bulk level, MAPK1, NFKBIA, and RAF1 emerged as a compact three-gene set with acceptable discrimination and an interpretable nomogram. At the cellular level, the same signal localized preferentially to discrete monocyte states characterized by

coupled oxidative-stress and inflammatory activity, trajectory progression, and NFKBIA-centered network stability. Recent consensus and cohort studies increasingly support viewing sepsis as a transcriptomically heterogeneous syndrome with reproducible immune-dysregulation states rather than a single uniform host-response pattern, which is consistent with the framework advanced here.[15-19]

The diagnostic performance of the model should be interpreted pragmatically. Training performance was good and external validation remained acceptable, but attenuation in the validation cohort indicates expected between-cohort heterogeneity rather than complete transportability across all settings. This is not surprising in sepsis, where sampling time, clinical severity, pathogen spectrum, organ support, assay platform, and special populations such as immunocompromised hosts can all reshape host-response readouts and pretest probabilities.[2,4-9,20] From a translational perspective, the value of the three-gene panel lies less in replacing clinical judgment than in offering a compact host-response signal that can be tested prospectively alongside routine markers and clinical scores.

A key conceptual advance provided by the new analyses is clarification of cellular context. The original bulk results highlighted neutrophil-associated pathways and broad immune dysregulation, which remain biologically plausible and consistent with whole-blood composition shifts.[2,3,12] However, recent single-cell studies have also shown that clinically relevant septic remodeling is concentrated within heterogeneous monocyte states with immunosuppressive, diagnostic, or tissue-specific features.[21-24] Our results align with this direction by localizing the oxidative stress-centered three-gene signal to specific monocyte states rather than to a diffuse pan-leukocyte program.

Among the three genes, NFKBIA is particularly relevant when the bulk and single-cell results are viewed together. NFKBIA encodes the classical negative-feedback brake of NF-kB signaling, and its localization within oxidative stress-high monocyte territories is consistent with a setting in which inflammatory induction and counter-regulatory restraint are engaged simultaneously.[25] The overlap between oxidative stress activity, NF-kB-related activity, and NRF2-related activity further supports a model of coupled inflammatory-redox remodeling rather than isolated oxidative injury.[10,19,26,27] In this context, NFKBIA may mark a

putative network position through which antioxidant defense and inflammatory signaling are associated, although direct perturbation experiments are needed.

The trajectory, ligand-prioritization, and perturbation analyses add another mechanistic layer. Rather than depicting septic monocytes as static end states, pseudotime inference suggests an ordered pathological evolution axis along which oxidative stress activity and NFKBIA vary dynamically. Candidate upstream ligands such as IL1B, TNF, IFNG, and CSF1 are biologically congruent with the inflammatory milieu of sepsis and provide a plausible extracellular framework for sustaining the monocyte program. The broad distribution of the reprogramming-related score indicates that this process is not restricted to a rare cell subset but affects a sizeable fraction of monocytes with variable intensity, which is consistent with the clinical heterogeneity and temporal fluidity of sepsis.[13,14,21-24]

These additions also sharpen the interpretation of the pathway results. Neutrophil degranulation, translation initiation, and oxidative phosphorylation-related signals identified in the bulk analysis remain relevant, but the single-cell data argue that at least part of the stress-response component is rooted in monocyte state remodeling rather than merely in shifts in overall cell proportions.[10,11,26] This distinction matters because it raises the possibility that the three-gene signature may reflect a disease process with cellular organization and regulatory structure, making it more amenable to mechanistic follow-up than a purely empirical classifier.

Several limitations remain. First, the diagnostic model was derived retrospectively from public bulk datasets and showed only moderate external validation, so prospective multicenter benchmarking against clinically relevant comparators, including procalcitonin, SOFA score, blood culture, and routine inflammatory markers, is still required. Second, the discovery and validation datasets differed in platform, specimen composition, sampling context, and clinical annotation, which may limit transportability. Third, the study lacks direct qPCR, ELISA, Western blot, phospho-signaling, or functional perturbation validation. Fourth, the single-cell analyses were incorporated as mechanistic support rather than as a formally independent clinical validation cohort and therefore should be interpreted as biologically reinforcing rather than definitive. Finally, model performance has not yet been tested across immunocompromised populations, distinct infection sources, pathogen categories, geographic

cohorts, or integrated omics settings.[20,28-30]

Overall, the present revision connects a compact diagnostic panel to a specific monocyte-centered disease program, providing a more cautious and biologically contextualized framework for future clinical and experimental testing.

Inflammatory ligand milieu	Monocyte signaling state	Oxidative-stress remodeling	Three-gene host-response signal	Diagnostic interpretation
IL1B, TNF, IFNG, and CSF1 candidate inputs	NF-kB/NRF2-related monocyte activity	Redox-inflammatory transcriptional remodeling	MAPK1, NFKBIA, and RAF1 signature	Sepsis risk stratification requiring prospective testing

Scheme 2. Proposed mechanistic interpretation linking inflammatory ligand input, monocyte NF-kB/NRF2-related activity, oxidative-stress remodeling, and the MAPK1/NFKBIA/RAF1 host-response signature. The scheme summarizes a hypothesis-generating framework and does not imply proven causal regulation.

5. Conclusion

The MAPK1/NFKBIA/RAF1 signature may provide an interpretable host-response signal for sepsis diagnosis, and the monocyte-centered single-cell findings support its biological plausibility. Prospective clinical validation and orthogonal experimental studies are required before clinical utility or causal relevance can be inferred.

Declarations

Ethics approval and consent to participate: Not applicable. This study analyzed only publicly available, de-identified datasets and did not involve direct recruitment, intervention, or specimen collection from human participants by the authors. Public data sources included GEO accessions GSE6535, GSE54514, GSE46955, and GSE290679.

Consent for publication: Not applicable.

Availability of data and materials: The datasets analysed during the current study are available in the Gene Expression Omnibus (GEO) repository: GSE6535

(<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6535>), GSE54514
(<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54514>), and GSE46955
(<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46955>). The complementary
single-cell RNA-seq dataset used for contextual analyses is available under GEO accession
GSE290679 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE290679>). Processed
analysis outputs are available from the corresponding author on reasonable request.

Competing interests: The authors declare that they have no competing interests.

Funding: No specific funding was declared for this study.

Authors' contributions: Kun Liu contributed to conceptualization, data curation, formal
analysis, methodology, visualization, and writing of the original draft. Qun Ye contributed to
methodology, validation, interpretation, and writing-review and editing. Kunhai Xiao
contributed to conceptualization, supervision, project administration, correspondence, and
writing-review and editing. All authors read and approved the final manuscript.

Acknowledgements: The authors thank the investigators who generated and shared the
public GEO datasets analyzed in this study.

References

1. Evans L, Rhodes A, Alhazzani W, et al. Surviving sepsis campaign: international
guidelines for management of sepsis and septic shock 2021. *Intensive Care Med.*
2021;47(11):1181-1247. doi:10.1007/s00134-021-06506-y.
2. Cajander S, Kox M, Scicluna BP, et al. Profiling the dysregulated immune response in
sepsis: overcoming challenges to achieve the goal of precision medicine. *Lancet Respir Med.*
2024;12(4):305-322. doi:10.1016/S2213-2600(23)00330-2.
3. Giamarellos-Bourboulis EJ, Aschenbrenner AC, Bauer M, et al. The pathophysiology of
sepsis and precision-medicine-based immunotherapy. *Nat Immunol.* 2024;25(1):19-28.
doi:10.1038/s41590-023-01660-5.
4. Hanson KE, Tsalik EL. Host immune response profiling for the diagnosis of infectious
diseases. *J Infect Dis.* 2025;232(5):1029-1039. doi:10.1093/infdis/jiaf428.

5. Scicluna BP, Cano-Gamez K, Burnham KL, et al. A consensus blood transcriptomic framework for sepsis. *Nat Med.* 2025;31(12):4119-4130. doi:10.1038/s41591-025-03964-5.
6. Balk R, Esper AM, Martin GS, et al. Rapid and robust identification of sepsis using SeptiCyte RAPID in a heterogeneous patient population. *J Clin Med.* 2024;13(20):6044. doi:10.3390/jcm13206044.
7. von der Forst M, Back L, Tourelle KM, et al. SeptAsTERS: SeptiCyte RAPID as assessment tool for early recognition of sepsis, a prospective observational study. *Infection.* 2025;53(3):953-965. doi:10.1007/s15010-024-02409-4.
8. Llitjos JF, Carrol ED, Osuchowski MF, et al. Enhancing sepsis biomarker development: key considerations from public and private perspectives. *Crit Care.* 2024;28(1):238. doi:10.1186/s13054-024-05032-9.
9. He RR, Yue GL, Dong ML, et al. Sepsis biomarkers: advancements and clinical applications-a narrative review. *Int J Mol Sci.* 2024;25(16):9010. doi:10.3390/ijms25169010.
10. Lin Y, Alhaskawi A, Chen L, et al. Recent advances in understanding oxidative stress in sepsis: pathogenic roles and antioxidant therapeutic prospects-a narrative review. *Front Pharmacol.* 2025;16:1695992. doi:10.3389/fphar.2025.1695992.
11. Hu D, Sheeja Prabhakaran H, Zhang YY, et al. Mitochondrial dysfunction in sepsis: mechanisms and therapeutic perspectives. *Crit Care.* 2024;28(1):292. doi:10.1186/s13054-024-05069-w.
12. Zhang J, Shao Y, Wu J, et al. Dysregulation of neutrophil in sepsis: recent insights and advances. *Cell Commun Signal.* 2025;23(1):87. doi:10.1186/s12964-025-02098-y.
13. Wang Z, Zhang W, Chen L, et al. Lymphopenia in sepsis: a narrative review. *Crit Care.* 2024;28(1):315. doi:10.1186/s13054-024-05099-4.
14. Ayala A. Mechanisms of immune suppression in sepsis/shock: one investigator's/lab group's experience (SLB 2024 legacy award presentation). *J Leukoc Biol.* 2025;117(8):qiaf108. doi:10.1093/jleuko/qiaf108.

15. Moore AR, Zheng H, Ganesan A, et al. A consensus immune dysregulation framework for sepsis and critical illnesses. *Nat Med.* 2025;31(12):4084-4096. doi:10.1038/s41591-025-03956-5.
16. Chenoweth JG, Colantuoni C, Striegel DA, et al. Gene expression signatures in blood from a West African sepsis cohort define host response phenotypes. *Nat Commun.* 2024;15(1):4606. doi:10.1038/s41467-024-48821-0.
17. Chenoweth JG, Brandsma J, Striegel DA, et al. Sepsis endotypes identified by host gene expression across global cohorts. *Commun Med (Lond).* 2024;4(1):120. doi:10.1038/s43856-024-00542-7.
18. Brandes-Leibovitz R, Riza A, Yankovitz G, et al. Sepsis pathogenesis and outcome are shaped by the balance between the transcriptional states of systemic inflammation and antimicrobial response. *Cell Rep Med.* 2024;5(11):101829. doi:10.1016/j.xcrm.2024.101829.
19. Gao X, Cai S, Li X, et al. Sepsis-induced immunosuppression: mechanisms, biomarkers and immunotherapy. *Front Immunol.* 2025;16:1577105. doi:10.3389/fimmu.2025.1577105.
20. Deinhardt-Emmer S, Chousterman BG, Schefold JC, et al. Sepsis in patients who are immunocompromised: diagnostic challenges and future therapies. *Lancet Respir Med.* 2025;13(7):623-637. doi:10.1016/S2213-2600(25)00124-9.
21. Yao RQ, Zhao PY, Li ZX, et al. Single-cell transcriptome profiling of sepsis identifies HLA-DR(low)S100A(high) monocytes with immunosuppressive function. *Mil Med Res.* 2023;10(1):27. doi:10.1186/s40779-023-00462-y.
22. Hua N, Kong L, Fan L, et al. Identification of CD38(high) monocyte as a candidate diagnostic biomarker and therapeutic target for sepsis. *Adv Sci (Weinh).* 2025;12(23):2500457. doi:10.1002/advs.202500457.
23. Sun L, Zhang P, Zhang H, et al. Single-cell transcriptomic profiles of peripheral blood immune cells reveal early monocyte and platelet activation in the transition from high-risk states to clinical sepsis. *Sci Rep.* 2025;15(1):32879. doi:10.1038/s41598-025-17078-y.

24. Ye Q, Lai X, Liu Y, et al. Single-cell multi-omic landscape reveals anatomical-specific immune features in adult and pediatric sepsis. *Nat Immunol.* 2026;27(1):150-165. doi:10.1038/s41590-025-02345-x.

25. Mukherjee T, Chatterjee B, Dhar A, Kundu M. The NF- κ B signaling system in the immunopathogenesis of infectious diseases. *Sci Signal.* 2024;17(856):eadh1641. doi:10.1126/scisignal.adh1641.

26. Yu X, Hu X, Wang D, et al. Macrophage S1PR2 drives sepsis-induced immunosuppression by exacerbating mitochondrial fragmentation. *Am J Respir Cell Mol Biol.* 2025;72(6):615-626. doi:10.1165/rcmb.2024-0161OC.

27. Yang L, Gao Q, Li Q, Guo S. PD-L1 blockade improves survival in sepsis by reversing monocyte dysfunction and immune disorder. *Inflammation.* 2024;47(1):114-128. doi:10.1007/s10753-023-01897-0.

28. Cummings MJ, Lutwama JJ, Tomoiaga AS, et al. Identification of transcriptomic sepsis endotypes in sub-Saharan Africa: derivation, validation, and global alignment in two Ugandan cohorts. *Intensive Care Med.* 2025;51(9):1573-1586. doi:10.1007/s00134-025-08047-0.

29. Jin X, Shen H, Zhou P, et al. Research progress on sepsis diagnosis and monitoring based on omics technologies: a review. *Diagnostics (Basel).* 2025;15(22):2887. doi:10.3390/diagnostics15222887.

30. Spottiswoode N, Neyton LP, Mick E, et al. Host-microbe multiomic profiling predicts mortality in sepsis. *Am J Respir Crit Care Med.* 2026;212(1):95-104. doi:10.1164/rccm.202410-1996OC.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FundingStatement.docx](#)
- [Response.docx](#)