

MFGRN: Multi-Algorithm Fusion for Gene Regulatory Network Inference in Sickle Cell Disease

Zhenhao Zan

Dali University

Zhangbin Chen

Dali University

Qian Lu

Dali University

Jiaxin Wang

Dali University

Junpeng Zhu

Dali University

Qing zhao

Dali University

Pinghong Wang

Dali University

Zidan Hu

Dali University

Shujie Fan

Dali University

Xuemei Wei

wxm201@dali.edu.cn

Dali University

Junpeng Zhang

Dali University

Article

Keywords: Gene regulatory network, Multi-algorithm fusion, Network modules, Sickle cell disease, Transcription factors

Posted Date: October 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10880995/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

MFGRN: Multi-Algorithm Fusion for Gene Regulatory Network Inference in Sickle Cell Disease

Zhenhao Zan¹, Zhangbin Chen¹, Qian Lu¹, Jiaxin Wang¹, Junpeng Zhu¹, Qing Zhao¹,
Pinghong Wang¹, Zidan Hu¹, Shujie Fan¹, Xuemei Wei^{1*}, Junpeng Zhang^{1*}

¹ School of Engineering, Dali University, No. 2 Hongsheng Road, Dali 671003, China

*Correspondence to: School of Engineering, Dali University, No. 2 Hongsheng Road,
Dali 671003, China.

E-mail addresses: wxm201@dali.edu.cn (Xuemei Wei); zjp@dali.edu.cn (Junpeng
Zhang).

Other authors' emails:

Zhenhao Zan: zanzhenhao@stu.dali.edu.cn

Zhangbin Chen: chenzhangbin@stu.dali.edu.cn

Qian Lu: luqian@stu.dali.edu.cn

Jiaxin Wang: wangjiaxin@stu.dali.edu.cn

Junpeng Zhu: zhujunpeng@stu.dali.edu.cn

Qing Zhao: zhaoqing@stu.dali.edu.cn

Pinghong Wang: wangpinghong@stu.dali.edu.cn

Zidan Hu: huzidan@stu.dali.edu.cn

Shujie Fan: fanshujie@stu.dali.edu.cn

Abstract

Gene expression is regulated by dynamic gene regulatory networks (GRNs), whose dysregulation may contribute to clinical heterogeneity in sickle cell disease (SCD). We developed MFGRN, a multi-algorithm fusion framework integrating regulatory relationships inferred by 3DCEMA, DeepFGRN, and DeepSEM. Using whole-blood bulk RNA-seq data, we reconstructed GRNs for healthy controls and patients with SCD before and after exercise. MFGRN consistently achieved higher AUROC values than individual methods across four primary conditions and two independent external datasets. The inferred GRNs exhibited pronounced small-world characteristics, consistent with common topological features of biological regulatory networks. Key transcription factors (TFs) were associated with hematopoiesis, immune regulation, development, and transcriptional control. In the pre-exercise SCD network, key TFs were enriched in rhythmic process and respiratory system development, while functional modules were enriched in oxidative stress, cytokine regulation, immune processes, cellular stress responses, and metabolic regulation. Overall, MFGRN provides an effective strategy for GRN inference and reveals SCD-associated regulatory dysregulation under resting and exercise conditions, while identifying candidate regulatory factors for experimental investigation.

Keywords: Gene regulatory network, Multi-algorithm fusion, Network modules, Sickle cell disease, Transcription factors

1. Introduction

Gene expression is governed by multiple regulators acting across different regulatory layers, forming dynamic gene regulatory networks (GRNs) that control diverse biological processes[1]. At the epigenetic level, DNA methylation and histone modifications regulate chromatin accessibility and influence transcriptional machinery. At the transcriptional level, transcription factors (TFs) bind to cis-regulatory regions, such as promoters and enhancers, to activate or repress downstream gene expression.

At the post-transcriptional level, non-coding RNAs, including microRNAs (miRNAs) and long non-coding RNAs (lncRNAs), regulate gene expression by affecting mRNA stability, translation, and RNA processing[2]. Through the interplay of these mechanisms, gene regulation functions as a complex and dynamic network rather than a simple linear process. Therefore, GRN construction and analysis provide an important approach for understanding the molecular mechanisms underlying human diseases.

Sickle cell disease (SCD) is a monogenic hemoglobinopathy caused by pathogenic variants in the *HBB* gene, most commonly the point mutation responsible for sickle hemoglobin production. Under deoxygenated conditions, abnormal hemoglobin polymerization causes red blood cells to adopt a sickled shape and contributes to hemolytic anemia and vaso-occlusion[3]. Despite carrying similar disease-causing variants, patients with SCD exhibit substantial clinical heterogeneity in disease severity, pain crisis frequency, and organ damage[4–6]. This heterogeneity suggests that SCD pathophysiology is influenced not only by erythrocyte sickling but also by broader molecular and regulatory alterations. Previous studies have identified numerous SCD-associated gene expression changes; however, analyses focused on individual genes or differential expression alone may not fully capture the coordinated reorganization of regulatory relationships[7]. With the growing availability of high-throughput omics data and advances in statistical learning, machine learning, and deep learning, systematic GRN inference from large-scale transcriptomic data has become an important strategy for investigating disease-associated regulatory mechanisms.

Over the past decade, GRN inference methods have progressed from traditional correlation-, mutual information-, Bayesian network-, and regression-based approaches to machine learning and deep learning frameworks[8]. Modern computational models offer greater capacity to capture nonlinear relationships, learn complex expression patterns, and integrate multiple sources of regulatory information. Representative methods, including CNNC[9], SCENIC+[10], LINGER[11], and DeepSEM[12], illustrate the diversity of computational strategies for reconstructing regulatory networks and identifying key regulatory factors and functional modules. However, GRN inference performance often varies across datasets and biological contexts

because of differences in model assumptions, data characteristics, and parameter settings[13]. Consequently, networks inferred by a single method may contain method-specific biases and may not consistently capture the full complexity of biological regulation.

To address these challenges, we developed MFGRN, a multi-algorithm fusion framework for GRN inference. MFGRN integrates regulatory relationships predicted by three complementary deep learning-based methods—3DCEMA, DeepSEM, and DeepFGRN—using a majority-voting strategy to reduce method-specific bias and improve the consistency of inferred regulatory relationships. Using whole-blood bulk RNA-seq data, we reconstructed condition-specific GRNs for healthy controls and patients with SCD before and after exercise and evaluated the framework in two independent external datasets. We further analyzed network topology, TF centrality, key regulatory factors, and functional modules to characterize SCD-associated regulatory dysregulation. This study provides a network-level framework for examining regulatory changes associated with SCD and exercise and for identifying candidate regulatory factors for further experimental investigation.

2. Materials and Methods

2.1 Data Collection and Preprocessing

The GSE244401 dataset was obtained from the Gene Expression Omnibus (GEO) and used as the primary dataset[14]. It contains whole-blood bulk RNA-seq profiles from 23 children with SCD and 17 healthy controls before and after exercise, grouped as CON_T1, CON_T2, SCD_T1, and SCD_T2. Gene identifiers were converted to standardized gene symbols, and genes with zero expression across all samples within each group were removed. Human TF annotations were obtained from HOCOMOCO v14[15], while TF–target relationships from hTFtarget were used as prior information for GRN construction[16]; only regulatory pairs whose TF and target were present in the corresponding expression matrix were retained. GSE254951[17] and GSE117221[18] were used for external evaluation. For GSE254951, only pre-treatment SCD samples were retained, whereas GSE117221 contains expression profiles from

patients with β -thalassemia and healthy controls. For both datasets, zero-expression genes were removed, and the 10,000 genes with the highest mean expression levels were retained for GRN construction.

2.2 MFGRN Framework

To reduce method-specific biases in single GRN inference approaches, we developed MFGRN, a multi-algorithm fusion framework for network construction and downstream analysis (Figure 1). Whole-blood bulk RNA-seq data, TF annotations, and prior regulatory information were used as inputs. Three deep learning-based methods were independently applied to infer GRNs:

1. DeepSEM[12], which combines a variational autoencoder with structural equation modeling to capture nonlinear gene regulatory relationships and directed dependency structures in the latent space;
2. 3DCEMA[19], a three-dimensional convolutional neural network-based approach that identifies nonlinear regulatory patterns by learning local joint gene-expression distributions;
3. DeepFGRN[20], which combines generative adversarial learning with directed graph embedding for GRN inference.

For each method, the top 10% of regulatory edges ranked by predicted confidence score were retained as candidate interactions. A majority-voting strategy was then applied, and regulatory relationships supported by at least two of the three algorithms were retained to construct the final high-confidence GRN. ChIP-seq-supported TF-target relationships were subsequently used to evaluate the inferred networks. Downstream analyses included network topology characterization, degree centrality analysis, key TF identification, functional enrichment analysis, and module detection. Together, these analyses enabled the characterization of SCD-associated regulatory features at the levels of global network architecture, regulatory nodes, and functional modules.

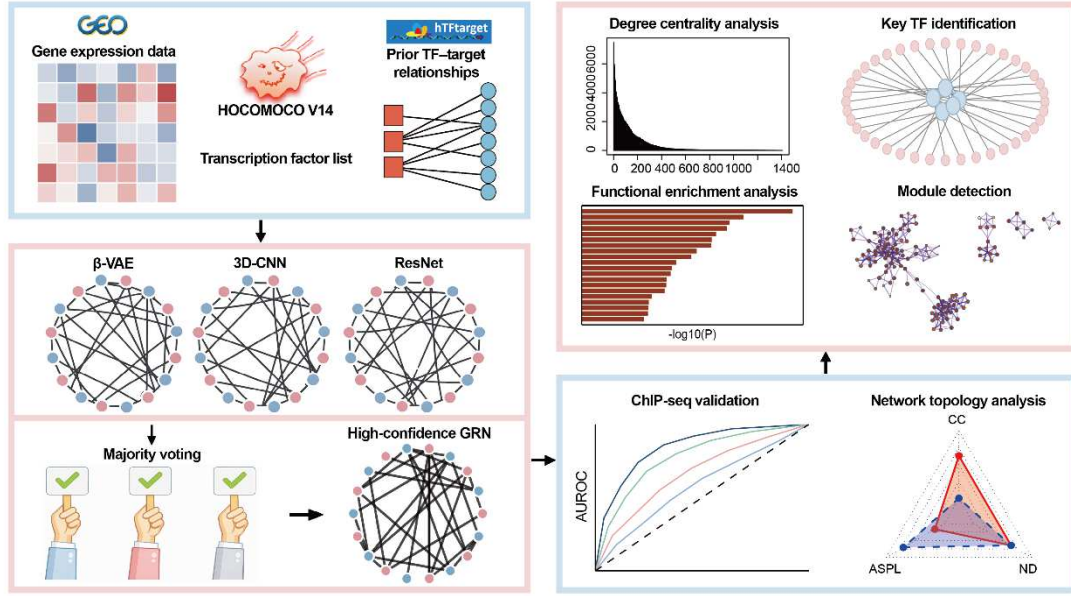


Figure 1. Overview of the MFGRN framework for GRN construction and systematic network analysis. β -VAE, beta-variational autoencoder; 3D-CNN, three-dimensional convolutional neural network; ResNet, residual network; CC, clustering coefficient; ASPL, average shortest path length; ND, network density.

2.3 Differential Expression Analysis

Differential expression analysis between SCD_T1 and CON_T1 was performed using edgeR and limma in R. Genes shared between the two groups were retained, and TFs were identified by matching standardized gene symbols to the human TF list. Raw counts were normalized using the trimmed mean of M-values (TMM) method and transformed with voom before linear modeling. Empirical Bayes moderation was applied, and P values were adjusted using the Benjamini–Hochberg procedure. TFs with $|\log_2\text{FC}| \geq 0.5$ and $\text{FDR} < 0.05$ were considered significantly differentially expressed and visualized using a volcano plot.

2.4 ChIP-seq Data Collection and Network Evaluation

ChIP-seq peak datasets for *BCL11A*, *GATA1*, and *TAL1* were obtained from GSE103445[21], GSE36985[22], and GSE70660[23], respectively. Peaks were annotated using ChIPseeker, and genes associated with ChIP-seq peaks located within 2 kb upstream to 0.5 kb downstream of their transcription start sites were defined as ChIP-seq-supported putative targets. For MFGRN, interactions supported by at least

two algorithms were retained, and their mean edge weights were used as fusion scores. Predicted TF–target pairs overlapping the corresponding ChIP-seq-supported pairs were labeled as positive and the remaining pairs as negative. AUROC was calculated using the pROC package in R. Overlap significance was assessed using a one-sided Fisher’s exact test.

2.5 Network Topology Analysis

To evaluate the global topological properties of the inferred GRNs, we performed network analysis using the igraph package in R. The calculated metrics included network density (ND), mean clustering coefficient (CC), and average shortest path length (ASPL).

To assess whether the inferred networks exhibited small-world characteristics, we generated random networks with the same numbers of nodes and edges as the corresponding observed networks as null models. The CC and ASPL of each random network were calculated, and their mean values were used as random-network references. The small-world index (σ), proposed by Humphries and Gurney[24], was calculated as follows:

$$\sigma = \frac{C/C_{\text{rand}}}{L/L_{\text{rand}}} \quad (1)$$

where C and L denote the mean clustering coefficient and average shortest path length of the observed network, respectively, and C_{rand} and L_{rand} denote the corresponding mean values obtained from the random networks. A network was considered to exhibit small-world properties when $\sigma > 1$.

2.6 Null Model Evaluation

To assess whether the overlap between inferred GRNs and ChIP-seq-supported targets could be explained by network size or random effects, we performed a null-model analysis. For each GRN, 1,000 random networks with the same numbers of nodes and edges were generated from the corresponding candidate TF–target background. Their overlaps with the ChIP-seq-supported target set were used to construct a null distribution, against which the observed overlap was compared. The empirical P value was calculated as

$$P_{\text{emp}} = \frac{1 + \sum_{b=1}^B I(R_b \geq R_{\text{obs}})}{B+1} \quad (2)$$

where R_{obs} denotes the observed overlap, R_b denotes the overlap obtained from the b -th random network, and $B=1000$ denotes the number of randomizations.

2.7 Identification of Key Transcription Factors

For each inferred GRN containing n nodes and m edges, the average out-degree (λ) was calculated as follows:

$$\lambda = \frac{m}{n} \quad (3)$$

Assuming that node out-degree in a random network approximately follows a Poisson distribution, a right-tailed test was performed for each candidate TF based on its observed out-degree. P values were adjusted using the Benjamini–Hochberg (BH) procedure. TFs were identified as key regulators if they met both criteria: $q < 0.05$ and an out-degree greater than or equal to the 99.9th percentile of the corresponding Poisson distribution[25–27].

To identify exercise-associated key TFs in SCD, key TFs identified across the four GRNs were combined, and their total degree centralities were calculated in CON_T1, CON_T2, SCD_T1, and SCD_T2. Using CON_T1 as a fixed healthy reference, the relative degree changes before and after exercise were calculated as

$$\Delta_{\text{before}} = \log_2 \left(\frac{D_{\text{SCD_T1}} + 1}{D_{\text{CON_T1}} + 1} \right), \Delta_{\text{after}} = \log_2 \left(\frac{D_{\text{SCD_T2}} + 1}{D_{\text{CON_T1}} + 1} \right) \quad (4)$$

where D denotes TF degree centrality. TFs were considered to show an exercise-associated shift toward the healthy state when $|\Delta_{\text{after}}| < |\Delta_{\text{before}}|$. To focus on TFs with substantial changes, exercise-associated key TFs were further required to satisfy $|\Delta_{\text{before}} - \Delta_{\text{after}}| > 1$.

2.8 Gene Ontology Enrichment Analysis

Gene Ontology (GO) enrichment analysis was performed using Metascape[28], with an adjusted P value < 0.01 considered significant. For visualization and functional interpretation, the top 20 GO terms from the key TF analysis and the top five from the module analysis were retained based on statistical significance.

2.9 Network Module Identification

Network modules were identified using the Module function in the miR spongeR[29] package with the Infomap community detection algorithm. Module detection was performed independently for each GRN, and modules containing 10–300 genes were retained. These modules were ranked by the number of key TFs, and the five highest-ranking modules in each GRN were selected for further analysis.

3. Results

3.1 Dysregulation of the Regulatory State in SCD

To explore the transcriptional regulatory characteristics of SCD, we compared filtered genes and TFs across the four experimental groups. Venn diagram analysis shows substantial overlap among the gene sets of CON_T1, CON_T2, SCD_T1, and SCD_T2, with few group-specific genes (Figure 2a). A similar pattern is observed for TFs, most of which are shared across the four groups (Figure 2b). Differential expression analysis further reveals significant TF expression changes between SCD_T1 and CON_T1 (Figure 2c). Most differentially expressed TFs are downregulated in SCD_T1, including *KLF2*, *KLF16*, *CEBPB*, *NR2F6*, and several ZNF family members, many of which are involved in hematopoietic development, immune regulation, and transcriptional control. In contrast, a small number of TFs, including *SOX6* and *NFIX*, are upregulated. These findings indicate that despite broadly similar gene and TF compositions across the four groups, marked TF expression changes occur in the pre-exercise SCD state, suggesting dysregulated transcriptional regulation in SCD.

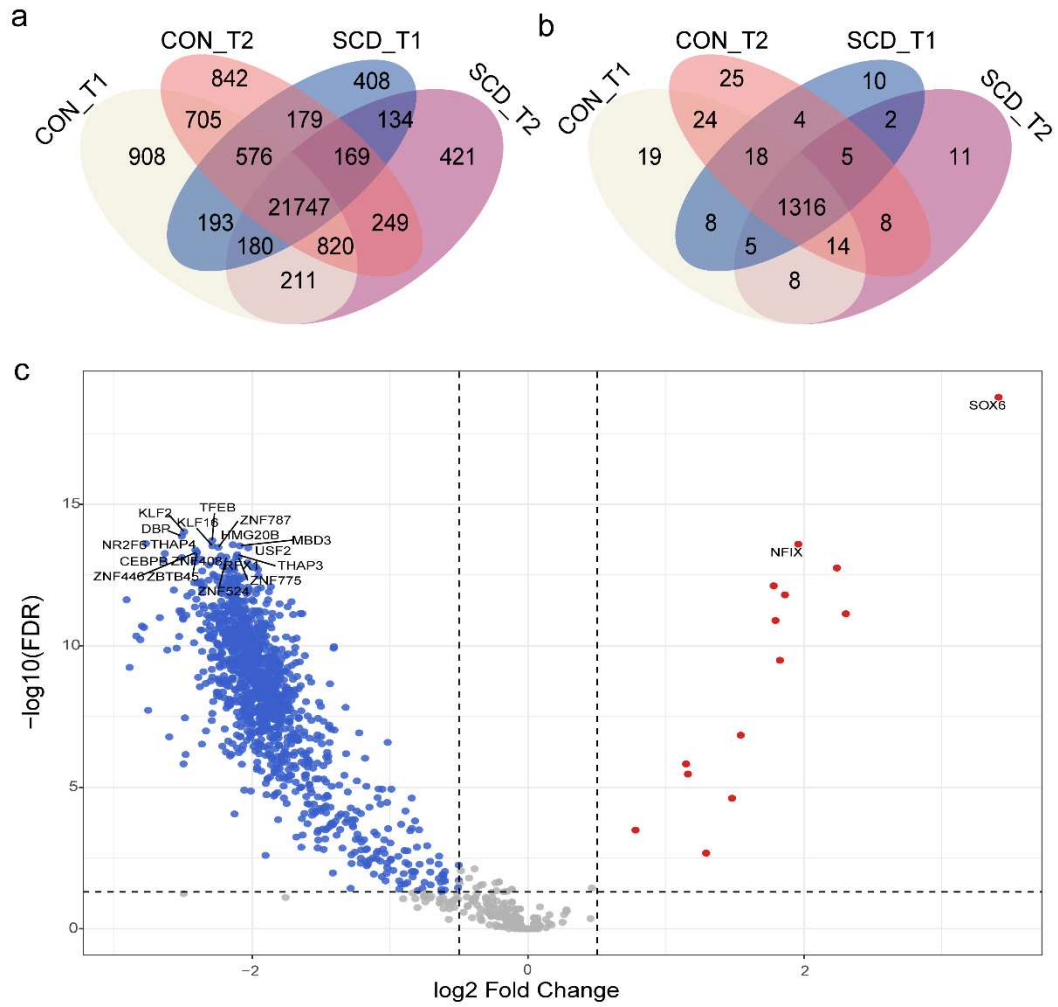


Figure 2. Shared gene and TF composition across the four groups and differentially expressed TFs in SCD. (a) Venn diagram showing the overlap among the filtered gene sets of CON_T1, CON_T2, SCD_T1, and SCD_T2. (b) Venn diagram showing the overlap among the filtered TF sets of the four groups. (c) Volcano plot showing differential TF expression between SCD_T1 and CON_T1. The x-axis represents $\log_2$ fold change ($\log_2\text{FC}$), and the y-axis represents $-\log_{10}(\text{FDR})$. Red dots indicate significantly upregulated TFs in SCD_T1, blue dots indicate significantly downregulated TFs, and gray dots indicate TFs without significant differential expression. Representative differentially expressed TFs are labeled.

3.2 MFGRN Shows Consistent GRN Inference Performance Across Independent Datasets

To comprehensively evaluate MFGRN, we constructed GRNs under four biological conditions using GSE244401 and further evaluated its performance in two external datasets, GSE254951 and GSE117221. Using experimentally supported ChIP-

seq data as the reference, AUROC was used to compare MFGRN with individual GRN inference algorithms.

In GSE244401, MFGRN achieves AUROC values of 0.7608, 0.6361, 0.7109, and 0.7851 for CON_T1, CON_T2, SCD_T1, and SCD_T2, respectively, exceeding the corresponding best individual methods (0.6598, 0.5891, 0.7097, and 0.7357; Figure 3a). In GSE254951, MFGRN achieves an AUROC of 0.6211 and outperforms all individual algorithms. In GSE117221, MFGRN achieves AUROC values of 0.5492 and 0.5481 for the control and disease groups, respectively, again outperforming the best individual methods (Figure 3b).

The best-performing individual algorithm varies across conditions and datasets. 3DCEMA performs best in CON_T1, DeepFGRN in CON_T2, SCD_T1, SCD_T2, and GSE254951, and DeepSEM in both groups of GSE117221 (Figures 3a–b), indicating that individual algorithm performance depends on biological context and dataset characteristics. In contrast, MFGRN achieves higher AUROC values than the best individual method across all seven evaluation tasks. A paired t test across these tasks shows a significant difference ($P = 0.006$; Supplementary Table S1), supporting the overall performance advantage of the multi-algorithm fusion framework.

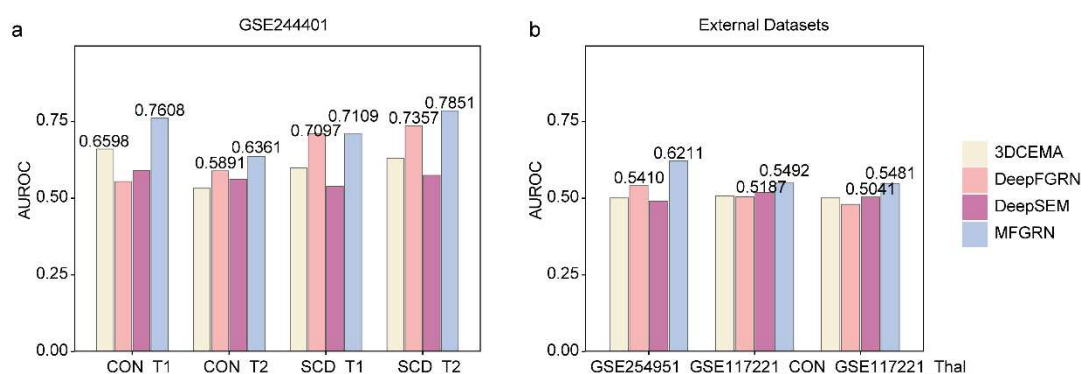


Figure 3. Comparison of GRN inference performance between MFGRN and individual algorithms. (a) AUROC values of MFGRN, 3DCEMA, DeepFGRN, and DeepSEM across the four biological conditions in the GSE244401 dataset. (b) AUROC values in the two external evaluation datasets, GSE254951 and GSE117221. Numerical labels above the bars indicate the AUROC values of MFGRN and the best-performing individual algorithm for each condition or dataset. MFGRN consistently achieves the highest AUROC across all seven evaluation tasks.

3.3 Contribution Analysis of MFGRN Components

To evaluate the contribution of each algorithm to MFGRN, we conducted an ablation study comparing individual algorithms, dual-algorithm fusion strategies, and the complete three-algorithm framework (Figure 4). Dual-algorithm fusion shows condition-dependent performance and does not consistently outperform the best individual method, indicating that partial integration alone is insufficient for consistent gains across datasets and biological states. In contrast, MFGRN consistently achieves the highest AUROC values under all evaluated conditions, supporting the benefit of integrating all three approaches. These results suggest that the three algorithms capture complementary regulatory information and that excluding any one may reduce overall performance consistency. By integrating evidence from distinct modeling strategies, MFGRN reduces method-specific limitations and improves the consistency and predictive performance of GRN inference across biological states and datasets.

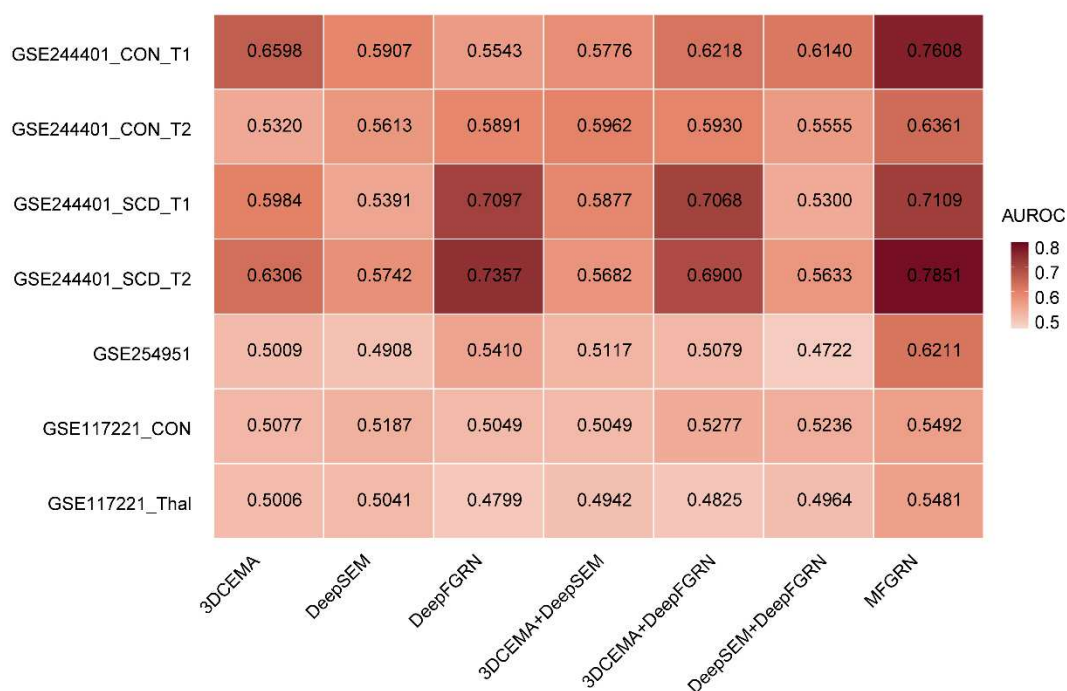


Figure 4. Ablation analysis of MFGRN inference performance. The heatmap shows the AUROC values of the individual algorithms (3DCEMA, DeepSEM, and DeepFGRN), dual-algorithm fusion variants, and the complete three-algorithm fusion framework (MFGRN) across different datasets and biological conditions. The x-axis represents the individual algorithms and fusion variants, whereas the y-axis represents the evaluated datasets and biological states. Numerical values within each cell indicate the corresponding AUROC scores, with higher values represented by greater color

intensity according to the scale on the right.

3.4 GRNs Inferred by MFGRN Exhibit Non-Random Topological Characteristics and Biological Plausibility

To evaluate the topological properties and experimental support for GRNs inferred by MFGRN, we assessed their topological organization and consistency with independent experimental evidence. First, GRNs from the four biological conditions were compared with random networks matched for node and edge counts (Figure 5a). The inferred networks consistently show higher CC values and shorter ASPLs than the corresponding random networks, indicating stronger local clustering and more efficient global connectivity, consistent with the small-world properties commonly observed in biological regulatory networks[30].

Second, three SCD-related TFs with publicly available ChIP-seq data—*GATA1*, *TALI*, and *BCL11A*—were selected for validation. Predicted targets from the four GRNs were compared with ChIP-seq-supported target genes. Substantial overlap is observed across the examined TFs and biological conditions, with most comparisons reaching statistical significance (Figure 5b; Supplementary Table S2).

To determine whether these overlaps could be explained by network size or chance, we further performed a null-model analysis. The observed overlaps consistently exceed the mean overlaps from matched random networks, with most comparisons remaining statistically significant (Figure 5c; Supplementary Table S3). These results indicate that the agreement between predicted regulatory relationships and ChIP-seq-supported targets is not attributable solely to random effects or network size. Collectively, the topological, ChIP-seq, and null-model results support the biological plausibility and interpretability of the inferred GRNs and provide a suitable basis for downstream identification of key regulatory factors and functional modules.

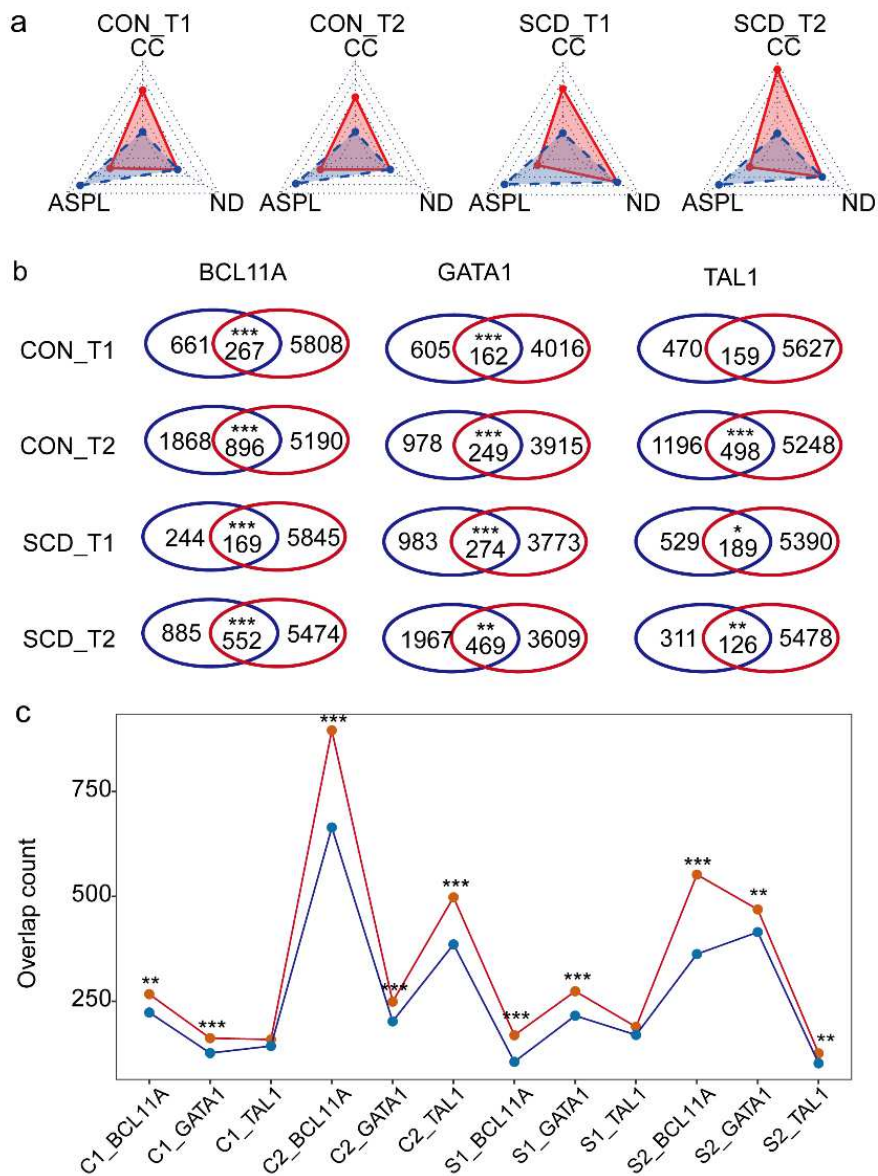


Figure 5. Network topological characteristics of the inferred GRNs and validation of predicted regulatory interactions. (a) Radar plots comparing the topological properties of the inferred networks (red) with those of the corresponding random networks (blue) across the four biological conditions. (b) Venn diagrams showing the overlap between ChIP-seq-supported target gene sets (red circles) and predicted target gene sets (blue circles) for *GATA1*, *TAL1*, and *BCL11A* across the four biological conditions. Numbers in the overlapping regions indicate the number of shared target genes. (c) Null-model analysis comparing the observed and random overlaps. The red line indicates the observed number of overlapping genes between MFGRN-predicted and ChIP-seq-supported target sets, whereas the blue line indicates the mean overlap obtained from 1,000 matched random networks. Asterisks indicate empirical significance in the null-model analysis: * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

3.5 Differences in Network Topology and TF Centrality Across Biological States

Comparing GRNs between SCD and healthy conditions helps identify key regulatory factors and characterize disease-associated network changes. We compared the GRNs across groups from two perspectives: global network architecture and TF centrality. Global topology differs substantially among the inferred GRNs. Compared with control networks, SCD networks show marked differences in CC, ASPL, and ND, which become more pronounced after exercise. Clear topological differences are also observed between SCD_T1 and SCD_T2 (Figure 6a), suggesting that exercise is associated with GRN remodeling in SCD, including changes in local connectivity and global path structure.

At the node level, TF degree centrality also varies across the four biological states. Some TFs show high centrality in CON_T1, whereas others exhibit increased centrality under SCD-related conditions (Figure 6b). Notably, several TFs that differ between SCD_T1 and CON_T1 shift toward control-like centrality levels after exercise, indicating that both disease status and exercise are associated with changes in TF-centered regulatory organization. However, these shifts do not necessarily represent complete restoration of the SCD network to the healthy state.

To explore the biological relevance of these changes, we selected the 100 TFs with the largest increases and decreases in degree centrality across group comparisons for functional enrichment analysis (Figure 6c). These TFs are enriched in transcription-related processes, including RNA polymerase II-mediated transcription, RNA and miRNA transcription, as well as processes involving cellular differentiation and fate determination, organogenesis, morphogenesis, immunity, and hematopoiesis. Collectively, these findings indicate that SCD-associated GRN remodeling involves coordinated changes in transcriptional regulation, development, immunity, and hematopoiesis.

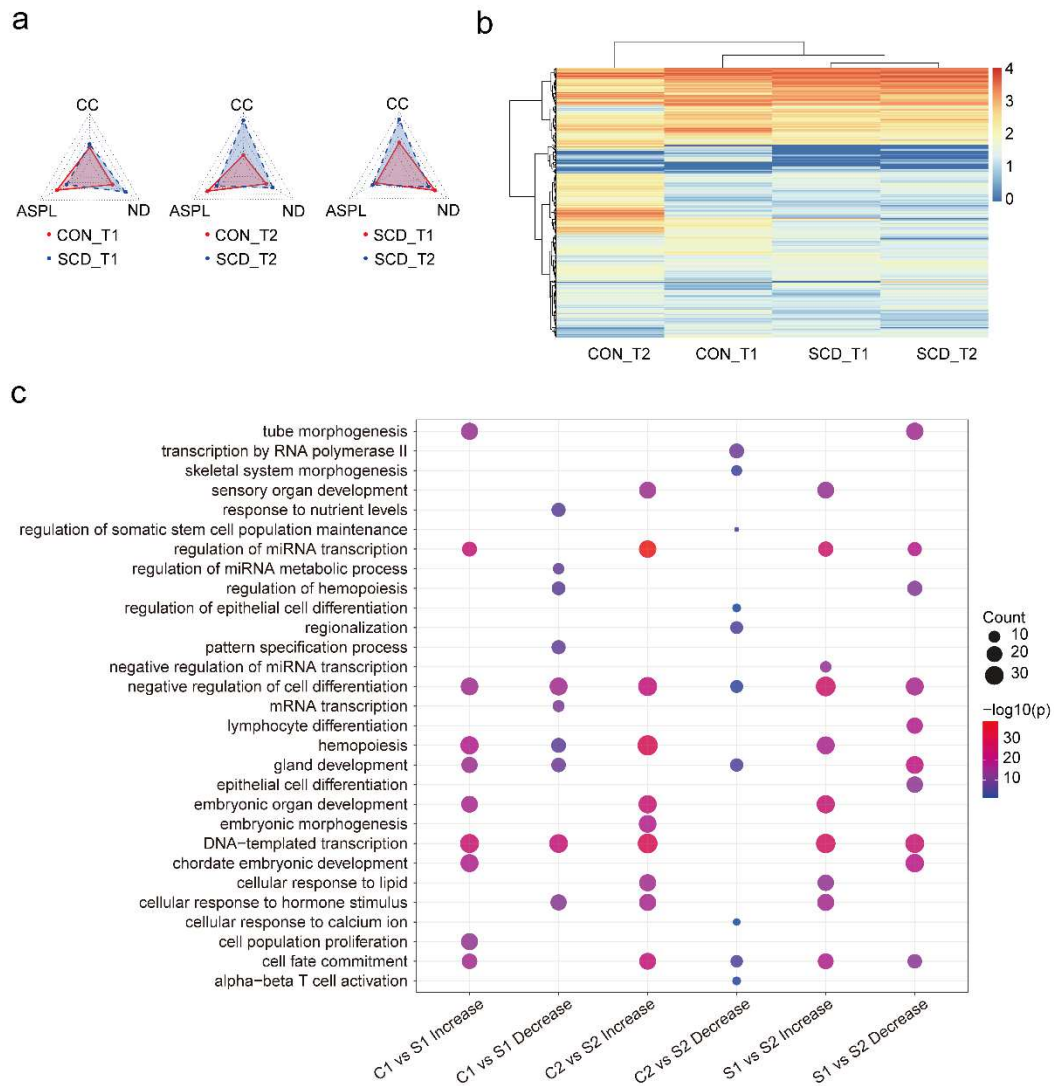


Figure 6. Comparison of network topology and TF degree centrality across biological states. (a) Comparison of network topological properties among the four groups. (b) Heatmap of TF degree centrality across the four inferred networks. Color indicates $\log_{10}$ -transformed degree centrality. The left dendrogram shows similarities among TFs based on their degree-centrality patterns, whereas the top dendrogram shows similarities among networks based on their TF degree-centrality profiles. (c) GO enrichment bubble plot for the 100 TFs showing the largest differences in degree centrality between groups. “Increase” indicates higher degree centrality in the latter group than in the former group, whereas “Decrease” indicates the opposite. Bubble size represents the number of TFs associated with each GO term, and bubble color represents enrichment significance measured by $-\log_{10}(P)$, with redder colors indicating greater statistical significance.

3.6 Key TF Analysis Reveals SCD-Related Regulatory Dysregulation and Exercise-Associated Network Remodeling

Based on TF out-degree distributions, we applied a Poisson model to identify key TFs in each group, yielding 593, 796, 423, and 402 key TFs in CON_T1, CON_T2, SCD_T1, and SCD_T2, respectively (Figure 7a). GO enrichment analysis shows that key TFs in SCD_T1 are prominently enriched in rhythmic process and respiratory system development (Figure 7b). Differential expression analysis was further performed for key TFs associated with these SCD_T1-enriched processes (Figure 7c–d).

Most of these TFs are significantly downregulated in SCD_T1, including circadian regulators such as *DBP*, *TEF*, *NFIL3*, *BHLHE40*, *BHLHE41*, *CLOCK*, and *NPAS2*, as well as stress- and inflammation-related regulators such as *STAT5A*, *STAT5B*, *NFKB2*, *RELB*, *SP1*, *CREB1*, and *ATF4*. Among them, *KLF2*, *HES1*, *SRF*, *RBPJ*, *RARA*, *RARG*, and *EPAS1* show particularly pronounced downregulation. These findings suggest that the pre-exercise SCD regulatory state involves extensive network remodeling related to circadian regulation, respiratory system development, and hypoxia adaptation.

Literature review indicates that several identified key TFs, including *TP53*[31], *STAT5A* and *STAT5B*[32], *NFKB2* and *RELB*[33], *KLF2*[34], *EPAS1*[35], *CREB1*[36], and *HMG42*[37], have been reported in SCD-related, erythroid, inflammatory, or hemoglobin-regulatory contexts. To our knowledge, direct experimental evidence linking *CLOCK*, *NPAS2*, *DBP*, *TEF*, *NFIL3*, *BHLHE40* and *BHLHE41*, *RORA* and *RORC*, *SOX11*, *HES1*, *GATA6*, *LEF1*, *RBPJ*, *SRF*, *GLI1* and *GLI3*, or *WT1* to SCD remains limited (Supplementary Table S4). These TFs may therefore represent candidate regulators for further functional investigation.

To examine exercise-associated regulatory changes in SCD, we focused on key TFs meeting the predefined exercise-associated shift criteria. Functional enrichment analysis of this subset reveals significant enrichment in cellular response to hormone stimulus, intracellular receptor signaling, hemopoiesis, and erythrocyte differentiation (Figure 7e). Overall, these changes suggest coordinated remodeling of TF-centered regulatory networks after exercise rather than complete restoration to the healthy state.

colors indicating greater statistical significance. (c, d) Differential expression of key TFs associated with rhythmic process and respiratory system development, respectively, between SCD_T1 and CON_T1. (e) GO enrichment analysis of key TFs meeting the predefined exercise-associated shift criteria.

3.7 GRNs Under Different Biological States Exhibit Distinct Modular Organization and Functional Profiles

Network module analysis improves the interpretability of complex regulatory networks by organizing local connectivity patterns into functional units[38]. To characterize GRN modular organization across biological states, we applied the Infomap community detection algorithm to the four GRNs. A total of 75, 45, 82, and 57 modules were identified in CON_T1, CON_T2, SCD_T1, and SCD_T2, respectively (Figure 8a). SCD_T1 contains the most modules, whereas CON_T2 contains the fewest, indicating that disease status and exercise are associated with changes in GRN modular organization. The larger number of modules in SCD_T1 may reflect greater reorganization of the disease-associated regulatory network.

To focus on modules containing core regulatory factors, the five modules with the largest numbers of key TFs from each group were selected, yielding 20 representative modules for downstream analysis. GO enrichment analysis reveals distinct functional patterns among the four groups (Figure 8b; Supplementary Table S5). CON_T1 modules are mainly enriched in basic transcriptional processes, including RNA polymerase II-mediated transcription, whereas CON_T2 modules additionally show enrichment in transcriptional regulation and immune-related processes, suggesting exercise-associated changes in the control network.

In contrast, SCD_T1 modules show stronger enrichment in disease-associated processes, including oxidative stress response, cytokine regulation, leukocyte-related processes, and lipid metabolism, consistent with the inflammatory and oxidative-stress environment of SCD. SCD_T2 modules retain immune- and stress-related features while also showing enrichment in metabolic regulation, cellular stress responses, and immune cell migration. These findings suggest that exercise is associated with functional remodeling of key TF-containing modules in SCD rather than complete

restoration to the control state. Overall, the distinct modular structures and enrichment patterns further support SCD-associated regulatory dysregulation at the network level.

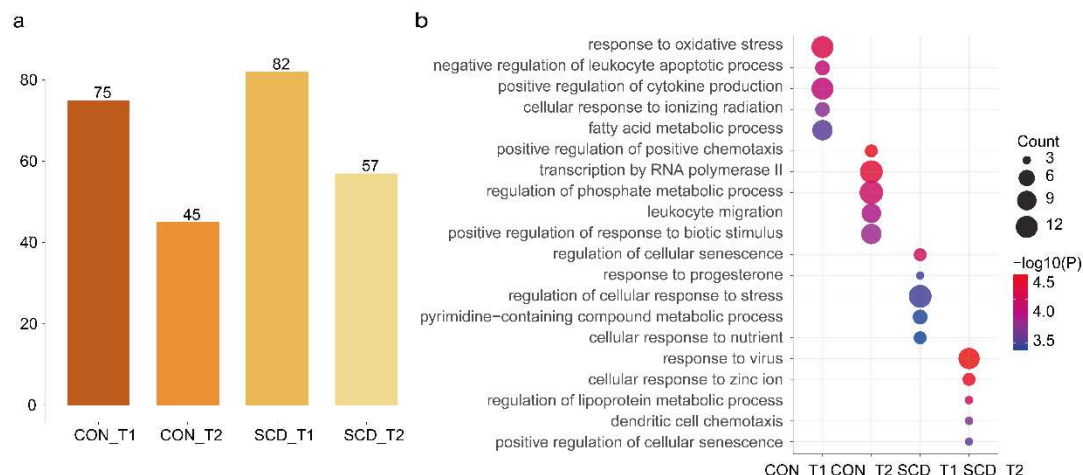


Figure 8. Identification of functional modules in the inferred GRNs and GO enrichment analysis. (a) Numbers of modules identified in the four GRNs using the Infomap algorithm after retaining modules containing 10–300 genes. (b) GO enrichment analysis of representative modules from the four groups. Bubble size represents the number of genes associated with each biological process, and bubble color represents enrichment significance measured by $-\log_{10}(P)$. The displayed GO terms represent selected enrichment results highlighting the major functional characteristics of the representative modules. Complete GO enrichment results for all representative modules are provided in Supplementary Table S5.

4. Discussion

To address the limitations of single GRN inference methods, which rely on specific modeling assumptions and may perform inconsistently across biological contexts, we developed a multi-algorithm fusion framework termed MFGRN. Using whole-blood bulk RNA-seq data, we reconstructed and evaluated GRNs across seven dataset-condition combinations in three datasets (GSE244401, GSE254951, and GSE117221). MFGRN integrates regulatory interactions inferred by DeepSEM, 3DCEMA, and DeepFGRN, three deep learning-based approaches that characterize gene regulation from complementary computational perspectives.

The improved performance of MFGRN may reflect the distinct inductive biases and information representations of these algorithms. DeepSEM incorporates structural equation modeling into a variational autoencoder framework to learn nonlinear

regulatory dependencies and latent structures through optimization of a sparse regulatory adjacency matrix, providing an interpretable representation of gene regulatory relationships[12]. In contrast, 3DCEMA constructs three-dimensional co-expression representations based on gene triplets and applies 3D convolutional neural networks to capture relative expression patterns, enabling detection of regulatory signals that may be missed by conventional pairwise methods[19]. DeepFGRN further complements these approaches by jointly modeling gene expression and directed network topology through graph embedding and adversarial learning, integrating molecular expression patterns with regulatory network structure to infer fine-grained regulatory relationships[20].

Although these methods share the goal of GRN reconstruction, their distinct modeling strategies capture partially overlapping but complementary regulatory information, including latent regulatory dependencies, high-dimensional expression patterns, and network structural features. From an ensemble-learning perspective, integrating models with diverse inductive biases and partially independent errors can reduce method-specific limitations and improve prediction consistency and generalizability[39,40]. This is particularly relevant to GRN inference because biological regulatory systems involve nonlinear interactions, context-dependent regulation, hierarchical organization, and sparse connectivity that are difficult to capture with a single framework[41]. By integrating complementary evidence from DeepSEM, 3DCEMA, and DeepFGRN through consensus-based fusion, MFGRN provides a more comprehensive reconstruction of regulatory landscapes and improves GRN inference across diverse biological states and independent datasets.

The GRNs inferred by MFGRN exhibit pronounced small-world characteristics, including high local clustering and efficient global connectivity, consistent with common properties of biological networks[42]. Because topology alone cannot validate individual regulatory interactions, we further evaluated the networks using ChIP-seq-supported TF–target relationships, external datasets, and random-network null models. In GSE117221, MFGRN achieves AUROC values of 0.5492 and 0.5481 in the control and disease groups, respectively, indicating modest absolute performance but

remaining above the random baseline. These relatively low values may reflect differences between the external thalassemia and primary SCD datasets, as well as biological heterogeneity, limited sample size, regulatory complexity, and incomplete reference networks. Similar performance limitations have been reported in benchmark evaluations of biological network inference; for example, DREAM5 methods achieve AUROC values of approximately 0.518–0.671 for in vivo networks, although direct comparisons should be interpreted cautiously because datasets and evaluation settings differ[43,44]. Importantly, MFGRN improves upon the best individual method by 0.0305 and 0.0440 in the two GSE117221 groups, respectively. In GSE244401, overlaps between MFGRN-predicted interactions and ChIP-seq-supported targets exceed those from matched random networks, indicating that the agreement is not explained solely by network size or chance. Collectively, the external evaluation, consistent improvement over individual algorithms, and independent ChIP-seq and null-model evidence support the biological plausibility and comparative predictive advantage of MFGRN while highlighting the challenges of GRN evaluation in heterogeneous real-world datasets.

Our findings reveal substantial abnormalities in network architecture and transcriptional regulation in patients with SCD at rest compared with healthy controls. Topological analyses show pronounced differences in clustering coefficient, average shortest path length, and network density, indicating extensive network remodeling under disease conditions. Biological networks typically exhibit strong local clustering and efficient information transfer, which are important for maintaining homeostasis and responding to perturbations[42]. Although SCD is caused by a single-gene mutation, widespread changes occur in network topology, key regulatory nodes, and functional modules. Together with the chronic inflammation, hypoxia, oxidative stress, vascular dysfunction, and immune dysregulation characteristic of SCD[45–48], these changes suggest that systemic regulatory network dysregulation may contribute to the broader molecular heterogeneity associated with SCD beyond the direct effects of the causative mutation. Exercise induces notable topological changes, but does not simply restore the network to a healthy state. Instead, these changes are more consistent with exercise-

associated remodeling of SCD regulatory networks within the chronic hypoxic, inflammatory, and oxidative-stress environment.

Key TF and functional module analyses provide further insight into SCD-associated network dysregulation. Across groups, key TFs are mainly involved in development and differentiation, hematopoiesis, immune regulation, and transcriptional control, consistent with persistent hematopoietic stress, immune activation, and chronic inflammation in SCD[49,50]. Notably, rhythmic process and respiratory system development are specifically enriched in SCD_T1. Circadian regulation is closely linked to inflammation, immune homeostasis, and cellular migration[51,52], while hypoxia can disrupt circadian regulation through altered clock gene expression[53–55]. Enrichment in respiratory system development is also consistent with chronic hypoxia, vaso-occlusion, and impaired tissue oxygen delivery in SCD[56–58]. These findings suggest that resting SCD is associated with network remodeling involving circadian regulation and adaptation to chronic hypoxia. At the module level, SCD_T1 modules are enriched in oxidative stress responses, cytokine regulation, leukocyte-related processes, and fatty acid metabolism, whereas SCD_T2 retains immune- and stress-related signatures while showing changes in metabolic regulation, cellular stress responses, and immune cell migration. These patterns are consistent with the coexistence of oxidative stress, chronic inflammation, immune dysregulation, and metabolic alterations in SCD[59]. Overall, exercise appears to reshape disease-associated regulatory networks through functional adjustments in key TF-driven modules rather than restoring them to a healthy state[60], and SCD-associated GRN dysregulation extends beyond hematopoiesis to circadian, hypoxia-related, inflammatory, oxidative-stress, and metabolic processes.

Despite progress in inferring SCD-related GRNs and performing systems-level analyses, several limitations remain. First, the GRNs were constructed from whole-blood bulk RNA-seq data, limiting the resolution of cell-type-specific regulatory relationships in heterogeneous blood populations. Second, although the inferred networks were evaluated using ChIP-seq data, external datasets, and null-model analyses, these validations remain largely computational and indirect, and the identified

key TFs and regulatory interactions require experimental confirmation. Third, the consensus-based edge selection strategy may exclude biologically relevant interactions detected by only one algorithm. Future integration of single-cell transcriptomics, ATAC-seq, and other multi-omics data may improve the fusion strategy and provide a more comprehensive understanding of SCD regulatory mechanisms.

5. Conclusion

MFGRN, the multi-algorithm fusion framework developed in this study, improves the predictive performance and consistency of GRN inference across the evaluated biological states and datasets. Using the reconstructed GRNs, we characterized SCD-associated regulatory changes under resting and exercise conditions. SCD involves widespread GRN dysregulation affecting network topology, key regulatory nodes, and functional modules. Exercise does not simply restore the SCD network to a healthy state but appears to induce regulatory remodeling through changes in key TFs and functional modules. Notably, key TFs in SCD_T1 are specifically enriched in rhythmic process and respiratory system development, suggesting potential involvement of circadian regulation and hypoxia-related adaptation. Overall, MFGRN provides an effective framework for GRN reconstruction and offers a network-level perspective on SCD-associated regulatory alterations while identifying candidate regulators for further experimental investigation.

Data availability

The gene expression datasets analyzed in this study are publicly available in the NCBI Gene Expression Omnibus (GEO): GSE244401, GSE254951, and GSE117221. The ChIP-seq datasets used for network evaluation are also publicly available in GEO: GSE103445, GSE36985, and GSE70660. The analysis code is publicly available in the MFGRN GitHub repository (<https://github.com/ZZH706/MFGRN>).

Funding

This work was supported by the Graduate Research Project of the Scientific Research

Fund of the Yunnan Provincial Department of Education (No. 2026Y1256 to Z.Z., No. 2026Y1260 to Q.L., and No. 2026Y1265 to J.Zhu), the Doctoral Research Support Fund of Dali University (No. KYBS2023002 to X.W.), and the Joint Special Fund for Basic Research of Local Universities in Yunnan Province (No. 202301AO070378 to X.W.).

Competing interests

The authors declare no competing interests.

Author contributions

X.W., J.Zhang, and Z.Z. conceived and designed the study. X.W. and J.Zhang acquired funding and provided technical support. Z.Z. developed the MFGRN framework, performed all data analyses, and drafted the manuscript. Z.C., Q.L., J.W., J.Zhu, Q.Z., P.W., Z.H., and S.F. contributed to the interpretation and discussion of the results. X.W. and J.Zhang reviewed and revised the manuscript. All authors read and approved the final manuscript.

References

1. Badia-I-Mompel, P. *et al.* Gene regulatory network inference in the era of single-cell multi-omics. *Nat Rev Genet* **24**, 739–754 (2023).
2. Kumar, V., Srivastava, A. K. & AbdElgawad, H. Transcriptional and Post-Transcriptional Regulation of Plant Growth, Development, and Stress Responses. *J Plant Growth Regul* **44**, 1317–1322 (2025).
3. Shah, F. & Dwivedi, M. Pathophysiology and recent therapeutic insights of sickle cell disease. *Ann Hematol* **99**, 925–935 (2020).
4. Aboderin, F. I., Oduola, T., Davison, G. M. & Oguntibeju, O. O. A Review of the Relationship between the Immune Response, Inflammation, Oxidative Stress, and

the Pathogenesis of Sickle Cell Anaemia. *Biomedicines* **11**, 2413 (2023).

5. Obeagu, E. I. Adhesion molecules in focus: mechanistic pathways and therapeutic avenues in sickle cell vaso-occlusion - a narrative review. *Ann Med Surg (Lond)* **87**, 5775–5783 (2025).
6. Trovati Maciel, T., Rignault, R., Allali, S. & Hermine, O. Hemolysis and innate immunity contribution to sickle cell disease pathophysiology. in *Current Practices in Sickle Cell Disease* (ed. Zakaria, M.) (IntechOpen, 2024). doi:10.5772/intechopen.1005751.
7. van Dam, S., Vösa, U., van der Graaf, A., Franke, L. & de Magalhães, J. P. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform* **19**, 575–592 (2018).
8. Hegde, A., Nguyen, T. & Cheng, J. Machine learning methods for gene regulatory network inference. *Briefings in Bioinformatics* **26**, bbaf470 (2025).
9. Yuan, Y. & Bar-Joseph, Z. Deep learning for inferring gene relationships from single-cell expression data. *Proc Natl Acad Sci U S A* **116**, 27151–27158 (2019).
10. Bravo González-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods* **20**, 1355–1367 (2023).
11. Yuan, Q. & Duren, Z. Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data. *Nat Biotechnol* **43**, 247–257 (2025).
12. Shu, H. *et al.* Modeling gene regulatory networks using neural network architectures. *Nat Comput Sci* **1**, 491–501 (2021).
13. Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A. & Murali, T. M. Benchmarking

algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods* **17**, 147–154 (2020).

14. Plaza-Florido, A., Liem, R. I., Haddad, F. & Radom-Aizik, S. Whole-blood transcriptome analysis reveals distinct gene expression signatures in paediatric patients with sickle cell anaemia before and after exercise. *Br J Haematol* **205**, 320–328 (2024).
15. Kulakovskiy, I. V. *et al.* HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res* **46**, D252–D259 (2018).
16. Zhang, Q. *et al.* hTFtarget: A Comprehensive Database for Regulations of Human Transcription Factors and Their Targets. *Genomics, Proteomics & Bioinformatics* **18**, 120–128 (2020).
17. Bhat, V., Potdar, A. A., Yu, G. K., Gibson, G. & Sheehan, V. A. Impact of hydroxycarbamide treatment on the whole-blood transcriptome in sickle cell disease. *Br J Haematol* **206**, 713–720 (2025).
18. Nanou, A. *et al.* Sex-specific transcriptional profiles identified in β -thalassemia patients. *Haematologica* **106**, 1207–1211 (2021).
19. Fan, Y. & Ma, X. Gene Regulatory Network Inference using 3D Convolutional Neural Network. *AAAI* **35**, 99–106 (2021).
20. Gao, Z. *et al.* DeepFGRN: inference of gene regulatory network with regulation type based on directed graph embedding. *Briefings in Bioinformatics* **25**, bbae143 (2024).

21. Martyn, G. E. *et al.* Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. *Nat Genet* **50**, 498–503 (2018).
22. Xu, J. *et al.* Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. *Dev Cell* **23**, 796–811 (2012).
23. Huang, J. *et al.* Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Dev Cell* **36**, 9–23 (2016).
24. Humphries, M. D. & Gurney, K. Network ‘Small-World-Ness’: A Quantitative Method for Determining Canonical Network Equivalence. *PLoS ONE* **3**, e0002051 (2008).
25. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **57**, 289–300 (1995).
26. Newman, M. E. J. Random graphs as models of networks. Preprint at <https://doi.org/10.48550/ARXIV.COND-MAT/0202208> (2002).
27. Saint-André, V. Computational biology approaches for mapping transcriptional regulatory networks. *Computational and Structural Biotechnology Journal* **19**, 4884–4895 (2021).
28. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019).
29. Zhang, J. *et al.* miRspongeR 2.0: an enhanced R package for exploring miRNA sponge regulation. *Bioinform Adv* **2**, vbac063 (2022).

30. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**, 101–113 (2004).
31. Cannas, G. *et al.* Sick cell disease and acute leukemia: one case report and an extensive review. *Ann Hematol* **102**, 1657–1667 (2023).
32. Han, Y. *et al.* Hemolysis-driven IFN α production impairs erythropoiesis by negatively regulating EPO signaling in sickle cell disease. *Blood* **143**, 1018–1031 (2024).
33. Yu, J. C., Lin, G., Ken, R., Field, J. J. & Linden, J. The Role Of NF- κ B In The Activation Of Human iNKT Cells In Sick Cell Disease Patients and In Vitro. *Blood* **122**, 2291–2291 (2013).
34. Enenstein, J. *et al.* Proinflammatory phenotype with imbalance of KLF2 and RelA: risk of childhood stroke with sickle cell anemia. *Am J Hematol* **85**, 18–23 (2010).
35. White, G., Nonaka, D., Chung, T.-T., Oakey, R. J. & Izatt, L. Somatic EPAS1 Variants in Pheochromocytoma and Paraganglioma in Patients With Sickle Cell Disease. *J Clin Endocrinol Metab* **108**, 3302–3310 (2023).
36. Sangerman, J. *et al.* Mechanism for fetal hemoglobin induction by histone deacetylase inhibitors involves gamma-globin activation by CREB1 and ATF-2. *Blood* **108**, 3590–3599 (2006).
37. de Vasconcellos, J. F. *et al.* HMGA2 Moderately Increases Fetal Hemoglobin Expression in Human Adult Erythroblasts. *PLoS One* **11**, e0166928 (2016).
38. Alcalá-Corona, S. A., Sandoval-Motta, S., Espinal-Enríquez, J. & Hernández-Lemus, E. Modularity in Biological Networks. *Front. Genet.* **12**, 701331 (2021).

39. Kuncheva, L. I. & Whitaker, C. J. Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. *Machine Learning* **51**, 181–207 (2003).
40. Dietterich, T. G. Ensemble Methods in Machine Learning. in *Multiple Classifier Systems* vol. 1857 1–15 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2000).
41. De Smet, R. & Marchal, K. Advantages and limitations of current network inference methods. *Nat Rev Microbiol* **8**, 717–729 (2010).
42. Watts, D. J. & Strogatz, S. H. Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442 (1998).
43. The DREAM5 Consortium *et al.* Wisdom of crowds for robust gene network inference. *Nat Methods* **9**, 796–804 (2012).
44. Haury, A.-C., Mordelet, F., Vera-Licona, P. & Vert, J.-P. TIGRESS: Trustful Inference of Gene REGulation using Stability Selection. *BMC Syst Biol* **6**, 145 (2012).
45. Obeagu, E. I. Hypoxia, Inflammation, and Cytokine Crosstalk in Sickle Cell Disease: From Mechanisms to Modulation- A Narrative Review. *Pediatric Health Med Ther* **16**, 217–225 (2025).
46. Li, W. *et al.* Inflammation and autoimmunity are interrelated in patients with sickle cell disease at a steady-state condition: implications for vaso-occlusive crisis, pain, and sensory sensitivity. *Front Immunol* **15**, 1288187 (2024).
47. Gollamudi, J. *et al.* A bone to pick-cellular and molecular mechanisms of bone pain in sickle cell disease. *Front Pain Res (Lausanne)* **4**, 1302014 (2023).

48. Zhang, D., Xu, C., Manwani, D. & Frenette, P. S. Neutrophils, platelets, and inflammatory pathways at the nexus of sickle cell disease pathophysiology. *Blood* **127**, 801–809 (2016).
49. Argueta, D. A. *et al.* Mast cell extracellular trap formation underlies vascular and neural injury and hyperalgesia in sickle cell disease. *Life Sci Alliance* **7**, e202402788 (2024).
50. Obadina, M., Wilson, S., Leverage, B. & Little, J. Uncovering Hypoxemia in Sickle Cell Disease. *The Hematologist* **21**, (2024).
51. Scheiermann, C., Gibbs, J., Ince, L. & Loudon, A. Clocking in to immunity. *Nat Rev Immunol* **18**, 423–437 (2018).
52. Druzd, D. *et al.* Lymphocyte Circadian Clocks Control Lymph Node Trafficking and Adaptive Immune Responses. *Immunity* **46**, 120–132 (2017).
53. Adamovich, Y., Ladeuix, B., Golik, M., Koeners, M. P. & Asher, G. Rhythmic Oxygen Levels Reset Circadian Clocks through HIF1 α . *Cell Metabolism* **25**, 93–101 (2017).
54. Peek, C. B. *et al.* Circadian Clock Interaction with HIF1 α Mediates Oxygenic Metabolism and Anaerobic Glycolysis in Skeletal Muscle. *Cell Metab* **25**, 86–92 (2017).
55. O’Connell, E. J., Martinez, C.-A., Liang, Y. G., Cistulli, P. A. & Cook, K. M. Out of breath, out of time: interactions between HIF and circadian rhythms. *American Journal of Physiology-Cell Physiology* **319**, C533–C540 (2020).
56. Rees, D. C., Williams, T. N. & Gladwin, M. T. Sickle-cell disease. *The Lancet* **376**,

2018–2031 (2010).

57. Manwani, D. & Frenette, P. S. Vaso-occlusion in sickle cell disease: pathophysiology and novel targeted therapies. *Blood* **122**, 3892–3898 (2013).
58. Caboot, J. B. & Allen, J. L. Hypoxemia in sickle cell disease: significance and management. *Paediatr Respir Rev* **15**, 17–23 (2014).
59. Irwin, D. C. *et al.* Understanding exercise (in)tolerance in sickle cell disease: impacts of hemolysis and exercise training on skeletal muscle oxygen delivery. *J Appl Physiol (1985)* **137**, 975–983 (2024).
60. Connes, P., Stauffer, E., Liem, R. I. & Nader, E. Exercise and training in sickle cell disease: Safety, potential benefits, and recommendations. *American J Hematol* **99**, 1988–2001 (2024).

Supplementary Files

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

- [SupplementaryTableS3.xlsx](#)
- [SupplementaryTableS1.xlsx](#)
- [SupplementaryTableS2.xlsx](#)
- [SupplementaryTableS4.xlsx](#)
- [SupplementaryTableS5.xlsx](#)