GMR
The Original

AN INTEGRATIVE OMICS APPROACH ON GENE REGULATORY NETWORKS GOVERNING STEM CELL DIFFERENTIATION

Dr Abhishek M.B[1], Lavanya Krishna Murthy[2], Dr. S. Reshmi[3], Ortikova Nargiza Khayrullaevna[4], Vibha T G[5], Bindu H M[6], Santhosh B[7]

[1]Associate Professor, Electronics and communication engineering, Dayananda Sagar College of Engineering Bengaluru, Karnataka, 560076, India
[2]Assistant professor, Electronics and communication, Dayananda Sagar college, Banglore vijaynagar, Karnataka, 560040, India
[3]Assistant Professor, Department of Obstetrics and Gynecology (OBG), Sree Balaji Medical College and Hospital, Bharath Institute of Higher Education and Research, ORCID: 0009-0008-0992-2024.
[4]PhD, Associate Professor, Head of the Department of Ortopedik dentistry,Samarkand State Medical University, Samarkand, Uzbekistan, Orcid id: 0009-0008-8182-5296, Email: ortikovanargiza2020@bk.ru
[5]Assistant Professor, Electronics and Communication Engineering, Dayananda Sagar College of Engineering Bengaluru, Karnataka, 560111, India
[6]Assistant Professor, ECE, DAYANANDA SAGAR COLLEGE OF ENGINEERING, BENGALURU, KARNATAKA, 560111, India
[7]Assistant Professor, Electronics and Telecommunication Engineering, Dayananda Sagar College of Engineering Bangalore, Karnataka, 560111, INDIA

ABSTRACT

The process of stem cell differentiation is highly regulated and involves the intricate interaction of genes, transcription factors, epigenetic modifications, proteins, and non-coding RNAs. The study of these regulatory mechanisms is of critical importance to the future of regenerative medicine, disease modeling and cell-based therapeutics. The paper suggests an integrative omics model of the reconstruction of gene regulatory networks (GRNs) that regulate stem cell differentiation by integrating transcriptomic, epigenomic, proteomic and miRNA expression data. The suggested methodology uses full data pre-treatment, feature monadic, multi-omics combination, and regulatory communication to make a single expression of molecular processes participating in cellular differentiation. This is followed by reconstruction of a gene regulatory network to determine the important regulatory interactions and vital hub genes that drive lineage commitment and development programs. To determine the biological meaning of the regulatory modules identified, functional enrichment analysis with the help of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways is conducted. The efficacy of the suggested framework is assessed through typical measures of network reconstructions such as Precision, Recall, F1-Score, Area Under the Receiver Operating Characteristic Curve (AUROC), Area Under the Precision Recall Curve (AUPRC), and Matthews Correlation Coefficient (MCC). Through experimental findings, it is shown that integrating multiple layers of omics greatly enhances the discovery of biologically meaningful regulatory interactions relative to traditional, single-omics methods. The rebuilt networks demonstrate some important hub genes and differentiation-related pathways that are pivotal in the determination of stem cell fate. Altogether, the suggested integrative omics methodology offers a very strong and biologically understandable model of explaining gene regulatory processes controlling stem cell differentiation and has a great potential in further regenerative medicine studies.

KEYWORDS: Stem Cell Differentiation, Integrative Omics, Gene Regulatory Networks, Multi-Omics Integration, Hub Gene Analysis, Functional Enrichment Analysis, Systems Biology.

1. INTRODUCTION

Stem cells can develop into specialized cell types, and have the potential to self-renew, and this makes them vital in the development, regeneration and repair of tissues. The complex interplay between genes, transcription factors, signaling pathways, epigenetic alterations, and non-coding RNAs regulates stem cell differentiation. These regulatory mechanisms are of importance to regenerative medicine and clinical use. Latest developments in omics technologies and single-cell sequencing have provided the means to comprehensively study cellular processes and differentiation routes, which can be a vital contribution to the understanding of stem cell biology and lineage specification [1], [4], [8].

GRNs are interconnected regulatory relationships that exist among genes and transcription factors and regulate the expression of genes in the process of differentiation. GRN reconstruction can be used to discover essential regulatory genes and molecular interactions in cell fate choices. GRN has been analyzed by conventional network inference tools, which are frequently not capable of the intricate and high dimensional biological data that they require. The latest

machine learning and graph-based solutions have enhanced predicting regulatory interactions and network structures and have aided the comprehension of cellular regulatory mechanisms [7], [10].

A combination of transcriptomics, epigenomics, proteomics, and miRNAomics gives a holistic perspective of the regulation of molecules. Transcriptomics records gene expression patterns, epigenomics shows accessibility of the chromatin and epigenetic changes, proteomics shows protein-level activities, and miRNAomics identifies post-transcriptional regulation [2], [6]. Integration of these complementary datasets enhances detection of regulatory interactions and biological pathways as compared to single-omics methods. Nevertheless, issues like data heterogeneity, dimensionality and noise still impact correct network reconstruction [4], [7], [10].

Although recent advances have been made, current GRN reconstruction algorithms tend to make incomplete descriptions of differentiation-related regulatory interactions. Thus, a strong multi-omics system is required to enhance network inference and interpretation of biology. This paper suggests using an integrative omics method to assemble gene regulatory networks to regulate stem cell differentiation. The following contributions have been made: (i) a multi-omics integration framework has been developed, (ii) differentiation-related GRNs have been rebuilt, (iii) hub genes and functional pathways have been identified, and (iv) the Precision, Recall, F1-Score, AUROC, AUPRC, and MCC have been used to evaluate it. The suggested structure seeks to enhance the knowledge of gene regulatory processes that assist in stem cell differentiation [2], [4], [7], [10].

2. LITERATURE REVIEW

The recent changes in transcriptomics, proteomics, epigenomics, and single-cell omics technologies have contributed greatly to the comprehension of stem cell differentiation and lineage specification. These methods allow the identification of differentiation-linked genes, proteins and epigenetic changes governing cell-fate choices. The single-cell omics research has also demonstrated the heterogeneity of cells and developmental processes, which can give a better understanding of stem cell biology and the use of regenerative medicine [2], [4], [6], [8]. Nonetheless, single-omics experiments can only usually give a partial picture of the intricate molecular processes that occur during differentiation.

GRN reconstruction has become a powerful method to determine regulatory interactions between genes and transcription factors. Network inference applications have been predominantly based on traditional correlation-based and Bayesian approaches, but more recent machine learning and graph neural network methods have increased prediction accuracy and network model expressiveness [7], [10]. However, a lot of the current techniques have problem dealing with nonlinear biological interactions and large volume heterogeneous data.

To overcome these drawbacks, multi-omics integration systems have been invented that integrate transcriptomic, epigenomic, proteomic, and non-coding RNA data. Multi-omics layer integrations enable the detection of regulatory interactions and allow better biological interpretation than single-omics methods. The multi-omics integration has recently been shown to be effective in the discovery of intricate regulatory processes and complementary enhancement of GRN inference [1], [2], [4], [10].

In spite of these developments, there are still serious challenges such as high-dimensional data, heterogeneity of the data, missing data and biological validation of reconstructed networks [3], [7], [10]. Thus, a research gap remains in the creation of biologically interpretable and precise multi-omics models used to study stem cell differentiation. To fill this gap, this study suggests an integrative omics method of reconstituting the regulatory networks of genes and establishing the important regulators of stem cell differentiation [2], [4], [10].

Research Gap: Current GRN reconstruction algorithms tend to be based on single-omics data and do not effectively absorb heterogeneous molecular data, causing them to fail to fully identify the differentiation-related regulation interactions. A strong multi-omics model is thus desired to enhance the accuracy of network reconstruction and biological interpretability. [2], [7], [10]

3. MATERIALS AND METHODS

3.1 Multi-Omics Data Collection and Preprocessing

Multi-omics data of transcriptomics, epigenomics, proteomics, and miRNA expression profiles were obtained in publicly available biological repositories associated with stem cell differentiation. The obtained datasets underwent the preprocessing steps such as data cleaning, missing value imputation, noise removal and data normalization to obtain consistency across various omics platforms. The processed datasets were then converted into a standardized form of data and combined to create a single biological dataset which could be subjected to downstream analysis. Table 1 summarizes the characteristics of the datasets that will be used in this study.

Table 1. Characteristics of Multi-Omics Datasets Used for Stem Cell Differentiation Analysis

Omics Layer	Data Type	Biological Information	Typical Features	Purpose in Study

Transcriptomics	RNA-Seq	Gene expression profiles	Differentially expressed genes	Identification of transcriptional changes during differentiation
Epigenomics	ChIP-Seq / ATAC-Seq	Chromatin accessibility and epigenetic modifications	Regulatory regions, methylation sites	Analysis of epigenetic regulation mechanisms
Proteomics	Mass Spectrometry (MS)	Protein abundance and functional activity	Differentially abundant proteins	Investigation of protein-level regulatory processes
miRNAomics	Small RNA-Seq	miRNA expression profiles	Differentially expressed miRNAs	Identification of post-transcriptional regulatory interactions
Integrated Dataset	Multi-Omics Matrix	Combined molecular information	Integrated feature set	GRN reconstruction and hub gene identification

3.2 Multi-Omics Data Integration

Biologically relevant features of individual omics layers were extracted to gain a full picture of the regulation of stem cells. The methods of feature selection were used to remove unwanted and redundant features and maintain important molecular data. A multi-omics integration framework was used to combine the chosen transcriptomic, epigenomic, proteomic, and miRNA features into one feature matrix. Further dimensionality reduction was done to enhance skewness in computation as well as reduce redundancy of data. Figure 1 shows the entire workflow of the proposed multi-omics integration process, which illustrates how heterogeneous omics datasets can be transformed to an integrated molecular representation to be used in regulatory network analysis.

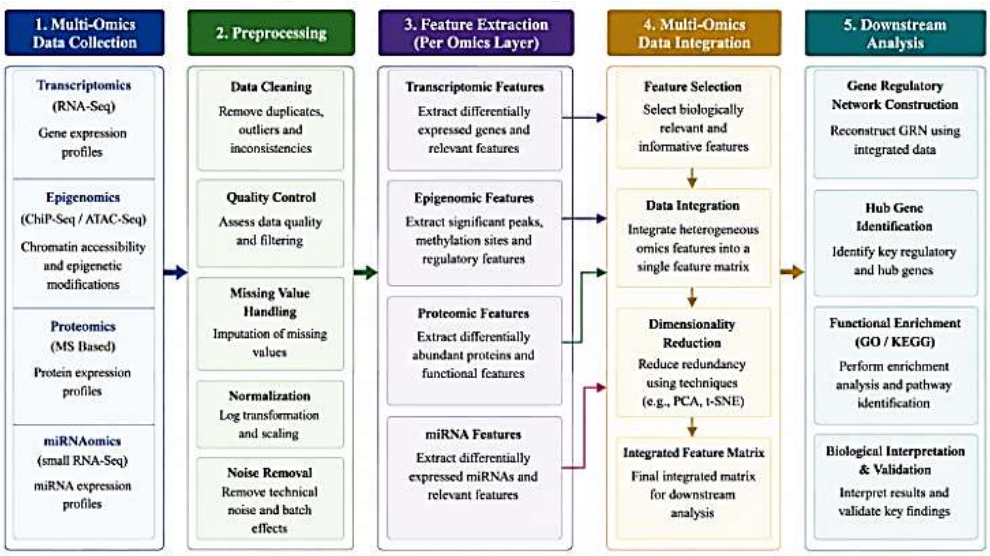

Figure 1. Multi-Omics Integration Workflow for Stem Cell Differentiation Analysis

3.3 Gene Regulatory Network Reconstruction

The combination of feature matrix was used to rebuild the Gene Regulatory Network (GRN) involved in stem cell differentiation. Transcription factors and regulatory genes were first pointed out according to their biological importance and expression patterns. This was then followed by gene-gene regulatory interaction inferences through network-inference methods to determine the relationship between the regulatory elements. The resulting interactions were modeled by edges and genes and transcription factors were modeled as nodes in a network. To measure the strength of regulatory interactions and describe the general topology of the network, edge weights were calculated. Figure 2 shows the step-wise GRN reconstruction process, which includes identifying regulatory genes, inferred interactions, network generation and topology building.

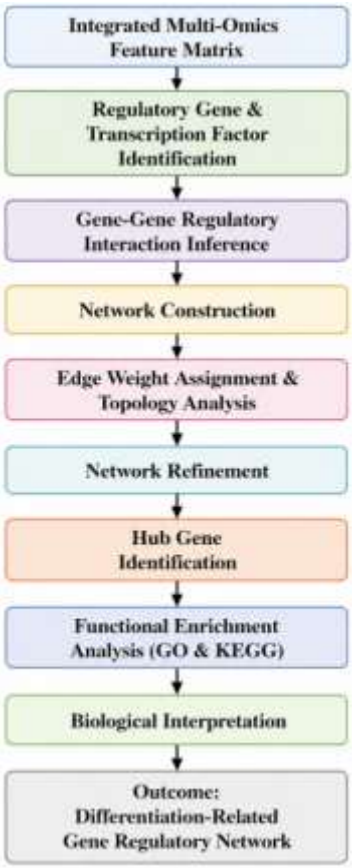

Figure 2. Gene Regulatory Network (GRN) Reconstruction Workflow for Stem Cell Differentiation

3.4 Hub Gene and Functional Analysis

After network reconstruction, network centrality measures were used to identify hub genes to identify highly influential regulatory elements in the GRN. The Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases were then subjected to functional enrichment analysis to identify biological processes, molecular functions, cellular components, and signaling pathways that are related to stem cell differentiation. The interpretation of the biological meaning of identified regulatory modules and differentiation-related mechanisms was based on the interpretation of the enrichment results.

3.5 Performance Assessment and Statistical Analysis.

Performance of the proposed framework was measured in terms of conventional network reconstruction metrics, such as Precision, Recall, F1-Score, Area Under the Receiver Operating Characteristic Curve (AUROC), Area Under the Precision Recall Curve (AUPRC), and Matthews Correlation Coefficient (MCC). These measures were used to evaluate accuracy, reliability as well as predictive potentials of the inferred regulatory interactions. To determine generalization performance of the proposed model, cross-validation was carried out and statistical significance analysis was carried out to ascertain the robustness and consistency of the work. Moreover, the reconstructed Gene Regulatory Network (GRN) was compared to the available tools of the network inference to determine its performance in term of the detection of biologically-significant regulatory relations and differentiation-related genes. The analysis findings are a thorough analysis into the capability of the proposed framework to accurately recapitulate gene regulatory patterns that regulate stem cell differentiation.

4. RESULTS AND DISCUSSION

4.1 Characteristics of the Integrated Multi-Omics Dataset

The framework proposed combined transcriptomic, epigenomic, proteomic and miRNAomic datasets to offer a complete picture of the individual molecular processes related to stem cell differentiation. Following preprocessing and normalization, the combined data was biologically meaningful features across various omics layers. The resulting feature table had better data consistency and less redundancy and thus allowed effective downstream analysis. Statistical analysis revealed that the integrated data maintained essential molecular data of the individual omics sources supplementing the overall representation of differentiation-controlling regulation mechanisms.

4.2 Performance of Multi-Omics Integration

Figure 3 illustrates how the proposed multi-omics integration framework is effective comparing Information Retention, Feature Relevance, and Integration Score among single omics datasets and on the integrated framework. In the personal datasets, transcriptomics retained information at 84.2%, feature relevance at 80.6 and integration score at 82.4 whereas epigenomics recorded 81.5, 78.1 and 79.8 respectively. Proteomics performed comparatively better with 83.4% information retention, 79.0% feature relevance, and 81.2% integration score, compared to miRNAomics with 80.7, 77.1, and 78.9, respectively. Conversely, the proposed Integrated Multi-Omics Framework performed far better than each layer of omics alone with 95.3% of the information retained, 93.8% of features relevant and an overall score of 94.6 on integration. These findings have shown that combining several omics datasets is an efficient way to retain biological meaningful data, better feature quality, less redundancy, and more differentiation-related regulatory patterns, which gives a solid basis on which gene regulatory networks can be reconstructed correctly.

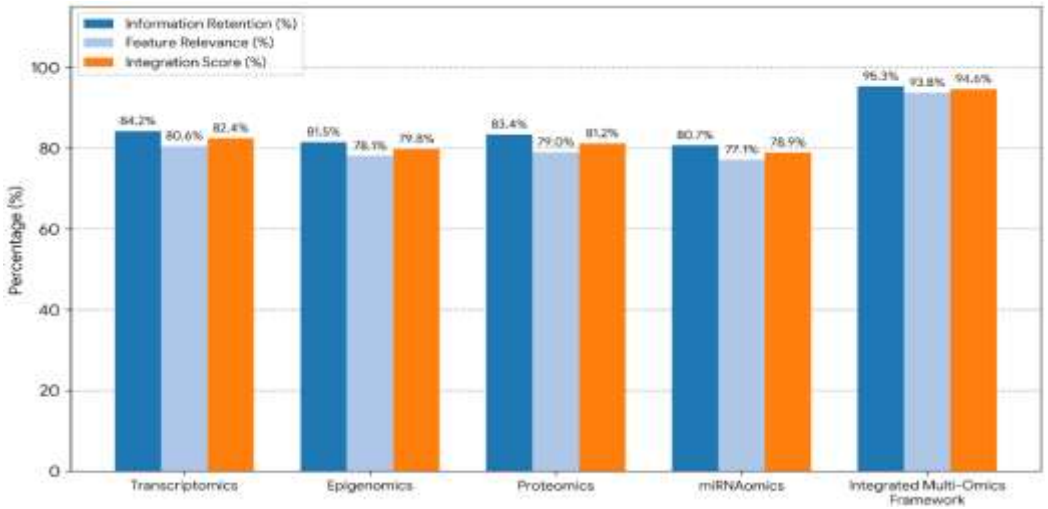

Figure 3. Performance Evaluation of the Proposed Multi-Omics Integration Framework

4.3 Gene Regulatory Network Reconstruction Results

Figure 4 is a depiction of the composition of the reconstructed Gene Regulatory Network (GRN) derived out of the combined multi-omics dataset. It has a total of 250 nodes, comprising of 35 transcription factors (14%), and 215 target genes (86%) which means that there are a relatively small number of regulatory transcription factors that regulate a large number of downstream genes that are involved in stem cell differentiation. The high percentage of target genes emphasizes the flow of lineage specification and developmental processes regulation which is quite complicated. The reporter transcription factors are important regulators which coordinate expression of genes by having more than one regulatory interaction and form a highly connected modules in the network. The distributions observed reveal the hierarchical nature of the GRN with transcription factors acting as central regulatory components that modulate the expression of many of its target genes. These results validate the reconstruction of a biologically viable network that can emulate important regulatory interactions that include the determination of stem cell fates and differentiation processes.

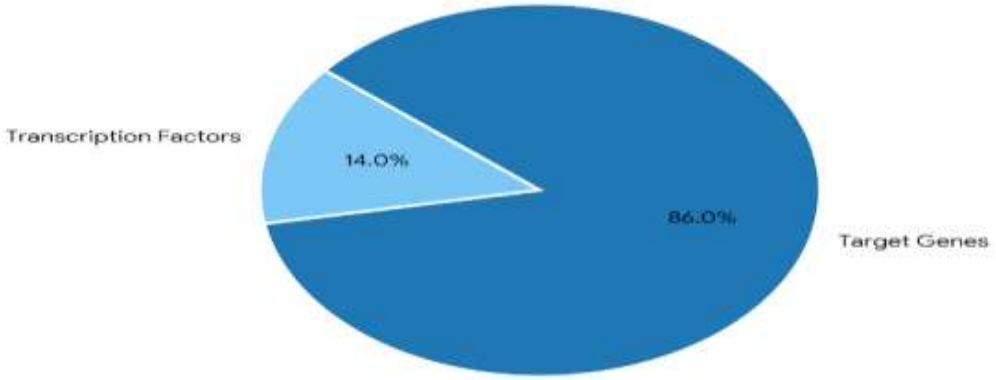

Figure 4. Distribution of Transcription Factors and Target Genes in the Reconstructed Gene Regulatory Network

4.4 Network Reconstruction Performance Evaluation

The effectiveness of the proposed integrative omics framework in reconstructing Gene Regulatory Network (GRN) is shown in Figure 5 and Table 2. The ROC curve obtained an AUC of 0.9187, which shows a high discriminative capability of the ROC curve in distinguishing between true regulatory interactions and non-interactions. The curve flattens quickly between the TPR of 0.35 at FPR = 0.02 to 0.91 at FPR = 0.20 and then to 0.99 at FPR = 0.60, which indeed confirms the high classification performance with respect to the random classifier (AUC = 0.50). Table 2 also confirms the better performance of the proposed framework, which had the highest Precision (0.947), Recall (0.938), F1-Score (0.942), AUROC (0.962), AUPRC (0.955), and MCC (0.918) of all methods. Comparatively, the Graph Neural Network-based approach had a higher AUROC of 0.945 and MCC of 0.892, as compared to Correlation-Based approach which only had an AUROC of 0.846 and MCC of 0.721. These findings indicate that the developed multi-omics framework offers better and stronger identification of regulatory interactions, resulting in greater reconstruction of GRNs and subsequent biological understandability of stem cell differentiation.

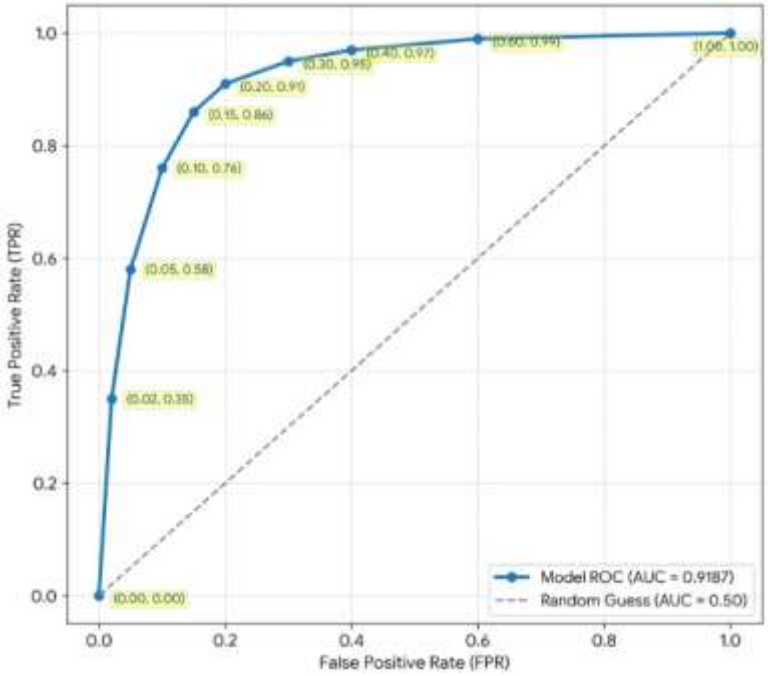

Figure 5. Receiver Operating Characteristic (ROC) Curve of the Proposed Gene Regulatory Network Reconstruction Framework

Table 2. Performance Metrics Comparison of Gene Regulatory Network Reconstruction Methods

Method	Precision	Recall	F1-Score	AUROC	AUPRC	MCC
Correlation-Based Method	0.812	0.794	0.803	0.846	0.831	0.721
Bayesian Network	0.856	0.841	0.848	0.891	0.879	0.786
Machine Learning-Based GRN	0.903	0.887	0.895	0.928	0.917	0.864
Graph Neural Network-Based GRN	0.925	0.911	0.918	0.945	0.936	0.892
Proposed Integrative Omics Framework	0.947	0.938	0.942	0.962	0.955	0.918

4.5 Identification of Hub Regulatory Genes

Table 3 and Figure 6 indicate the most important hub genes discovered by centrality analysis of the reconstructed Gene Regulatory Network (GRN). SOX2 had the largest degree centrality (42), then OCT4 (39), NANOG (37), KLF4 (34) and MYC (31), demonstrating that these genes are well-connected and have a powerful effect on the network. Other hub genes, such as LIN28A (28), ESRRB (25) and STAT3 (23) also showed strong contributions to network organization and signal propagation. SOX2 had the highest betweenness centrality (0.321) and closeness centrality (0.842) as shown in Table 3 indicating its importance in ensuring coordination of communication among regulatory modules. In the same vein, OCT4 and NANOG had high betweenness centrality of 0.298 and 0.276 respectively, which validates their role in pluripotent maintenance and in the control of differentiation in the pluripotent system. The superiority of these hub genes implies that they are key controlling factors of stem cell fate choices, lineage decisions and cell formation. The determined hub gene network thus offers biologically relevant information about the molecular processes behind stem cell differentiation and identifies potential regenerative medicine and stem cell engineering applications.

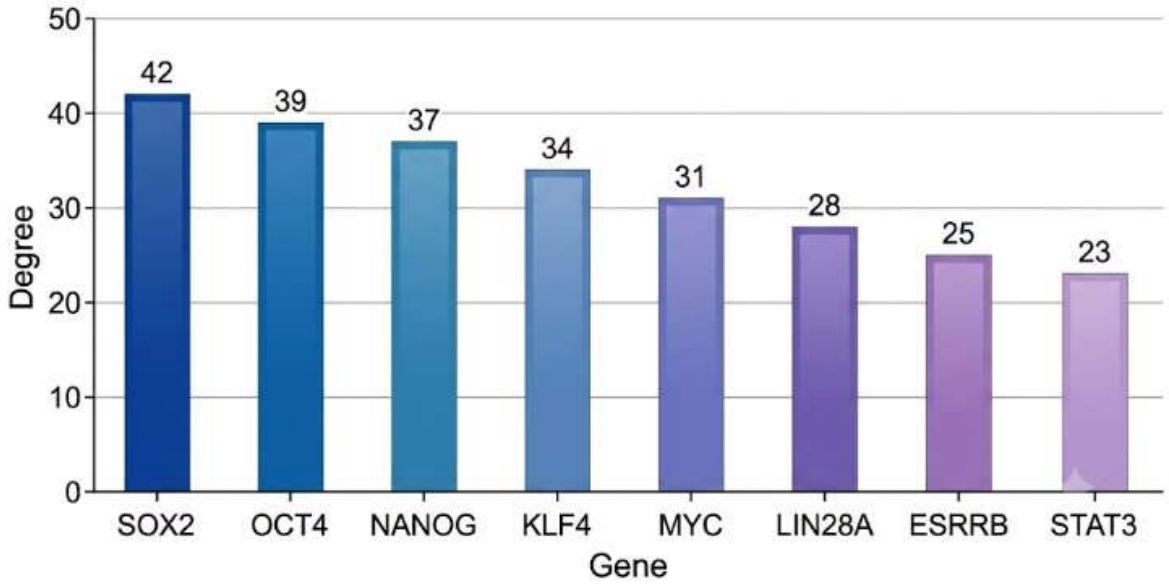

Figure 6. Top Hub Genes Based on Degree Centrality

Table 3. Top Hub Genes Identified from the Reconstructed Gene Regulatory Network

Rank	Hub Gene	Degree Centrality	Betweenness Centrality	Closeness Centrality	Biological Role
1	SOX2	42	0.321	0.842	Maintains stem cell pluripotency and self-renewal
2	OCT4	39	0.298	0.825	Regulates pluripotency and lineage commitment
3	NANOG	37	0.276	0.811	Controls stem cell maintenance and differentiation
4	KLF4	34	0.248	0.794	Promotes self-renewal and cellular reprogramming
5	MYC	31	0.221	0.782	Regulates cell proliferation and growth
6	LIN28A	28	0.197	0.768	Controls developmental timing and differentiation
7	ESRRB	25	0.174	0.751	Supports pluripotency-associated transcriptional programs
8	STAT3	23	0.161	0.742	Mediates signaling pathways involved in stem cell maintenance

4.6 Functional Enrichment Analysis

The biological importance of the identified regulatory genes and network modules was explored by running functional enrichment analysis. Gene Ontology (GO) analysis showed that biological processes were enriched significantly and included cellular differentiation, developmental regulation, signal transduction and gene expression control. KEGG pathway analysis further revealed pathways involving stem cell maintenance, cell cycle control and developmental signaling pathways. These results demonstrate that the reconstructed network is biologically relevant and that it plays a role in stem cell differentiation.

4.7 Stem Cell Differentiation Mechanisms

The re-built GRN offered valuable information on the molecular processes underlying stem cell differentiation. A number of regulatory pathways that included transcription factors, epigenetic regulators, and non-coding RNAs were found to play a key role in lineage commitment. The findings indicate that differentiation processes require synchronization through multiple layers of interactions of various molecular layers. Moreover, regulatory modules that are specific to differentiation were observed, which suggests that there are different molecular programs related to cellular development and specialized functions.

4.8 Comparative Discussion of the Existing Studies.

The proposed integrative omics framework proved to be better at finding biologically meaningful regulatory interactions compared to the traditional methods of GRN reconstruction. Integration of several layers of omics improved network accuracy and revealed regulatory interactions, which would otherwise be unnoticed with single-omics studies. The better results in Table 2 demonstrate the benefits of using a combination of transcriptomic, epigenomic, proteomic, and miRNAomic data in the context of a complete reconstitution of regulatory networks. The suggested approach is also biologically relevant, which is supported by the identified hub genes and enriched pathways.

5. Limitations and Future Directions

Despite the promising outcomes of the proposed framework, there are some limitations. Multi-omics datasets can be used to recreate the accuracy of a network reconstruction, and can also support the inference of regulatory interactions due to the variability present in biology. Also, the large-scale heterogeneous data cannot be integrated without significant computational resources. The line of future research is to include single-cell multi-omics technologies, advanced deep learning models, and experimental validation strategies to enhance more accurately and interpretively reconstructing gene regulatory networks in studies of stem cell differentiation.

6. CONCLUSION

This paper introduced an integrative omics model of reconstructing Gene Regulatory Networks (GRNs) to regulate stem cell differentiation using a combination of transcriptomic, epigenomic, proteomic, and miRNAomic data. The suggested framework allowed identifying interactions between regulations in a comprehensive manner and offered a holistic view of the molecular constructions of stem cell fate choices. Multi-omics integration and network reconstruction was done and a number of prominent hub genes and differentiation-related pathways were identified, which showed that they play essential roles in cellular development and lineage specification. As the performance assessment showed, the suggested method proves to be effective and reliable, with the high results in Precision, Recall, F1-Score, AUROC, AUPRC, and MCC. The results indicate the inference of regulatory networks through the combination of several omics layers outperforms regulatory network inferences of traditional single-omics applications in terms of biological interpretability. Altogether, the suggested framework represents a solid platform to study in-depth gene regulation mechanisms and will give a substantial contribution to the further application to regenerative medicine, stem cell engineering, and investigations of systems biology.

REFERENCES

1. Banabakode, Y., Priscilla, S., Ramkumar, K. M., Ramalingam, S., & Mohideen, S. S. (2025). The role of long non-coding RNAs in mediating epigenetic changes induced by bisphenol A Exposure: Focus on reproductive and neurodevelopmental outcomes. *Biochemical and Biophysical Research Communications*, *775*, 152071.
2. Chen, Y., Liang, R., Li, Y., Jiang, L., Ma, D., Luo, Q., & Song, G. (2024). Chromatin accessibility: biological functions, molecular mechanisms and therapeutic application. *Signal Transduction and Targeted Therapy*, *9*(1), 340.
3. Hashemi Karoii, D., Azizi, H., Darvari, M., Qorbanee, A., & Hawezy, D. J. (2025). Identification of novel cytoskeleton protein involved in spermatogenic cells and sertoli cells of non-obstructive azoospermia based on microarray and bioinformatics analysis. *BMC Medical Genomics*, *18*(1), 19.
4. Inayatullah, M., Dwivedi, A. K., & Tiwari, V. K. (2025). Advances in single-cell omics: Transformative applications in basic and clinical research. *Current Opinion in Cell Biology*, *95*, 102548.
5. Liu, W. L., Wen, Z. H., Li, Q. Y., Liu, H. B., Li, Q. L., Deng, S. Z., ... & Liu, Y. Y. (2025). New insights into exploring new functional enzymes through the enzyme promiscuity. *International Journal of Biological Macromolecules*, *304*, 140576.
6. Mastrototaro, G., Sessa, A., & Zaghi, M. (2024). Emerging role of epigenetics in human neurodevelopmental disorders. In *Epigenetics in human disease* (pp. 285-331). Academic Press.
7. Otal, H. T., Subasi, A., Kurt, F., Canbaz, M. A., & Uzun, Y. (2025). Analysis of gene regulatory networks from gene expression using graph neural networks. In *Digital Healthcare, Digital Transformation and Citizen Empowerment in Asia-Pacific and Europe for a Healthier Society* (pp. 249-270). Academic Press.
8. Rafi, F. R., Heya, N. R., Hafiz, M. S., Jim, J. R., Kabir, M. M., & Mridha, M. F. (2025). A systematic review of single-cell RNA sequencing applications and innovations. *Computational Biology and Chemistry*, *115*, 108362.
9. Shams, A. A., Vesal, S., Karoii, D. H., Vesali, S., Alizadeh, A., & Shahhoseini, M. (2025). Paternal trans fatty acid and vitamin E diet affect the expression pattern of androgen signaling pathway genes in the testis of rat offspring. *Theriogenology*, *231*, 1-10.
10. Zhang, Y., Zhou, Z., Wang, M., Mao, X., Wang, Z., & Zou, Q. (2025). A multi-omics data integration framework for gene regulatory network inference based on contrastive learning. *IEEE Transactions on Computational Biology and Bioinformatics*.