

RESEARCH

Open Access



Gene2role: a role-based gene embedding method for comparative analysis of signed gene regulatory networks

Xin Zeng^{1,2†}, Shu Liu^{3*†}, Bowen Liu^{1,2}, Weihang Zhang^{1,2}, Wanzhe Xu^{1,2}, Fujio Toriumi³ and Kenta Nakai^{1,2*}

[†]Xin Zeng and Shu Liu have contributed equally to this work.

*Correspondence:
liu-shu627@g.ecc.u-tokyo.ac.jp;
knakai@ims.u-tokyo.ac.jp

¹ Department of Computational Biology and Medical Sciences, The University of Tokyo, Kashiwa 277-8563, Japan

² Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan

³ Department of Systems Innovation, The University of Tokyo, Tokyo 113-8654, Japan

Abstract

Background: Understanding the dynamics of gene regulatory networks (GRNs) across various cellular states is crucial for deciphering the underlying mechanisms governing cell behavior and functionality. However, current comparative analytical methods, which often focus on simple topological information such as the degree of genes, are limited in their ability to fully capture the similarities and differences among the complex GRNs.

Results: We present Gene2role, a gene embedding approach that leverages multi-hop topological information from genes within signed GRNs. Initially, we demonstrated the effectiveness of Gene2role in capturing the intricate topological nuances of genes using GRNs inferred from four distinct data sources. Then, applying Gene2role to integrated GRNs allowed us to identify genes with significant topological changes across cell types or states, offering a fresh perspective beyond traditional differential gene expression analyses. Additionally, we quantified the stability of gene modules between two cellular states by measuring the changes in the gene embeddings within these modules.

Conclusions: Our method augments the existing toolkit for probing the dynamic regulatory landscape, thereby opening new avenues for understanding gene behavior and interaction patterns across cellular transitions.

Keywords: Graph representation learning, Gene module analysis, Gene regulatory network analysis

Background

Gene expression, a meticulously precise and intricately regulated process, is pivotal in maintaining cellular identity and facilitating differentiation. Within this complex regulatory framework lies a network of gene-to-gene interactions, known as the gene regulatory network (GRN). This network comprises nodes, representing genes, and edges, denoting the regulatory relationships between these genes. The nature of these regulatory relationships, whether activation or inhibition, is indicated by the sign of the edges.



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

GRNs can be constructed using a variety of methods based on diverse data sources [1]. First, GRNs can be built by collecting validated gene regulatory information from the literature. Due to its reliance on manually curated and experimentally verified data, this approach tends to produce relatively small networks, thereby limiting our exploration of the comprehensive regulatory relationships inside cells. Single-cell RNA sequencing (scRNA-seq) advancements have facilitated the reconstruction of GRNs encompassing thousands of genes by identifying gene co-expression patterns across various conditions or tissues [2]. For instance, EEISP [3] constructs GRNs from scRNA-seq data based on co-dependency and mutual exclusivity of gene expression. However, this co-expression-based method struggles to differentiate direct from indirect gene regulatory relationships, complicating the accurate depiction of cellular dynamics. More recently, progress in single-cell multi-omics technology has further enhanced our ability to construct GRNs [4]. For example, CellOracle [5] integrates scATAC-seq and scRNA-seq data, leveraging transcription factor (TF) binding motifs and co-expression information to infer GRNs, providing a more detailed and comprehensive understanding of gene regulatory mechanisms.

Downstream analysis of Gene Regulatory Networks (GRNs) is pivotal in uncovering gene functions. Prevalent analytical approaches emphasize the topology of cell type-specific GRNs to identify key transcription factors [6, 7] and gene modules [8, 9]. However, these typical approaches, which focus on analyzing a single network, often overlook the comparative analysis between GRNs across different cell states or types, thereby missing critical insights into the dynamics of regulatory mechanisms. Although methods exist for comparing GRNs between cell states [5, 10], they often focus solely on the direct topological information of genes, overlooking deeper structural connections (e.g., 1-hop and 2-hop neighbors), resulting in a shallow understanding of the complexity inherent in GRNs. To overcome this challenge, graph embedding techniques, which consider multi-hop connectivity, have been developed to project genes into an embedding space. These approaches preserve a richer representation of the original GRN information, thereby enabling a more precise quantification of gene distances. However, current graph embedding approaches for GRNs rely on proximity principles [11–13], hindering the projection of genes from separate networks into closely positioned spaces for comparative analysis. Role-based network embedding methods such as struc2vec [14] and SignedS2V [15] have introduced an advanced perspective by constructing a multi-layer weighted graph that reflects structural similarities among nodes at various depths. These methods facilitate embedding diverse networks into a unified space, allowing for nuanced comparisons of topological similarities across networks. Applying such advanced methodologies to compare GRNs among different cell states or types could significantly enhance our understanding of GRN dynamics.

In this article, we introduce Gene2role, the first method to apply role-based graph embedding approaches for signed GRNs, employing the frameworks from struc2vec [14] and SignedS2V [15]. We conducted experiments on GRNs generated from one simulated network, manually curated networks, single-cell co-expression networks and single-cell multi-omics networks. These experiments demonstrated the ability of Gene2role to capture the topological information of GRNs. Additionally, we used

Gene2role embeddings to analyze genes that exhibit structural variations across multiple GRNs and assessed the stability of gene modules across different cell states.

Materials and methods

The overall conceptual framework, shown in Fig. 1, consists of three major parts: network construction, embedding generation, and downstream analysis. Firstly, we introduce the network construction process from four data sources in “[Network preparation](#)” section. Then, we introduce the embedding generation framework in “[Gene topological representation in signed gene regulatory network \(GRN\)–Gene topological similarity calculation](#)” sections, which applies SignedS2V to represent the topological features of genes in the signed GRN and to calculate the similarity between genes. Next, we explain the details of the gene embedding procedure in “[Multilayer graph construction–Embedding learning](#)” sections that adopts the struc2vec framework. The hyperparameters for experiments in our study are shown in “[Hyperparameter for GRNs embedding experiments](#).” section. We showcase the downstream analysis based on gene embeddings at the gene and module levels, as depicted in “[Identification of differentially topological genes \(DTGs\)](#)” and “[Gene module stability analysis](#)” sections, respectively. Finally, we provide information about

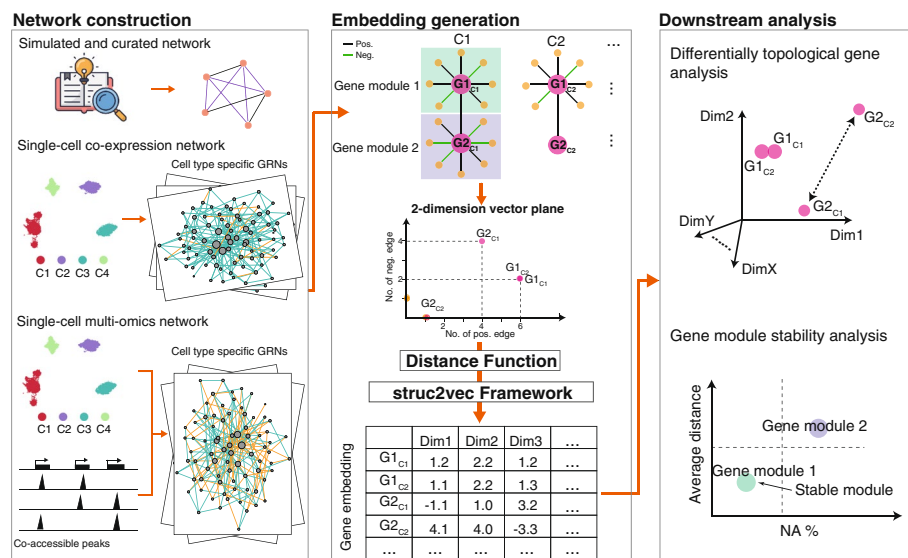


Fig. 1 Overview of the Gene2role framework. Gene2role is a multi-scale analysis framework for GRNs using a role-based graph embedding approach, which includes three components: network construction (left), embedding generation (middle), and downstream analysis (right). In network construction, one or more GRNs inferred by various methods are built, such as single-cell co-expression networks and single-cell multi-omics networks. In embedding generation, first, each gene is mapped to a 2-dimensional vector degree representing the number of positive and negative links. Next, the similarity between genes is evaluated using a series of distance functions that integrate multi-hop local topology information around the genes. Finally, the embedding is learned using the struc2vec framework. In downstream analysis, gene-level and gene-module-level analyses are performed. In gene-level analysis, differentially topological genes (such as G2) are extracted by comparing the distances of embeddings between GRNs. In gene-module-level analysis, the stability analysis of gene modules is performed by comparing the average distance between two GRNs and the proportion of genes that exist only in one of the GRNs (NA%) for pre-extracted gene modules. GRN, Gene regulatory network

baseline methods and evaluation metrics in “Baseline methods” and “Evaluation metrics” sections.

Network preparation

We constructed a simple simulated network to mimic the scale-free characteristics of GRNs, which comprises 31 genes. The four curated networks—hematopoietic stem cell (HSC) [16], mammalian cortical area development (mCAD) [17], ventral spinal cord (VSC) [18], and gonadal sex determination (GSD) [19]—containing between 5 and 19 genes, were downloaded from BEELINE [20].

For single-cell RNA-seq data, we utilized the count matrices and cell type annotation data from two previous studies. Specifically, the dataset for human glioblastoma, as reported by [3], was collected at two distinct stages: the glioblastoma stem-like cells (0-h) stage and the serum-induced differentiated (12-h) stage. The datasets for human bone marrow mononuclear cells (BMMC) and human peripheral blood mononuclear cells (PBMC) were collected from [21]. In the BMMC dataset, only Granulocyte–Macrophage Progenitors (GMPs) and CD14+ monocytes were kept, while in the PBMC dataset, all ten cell types were retained. For each cell type, count matrices were generated using 2000 highly variable genes. Subsequently, cell type-specific GRNs were constructed utilizing the EEISP [3] and Spearman correlation (Supplementary Note 1).

Single-cell multi-omics networks were obtained from CellOracle [5], which inferred them by integrating scRNA-seq data [22] and sci-ATAC-seq data [23] derived from differentiating mouse myeloid progenitors. In brief, this dataset encompasses the differentiation of myeloid progenitors across 24 cell states, primarily highlighting the process of megakaryocyte and erythroid progenitors (MEPs) differentiating into erythrocytes, as well as GMPs differentiating into granulocytes. Within these networks, only connections exhibiting a p -value less than 0.01 were considered. Moreover, a selection criterion was applied to maintain only the top 2000 edges, chosen based on their highest absolute coefficient values. Consequently, these networks were composed of between 521 and 642 genes. The sign of edges was established based on the positive or negative values of their coefficient.

The detailed information on the networks utilized in the experiments described in this paper can be found in (Supplementary Table 1).

Gene topological representation in signed gene regulatory network (GRN)

Given a signed GRN represented as $G = (V, E^+, E^-)$, where $V = \{v_1, v_2, \dots, v_n\}$ denotes the set of genes. The sets $E^+ = \{e_i | e_i \in V \times V\}$ and $E^- = \{e_i | e_i \in V \times V\}$ represent the positive and negative interactions between genes, respectively. To capture the topological nuances of each gene within the GRN, we introduce the concept of the signed-degree \mathbf{d} , which is a 2-dimensional vector defined as:

$$\mathbf{d} = [d^+, d^-] \quad (1)$$

where d^+ and d^- are the positive and negative degrees, respectively. By adopting the signed-degree \mathbf{d} , we map each gene from the signed GRNs to a point on the plane.

Gene topological similarity calculation

To quantify the topological similarity between genes within a GRN, we introduced the distance function named Exponential Biased Euclidean Distance (EBED), which evaluates the zero-hop distance (D_0) between the signed-degrees of two genes, \mathbf{d}_u^C and \mathbf{d}_v^C , as follows:

$$D_{0(u,v)} = \text{EBED}(\mathbf{d}_u, \mathbf{d}_v) \quad (2)$$

$$\text{EBED}(\mathbf{d}_u, \mathbf{d}_v) = \exp\left(\sqrt{\left(\log \frac{d_u^+ + 1}{d_v^+ + 1}\right)^2 + \left(\log \frac{d_u^- + 1}{d_v^- + 1}\right)^2}\right) \quad (3)$$

The rationale behind using EBED stems from the observation that GRNs are scale-free networks, and the degrees of their genes often follow a power-law distribution [24]. The EBED function initially employs a logarithmic transformation of the degrees to mitigate the effects of this distribution, then computes the Euclidean distance. An exponential function is subsequently applied to counterbalance the log transformation, thereby preserving the original proportionality of distances.

The topological identity of a gene is influenced not only by its direct connections but also by the broader topology that includes multi-hop neighborhoods. We define $R_k(u)$ as the sorted sequence of degrees for genes that are k hops away from gene u . Since the length of sorted sequence can be different for two genes, we employ dynamic time warping (DTW) [25] to calculate the k -hop distance ($D_k, k > 0$) between gene u and v using EBED as distance function:

$$D_k(u, v) = \text{DTW}(\text{EBED}, R_k(u), R_k(v)) \quad (4)$$

Then, the k -hop topological similarity between gene u and v can be calculated using the recursive formulation:

$$f_0(u, v) = D_0(u, v); f_k(u, v) = f_{k-1}(u, v) + D_k(u, v) \quad (5)$$

In this context, a lower value of $f_k(u, v)$ indicates a higher degree of similarity in the topology between gene u and v .

Multilayer graph construction

Next, we construct a multilayer weighted graph that encodes the topological information between genes. Each layer contains $|V|$ genes, where the weight $w_k(u, v)$ for a link in the k -th layer ($k > 0$) between gene u and v is computed as:

$$w_{k(u,v)} = e^{-f_k(u,v)}, \quad (6)$$

where a smaller value of $f_k(u, v)$ indicates a higher topological similarity between gene u and v , thereby resulting in a greater weight for their link. For inter-layer connections, only the same genes are connected through directed links, with the weights of these links defined by:

$$w(u_k, u_{k+1}) = \log(\Gamma k(u) + e), k = 0, \dots, k^{\max-1}, \quad (7)$$

$$w(u_k, u_{k-1}) = 1, k = 1, \dots, k^{\max}, \quad (8)$$

where $w(u_k, u_{k+1})$ corresponds to the weight between u_k and u_{k+1} , and $w(u_k, u_{k-1})$ corresponds to the weight between u_k and u_{k-1} . $\Gamma k(u)$ denotes the number of links from gene u with weight exceeding the average link weight in the k -th layer. A high number of genes within the k -th layer share topological similarity with gene u leads to an increased $\Gamma k(u)$, subsequently increasing the probability for searching topologically similar genes in the deeper layer. By setting $w(u_k, u_{k+1})$ to be greater than $w(u_k, u_{k-1})$, we encourage the exploration path in random walks to delve deeper rather than merely skimming the surface, aiming to uncover a wide range of topological relations between genes.

Sequence generation by random walk

Context sequences are generated using random walk on the weighted multilayer graph. The transition probability from gene u to v on layer k , denoted as $p_k(u, v)$, is determined by dividing $w_k(u, v)$ by the sum of the weights of all links connected to u within the layer:

$$p_k(u, v) = q \frac{w_k(u, v)}{\sum_{t \in V, t \neq u} w_k(u, t)} \quad (9)$$

Here, q represents the intra-layer transition probability, set at 0.8. The inter-layer transition probability is set to $1 - q$, with the probabilities for upward and downward transitions being defined respectively as:

$$p_k(u_k, u_{k+1}) = (1 - q) \frac{w(u_k, u_{k+1})}{w(u_k, u_{k+1}) + w(u_k, u_{k-1})} \quad (10)$$

$$p_k(u_k, u_{k-1}) = (1 - q) \frac{w(u_k, u_{k-1})}{w(u_k, u_{k+1}) + w(u_k, u_{k-1})} \quad (11)$$

The transition probabilities ensure that every possible path originating from gene u is considered, with their probabilities summing to 1, reflecting the comprehensive exploration of connections both within and across layers.

Embedding learning

The Skip-Gram model [26] is finally used to create gene embeddings from the generated context sequences. The fundamental principle of this model is that the meaning of a word is determined by the words located closer to it in sentences. By leveraging this model, we can learn embeddings that encapsulate rich topological information, as context sequences consist of genes that are topologically similar. Importantly, this results in genes with similar topological structures being projected into closely situated points in the embedding space.

Hyperparameter for GRNs embedding experiments.

For each experiment, we adjusted the parameters based on its complexity and the number of integrated networks (Supplementary Note 2).

Identification of differentially topological genes (DTGs)

Gene2role emphasizes the topological information of genes, allowing for genes that are distant or even unconnected from each other in a GRN to be positioned closely in the embedding space. To analyze the embedding of the same gene u from C cell types, we first determine the center of embeddings denoted as cen_u . For each cell type, we calculate the Euclidean distance d_u^c from the embedding emb_u^c of gene u to the cen_u :

$$\text{cen}_u = \frac{1}{C} \sum_{c=1}^C \text{emb}_u^c \quad (12)$$

$$d_u^c = \|\text{emb}_u^c - \text{cen}_u\|_2 \quad (13)$$

We then determine the average distance \overline{d}_u , to provide a summary measure of the distance of gene u embeddings from the centroid across all cell types.

$$\overline{d}_u = \frac{1}{C} \sum_{c=1}^C d_u^c \quad (14)$$

For $C=2$, genes with the largest distances ranking in the top 10% are defined as DTGs. For $C>2$, we further calculate the standard deviation σ_u to quantify the topological variability of gene u :

$$\sigma_u = \sqrt{\frac{1}{C} \sum_{c=1}^C (d_u^c - \overline{d}_u)^2} \quad (15)$$

Genes located in the top 5th percentile of average distances are classified as DTGs. Additionally, genes whose standard deviation exceeds a certain empirical threshold are also identified as DTGs. The high average distances indicate that these genes consistently exhibit significant differences across all cell states, suggesting a universal role change. Conversely, a high standard deviation points to substantial variability in specific cell states, implying that these genes may only undergo significant role changes under certain conditions, while remaining relatively stable in others.

When comparing DTGs with differentially expressed genes (DEGs), we identified DEGs using Seurat [27] pipeline (Supplementary Note 3).

Gene module stability analysis

To quantify the stability of gene modules between two cell types or states, we initially define an anchor cell type (c_1). Then, we use the Louvain algorithm, a proximity-based clustering method, to organize the GRN of the anchor cell type into gene modules; Within these modules, genes typically collaborate in certain specific functions [9]. Gene modules containing fewer than 10 genes will be excluded. For each remaining gene module m , we calculate the average distance ($\overline{d}_{c_1^m}$) of genes between the two cell types using the following equation:

$$\overline{d}_{c_1^m} = \frac{1}{|G_{c_1^m} \cap G_{c_2}|} \sum_{u \in G_{c_1^m} \cap G_{c_2}} \|\text{emb}_u^{c_1} - \text{emb}_u^{c_2}\|_2 \quad (16)$$

$G_{c_1^m}$ is the gene set for module m in cell type c_1 , and G_{c_2} is the whole gene set in cell type c_2 . $|\cdot|$ represents the size of set.

Moreover, the absence of gene embeddings in cell types other than the anchor may indicate significant changes in their roles. To quantify these changes, we define $\text{NA}\%_{c_1^m}$ as the percentage of genes within a gene module m that only occur in the anchor cell type c_1 .

$$\text{NA}\%_{c_1^m} = 1 - \frac{|G_{c_1^m} \cap G_{c_2}|}{|G_{c_1^m}|} \quad (17)$$

We combine the average gene distance and NA% as a measure to assess the overall change in roles of gene modules that were initially clustered together in anchor cell type when observed in another cell type.

To assess the biological functions of a gene module, we employed the *compareCluster* function from the clusterProfiler package [28]. This analysis employed the *enrichGO* function (Supplementary Note 6), which targets biological processes (BPs) using org.Hs.eg.db and org.Mm.eg.db as the gene ontology database for dataset from human and mouse, respectively. We apply the Benjamini–Hochberg method to adjust p -values, with a q -value cutoff set at 0.05 to determine significance. The top five significant Gene Ontology (GO) terms for each gene module were retained for further analysis and visualization.

Baseline methods

We compared Gene2role to two graph embedding methods: struc2vec [14] and BESIDE [29], as well as the well-established benchmark end-to-end approach: Graph Neural Networks (GNNs). In brief, struc2vec is a role-based embedding method that does not consider sign information, whereas BESIDE considers the sign information and is based on proximity embedding. We selected SDGNN [30] for signed GRNs and GCN [31] for unsigned GRNs, and extracted embeddings from a topology structure-related task.

Evaluation metrics

The topological features of a gene in a signed GRN were evaluated using the following metrics: degree centrality (+), degree centrality (−), betweenness centrality, eigenvector centrality, degree assortativity (+), degree assortativity (−), clustering coefficient (+), clustering coefficient (−) (Supplementary Note 4).

Results

Gene2role captures the topological information of GRNs

To verify that Gene2role accurately captures the topological information of genes within GRNs, we analyzed GRNs derived from four distinct data sources: 1. one simulated network, and 2. four curated networks based on experimentally validated interactions (“Simulated and curated GRN” section), 3. co-expression networks generated

from single-cell RNA sequencing data (“Single-cell co-expression network” section), 4. multi-omics networks derived from multi-omics data (“Single-cell multi-omics network” section).

Simulated and curated GRN

To provide a clear overview of our network embedding study, we selected one simulated network (Fig. 2A) and four simple curated networks (Fig. 2C, Supplementary Fig. 1A, C, E) for analysis. We compared Gene2role to two other methods, struc2vec [14] and BESIDE [29], by setting the embedding dimension to 2 for all networks (Fig. 2B, D, Supplementary Fig. 1B, D, F). Overall, Gene2role effectively positioned genes with similar connectivity patterns—marked by both positive and negative edges—within proximity in the embedding space. In contrast, the disregard of edge sign in struc2vec limited its ability to accurately capture the topological nuances of genes. BESIDE, despite considering edge signs, its proximity-based embedding strategies, which bring genes with

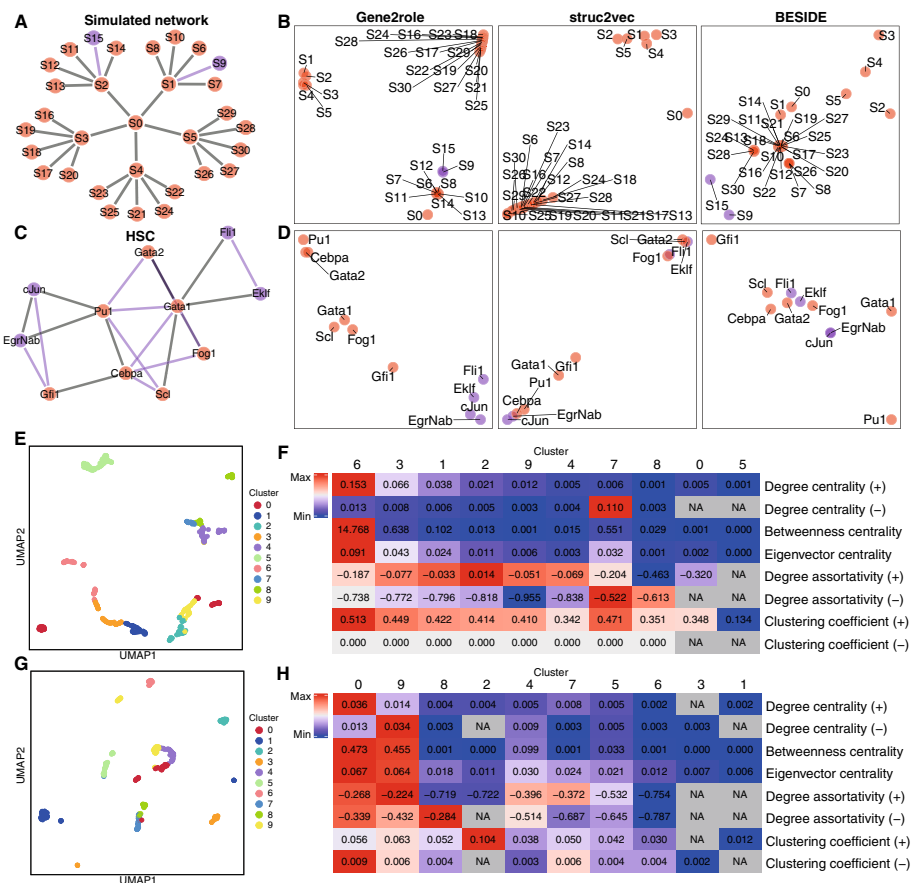


Fig. 2 Analysis of GRN embeddings from networks derived through various methods. **A, C** Simulated tree GRN (**A**) and HSC network (**C**). **B, D** 2D embeddings for networks from **A** and **C**, respectively. The embeddings were generated by Gene2role, struc2vec, and BESIDE, respectively. **E, G** K-means (K = 10) clustering of embeddings from B cells in human PBMC dataset (**E**) and Ery_0 stage in multi-omics dataset (**G**) displayed in UMAP. **F, H** Heatmap displaying the average values of 8 network feature metrics for the 10 clusters of genes from **E** and **G** within the GRN. The color scaling within each row is determined by the maximum and minimum values of that row. UMAP, Uniform manifold approximation and projection

positive connections closer together while pushing those with negative connections farther apart, failed to capture the topological information of genes. For example, in the simulated network, genes S9 and S15 were closely positioned by Gene2role due to their single negative edge. Gene S0 was projected near genes S6-S8 and S10-S14, reflecting their equal distance to negative edges. Conversely, although struc2vec closely aligned genes with similar positive topological information, it overlooked the S9 and S15 due to its exclusive focus on positive connections. BESIDE placed S9 far from S1 and S15 far from S2 due to their respective negative connections, clearly demonstrating how the method handles negative links between genes. Additionally, in the HSC network, despite their spatial separation, Gene2role clustered Fli1, Eklf, cJun, and EgrNab together due to their shared configuration of one negative and one or two positive edges. Struc2vec, however, placed cJun and EgrNab, each with a single positive edge, distantly from Fli1 and Eklf, which both had two positive edges, indicating a disregard for edge sign complexity. In contrast, although BESIDE grouped Fli1 and Eklf together, it also mixed in other genes, demonstrating its inability to effectively capture topological information. Furthermore, in the GSD network, Gene2role positioned DHH and PGD2 adjacent to each other because of their only positive linkage to a single gene and without any negative links. Struc2vec similarly grouped FGF9 with DHH and PGD2 solely based on their single positive edge, overlooking the negative edge. Conversely, BESIDE, by considering network proximity and edge signs, positioned AMH and FGF9 near PGD2 and DHH.

We also compared Gene2role to GNNs and found that both approaches exhibit similar tendencies on curated networks (Supplementary Fig. 1G). However, in the simulated network, GNNs failed to accurately capture the distinct connection patterns, as demonstrated by their inability to effectively differentiate between genes S0 and S1. This limitation might arise from the inherent nature of message passing in GNNs, where topological information can only propagate through a specific number of hops. Consequently, GNNs cannot effectively compare genes that are not directly connected.

Single-cell co-expression network

Having demonstrated the effectiveness of Gene2role in simple networks, we extended our exploration to more complex GRNs derived from single-cell RNA-seq data. We first applied Gene2role to a GRN that was inferred from B cells in the human PBMC dataset using EEISP. To explore the topological characteristics of neighboring genes within the gene embeddings, we empirically clustered the gene embeddings into ten groups using the K-means algorithm (Fig. 2E, Supplementary Note 5). We then calculated the average value of eight topological metrics for the genes within each cluster, reflecting their interconnectedness within the original GRN. The heatmap revealed unique metric enrichment across the clusters (Fig. 2F), highlighting the distinct topological attributes inherent to each cluster. Notably, genes in cluster 6 exhibited significant prominence in positive degree centrality, indicating that they were central genes with a high density of positive interactions. In contrast, cluster 7 was distinguished by its considerable negative degree centrality and minimal positive connections, suggesting a group of genes primarily linked by negative edges. Cluster 2, enriched with high positive degree assortativity but minimal negative degree centrality, predominantly comprised genes positioned on the outer edges with positive links. Furthermore, cluster 9, characterized by the lowest

negative degree assortativity and minimal negative degree centrality, was identified as a peripheral cluster primarily connected through negative relationships.

To test the performance of our method across different GRN inference approaches, we applied Gene2role to the GRN inferred from B cells in the human PBMC dataset using Spearman correlation, a widely used baseline method for GRN construction [32, 33]. The gene distribution across clusters demonstrated a consistent topological structure, mirroring the patterns discerned in the EEISP results (Supplementary Fig. 1H, I). For instance, the characteristics of cluster 7 were similar to those of cluster 6 in the EEISP-derived clusters, delineating a hub of genes with densely packed positive linkages. Similarly, cluster 5 corresponded to the previously identified cluster 7, characterized by a richness in negative interactions. Additionally, cluster 0 shared traits with cluster 2, encompassing genes predominantly linked through positive edges and situated at the periphery of the GRN. Finally, the genes within cluster 1 resembled those in cluster 9, forming a peripheral group predominantly characterized by negative edges.

Single-cell multi-omics network

We further explored more sophisticated GRNs that were inferred by integrating single-cell RNA-seq and single-cell ATAC-seq data. We used Gene2role to analyze the Ery_0 stage of the multi-omics GRNs, which were collected from CellOracle. We segmented genes into ten groups using K-means and calculated the average value of eight key metrics for each cluster (Fig. 2G, H). Genes in cluster 0 functioned as a hub predominantly associated with positive edges, whereas genes in cluster 9 functioned as their negative-edge counterpart. Cluster 8, characterized by high negative degree assortativity and low centrality, represented a group of genes on the periphery of the negative network. Conversely, cluster 2, lacking negative connections and possessing the highest positive clustering coefficient, comprised a cohesive group of genes with similar functions and low degrees of connectivity.

Identification of DTGs between paired states using Gene2role embeddings

Given the ability of Gene2role to group genes by topological patterns, we merged GRNs from two cell types to explore the gene role changes across these networks. We applied Gene2role to the human glioblastoma dataset, which comprised GRNs from 0-h and 12-h stages. We computed the pairwise distances for each gene and extracted the DTGs by the top 10% largest distances (Fig. 3A). 66 DTGs were identified and 50 of them were overlapped with DEGs (Fig. 3B). For instance, there was a significant drop in CD164 expression and its network connections at 12-h stage (Supplementary Fig. 2A, B), which aligns with its known role in glioblastoma proliferation [34]. Additionally, we observed that although 16 genes exhibited minor expression differences between the two cell types, their connection patterns within GRNs underwent significant changes. Specifically, the expression levels of DKK3 remained consistent between both cell types (Fig. 3C); however, its network connectivity diminished at the 12-h stage (Fig. 3D). Moreover, we identified that 397 genes that underwent significant changes in expression levels, yet their topological structures within the GRNs remained unchanged. For instance, while the expression level of EGR1 decreased, its network connectivity was maintained (Supplementary Fig. 2C, D).

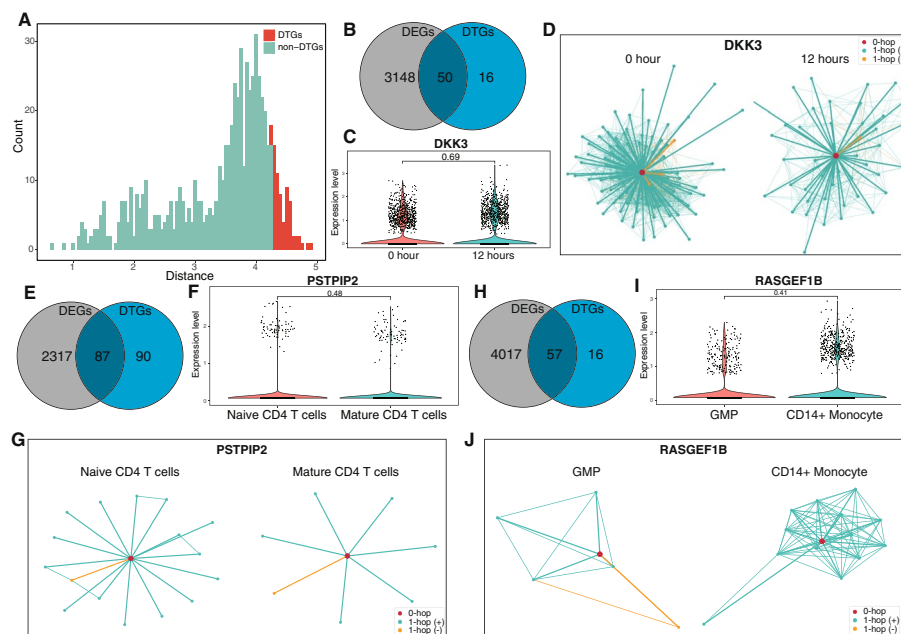


Fig. 3 Distinctive analysis of DTGs in comparison with DEGs across two cell types. **A** histogram of the frequency distribution of Gene2role embedding pair distances in GRNs at 0-h and 12-h stage within a human glioblastoma dataset, with the top 10th percentile of average distances designated as DTGs. **B**, **E**, and **H** Venn diagrams depict the intersection between DEGs and DTGs identified in human glioblastoma at 0-h and 12-h stage (**B**), in human PBMCs comparing naïve CD4 T cells with mature CD4 T cells (**E**), and in human BMMC comparing GMPs with CD14+ Monocytes (**H**), respectively. **C**, **F**, and **I** Examples of genes that are exclusively DTGs not overlapping with DEGs from the intersecting sets in **B**, **E**, and **H**, respectively. **D**, **G**, **J** 1-hop network structures corresponding to the DTG shown in **C**, **F**, and **I**. DTGs, differentially topological genes; DEGs, differentially expressed genes; PBMC, peripheral blood mononuclear cells; BMMC, human bone marrow mononuclear cells; GMPs, granulocyte-macrophage progenitors

We extended our application of Gene2role to analyze the topological shifts of genes between naïve CD4 T cells and mature CD4 T cells within the human PBMC dataset. We discovered that 87 DTGs overlapped with DEGs, indicating a robust pattern across cell types (Fig. 3E). For example, CARD16 showcased an increase in both expression level and connectivity when comparing naïve to mature CD4 T cells (Supplementary Fig. 2E, F). Furthermore, despite 90 genes showing only minor expression differences between these two cell types, significant alterations in their network connectivity were evident. A case in point is PSTPIP2, which maintained consistent expression levels across both cell types (Fig. 3F), yet its network connectivity notably decreased in mature CD4 T cells (Fig. 3G). In addition, an analysis revealed that 716 genes experienced significant shifts in expression levels without corresponding changes in their topological structures within the GRNs. SP140, for instance, exhibited an increased expression level, while its network position remained stable (Supplementary Fig. 2G, H).

In the analysis of GMPs and CD14+ Monocytes from the human BMMC dataset, we identified 57 DTGs that overlapped with DEGs (Fig. 3H), including genes such as ETS1 that displayed significant alterations in both expression levels and topological structures (Supplementary Fig. 2I, J). Conversely, 16 DTGs, which were not DEGs, such as RAGEF1B, showed significant topological changes without a corresponding

shift in expression levels (Fig. 3I, J). Moreover, within the DEGs, 524 genes, exemplified by RETN, underwent substantial changes in expression while their topological configurations were preserved (Supplementary Fig. 2K, L).

Identification of DTGs among cell types

Having analyzed gene role changes between two cell types, we extended our study to merged GRNs from multiple cell types. We applied Gene2role to the integrated GRNs from the human PBMC dataset, which comprise 10 cell types. Initially, we identified 33 DTGs by focusing on the top 5th percentile of average distances (Fig. 4A). These genes exhibited significant variability across the 10 cell types. For instance, the gene

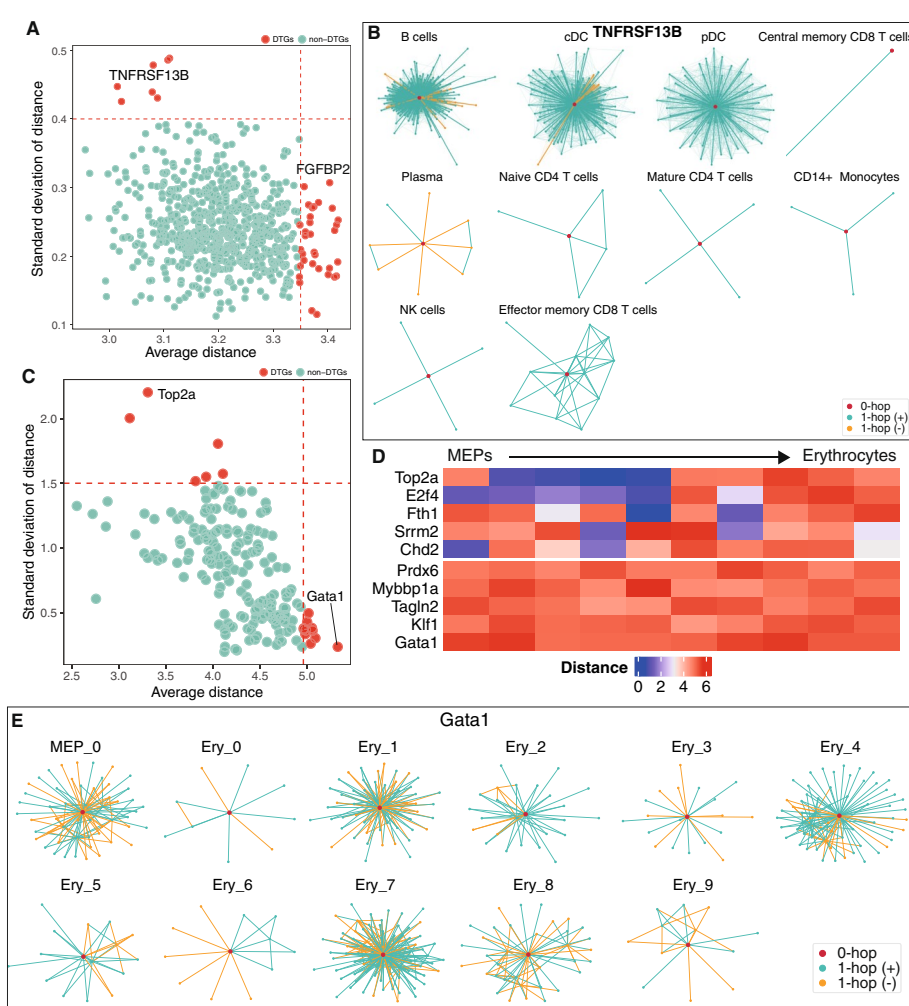


Fig. 4 Comparative analysis of gene distance across multiple cell types. **A** Scatter plot of the average and standard deviation of gene distances within human PBMC dataset across 10 cell types, with genes in the top 5th percentile of average distances or standard deviation greater than 0.4 were categorized as DTGs. **B** 1-hop network structures of TNFRSF13B in ten GRNs from human PBMC dataset. **C** Scatter plot of the average and standard deviations of gene distances between each pair of adjacent cell types in the differentiation trajectory of MEPs. **D** Heatmap of two patterns of gene role changes during MEP differentiation: the top heatmap represents genes with high variance and low average distance, while the bottom heatmap depicts genes with low variance but high average distance. **E** 1-hop network structures of TNFRSF13B in eleven GRNs from multi-omics dataset. MEPs, Megakaryocyte-erythroid progenitors

FGFBP2 had a higher count of negative edges in NK cells and central memory CD8 T cells compared to other cell types, while it showed the fewest positive edges in B cells (Supplementary Fig. 3A). Moreover, analysis revealed that seven genes exhibited a standard deviation in topological distance exceeding 0.4, yet an average distance below 3.1, indicative of substantial structural dispersion among these cell types. For example, TNFRSF13B functioned as a hub gene in B cells, conventional dendritic cells (cDCs), and plasmacytoid dendritic cells (pDCs), but maintained fewer connections in other cell types (Fig. 4B).

Subsequently, we used Gene2role embeddings to investigate significant topological changes in genes during the differentiation of MEPs and GMPs in the multi-omics dataset. For each gene, we calculated the average and standard deviation of distance between adjacent cellular state across 11 sequential stages during erythrocyte differentiation. By focusing on the top 5th percentile of genes in terms of average distance or genes with high standard deviation of distance, we identified two distinct patterns of DTGs (Fig. 4C, D). The first pattern, characterized by a high standard deviation in distance but a relatively small average distance, suggests that the roles of these genes may remain unchanged during certain developmental stages, only to shift abruptly thereafter. For instance, Top2a exhibited no significant topological changes from the Ery_0 to Ery_4 stages but underwent noticeable changes post-Ery_5 (Supplementary Fig. 3B), suggesting a role shift after this stage. The second pattern is characterized by consistently large changes in role distance throughout development, indicating dynamic and substantial fluctuations in role of genes. For example, the gene Gata1 exhibited significant periodic fluctuations in the number of its connections (Fig. 4E). Previous studies, such as [35], have shown that Gata1 promotes the differentiation of erythroid cells, and our findings corroborate its dynamic involvement in this process.

Similarly, we calculated the average distance and standard deviation of each gene between seven adjacent cellular states during granulocyte differentiation. We observed topological shift patterns similar to those identified in erythrocyte differentiation (Supplementary Fig. 4A, B). For example, structural variations in the gene Smarcc1 were primarily observed before the Gran_0 stage and stabilized between Gran_0 and Gran_2, suggesting a temporal shift in its topological significance (Supplementary Fig. 4C). Additionally, Eif3g exhibited continuous changes throughout its development, reinforcing its active role in the differentiation process (Supplementary Fig. 4D).

Evaluation of gene module stability

After analyzing topological variations of the same gene across different GRNs, we focused on the overall topological changes in a single gene module between two GRNs. In the human glioblastoma dataset, we designated the 0-h stage as the anchor cell type, clustered the GRN into seven gene modules, and calculated the average pairwise gene embedding distance and NA% between the 0-h and 12-h stages. Each gene module, characterized by distinct NA% and average distances, suggests shifts in the roles of these gene modules during differentiation (Fig. 5A). For example, gene module 5 exhibited a relatively large mean distance and high NA%, suggesting substantial changes in the topological structure of the genes within this module. GO analysis for gene module 5 highlighted metabolic processes, notably cholesterol and

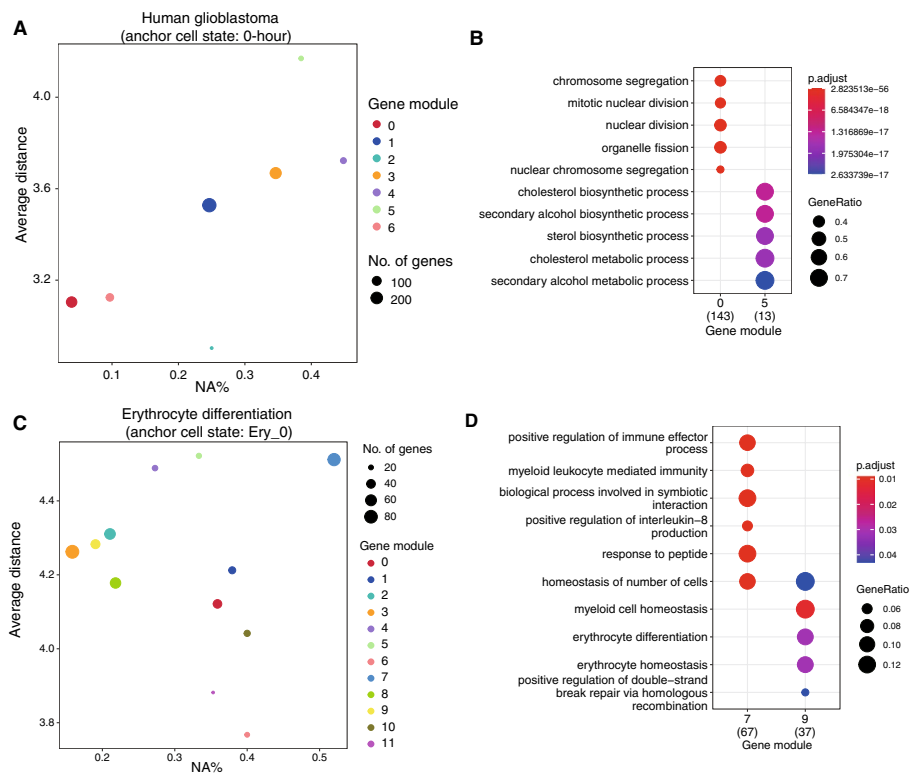


Fig. 5 Gene module stability analysis using Gene2role embeddings. **A** Scatter plot depicting the mean distance and the percentage of genes exclusively found in the anchor cell type (NA%) for seven gene modules in a human glioblastoma dataset, using the 0-h stage as the anchor cell type and comparing against the 12-h stage. **B** Dot plot of the top 5 significant biological processes from the GO analysis for gene modules 0 and 5 in **A**. **C** Scatter plot for twelve gene modules during MEP differentiation, with the Ery_0 stage as the anchor cell type and comparing against the Ery_9 stage, illustrating the mean distance and percent of genes unique to the anchor cell type (NA%). **D** Dot plot for the top 5 significant biological processes from the GO analysis of gene modules 7 and 9 in **C**

sterol biosynthesis (Fig. 5B). Consequently, we infer that the changes in the topological structure of gene module 5 may indicate reduced sterol biosynthesis capabilities within the cells at the 12-h stage. Previous research has demonstrated that patient-derived glioblastoma stem cells (GSCs) activate cholesterol biosynthesis more extensively than differentiated glioblastoma cells [36]. Conversely, gene module 0 exhibited a lower mean distance and NA%, indicating minimal changes in the topological structure of the genes within this module. GO analysis revealed that gene module 0 was enriched in fundamental cellular processes, including chromosome segregation and nuclear division (Fig. 5B). This observation suggests that the functions of chromosome segregation and nuclear division are relatively stable between the 0-h and 12-h stages. Additionally, when we designated the 12-h stage as the anchor cell type, the clustering resulted in five main gene modules. We observed that gene module 1 was relatively stable between the 0-h and 12-h stages (Supplementary Fig. 5A), exhibiting similar biological processes to those of gene module 0 when 0-h served as the anchor cell type (Supplementary Fig. 5B). This observation suggests that they may represent the same gene module, consistently maintaining a stable topological structure and biological function throughout the differentiation process. In contrast, gene

module 9, characterized by its enrichment in biological functions related to extracellular matrix and structure organization, was relatively unstable.

To explore gene module stability during erythrocyte differentiation, we established Ery_0 stage as the anchor cell type and identified 12 gene modules. We calculated the average distances and NA% for Ery_3 (Supplementary Fig. 5C), Ery_6 (Supplementary Fig. 5D), and Ery_9 (Fig. 5C) in relation to the Ery_0 stage across the 11 gene modules. We observed that several gene modules, including gene module 9, maintained stable positions throughout the development process. GO analysis indicated that gene module 9 was enriched in biological processes related to erythrocyte differentiation homeostasis, suggesting a continuous role in driving erythrocyte differentiation (Fig. 5D). In contrast, gene module 7 experienced significant role changes and was involved in biological processes related to immune regulation, suggesting that its immunomodulatory functions may be diminished during differentiation.

By setting GMP_0 as the anchor cell type and computing the average distances and NA% from Gran_3 stage, we further explored the stability of 11 gene modules during granulocytes differentiation (Supplementary Fig. 5E). Similarly, we observed that gene module 2, which was enriched in biological processes related to mononuclear cell differentiation, was unstable during granulocytes differentiation (Supplementary Fig. 5F). This observation suggests that roles of genes driving mononuclear cell differentiation has changed during granulocyte differentiation. Gene module 6, which was stable and enriched in fundamental metabolic processes, implies that this module provides essential metabolites during differentiation.

Discussion

In this study, we introduced Gene2role, the first gene embedding method that utilizes the topological attributes of genes within gene regulatory networks (GRNs). Our findings demonstrate that Gene2role effectively captures the topological information within GRNs constructed from four distinct data sources. By facilitating the integration of GRNs across various cell states and types, Gene2role enables robust comparative analysis at two levels: first, it identifies genes exhibiting significant topological discrepancies among cell types; secondly, it evaluates the stability of gene modules across two cellular states. It is worth noting that the quality of our graph embeddings heavily depends on the quality of the GRNs. As improved GRN construction methods become available, Gene2role can generate more precise embeddings, enabling a broader range of downstream analyses.

Our analysis of DTGs between two cell types across multiple datasets consistently revealed three gene patterns: those exhibiting significant changes in both expression and topology, those with altered expression but stable topology, and those with stable expression but significant topological changes. These observations suggest a relative independence between changes in gene roles within cellular networks and alterations in their expression level. Notably, genes with stable expression but significant topological changes may undergo functional changes that traditional differential gene expression analyses often overlook. Thus, our method provides a valuable complementary perspective by focusing on topological shifts within GRNs. One limitation of our method is that it identifies only those genes with structural variations that maintain at least

one connection in the networks. Consequently, it may miss changes in genes that are embedded in some cell types but absent in others. Future research could focus on inductively quantifying distances for genes selectively connected across different cell types to enhance our comprehension of how gene role changes across cellular contexts.

We examined topological shifts within gene modules to assess their stability across two cellular states. Combined with GO analysis, this stability assessment enables a nuanced understanding of the relationship between the functional dynamics of gene modules and shifts in cellular states. For instance, in analyzing erythrocyte differentiation from the multi-omics dataset, we observed that gene module 9 maintained a low average distance and NA% between Ery_0 and Ery_9, and it was implicated in the promotion of erythrocyte differentiation by GO analysis (Fig. 5D). Hence, this gene module may play a stable role in driving erythrocyte differentiation. In our stability analysis of a gene module between two cellular states, we focused on the proximity structures with the anchor cell type. However, a deeper investigation into the proximity information in the non-anchor cell type could reveal proximity-based shifts between cell types. By comprehensively investigating the topological and proximity shifts between cell types, we can further discern the coherence or dysfunction of gene modules, thus providing specific insights into their stability.

Although our method has been tested only on GRNs inferred from single-cell RNA-seq and multi-omics datasets, it is equally applicable to spatial transcriptomics data [37]. Recent analyses of spatial transcriptomics have concentrated on inferring spatial domains by integrating expression, spatial, and histological information [38, 39]. These spatial domains can serve as a basis for identifying spatial GRNs as inputs for our method, thereby enabling a deeper understanding of topological variations among GRNs in spatial contexts.

Conclusions

We introduced Gene2role, a method that captures the topological information of genes within GRNs and provides insights into gene function changes that may be missed by traditional expression analyses. Our method successfully identified genes with significant topological shifts across various cell types and states, revealing functional dynamics beyond changes in expression. In the future, Gene2role could be extended to analyze biological networks generated by other data types, such as spatial transcriptomics, to further explore gene regulation in diverse biological contexts.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-025-06128-x>.

Supplementary material 1 (DOCX 11562 KB)

Acknowledgements

Computational resources were provided by the supercomputer system SHIROKANE at the Human Genome Center, Institute of Medical Science, the University of Tokyo.

Author contributions

Conceptualization, X.Z., S.L.; Data Curation, X.Z.; Formal Analysis, X.Z.; Investigation, X.Z., S.L.; Algorithm Optimization, X.Z., S.L., B.L., W.Z., W.X.; Writing – Original Draft, X.Z. and S.L.; Writing – Review & Editing, X.Z., S.L., F.T., and K.N.; Resources, F.T., K.N.; Supervision, F.T. and K.N.

Availability of data and materials

The edgelist of curated networks were downloaded from BEELINE [20]. For single-cell RNA-seq data, the count matrix and metadata of human glioblastoma, human PBMC, and human BMMC were collected from GEO dataset (GSE144623, GSE139369). The edgelist generated from multi-omics data were collected from CellOracle [5]. Codes and processed data generated for this project is available on GitHub (<https://github.com/liushu2019/Gene2Role>) and Figshare (<https://figshare.com/articles/dataset/data/25852915?file=46401100>)

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 15 September 2024 Accepted: 1 April 2025

Published online: 24 May 2025

References

- Emmert-Streib F, Dehmer M, Haibe-Kains B. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Front Cell Dev Biol*. 2014. <https://doi.org/10.3389/fcell.2014.00038/abstract>.
- Akers K, Murali TM. Gene regulatory network inference in single-cell biology. *Curr Opin Syst Biol*. 2021;26:87–97.
- Nakajima N, Hayashi T, Fujiki K, Shirahige K, Akiyama T, Akutsu T, et al. Codependency and mutual exclusivity for gene community detection from sparse single-cell transcriptome data. *Nucl Acids Res*. 2021;49(18):e104–e104.
- Badia-i-Mompel P, Wessels L, Müller-Dott S, Trimbou R, Ramirez Flores RO, Argelaguet R, et al. Gene regulatory network inference in the era of single-cell multi-omics. *Nat Rev Genet*. 2023;24(11):739–54.
- Kamimoto K, Stringa B, Hoffmann CM, Jindal K, Solnica-Krezel L, Morris SA. Dissecting cell identity via network inference and in silico gene perturbation. *Nature*. 2023;614(7949):742–51.
- Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods*. 2017;14(11):1083–6.
- Zeng X, Gyoja F, Cui Y, Loza M, Kusakabe TG, Nakai K. Comparative single-cell transcriptomic analysis reveals key differentiation drivers and potential origin of vertebrate retina. *bioRxiv*; 2023. <https://doi.org/10.1101/2023.12.03.569795>.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 2008;9(1):559.
- Lemoine GG, Scott-Boyer MP, Ambroise B, Périn O, Droit A. GWENA: gene co-expression networks analysis and extended modules characterization in a single Bioconductor package. *BMC Bioinform*. 2021;22(1):267.
- Duren Z, Lu WS, Arthur JG, Shah P, Xin J, Meschi F, et al. Sc-compReg enables the comparison of gene regulatory networks between conditions using single-cell data. *Nat Commun*. 2021;12(1):4763.
- Wang J, Ma A, Ma Q, Xu D, Joshi T. Inductive inference of gene regulatory network using supervised and semi-supervised graph neural networks. *Comput Struct Biotechnol J*. 2020;18:3335–43.
- Wu YH, Huang YA, Li JQ, You ZH, Hu PW, Hu L, et al. Knowledge graph embedding for profiling the interaction between transcription factors and their target genes. *PLoS Comput Biol*. 2023;19(6):1011207.
- Gao Z, Su Y, Xia J, Cao RF, Ding Y, Zheng CH, et al. DeepFGRN: inference of gene regulatory network with regulation type based on directed graph embedding. *Brief Bioinform*. 2024;25(3):143.
- Ribeiro LFR, Savarese PHP, Figueiredo DR. struc2vec: learning node representations from structural identity. In: *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*; 2017. pp. 385–94.
- Liu S, Toriumi F, Zeng X, Nishiguchi M, Nakai K. SignedS2V: structural embedding method for signed networks. In: Cherifi H, Mantegna RN, Rocha LM, Cherifi C, Micciché S, editors. *Complex networks and their applications XI. Studies in computational intelligence*, vol. 1077. Cham: Springer; 2023. pp. 337–49. https://doi.org/10.1007/978-3-031-21127-0_28.
- Krumsiek J, Marr C, Schroeder T, Theis FJ. Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network. *PLoS ONE*. 2011;6(8):e22649.
- Giacomantonio CE, Goodhill GJ. A Boolean model of the gene regulatory network underlying mammalian cortical area development. *PLoS Comput Biol*. 2010;6(9):e1000936.
- Lovrics A, Gao Y, Juhász B, Bock I, Byrne HM, Dinnyés A, et al. Boolean modelling reveals new regulatory connections between transcription factors orchestrating the development of the ventral spinal cord. *PLoS ONE*. 2014;9(11):e111430.
- Ríos O, Frías S, Rodríguez A, Kofman S, Merchant H, Torres L, et al. A Boolean network model of human gonadal sex determination. *Theor Biol Med Model*. 2015;12(1):26.
- Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods*. 2020;17(2):147–54.

21. Granja JM, Klemm S, McGinnis LM, Kathiria AS, Mezger A, Corces MR, et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol.* 2019;37(12):1458–65.
22. Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell.* 2015;163(7):1663–77.
23. Cusanovich DA, Hill AJ, Aghamirzaie D, Daza RM, Pliner HA, Berletch JB, et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell.* 2018;174(5):1309–1324.e18.
24. Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 2004;5(2):101–13.
25. Salvador S, Chan P. FastDTW: toward accurate dynamic time warping in linear time and space. 2007.
26. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. 2013. <http://arxiv.org/abs/1301.3781>.
27. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell.* 2021;184(13):3573–87.
28. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS J Integr Biol.* 2012;16(5):284–7.
29. Chen Y, Qian T, Liu H, Sun K. "Bridge": enhanced signed directed network embedding. In: Proceedings of the 27th ACM international conference on information and knowledge management. Torino: ACM; 2018. pp. 773–82. <https://doi.org/10.1145/3269206.3271738>.
30. Huang J, Shen H, Hou L, Cheng X. SDGNN: learning node representation for signed directed networks. *AAAI.* 2021;35(1):196–203.
31. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. 2017. <http://arxiv.org/abs/1609.02907>.
32. Vandenbon A. Evaluation of critical data processing steps for reliable prediction of gene co-expression from large collections of RNA-seq data. *PLoS ONE.* 2022;17(1):e0263344.
33. Algabri YA, Li L, Liu ZP. scGENA: a single-cell gene coexpression network analysis framework for clustering cell types and revealing biological mechanisms. *Bioengineering.* 2022;9(8):353.
34. Wang CC, Hueng DY, Huang AF, Chen WL, Huang SM, Yi-Hsin CJ. CD164 regulates proliferation, progression, and invasion of human glioblastoma cells. *Oncotarget.* 2019;10(21):2041–54.
35. Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci USA.* 1996;93(22):12355–8.
36. Gu D, Zhou F, You H, Gao J, Kang T, Dixit D, et al. Sterol regulatory element-binding protein 2 maintains glioblastoma stem cells by keeping the balance between cholesterol biosynthesis and uptake. *Neuro Oncol.* 2023;25(9):1578–91.
37. Larsson L, Frisén J, Lundeberg J. Spatially resolved transcriptomics adds a new dimension to genomics. *Nat Methods.* 2021;18(1):15–8.
38. Long Y, Ang KS, Li M, Chong KLK, Sethi R, Zhong C, et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat Commun.* 2023;14(1):1155.
39. Yang Y, Cui Y, Zeng X, Zhang Y, Loza M, Park SJ, et al. STAIG: spatial transcriptomics analysis via image-aided graph contrastive learning for domain exploration and alignment-free integration. *Bioinformatics.* 2023. <https://doi.org/10.1101/2023.12.18.572279>.

Publisher's Note

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