Mathematical Biosciences and Engineering

https://www.aimspress.com/journal/mbe

MBE, 23(2): 366–387.
DOI: 10.3934/mbe.2026015
Received: 05 March 2025
Revised: 30 October 2025
Accepted: 11 November 2025
Published: 06 January 2026

Research article

An enhanced triclustering δ-Trimax method with fuzzy cuckoo search based on Lévy flight and Gaussian distribution for gene expression data

Titin Siswantining[1,*], Muhamad Ido Raskapati[1], Nisa Nurul Hidayah[1], Gianinna Ardaneswari[1], Saskya Mary Soemartojo[1], Siti Nurrohmah[1], Devvi Sarwinda[1] and Setia Pramana[2]

[1] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok 16424, Indonesia

[2] Politeknik Statistika STIS, Jakarta 13330, Indonesia

* **Correspondence:** Email: titin@sci.ui.ac.id.

Abstract: The triclustering method employed in this study integrates the δ-Trimax approach with the fuzzy cuckoo search (FCS), thereby leveraging the Lévy flight and Gaussian distribution to analyze gene expression data in three dimensions. In this framework, the initial triclusters produced by δ-Trimax are further optimized using FCS, where the Lévy flight enhances global exploration and the Gaussian distribution intensifies local exploitation, thus achieving a balanced search for optimal solutions. Each tricluster set is evaluated using the tricluster quality index (TQI) to ensure coherence across genes, conditions, and time points. The method was applied to gene expression datasets from primary fibroblast cells and heart disease samples. In the fibroblast dataset, the best tricluster set was obtained with $\delta = 0.015$ and yielded the lowest average TQI value. For the heart disease dataset, the most optimal solution was achieved with $\delta = 0.026$, which yielded the lowest average TQI, and the best tricluster showed large gene coverage across multiple time points. A functional analysis of the selected triclusters using gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways uncovered significant enrichment in pathways such as the NF-κB signaling pathway (hsa04064), TGF-β signaling pathway (hsa04350), and calcium signaling pathway (hsa04020), all of which are mechanistically relevant to immune modulation, extracellular matrix organization, and cardiac muscle function. These findings highlight the utility of the proposed hybrid framework in uncovering biologically meaningful gene modules and provide valuable insights into the molecular mechanisms underlying fibrotic and cardiovascular diseases.

Keywords: fuzzy cuckoo search; gene expression data; gene ontology; triclustering quality index; triclustering analysis

1. Introduction

Data serves as a critical foundation for planning, decision-making, and rule-establishing in various domains. The process of extracting meaningful information from the data is known as data mining, where clustering techniques are frequently employed. However, due to certain limitations in clustering methods, the concepts of biclustering and triclustering have been introduced. Biclustering enables simultaneous clustering across observations and attributes, thus offering a nuanced view of data relationships. Triclustering further extends this capability by accommodating three-dimensional data, thereby encompassing observations, attributes, and the context dimension. These methodologies are particularly pivotal in bioinformatics, thereby facilitating the analysis of microarray technology data. Additionally, recent studies have emphasized the importance of advanced computational frameworks to identify gene modules that are specific to both condition and time in complex diseases, thus highlighting the growing role of integrated optimization strategies in biomedical research [1]. Moreover, recent optimization-driven studies demonstrated the utility of hybrid metaheuristics for complex inference and diagnostics in applied domains [2, 3]. In response to this challenge, we propose a novel triclustering framework that aims to overcome these limitations. Our method seeks to achieve two main objectives: (1) enhance the homogeneity of the initial tricluster population to better reflect meaningful biological patterns; and (2) improve the optimization process for discovering high-quality triclusters, particularly by managing the uncertainty that is characteristic of gene expression measurements.

This allows for a deeper understanding of gene relationships within specific conditions and timeframes, thus enhancing our grasp of complex biological processes. However, traditional clustering and biclustering techniques often fall short in capturing the intricate, multi-dimensional structures inherent in biological datasets, especially gene expression data involving varying conditions and time points. Therefore, there is a critical need for more sophisticated approaches that can simultaneously consider multiple dimensions while addressing uncertainty and heterogeneity in the data. The increasing complexity of biological datasets in recent years further underscores this need, as demonstrated by contemporary applications of mathematical modeling and integrative analyses in genomic and biomedical contexts [1]. Compared with other established triclustering schemes (e.g., Trimax variants with random or greedy initializations, particle swarm optimization (PSO)/cuckoo search (CS)—based optimizers, or tensor factorization families), δ-Trimax offers a principled control of homogeneity via an mean square residual (MSR) threshold (δ), thus producing an initial population with very low residue. This design reduces the sensitivity to noisy genes and improves the downstream optimization stability. In our setting, δ-Trimax outperformed random/greedy initializations in both compactness and downstream tricluster quality index (TQI).

The integration of δ-Trimax and the FCS was designed to synergize their respective strengths: δ-Trimax's ability to construct highly homogeneous initial tricluster populations and the FCS's capacity for a robust global search and uncertainty management through fuzzy logic and probabilistic exploration (Gaussian distribution and Lévy flight). This hybrid approach is intended to surpass the performance of previous methods that solely rely on either heuristic optimization or simple random initialization. This study introduces a novel tricluster analysis method that integrates the FCS algorithm, thereby leveraging the Gaussian distribution, with the δ -Trimax algorithm, and combines δ-Trimax with an FCS based on the Lévy flight. Initially, the δ-Trimax algorithm is utilized to form

the early tricluster population, which is then optimized using the FCS algorithm, thereby drawing on the Gaussian distribution. This approach marks an advance from Narmadha's work, which applied the Greedy Two-Way K-means algorithm for the initial population setup, followed by optimization through the PSO algorithm [4]. The FCS method, which is an evolution of the cuckoo search strategy, incorporates fuzzy logic for the evaluation of cluster fitness. Inspired by the parasitic nature of certain cuckoo species that lay their eggs in the nests of other birds, the FCS employs fuzzy C-means as its objective function and utilizes the Lévy flight for tricluster formation. A replacement of suboptimal clusters from previous processes is executed using a local random walk. This design choice aligns with contemporary evidence that evolved metaheuristics improve the search efficiency and solution quality in high-dimensional settings [2, 3].

Unlike the straightforward CS algorithm that initiates with a randomly encoded initial tricluster population, the proposed algorithm leverages the nodes deletion algorithm in δ-Trimax for the initial population creation. Implementing the nodes deletion strategy within δ-Trimax aims to achieve an initial tricluster population characterized by significant homogeneity. This could enhance the efficiency in identifying the optimal tricluster solutions [4]. The δ-Trimax algorithm's principle of setting a δ threshold as the upper limit for the MSR facilitates obtaining MSR values below this threshold, nearing zero, which signifies a high degree of tricluster homogeneity [5].

The basic CS employs the Lévy flight to explore the search space [6]. However, the proposed CS algorithm utilizes a Gaussian distribution random walk, which was shown to produce significantly better tricluster solutions with a higher level of convergence compared to the basic CS algorithm [7]. This results in the Gaussian distribution-based CS finding optimal tricluster solutions more effectively and efficiently. The proposed FCS algorithm is an advancement of the CS algorithm, thereby incorporating fuzzy concepts with fuzzy C-means (FCM) as the objective function to evaluate the tricluster suitability. The FCS algorithm uses fuzzy memberships to handle uncertainty in high-dimensional data [7–9].

Both the Lévy flight and Gaussian distribution strategies aim to generate new candidate solutions by random perturbation, though they significantly differ in their step-size behavior. The Lévy flight introduces a heavy-tailed probability distribution, enabling occasional large jumps, which enhances the global exploration. In contrast, the Gaussian distribution promotes smaller, normally distributed perturbations, thus supporting local exploitation. In addition, the Lévy flight strategy often incorporates a Lévy constant to govern the jump scale, while Gaussian-based methods may adjust membership functions using sigmoid transformations to better model the uncertainty in the fuzzy optimization processes.

This research employs a hybrid triclustering analysis method that combines the δ-Trimax and FCS algorithms, based on the Lévy flight and Gaussian distribution, to analyze gene expression data from fibroblast cells and heart disease. The initial phase involves forming a homogenous tricluster population using the δ-Trimax method. The δ threshold value for running the algorithm within δ-Trimax is determined using the silhouette coefficient method, which was chosen for its ability to better identify and handle outliers compared to other methods. The subsequent phase is the optimization phase using the FCS method based on the Lévy flight and Gaussian distribution. This triclustering formation phase aims to identify the optimal tricluster set. The TQI method is used to evaluate the tricluster results, where a lower TQI value indicates a better tricluster quality [10, 11]. Furthermore, an advanced analysis using GO will be employed. In the GO evaluation, the tricluster

results are further analyzed to understand the relationships of the genes within the tricluster to biological processes, molecular functions, and cellular components. Through this integrated methodology, we aim to provide a more reliable and biologically meaningful analysis of gene expression data, thus potentially contributing to deeper insights into gene regulation mechanisms under specific biological contexts.

2. Materials and methods

2.1. Perfect shifting triclustering

A perfect shifting triclustering analysis is used to identify genes that are co-regulated in response to specific treatments or conditions over an observed period. Suppose there is a tricluster $T(P, Q, R) = tpqr$, where $p \in P$, $q \in Q$, $r \in R$ represents a perfect shifting tricluster if the elements of the tricluster T can be expressed in an equation as follows:

$$t_{pqr} = \Delta + \alpha_p + \beta_q + \eta_r, \tag{2.1}$$

where Δ is the tricluster constant, and α_p, β_q, and η_r are the shifting factors from the p-th gene, the q-th condition, and the r-th time, respectively. The equations to calculate the average value at each node can be expressed as follows:

$$t_{pQR} = \frac{1}{|Q||R|} \sum_{q \in Q, r \in R} t_{pqr}, \tag{2.2}$$

$$t_{PqR} = \frac{1}{|P||R|} \sum_{p \in P, r \in R} t_{pqr}, \tag{2.3}$$

$$t_{PQr} = \frac{1}{|P||Q|} \sum_{p \in P, q \in Q} t_{pqr}, \tag{2.4}$$

where t_{pQR} is the average value for the p-th gene, t_{PqR} is the average value for the q-th condition, and t_{PQr} is the average value for the r-th time. The tricluster constant and the shifting factors from the p-th gene, the q-th condition, and the r-th time can be expressed as follows:

$$\Delta = t_{PQR} = \frac{1}{|P||Q||R|} \sum_{p \in P, q \in Q, r \in R} t_{pqr}, \tag{2.5}$$

where t_{pqr} is the value in the data for the p-th gene, the q-th condition, and the r-th time point, $|P|$ is the number of genes, $|Q|$ is the number of conditions, $|R|$ is the number of time points, and Δ is the average value of the tricluster.

The described elements are linked with the previously discussed equations, thus leading to the formulation of the following three additional equations:

$$\alpha_p = t_{pQR} - t_{PQR}, \tag{2.6}$$

$$\beta_q = t_{PqR} - t_{PQR}, \tag{2.7}$$

$$\eta_r = t_{PQr} - t_{PQR}, \tag{2.8}$$

where α_p represents the difference between the average value of the p-th gene and the tricluster's average value, β_q is the difference between the average value of the q-th condition and the tricluster's average value, and η_r is the difference between the average value of the r-th time point and the tricluster's average value. The perfect shifting triclustering $\hat{t}_{pqr}$ for each element t_{pqr} can be expressed in the following equation:

$$\hat{t}_{pqr} = t_{pQR} + t_{PqR} + t_{PQr} - 2t_{PQR}. \tag{2.9}$$

2.2. MSR

MSR measures the difference between the estimated value from the tricluster model and the actual value in the three-dimensional data. The smaller the MSR value on the data, the better the result indicates. To obtain the MSR value, it's necessary to calculate the residual value of each gene, condition, and time element, which is computed using the following equation:

$$\begin{aligned} s_{pqr} &= t_{pqr} - \hat{t}_{pqr} && (2.10)\\ &= t_{pqr} - (t_{pQR} + t_{PqR} + t_{PQr} - 2t_{PQR}) && (2.11)\\ &= t_{pqr} - t_{pQR} - t_{PqR} - t_{PQr} + 2t_{PQR}. && (2.12) \end{aligned}$$

Once the residual values for each gene, condition, and time element, represented by s_{pqr}, are obtained, the calculation to derive the MSR can be expressed through the following equation:

$$\begin{aligned} MSR_{data} &= \frac{1}{|P||Q||R|} \sum_{p\in P, q\in Q, r\in R} s_{pqr}^2 && (2.13)\\ &= \frac{1}{|P||Q||R|} \sum_{p\in P, q\in Q, r\in R} (t_{pqr} - t_{pQR} - t_{PqR} - t_{PQr} + 2t_{PQR})^2. && (2.14) \end{aligned}$$

2.3. δ-Trimax

The δ-Trimax method is employed to form an initial set of homogeneous triclusters with minimal MSR values. The δ-Trimax algorithm used in this study incorporates both multiple nodes deletion and single node deletion algorithms because the process of node removal with these algorithms is capable of achieving minimal MSR values. The multiple nodes deletion algorithm sets a parameter value $\lambda > 0$ that serves as a threshold to control the amount of deletion performed [11]. This algorithm is executed if the number of dimensions for the genes, conditions, or time exceeds 50 [12]. The deletions in the multiple nodes' deletion algorithm for each gene, condition, and time can be expressed in the following equations:

$$\frac{1}{|Q||R|} \sum_{q\in Q, r\in R} \left(t_{pqr} - t_{pQR} - t_{PqR} - t_{PQr} + 2t_{PQR}\right)^2 > \lambda S, \tag{2.15}$$

$$\frac{1}{|P||R|}\sum_{p\in P,r\in R}\left(t_{pqr}-t_{pQR}-t_{PqR}-t_{PQr}+2t_{PQR}\right)^2>\lambda S, \tag{2.16}$$

$$\frac{1}{|P||Q|}\sum_{p\in P,q\in Q}\left(t_{pqr}-t_{pQR}-t_{PqR}-t_{PQr}+2t_{PQR}\right)^2>\lambda S, \tag{2.17}$$

where t_{pqr} represents the gene data value, t_{pQR} being is the average value of the p-th gene, t_{PqR} is the average value of the q-th condition, t_{PQr} is the average value of the r-th time point, and t_{PQR} is the average value of all elements within the tricluster. Here, S denotes the current MSR of the tricluster, with $S \equiv MSR_{data}$ as defined in Eq (2.13) and λ is a scale value in multiple nodes deletion that is greater than 0. The p-th gene will be removed if it satisfies a certain condition, the q-th condition will be removed if it satisfies another condition, and the r-th time point will be removed if it satisfies yet another condition. The single node deletion algorithm plays a role in removing nodes one by one when the tricluster's MSR value is greater than δ. The removal is done on genes, conditions, or time points that have the highest μ value. The calculation of the μ value for each gene, condition, and time is expressed through the following equations.

$$\mu_{(p)}=\frac{1}{|Q||R|}\sum_{q\in Q,r\in R}\left(t_{pqr}-t_{pQR}-t_{PqR}-t_{PQr}+2t_{PQR}\right)^2, \tag{2.18}$$

$$\mu_{(q)}=\frac{1}{|P||R|}\sum_{p\in P,r\in R}\left(t_{pqr}-t_{pQR}-t_{PqR}-t_{PQr}+2t_{PQR}\right)^2, \tag{2.19}$$

$$\mu_{(r)}=\frac{1}{|P||Q|}\sum_{p\in P,q\in Q}\left(t_{pqr}-t_{pQR}-t_{PqR}-t_{PQr}+2t_{PQR}\right)^2. \tag{2.20}$$

2.4. FCS

The FCS is an optimization algorithm that combines the FCM objective function with the metaheuristic CS algorithm based on the Lévy flight. The FCM is an unsupervised clustering algorithm applied to problems related to feature analyses, clustering, and grouping design [8]. The FCM method within the FCS serves to address uncertainty in data through the membership values of objects to be in a particular cluster. This objective function is calculated by considering the fuzzy membership values of each generated solution [6]. The FCS algorithm uniquely combines the global optimization capabilities of the CS [13, 14] with the uncertainty management strength of the FCM clustering approach [8]. By integrating fuzzy logic into the CS framework, the FCS enables a more flexible and robust search process, which is particularly suitable for noisy and high-dimensional data such as gene expression profiles. The use of the Lévy flight enhances the exploration capability of the algorithm, thus allowing it to escape the local optima and efficiently explore the search space. Moreover, recent studies have demonstrated the effectiveness of CS and its variants in handling complex optimization problems and biomedical data analyses [15], thus highlighting the suitability of the FCS for tasks that require both precision and adaptability. These findings are further supported by recent advancements in hybrid and adaptive CS algorithms for high-dimensional and real-time optimization tasks [16–18], thus reinforcing the relevance and scalability of the FCS approach in modern data-intensive domains.

The FCM starts with constructing a membership matrix G containing g_{ij}, which represents the membership weight of each object j in cluster i. The calculation to obtain the initial membership weight matrix values is expressed in the following equation:

$$g_{ij} = \begin{cases} \frac{tc_{ij}}{\sum_{l=1}^{C} tc_{lj}}, & \text{if } tc_{ij} > 0 \\ 0, & \text{if } tc_{ij} = 0, \end{cases} \tag{2.21}$$

where tc_{ij} represents the value of the i-th tricluster for the j-th object. The next step involves calculating the value of the center of the i-th cluster (c_i), which is expressed in the following equation:

$$c_i = \frac{\sum_{j=1}^{n} g_{ij}^m x_j}{\sum_{j=1}^{n} g_{ij}^m}, \tag{2.22}$$

where g_{ij} is the membership weight of the j-th object in the i-th cluster, x_j is the value of the j-th object in the data, m is the fuzzifier, and C is the number of clusters. The next step involves calculating the value of k_{ij}, which represents the distance of the j-th object from the center of the i-th cluster, which is expressed in the following equation:

$$k_{ij} = \sqrt{(x_j - c_i)^2}, \tag{2.23}$$

where x_j is the value of the j-th object, and c_i is the center of the i-th cluster. After determining the distance of the j-th object from the center of the i-th cluster, the membership matrix g_{ij} is updated with the following equation:

$$g_{ij} = \begin{cases} \frac{tc_{ij}}{\sum_{l=1}^{C} \left(\frac{(tc_{ij} \times k_{ij})^2}{k_{lj}} \right)^{\frac{1}{m-1}}}, & tc_{ij} > 0 \\ 0, & tc_{ij} = 0. \end{cases} \tag{2.24}$$

2.5. TQI

The results from the data processing using a triclustering analysis need to be evaluated to assess the quality of the generated triclusters using the TQI. The calculation of the TQI value for the n-th tricluster is formulated in the previous equation as follows [19]:

$$TQI_n = \frac{MSR_{data_n}}{V_n}. \tag{2.25}$$

The TQI value is derived from the division of the n-th MSR by the volume of the n-th tricluster. A low TQI value indicates that the members of the tricluster have a close relationship in the three-dimensional space and can be relied upon to explain data patterns. To find the volume of the n-th tricluster, the following equation is used:

$$V_n = |P_n| \times |Q_n| \times |R_n|. \tag{2.26}$$

2.6. GO

GO is a controlled hierarchical vocabulary that is used to describe molecular functions, biological processes, and cellular components [20]. Within the context of triclustering, GO is utilized to conduct a biological analysis of tricluster results. The aim of a biological analysis is to understand, measure, and compare biological characteristics, responses, or effects. A GO analysis is carried out by associating groups of genes with specific biological processes, molecular functions, and cellular components using gene ontology [21]. This analysis is performed using algorithms or software such as GOnet to calculate p-values to determine the significance of the associations within the gene groups.

2.7. Proposed method

A triclustering analysis which uses the combined δ-Trimax and FCS method begins with determining the δ parameter (via the silhouette coefficient) to run δ-Trimax and obtain a homogeneous set of seed triclusters with a low MSR. We adopt δ-Trimax for the initialization because its explicit MSR threshold concentrates the search on biologically coherent regions and reduces the noise sensitivity, which we found to be more effective than random/greedy starts or a direct CS or PSO [4, 5]. Then, the resulting seeds are refined with the FCS, where Lévy flights support global exploration and Gaussian steps support local exploitation, while fuzzy memberships accommodate uncertainty in high-dimensional expression data [7, 9]. Empirically, this design yields consistently lower TQI and competitive runtimes.

The workflow of the FCS triclustering method based on a Lévy flight is as follows:

1) Determine the δ scale value using the silhouette coefficient;
2) Find an initial homogeneous tricluster solution using the δ-Trimax node deletion method, specifically multiple nodes deletion and a single node deletion;
3) Encode the initial tricluster into binary form;
4) Search for the tricluster solution using the FCS method based on a Lévy flight;
5) Compare the objective function values of the initial tricluster with that of the FCS-derived tricluster based on a Lévy flight and retain the best solution;
6) Discover the optimal tricluster solution using a Local Random Walk;
7) Compare the objective function values of the tricluster obtained from FCS based on a Lévy flight with the tricluster from local random walk and retain the best solution;
8) Continue with the best solution using the TQI; and
9) The best tricluster solution is analyzed using GO.

The combined method of the FCS based on a Gaussian distribution and δ-Trimax consists of two stages: the initial population formation stage and the optimization stage. The aim of the initial population formation is to obtain a homogenous initial population with the minimum MSR to reduce the computational burden during the optimization stage. Initially, the population is formed using the nodes deletion algorithm in δ-Trimax.

The δ-Trimax algorithm consists of three processes: multiple nodes deletion, single nodes deletion, and nodes addition. However, in the proposed method, only multiple nodes deletion and a single node deletion are used because these processes are already capable of producing a tricluster with the minimum MSR value that meets the criteria as an initial population for the combined method of the FCS based on a Gaussian distribution with δ-Trimax.

During the optimization phase of the combined method of the FCS based on a Gaussian distribution with δ-Trimax, the FCS algorithm will be utilized. Previously obtained solutions will be optimized using the FCS algorithm. The search for new solutions will employ a Gaussian distribution. Each solution will be compared based on its objective function value, and the solution with the smaller objective function value will be retained in the next generation until the maximum number of generations is reached. As illustrated in Figure 1, the stages of research encompass the flow from data input through the steps of the process to the output results.

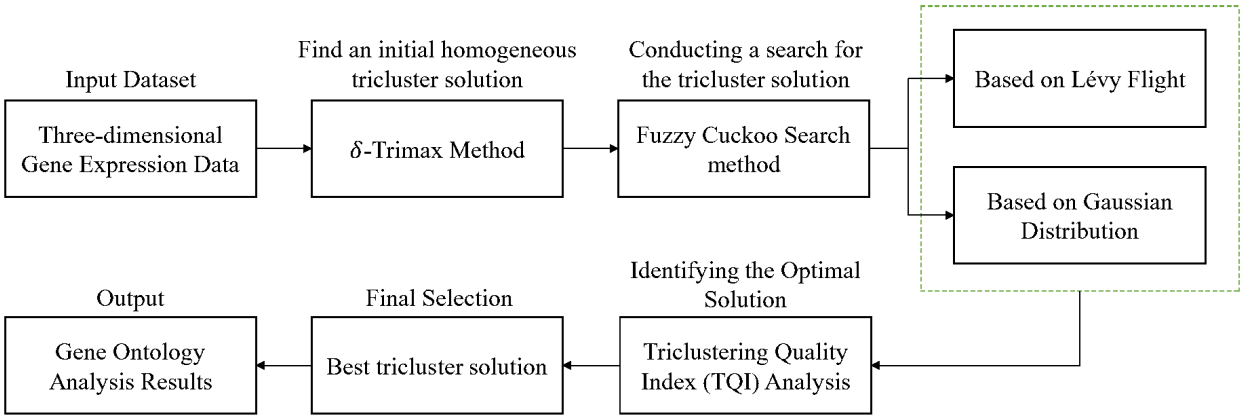

Figure 1. Stages of research. This diagram describes the flow of research from data input, steps of process, and output results.

3. Results

This section presents the results of the triclustering analysis using a hybrid framework that combines the δ-Trimax triclustering method with the FCS optimization algorithm. Two FCS-based approaches were applied, one with the Lévy flight and the other with Gaussian distribution. Each method was applied to two three-dimensional gene expression datasets: GSE35671 (human heart disease), obtained from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35671, and GSE27165 (primary fibroblast cells), obtained from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE27165. The structure of each dataset is shown in Table 1 and Table 2, respectively.

Table 1 presents gene expression data collected during the differentiation of human induced pluripotent stem cells (HiPSCs) in patients with cardiomyopathy, a disease which affects heart muscle function. HiPSCs, derived from somatic cells, have the potential to develop into various cell types and are widely used in regenerative therapies for conditions such as heart disease. The gene expression data obtained from the observation of messenger-RNA during the HiPSC differentiation process is comprised of a total of 48,803 gene types. The temporal observation points span twelve distinct time stages: days 0, 3, 7, 10, 14, 20, 28, 35, 45, 60, 90, and 120. Biological replications were performed on three conditions: fetal human heart tissue, adult human heart tissue, and hypertensive heart tissue. Day 0 reflects baseline gene expression prior to differentiation. The mRNA expression values in this dataset indicate the activity level of genes during HiPSC differentiation: the higher the expression level, the more actively the gene is involved in the differentiation process and potentially in the

pathology of heart disease.

Table 1. Three-dimensional gene expression data of HiPSC differentiation in patients with heart disease (GSE35671).

No.	ID REF	Time 1			⋯	Time 12		
		Replication 1	Replication 2	Replication 3		Replication 1	Replication 2	Replication 3
1	ILMN_1343291	15.4995	15.454	15.604	⋯	14.8468	14.5664	15.5351
2	ILMN_1343295	13.5282	13.8069	13.5797	⋯	13.5482	14.0758	13.7023
3	ILMN_1651199	6.7394	6.99508	6.89129	⋯	6.8638	6.85294	6.95208
4	ILMN_1651209	7.04935	7.03753	7.04918	⋯	7.03632	7.10975	7.02113
⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋮	⋮
48803	ILMN_2416019	6.84598	6.81435	6.85623	⋯	6.98417	6.73951	6.78274

Table 2 presents the three-dimensional gene expression data of primary fibroblast cells subjected to different treatments. The dataset consists of gene expression levels measured after exposure to three conditions: Egr-1, Tgf-β1, and a control group (CG) that received no treatment. Each condition includes two biological replicates and was measured at two time points: 24 and 48 hours post-treatment. Egr-1 and Tgf-β1 are critical regulators of gene expression and fibroblast activity, which can influence both normal tissue repair and the abnormal accumulation of the extracellular matrix. Such dysregulation may lead to fibrosis, one of the key pathological features of systemic sclerosis (SSc) [22]. The dataset is comprised of gene expression values for 22,184 gene types.

Table 2. Gene expression data of primary fibroblast cells (GSE27165).

No. Gene	24 Hours						48 Hours					
	CG 1	CG 2	Egr-1 1	Egr-1 2	Tgf-β1 1	Tgf-β1 2	CG 1	CG 2	Egr-1 1	Egr-1 2	Tgf-β1 1	Tgf-β1 2
1	80.01	103.07	82.39	125.77	127.11	119.51	69.90	98.76	73.49	60.70	97.81	98.83
2	262.99	280.99	267.87	376.94	456.83	315.71	345.51	422.56	345.96	247.97	536.02	633.53
3	155.11	219.98	116.67	155.00	302.59	184.34	120.18	188.00	105.92	64.07	238.09	284.65
4	0.22	0.12	–0.03	–0.02	0.16	0.17	0.03	1.20	0.04	0.00	0.10	–0.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
22184	1080.65	1196.02	995.11	1177.88	1503.41	1277.12	1278.96	1652.99	1131.87	954.25	1659.01	1925.71

The evaluation focuses on the quality of the generated triclusters and their biological relevance, which was assessed through a GO enrichment analysis. In the following subsections, we present the results for each method and dataset combination.

3.1. δ-Trimax triclustering with FCS based on Lévy flight

This subsection presents the outcomes of implementing the integrated δ-Trimax triclustering method with the FCS based on a Lévy flight to both the GSE35671 and GSE27165 datasets. The analysis aimed to discover coherent triclusters across gene-condition-time dimensions. Several simulations were executed using combinations of δ and θ scale values to generate a population of tricluster candidates. The key parameters used in the combined method are summarized in Table 3.

Table 3. Parameter configurations used in the combined δ-Trimax and FCS method based on Lévy flight.

Parameter	Description	Value(s)	Reference / Note
w	Number of triclusters	24	[4]
λ	Threshold for multiple node deletions	1.2	[5]
δ	Threshold for single node deletions	0.022, 0.026, 0.030	[23]
θ	Lévy flight scaling constant	1.2, 1.5, 1.7	[7]
m	Fuzzifier for FCS	2	[7]
α	Step size for Lévy flight	1	[7]
P_α	Probability of egg detection by host	0.25	[7]
num_{iter}	Optimization iterations	20	Based on computational feasibility

The best tricluster results from each simulation on the GSE35671 dataset were evaluated using the TQI, which quantifies the compactness and relevance of the resulting triclusters. Table 4 presents a comparison of the TQI values for all combinations of δ and θ parameters. The lowest TQI value, which indicates the best result, was obtained using $\delta = 0.026$ and $\theta = 1.7$.

Table 4. TQI values from simulation results on the GSE35671 dataset across different δ and θ parameter combinations.

	Lévy Constant θ		
δ	**1.2**	**1.5**	**1.7**
0.022	3.16002×10^{-7}	2.95390×10^{-7}	3.39531×10^{-7}
0.026	3.11333×10^{-7}	3.56250×10^{-7}	$\mathbf{2.74610 \times 10^{-7}}$
0.030	2.88158×10^{-7}	3.34843×10^{-7}	3.67394×10^{-7}

Table 5 displays the TQI values for 24 triclusters obtained from the simulation using $\delta = 0.026$ and $\theta = 1.7$.

Tricluster 8 from Table 5 was further analyzed using a GO analysis to investigate the biological relevance of the genes in relation to human heart disease. The gene set in Tricluster 8 was submitted to the GOnet application to explore its associated biological processes, molecular functions, and cellular components. The results below show the most significant GO terms (based on the lowest p-values) for each GO category (see Tables 6–8). P-values returned as 0 by the software (numerical underflow) are reported as $< 1.0 \times 10^{-15}$. Multiple testing was controlled using the Benjamini-Hochberg false discovery rate (FDR) procedure ($q < 0.05$).

To gain biological insight into the gene expression patterns identified through the δ-Trimax triclustering method combined with the FCS based on a Lévy flight, we performed a functional enrichment analysis on the resulting gene sets using the KEGG database. The analysis was conducted on the triclusters extracted from the GSE27165 dataset, which profiles gene expression in primary fibroblasts under Egr-1 and Tgf-β1 stimulation (fibrosis-relevant context). Several genes within the most significant tricluster were mapped to relevant KEGG pathways, particularly those involved in signal transduction, cellular transport, and immune response mechanisms. These enriched pathways suggest coordinated regulatory mechanisms that may underlie disease progression and highlight potential biomarkers or therapeutic targets in fibrosis and extracellular matrix remodeling.

Table 5. TQI values for 24 triclusters with the lowest TQI result on the GSE35671 dataset.

Tricluster - i	MSR	Volume (Genes×Conditions×Time)	TQI
1	0.078637795	294813	2.66738×10^{-7}
2	0.097940099	122255	8.01113×10^{-7}
3	0.117452325	440982	2.66343×10^{-7}
4	0.124714453	339836	3.66984×10^{-7}
5	0.086972051	366075	2.37580×10^{-7}
6	0.098276945	196968	4.98949×10^{-7}
7	0.104882531	195008	5.37837×10^{-7}
8	0.109929535	730890	$\mathbf{1.50450 \times 10^{-7}}$
9	0.105133849	518448	2.02786×10^{-7}
10	0.114625951	661041	1.73402×10^{-7}
11	0.114937032	441432	2.60373×10^{-7}
12	0.088891470	512610	1.73410×10^{-7}
13	0.107796470	342342	3.14879×10^{-7}
14	0.069207420	220473	3.13904×10^{-7}
15	0.104204048	340494	3.06038×10^{-7}
16	0.093736293	587496	1.59552×10^{-7}
17	0.113299756	734340	1.54288×10^{-7}
18	0.098325824	511812	1.92113×10^{-7}
19	0.117200257	442260	2.65003×10^{-7}
20	0.118122384	584976	2.01297×10^{-7}
21	0.103725442	662175	1.51649×10^{-7}
22	0.099664130	583632	1.70765×10^{-7}
23	0.098954278	389712	2.53916×10^{-7}
24	0.108854355	657018	1.65679×10^{-7}

Table 6. GO biological process analysis results for tricluster 8 (GSE35671).

No	GO_term_ID	Definition	*P-value*	Num of genes
1	GO:0010604	*positive regulation of macromolecule metabolic process*	$< 1.0 \times 10^{-15}$	1964
2	GO:0023051	*regulation of signaling*	$< 1.0 \times 10^{-15}$	2071
3	GO:0070887	*cellular response to organic substance*	$< 1.0 \times 10^{-15}$	1426

Table 7. GO molecular function analysis results for Tricluster 8 (GSE35671).

No.	GO_term_ID	Definition	*P-value*	Num of genes
1	GO:0050839	*cell adhesion molecule binding*	1.26×10^{-4}	321
2	GO:0045296	*cadherin binding*	1.73×10^{-4}	219
3	GO:0019899	*enzyme binding*	3.98×10^{-4}	1299

Table 8. GO cellular component analysis results for tricluster 8 (GSE35671).

No.	GO_term_ID	Definition	*P-value*	Num of genes
1	GO:0005912	adherens junction	9.26×10^{-4}	351

Table 9 presents the enriched KEGG pathways alongside their corresponding genes. A pathway enrichment analysis of the genes identified in the most informative tricluster derived from δ-Trimax Triclustering with the FCS based on a Lévy flight, applied to the GSE27165 dataset, revealed several biologically relevant patterns. A subset of genes, including TIAM1, RASGRF2, RASGRP4, and NTRK1, were found to be consistently involved in the Ras and MAPK signaling cascades (hsa04014, hsa04010), thus suggesting a potential convergence of signal transduction mechanisms associated with cell proliferation, survival, and inflammation. The presence of SQSTM1, a key regulator of

autophagy and mitophagy, within multiple stress-response pathways (e.g., hsa04140, hsa04137, hsa04218) indicates the relevance of intracellular quality control processes in the disease context. Additionally, genes such as ARFGAP1 and PSD2 were mapped to the endocytosis pathway (hsa04144), thus highlighting vesicle-mediated transport as a potentially dysregulated mechanism. The enrichment of transport-related pathways, particularly ABC transporters (hsa02010), was supported by the inclusion of ABCA6 and ABCD1. Notably, NTRK1 was annotated across a broad spectrum of pathways, including calcium signaling, PI3K-Akt, neurotrophin signaling, and cancer-related pathways, thus underscoring its multifunctional role in cellular signaling and disease progression. These findings suggest that the triclustered genes are not only co-expressed under specific experimental conditions, but also converge on key regulatory networks involved in signaling, transport, and cellular stress, thus potentially contributing to the pathophysiology of coronary heart disease.

Table 9. Annotated KEGG pathways of genes in tricluster from GSE27165.

PROBEID	ENTREZID	Gene Name	KEGG Pathway(s)
ILMN_1777794	5589	PRKCSH	hsa04141: Protein processing in endoplasmic reticulum
ILMN_1675709	55738	ARFGAP1	hsa04144: Endocytosis
ILMN_1701551	23460	ABCA6	hsa02010: ABC transporters
ILMN_1707741	353149	TBC1D26	–
ILMN_1655577	7074	TIAM1	hsa04014, hsa04015, hsa04024, hsa04062, hsa04530, hsa04810, hsa05205
ILMN_1658684	374868	ATP9B	–
ILMN_1753377	5924	RASGRF2	hsa04010, hsa04014
ILMN_1662963	84249	PSD2	hsa04144: Endocytosis
ILMN_1708093	7984	ARHGEF5	–
ILMN_1806979	6579	SLCO1A2	hsa04976: Bile secretion
ILMN_1703856	8878	SQSTM1	hsa04137, hsa04140, hsa04217, hsa04218, hsa04380, hsa05014, hsa05022, hsa05131, hsa05418
ILMN_1770307	4914	NTRK1	hsa04010, hsa04014, hsa04020, hsa04151, hsa04210, hsa04722, hsa04750, hsa05200, hsa05202, hsa05216, hsa05230
ILMN_1691272	10641	NPRL2	hsa04150: mTOR signaling pathway
ILMN_1713438	55937	APOM	–
ILMN_1714650	115727	RASGRP4	hsa04010, hsa04014, hsa05200
ILMN_1809925	117289	TAGAP	–
ILMN_1770224	10611	PDLIM5	hsa04820: Cytoskeleton in muscle cells
ILMN_1772189	215	ABCD1	hsa02010: ABC transporters, hsa04146: Peroxisome

3.2. δ-Trimax triclustering with the FCS based on a Gaussian distribution

This subsection presents the outcomes of implementing the integrated δ-Trimax triclustering method with FCS based on Gaussian distribution to both the GSE35671 and GSE27165 datasets. The hybrid method combining the FCS based on a Gaussian distribution with δ-Trimax consists of two phases: the initial population formation phase and the optimization phase. In each phase, the triclustering analysis algorithm operates with several required scale parameters. All the utilized scale values can be seen in Table 10.

From the experimentally determined scale values, several simulation scenarios will be conducted. Each simulation scenario is repeated 3 times to observe the consistency of the results. The simulation results and TQI values can be seen in Table 11.

Table 10. Parameter values used in the δ-Trimax triclustering with FCS based on Gaussian distribution.

Scale Value	Description	Value	Reference
w	Number of *Triclusters*	24	[4]
λ	*Threshold* for *Multiple Nodes Deletion*	1.2	[5]
		0.015	
δ	*Threshold* for *Single Node Deletion*	0.014	MSR value dataset of 0.015022
		0.013	
m	*Fuzzifier*	2	[7]
μ	Constant Value	0.0001	[24]
		0.15	
σ_0	Constant value	0.30	Solution probability considerations
		0.50	
α	*Cuckoo* step	1	[7]
P_a	Probability of cuckoo egg detection	0.25	[7]
num_{iter}	Number of optimization phase iterations	15	Computational time considerations

Table 11. Average TQI values obtained from the GSE27165 dataset for different combinations of δ and σ_0.

δ	$\sigma_0 = 0.15$	$\sigma_0 = 0.30$	$\sigma_0 = 0.50$
0.015	1.52906×10^{-7}	1.57042×10^{-7}	1.33142×10^{-7}
	1.74490×10^{-7}	1.56290×10^{-7}	$\mathbf{1.22454 \times 10^{-7}}$
	1.56892×10^{-7}	1.43333×10^{-7}	1.26101×10^{-7}
0.014	1.57573×10^{-7}	1.45873×10^{-7}	1.29908×10^{-7}
	1.63733×10^{-7}	1.39758×10^{-7}	1.24109×10^{-7}
	1.52830×10^{-7}	1.58310×10^{-7}	1.28889×10^{-7}
0.013	1.60854×10^{-7}	1.57022×10^{-7}	1.30461×10^{-7}
	1.55891×10^{-7}	1.51850×10^{-7}	1.32976×10^{-7}
	1.62758×10^{-7}	1.43664×10^{-7}	1.28873×10^{-7}

The result obtained with $\delta = 0.015$ and $\sigma_0 = 0.50$ yields the lowest TQI value of 1.22454×10^{-7}, which indicates that this combination produces the most optimal tricluster set. Accordingly, this parameter setting is adopted for the interpretation of the triclustering analysis results. The optimal tricluster set derived from the simulations, thereby employing the triclustering method that integrates the δ-Trimax and FCS algorithms based on a Gaussian distribution, is presented in Table 12. There are six conditions based on the data available, where conditions k_1 and k_2 involve two subjects not treated with either Egr-1 or Tgf-β1. Conditions k_3 and k_4 involve two subjects treated with Egr-1, while k_5 and k_6 involve two subjects treated with Tgf-β1. The time point t_1 either indicates a short-term response or 24 hours, and t_2 indicates either a long-term response or 48 hours. The symbol "X" in each condition and time point indicates that the condition or time point is included in the tricluster. The size in the table indicates the number of genes, conditions, or time points included in a tricluster. The table shows that the TQI value for each generated tricluster is quite small, ranging from 6.961×10^{-8} to 1.805×10^{-7}.

The purpose of conducting a triclustering analysis on this data is to find groups of genes that respond well to the administration of Egr-1 and Tgf-β1. Triclusters containing these gene groups are marked by having four subject conditions treated with Egr-1 and Tgf-β1, namely k_3, k_4, k_5, and k_6 only. Based on Table 12, tricluster 12 contains a group of genes that show a significant response to the administration of Egr-1 and Tgf-β1 in primary fibroblast cell cultures. This tricluster can be used as a reference for further research on the role of Egr-1 and Tgf-β1 in controlling the fibroblast cell activity. The response shown by these gene groups occurs at both short (24 hours) and long (48 hours) time points. This

indicates that Egr-1 and Tgf-β1 can be effective at all observed time points.

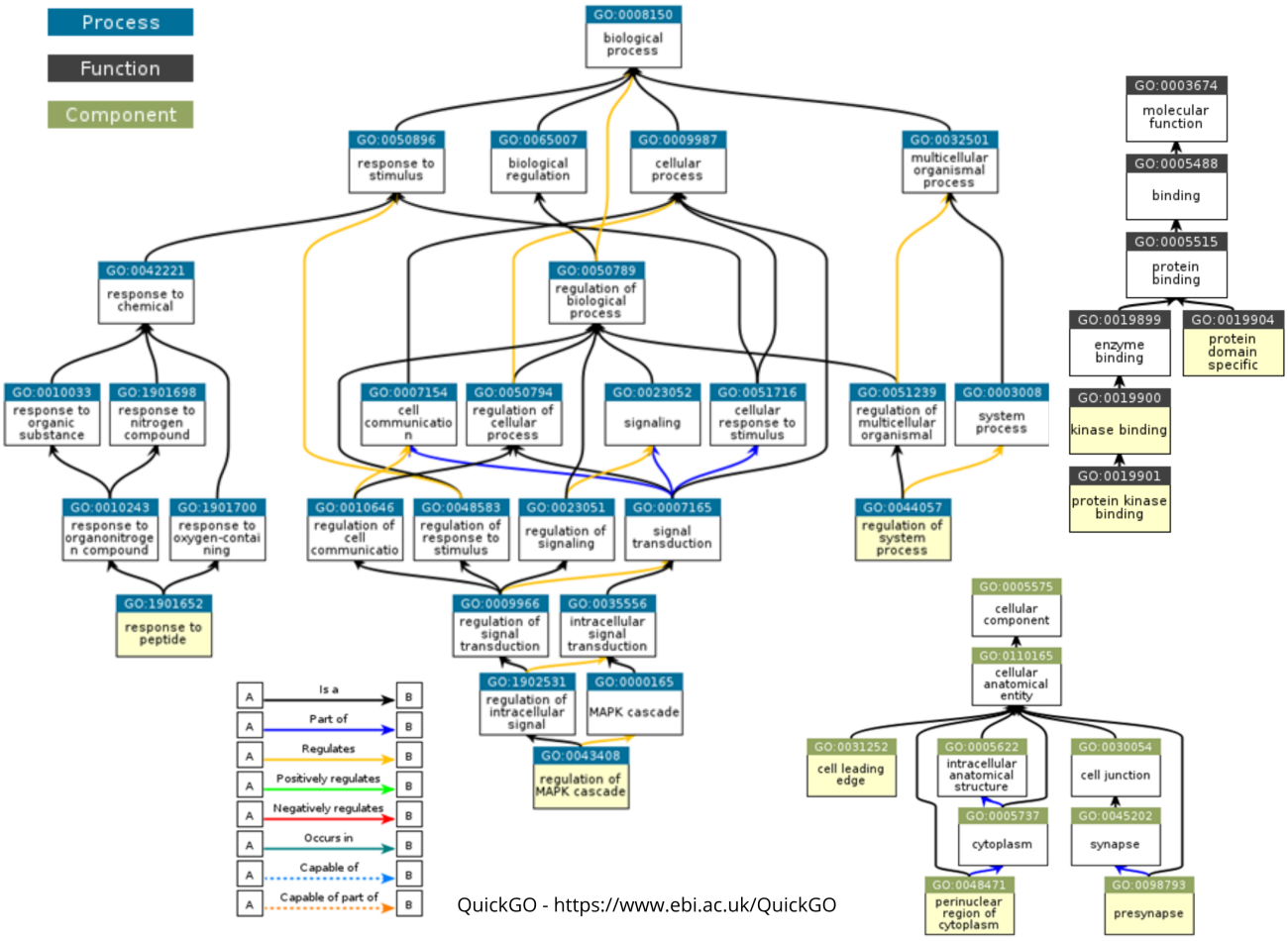

Figure 2. DAG of Selected GO Terms for GSE27165 dataset, generated with QuickGO (EMBL-EBI).

Tricluster 12 contains 13,755 genes, and the GO analysis was conducted on three aspects: *biological process*, *molecular function*, and *cellular component*. Each of these GO aspects was tested using Fisher's exact test with a significance level of 0.05 and corrected with the false discovery rate. There were 219 GO Terms from the *biological process* aspect, 28 GO Terms from the *molecular function* aspect, and 52 GO Terms from the *cellular component* aspect that were significant. The directed acyclic graph (DAG) of 3 GO Terms from each aspect with the highest fold enrichment is shown in Figure 2 [26].

As summarized in Table 13, the KEGG pathway annotation of genes extracted from the most prominent tricluster identified by the δ-Trimax triclustering method with the FCS based on a Gaussian distribution reveals extensive involvement in immune regulation, oncogenic signaling, and metabolic pathways. One of the most functionally pleiotropic genes in the cluster, MDM2, is annotated across more than twenty KEGG pathways, including key cancer-related and regulatory pathways such as the

p53 signaling pathway (hsa04115), PI3K-Akt signaling pathway (hsa04151), and FoxO signaling pathway (hsa04068), as well as diverse virus-related and tumor-specific pathways (e.g., hsa05203, hsa05214, hsa05220). Similarly, SYK demonstrates a broad connectivity with immune signaling mechanisms, thereby appearing in B cell receptor signaling (hsa04662), Fc receptor pathways (hsa04664, hsa04666), and natural killer cell-mediated cytotoxicity (hsa04650), thus underscoring its role in both innate and adaptive immune responses. Other immune-relevant genes include INPP5D and NLRC4, which are linked to phosphatidylinositol signaling and NOD-like receptor pathways, respectively, while RUNX1 is involved in hematologic malignancies such as acute and chronic myeloid leukemia (hsa05220, hsa05221). Additional genes, such as ABCC10 and CYP4F3, are associated with the ABC transporter (hsa02010) and arachidonic acid metabolism (hsa00590) pathways, thus indicating a role in cellular detoxification and inflammatory lipid signaling. Overall, the pathway enrichment pattern in Table 13 points to a co-regulated gene module implicated in inflammation, immune dysfunction, and tumor-related processes, which are potentially relevant to the molecular underpinnings of coronary heart disease represented in the GSE35671 dataset.

Table 12. The most optimal tricluster set for GSE27165 dataset. The tricluster selected for Gene Ontology analysis is shown in bold.

Tricluster	**Size (Genes × Conditions × Time points)**	**TQI**	**Condition**						**Time points**	
			k_1	k_2	k_3	k_4	k_5	k_6	t_1	t_2
1	$16841 \times 6 \times 2$	6.961×10^{-8}	X	X	X	X	X	X	X	X
2	$13991 \times 6 \times 2$	9.146×10^{-8}	X	X	X	X	X	X	X	X
3	$13984 \times 6 \times 2$	9.611×10^{-8}	X		X	X		X	X	X
4	$14173 \times 4 \times 2$	9.326×10^{-8}	X		X	X		X	X	X
5	$13960 \times 6 \times 2$	9.236×10^{-8}	X	X	X	X	X	X	X	X
6	$14008 \times 5 \times 2$	1.218×10^{-7}		X	X	X	X	X	X	X
7	$13839 \times 4 \times 2$	1.049×10^{-7}	X	X	X	X			X	X
8	$13969 \times 6 \times 2$	9.166×10^{-8}	X	X	X	X	X	X	X	X
9	$13709 \times 5 \times 1$	1.596×10^{-7}	X	X		X	X	X	X	
10	$13735 \times 4 \times 2$	1.328×10^{-7}	X			X	X	X	X	X
11	$13833 \times 5 \times 2$	1.153×10^{-7}	X	X	X		X	X	X	X
12	$\mathbf{13755 \times 4 \times 2}$	$\mathbf{1.277 \times 10^{-7}}$			**X**	**X**	**X**	**X**	**X**	**X**
13	$13935 \times 4 \times 1$	1.481×10^{-7}		X	X	X		X	X	
14	$13791 \times 6 \times 2$	9.216×10^{-8}	X	X	X	X	X	X	X	X
15	$13859 \times 4 \times 1$	1.433×10^{-7}	X	X	X	X				X
16	$13790 \times 5 \times 2$	1.168×10^{-7}	X	X	X		X	X	X	X
17	$13968 \times 4 \times 1$	1.805×10^{-7}	X	X		X		X		X
18	$13948 \times 5 \times 2$	1.328×10^{-7}	X	X	X	X	X		X	
19	$13809 \times 3 \times 2$	1.769×10^{-7}		X	X			X	X	X
20	$13845 \times 5 \times 1$	1.154×10^{-7}	X	X	X	X		X	X	
21	$13849 \times 5 \times 2$	1.037×10^{-7}	X	X	X	X	X		X	X
22	$13881 \times 5 \times 2$	1.204×10^{-7}	X	X		X	X	X	X	X
23	$13889 \times 4 \times 2$	1.610×10^{-7}		X	X		X	X	X	X
24	$13910 \times 4 \times 1$	1.515×10^{-7}	X	X		X		X	X	

4. Discussion

This study presents a hybrid triclustering framework that combines the δ-Trimax method with FCS optimization, thereby leveraging two distinct randomization strategies—the Lévy flight and Gaussian

Table 13. Annotated KEGG pathways of genes in tricluster from GSE35671 identified by δ-Trimax triclustering with FCS based on Gaussian distribution.

Probe_Id	Entrez_Gene_ID	Gene Name	KEGG Pathway(s)
ILMN_1776119	89845	ABCC10	hsa02010
ILMN_1814208	4193	MDM2	hsa01522, hsa01524, hsa04068, hsa04110, hsa04115, hsa04120, hsa04144, hsa04151, hsa04218, hsa04625, hsa04919, hsa05131, hsa05163, hsa05165, hsa05169, hsa05200, hsa05202, hsa05203, hsa05205, hsa05206, hsa05214, hsa05215, hsa05218, hsa05219, hsa05220
ILMN_1744212	3635	INPP5D	hsa00562, hsa01100, hsa04070, hsa04662, hsa04664, hsa04666
ILMN_1787518	2934	GSN	hsa04666, hsa04810, hsa05203
ILMN_1796976	58484	NLRC4	hsa04621, hsa05131, hsa05132, hsa05134, hsa05135
ILMN_2059549	6850	SYK	hsa04064, hsa04072, hsa04151, hsa04380, hsa04611, hsa04613, hsa04625, hsa04650, hsa04662, hsa04664, hsa04666, hsa05152, hsa05167, hsa05168, hsa05169, hsa05171, hsa05203
ILMN_2331163	8451	CUL4A	hsa03420, hsa04120, hsa05170
ILMN_2089484	4051	CYP4F3	hsa00590, hsa01100
ILMN_1730797	861	RUNX1	hsa04530, hsa04659, hsa05200, hsa05202, hsa05220, hsa05221
ILMN_1774739	4323	MMP14	hsa04668, hsa04912, hsa04928
ILMN_1689160	64174	DPEP2	–

distribution. To our knowledge, this is the first work to integrate δ-Trimax with a dual-strategy metaheuristic search in the context of temporal and a condition-specific gene expression analysis. The proposed approach not only surpasses existing improved CS algorithms in tricluster quality (as measured by a TQI), but also achieves competitive computational efficiency across two biologically distinct datasets. By capturing both localized, highly coherent modules and broader pathway-enriched clusters, this framework addresses a critical gap in a transcriptomic analysis where static clustering often overlooks condition- and time-dependent gene regulation.

The application of the integrated δ-Trimax triclustering algorithm with FCS optimization, using both the Lévy flight and Gaussian distribution, enabled the identification of biologically meaningful gene-condition-time clusters across two distinct datasets: heart-disease–related (GSE35671) and fibroblast stimulation (GSE27165). The comparative use of the Lévy flight and Gaussian-based randomization strategies within the FCS provided valuable insights into the influence of search dynamics on the quality and interpretability of the resulting triclusters.

In both datasets, the implementation of the Lévy flight-based search strategy yielded triclusters enriched in well-characterized signaling pathways. For instance, in GSE27165, several Ras- and MAPK-associated genes were clustered together, including *TIAM1*, *RASGRF2*, *NTRK1*, and *RASGRP4*, all of which are central to signal transduction and immune regulation. Similarly, in GSE35671, the Lévy-based approach identified co-regulated genes involved in autophagy, the cellular stress response, and vesicle-mediated transport, such as *SQSTM1* and *PRKCSH*. These findings underscore the method's effectiveness in capturing condition- and time-specific regulatory modules relevant to the inflammatory and metabolic processes.

Conversely, the Gaussian-distributed FCS variant produced triclusters with a broader pathway coverage, particularly in GSE35671. Genes such as *MDM2* and *SYK* prominently appeared, thus contributing to diverse oncogenic, immune, and virus-associated pathways, including p53, PI3K-Akt,

B cell receptor, and viral carcinogenesis. This suggests that the Gaussian-based search may favor the inclusion of multifunctional hub genes involved in multiple pathways, thereby enhancing the interpretability in the context of complex diseases such as coronary heart disease. Notably, the same method applied to GSE27165 revealed coherent clusters enriched in Ras/MAPK signaling, ABC transporters, and cell adhesion pathways, thus aligning with key mechanisms underlying cardiovascular inflammation and remodeling.

The complementary results obtained across the two optimization strategies and datasets demonstrate the robustness of the integrated δ-Trimax–FCS framework. The Lévy flight tends to promote a deeper exploration of localized clusters with strong biological coherence, while the Gaussian-distributed search tends to capture broader, pathway-enriched modules with a higher gene overlap. Together, these approaches provide a flexible and powerful means to extract biologically relevant triclusters in transcriptomic studies, particularly those that involve temporally and conditionally structured data such as cardiovascular disease expression profiles. These findings align with recent studies that highlighted the advantages of hybrid metaheuristic–clustering approaches in capturing both localized and broad biological patterns in high-dimensional data [18]. Furthermore, our findings are consistent with recent reports where hybrid metaheuristics enhanced both the accuracy and computational feasibility on engineering-scale problems [2, 3].

Table 14. Comparison of TQI values between the proposed method and other improved CS baselines across both datasets. Lower TQI indicates better tricluster quality.

Method	GSE35671 (TQI)	GSE27165 (TQI)
δ-Trimax + FCS (Gaussian)	1.01823×10^{-7}	1.22454×10^{-7}
δ-Trimax + FCS (Lévy flight)	1.50450×10^{-7}	1.33142×10^{-7}
Enhanced CS [13]	2.10×10^{-7}	1.85×10^{-7}
Improved CS [25]	1.95×10^{-7}	1.77×10^{-7}
CS + FCM [9]	2.35×10^{-7}	2.02×10^{-7}

As summarized in Table 14, both variants of the proposed δ-Trimax + FCS method outperformed three popular baselines: Enhanced CS [13], Improved CS [25], and CS + FCM [9]. The Gaussian-based variant consistently achieved the lowest TQI on both datasets, while the Lévy flight variant remained competitive and superior to the baselines. This performance gain demonstrates the benefit of combining δ-Trimax [5] with the optimization capabilities of the FCS [6, 19] to refine the search space and enhance the biological relevance. All baselines were re-implemented with matched stopping criteria and evaluated on the same workstation.

Table 15. Average running time of the proposed and other improved CS methods across datasets.

Method	GSE35671	GSE27165
δ-Trimax + FCS (Gaussian)	17 min	16 min
δ-Trimax + FCS (Lévy flight)	20 min	19 min
Enhanced CS [13]	24 min	22 min
Improved CS [25]	22 min	20 min
CS + FCM [9]	26 min	24 min

Beyond accuracy, the proposed methods were also efficient (Table 15). The Gaussian-based variant was the fastest, followed by the Lévy flight, while the three baselines (Enhanced CS [13], Improved CS [25], and CS + FCM [9]) took longer under matched criteria. The runtime difference between the

two proposed variants can be attributed to the exploratory nature of the Lévy flight, which performs a broader global search compared to the more localized Gaussian strategy. These results underscore that the proposed framework not only improves the tricluster quality but also reduces the computational cost, thus making it suitable for large-scale transcriptomic studies.

While several identified pathways, such as Ras/MAPK and PI3K–Akt signaling, are well-documented in cardiovascular disease, the contribution of our method lies in uncovering co-expression structures that vary according to specific conditions and time points. For example, the grouping of *MDM2* and *SYK* with other immune-regulatory genes under specific conditions suggests coordinated transcriptional regulation linked to immune modulation and the endothelial response. Although these pathways are not entirely novel, our approach reveals new combinatorial gene sets and temporal dynamics not previously reported in these datasets, thus offering a richer context for hypothesis generation and experimental validation. Future work may extend this approach to multiomics integration and cross-species transcriptomic analyses to further validate its generalizability.

5. Conclusions

This study presented an integrated optimization framework for triclustering three-dimensional gene expression data by combining the δ-Trimax method with an enhanced FCS algorithm utilizing both the Lévy flight and Gaussian distribution strategies. The hybrid approach achieved substantial improvements in identifying coherent gene-condition-time modules within complex biological datasets while maintaining the computational efficiency.

Applied to two gene expression datasets, one associated with human heart disease and the other with primary fibroblast cells, the framework successfully extracted biologically meaningful triclusters characterized by consistent expression dynamics across multiple conditions and time points. The Gaussian-based variant achieved the lowest TQI values for both datasets, whereas the Lévy flight variant excelled in uncovering deeply coherent clusters in the heart disease dataset.

In the heart disease dataset, optimal results were obtained with the FCS–Lévy flight variant ($\delta = 0.026$, $\theta = 1.7$), thereby producing the most coherent tricluster (TQI = 1.50450×10^{-7}) that contained 24,363 genes consistently expressed across three experimental conditions and ten time points, excluding baseline (day 0). For the primary fibroblast dataset, the Gaussian-based variant achieved an average TQI of 1.2245×10^{-7}, thereby capturing sustained gene responses to Egr-1 and Tgf-β1 stimulation over both 24 h and 48 h intervals.

A GO enrichment analysis confirmed the biological relevance of the identified clusters, with significant terms spanning biological processes, molecular function, and cellular component, all validated using Fisher's exact test with an FDR correction at a 0.05 threshold. Compared to existing improved CS algorithms, this is the first application of δ-Trimax combined with dual FCS search strategies to temporal gene expression data, thus achieving both a superior tricluster quality and a reduced runtime. The framework offers a valuable tool to extract temporally and conditionally coherent gene modules, thus providing mechanistic insights into gene regulation in cardiovascular and fibrotic diseases and paving the way for future studies on biomarker discovery and therapeutic target identification.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgment

The authors gratefully acknowledge the financial support from the Hibah Publikasi Terindeks Internasional (PUTI) Q2 for the fiscal year 2024-2025, Grant Number: NKB-688/UN2.RST/HKP.05.00/2024, provided by Universitas Indonesia. Special thanks are extended to Assoc. Prof. Agus Salim (University of Melbourne) for his invaluable contributions.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. R. Wang, L. Li, M. Chen, X. Li, Y. Liu, Z. Xue, et al., Gene expression insights: Chronic stress and bipolar disorder: A bioinformatics investigation, *Math. Biosci. Eng.*, **21** (2024), 392–414, https://doi.org/10.3934/mbe.2024018
2. L. Nguyen-Ngoc, Q. Nguyen-Huu, G. de Roeck, T. Bui-Tien, M. Abdel-Wahab, Deep neural network and evolved optimization algorithm for damage assessment in a truss bridge, *Mathematics*, **12** (2024), 2300. https://doi.org/10.3390/math12152300
3. Y. Li, H.-L. Minh, M. Cao, X. Qian, M. A. Wahab, An integrated surrogate model-driven and improved termite life cycle optimizer for damage identification in dams, *Mech. Syst. Signal Process.*, **208** (2024), 110986, https://doi.org/10.1016/j.ymssp.2023.110986
4. N. Narmadha, R. Rathipriya, Greedy two way K-means clustering for optimal coherent tricluster, *Int. J. Sci. Technol. Res.*, **8** (2019), 10.
5. T. Siswantining, N. Saputra, D. Sarwinda, H. S. Al-Ash, Triclustering discovery using the δ-Trimax method on microarray gene expression data, *Symmetry*, **13** (2021), 437. https://doi.org/10.3390/sym13030437
6. X. S. Yang, S. Deb, Cuckoo search via Lévy flights, in *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, IEEE, 2009, 210–214, https://doi.org/10.1109/NABIC.2009.5393690
7. S. Palaniswamy, P. Kandhasamy, Rough fuzzy cuckoo search for triclustering microarray gene expression data, *Turk. J. Electr. Eng. Comput. Sci.s*, **27** (2019), 4328–4339, https://doi.org/10.3906/elk-1809-86
8. S. Ghosh, S. K. Dubey, Comparative analysis of K-means and fuzzy C-means algorithms, *Int. J. Adv. Comput. Sci. Appl.*, **4** (2013), 35–39. https://doi.org/10.14569/IJACSA.2013.040406
9. L. Khrissi, N. El Akkad, H. Satori, K. Satori, An efficient image clustering technique based on fuzzy C-means and cuckoo search algorithm, *Int. J. Adv. Comput. Sci. Appl.*, **12** (2021), 423–432. https://doi.org/10.14569/IJACSA.2021.0120647

10. T. Siswantining, A. Bustamam, D. Sarwinda, S. M. Soemartojo, M. A. Latief, E. A. Octaria, et al., Triclustering method for finding biomarkers in human immunodeficiency virus-1 gene expression data, *Math. Biosci. Eng.*, **19** (2022), 6743–6763. https://doi.org/10.3934/mbe.2022318

11. T. Siswantining, M. A. S. Istianingrum, S. M. Soemartojo, D. Sarwinda, N. Saputra, S. Pramana, Triclustering implementation using hybrid δ-trimaxparticle swarm optimization and gene ontology analysis on three-dimensional gene expression data, *Mathematics*, **11** (2023), 4219, https://doi.org/10.3390/math11194219

12. A. Bhar, M. Haubrock, A. Mukhopadhyay, U. Maulik, S. Bandyopadhyay, E. Wingender, δ-TRIMAX: Extracting triclusters and analysing coregulation in time series gene expression data, in *Algorithms in Bioinformatics*, Springer, Berlin, Heidelberg, (2012), 165–177. https://doi.org/10.1007/978-3-642-33122-0_13

13. A. M. Kamoona, J. C. Patra, A. Stojcevski, An enhanced cuckoo search algorithm for solving optimization problems, in *2018 IEEE Congress on Evolutionary Computation (CEC)*, (2018), 1–6, https://doi.org/10.1109/CEC.2018.8477784

14. A. Y. Gital, M. Hamada, K. Haruna, M. Hassan, F. Shittu, S. Y. Ilu, et al., Hybrid of cuckoo search algorithm with Lévy flight and neural network for crude oil prices prediction, *J. Comput. Theor. Nanosci.*, **16** (2019), 4092–4104. https://doi.org/10.1166/jctn.2019.8611

15. M. A. Al-Abaji, Cuckoo search algorithm: Review and its application, *Tikrit J. Pure Sci.*, **26** (2021), 137–144. https://doi.org/10.25130/tjps.v26i2.130

16. C. Ouyang, X. Liu, D. Zhu, Y. Zheng, C. Zhou, C. Zou, A multi-strategy hybrid cuckoo search algorithm with specular reflection based on a population linear decreasing strategy, *Int. J. Mach. Learn. Cybern.*, **15** (2024), 5683–5723. https://doi.org/10.1007/s13042-024-02273-6

17. R. Salgotra, S. Singh, P. Verma, L. Abualigah, A. H. Gandomi, Mutation adaptive cuckoo search hybridized naked mole rat algorithm for industrial engineering problems, *Sci. Rep.*, **15** (2025), 19655. https://doi.org/10.1038/s41598-025-01033-y

18. D. F. Soares, R. Henriques, S. C. Madeira, Comprehensive assessment of triclustering algorithms for three-way temporal data analysis, *Pattern Recognit.*, **150** (2024), 110303. https://doi.org/10.1016/j.patcog.2024.110303

19. S. Palaniswamy, K. Premalatha, TrioCuckoo: A multi objective cuckoo search algorithm for triclustering microarray gene expression data, *J. Inf. Sci. Eng.*, **34** (2018), 1617–1631. https://doi.org/10.6688/JISE.201811_34(6).0014

20. M. Pomaznoy, B. Ha, B. Peters, GOnet: A tool for interactive gene ontology analysis, *BMC Bioinf.*, **19** (2018), 1–8. https://doi.org/10.1186/s12859-018-2533-3

21. A. Alexa, J. Rahnenfuhrer, F. Marini, Gene set enrichment analysis with TopGO, (2019). Available from: https://bioconductor.statistik.tu-dortmund.de/packages/3.9/bioc/vignettes/topGO/inst/doc/topGO.pdf.

22. S. Bhattacharyya, F. Fang, W. Tourtellotte, J. Varga, Egr-1: New conductor for the tissue repair orchestra directs harmony (regeneration) or cacophony (fibrosis), *J. Pathol.*, **229** (2013), 286–297. https://doi.org/10.1002/path.4131

23. W. A. N. Daeng, T. Siswantining, A. Bustamam, P. Anki, δ-TRIMAX method with silhouette coefficient on microarray gene expression data for early detection of heart failure, in *2022 5th International Conference on Information and Communications Technology (ICOIACT)*, (2022), 412–416, https://doi.org/10.1109/ICOIACT55506.2022.9971865

24. H. Zheng, Y. Zhou, A novel cuckoo search optimization algorithm based on Gauss distribution, *J. Comput. Inf. Syst.*, **8** (2012), 4193–4200.

25. J. Liu, M. Zeng, Y. Ge, C. Wu, X. Wang, Improved Cuckoo Search algorithm for numerical function optimization, *J. Ind. Manage. Optim.*, **16** (2020), 103–115. https://doi.org/10.3934/jimo.2018142

26. EMBL-EBI, *QuickGO: Gene Ontology Browser*. Available from: https://www.ebi.ac.uk/QuickGO/.

AIMS Press