

Article

# Triclustering Implementation Using Hybrid $\delta$ -Trimax Particle Swarm Optimization and Gene Ontology Analysis on Three-Dimensional Gene Expression Data

Titin Siswantining <sup>1,\*</sup>, Maria Armelia Sekar Istianingrum <sup>1</sup>, Saskya Mary Soemartojo <sup>1</sup>, Devvi Sarwinda <sup>1</sup>, Noval Saputra <sup>1</sup>, Setia Pramana <sup>2</sup> and Rully Charitas Indra Prahmana <sup>3</sup>

<sup>1</sup> Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok 16424, Indonesia

<sup>2</sup> Politeknik Statistika STIS, Jakarta 13330, Indonesia; setia.pramana@stis.ac.id

<sup>3</sup> Mathematics Education Department, Universitas Ahmad Dahlan, Yogyakarta 55166, Indonesia; rully.indra@mpmat.uad.ac.id

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

**Abstract:** Triclustering is a data mining method for grouping data based on similar characteristics. The main purpose of a triclustering analysis is to obtain an optimal tricluster, which has a minimum mean square residue (MSR) and a maximum tricluster volume. The triclustering method has been developed using many approaches, such as an optimization method. In this study, hybrid  $\delta$ -Trimax particle swarm optimization was proposed for use in a triclustering analysis. In general, hybrid  $\delta$ -Trimax PSO consist of two phases: initialization of the population using a node deletion algorithm in the  $\delta$ -Trimax method and optimization of the tricluster using the binary PSO method. This method, when implemented on three-dimensional gene expression data, proved useful as a Motexafin gadolinium (MGd) treatment for plateau phase lung cancer cells. For its implementation, a tricluster that potentially consisted of a group of genes with high specific response to MGd was obtained. This type of tricluster can then serve as a guideline for further research related to the development of MGd drugs as anti-cancer therapy.



**Citation:** Siswantining, T.; Istianingrum, M.A.S.; Soemartojo, S.M.; Sarwinda, D.; Saputra, N.; Pramana, S.; Prahmana, R.C.I. Triclustering Implementation Using Hybrid  $\delta$ -Trimax Particle Swarm Optimization and Gene Ontology Analysis on Three-Dimensional Gene Expression Data. *Mathematics* **2023**, *11*, 4219. <https://doi.org/10.3390/math11194219>

Received: 5 June 2023

Revised: 21 July 2023

Accepted: 11 August 2023

Published: 9 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** mean square residue; optimization; triclustering quality index; microarray

**MSC:** 68T45

## 1. Introduction

Over time, technology has developed increasingly quickly. The outputs produced from such technologies are also becoming increasingly complex, so the term “big data” has appeared. According to Riahi et al. [1], big data refers to a high volume of data produced at a high velocity and with great variety. Therefore, data have become an essential and valuable thing that can be analyzed to determine historical patterns, to observe the characteristics or tendencies of cases, or even to predict future trends. Then, the term “data mining” appeared, which refers to the process of extracting valuable and exciting information from data that can be interpreted and implemented [2].

One of the mostly commonly used data mining techniques is the clustering technique. The unsupervised-learning technique is one such technique and analyzes data without defining a target or outcome. Clustering can group data based on one dimension, that is, based on its rows or based on its columns. Because of this limitation, the biclustering method was developed, which can group data based on two dimensions, both its rows and columns, simultaneously [3]. The output obtained from the biclustering method is biclusters, a submatrix from an initial 3D matrix, also called subspace clustering.

In fact, the development of data is not limited to two dimensions but can be up to three dimensions or even more, so a more complex method is needed for its analysis.

One of the analytical methods used for three-dimensional data is the triclustering method. Triclustering is a derivation of clustering and biclustering, in which this method can simultaneously group data based on three dimensions based on its observations, attributes, and contexts. The output obtained from the triclustering method is triclusters, which are also submatrices from the initial 3D matrix data. The triclustering method was first introduced by Zhao and Zaki [4], with TriCluster as the method name. After that, another triclustering method was developed from another approach.

Bhar et al. [5] used  $\delta$ -Trimax for a triclustering analysis, which included the greedy approach and had deletion and addition nodes in its method. It obtained triclusters with a mean square residue (MSR) value below the  $\delta$  threshold. This method was then rediscussed by Siswantining et al. [6,7] and improved. Siswantining et al. [8,9] also developed several other approaches, including an order-preserving tricluster using similar gene expression patterns at the same time points and under the same conditions and the THD tricluster using a biclustering approach. Over time, other optimization methods (i.e., particle swarm optimization (PSO), cuckoo search (CS), etc.) have also been used for triclustering analyses.

SwathyPriyadharsini et al. [3] used the PSO method for a triclustering analysis with a random initial population and a minimum MSR value as the goals of the fitness function. Then, Narmadha et al. [10] proposed an improvement method, hybrid PSO, which combines greedy two-way K-means for initializing the population and binary PSO for optimizing the tricluster with a maximum tricluster volume as the goals of the fitness function.

Bioinformatics is one field that implements triclustering analyses, specifically genetics research. In the development of triclustering analyses, studies realized that it is inefficient to analyze thousands of genes using conventional methods [11], so a microarray technology was developed. This technology can store thousands of gene sequences and can measure gene expression values [12]. The output is two-dimensional gene expression data, in which rows represent the genes and columns represent the experimental conditions. In addition, three-dimensional gene expression data consisting of genes, experiment conditions, and time points can be obtained. From the three-dimensional data, a triclustering analysis is used to obtain a group of genes with significant responses to a subset of conditions and a subset of time points.

In this study, the proposed hybrid  $\delta$ -Trimax particle swarm optimization algorithm consists of two combinations: a node-deletion algorithm in the  $\delta$ -Trimax method for initializing the population and binary particle swarm optimization (binary PSO) for optimizing the triclusters. The tricluster volume is maximized and used as a fitness function for evaluation. The PSO method is used because it is easier to implement and sets better parameters than other optimization methods. In addition, a homogeneous initial population tends to produce better results than a random initial population [13].

The hybrid  $\delta$ -Trimax PSO method is implemented on three-dimensional gene expression data for chemotherapy drugs given to lung cancer cells. Cells were administered two treatments: 50  $\mu$ M Motexafin Gadolinium (MGd) and 5% mannitol (control group). Furthermore, observations were carried out for 4 h, 12 h, and 24 h after the drug was given. Using a triclustering analysis, a group of genes with a specific response to MGd was obtained, and this group can serve as a guideline for future research on MGd development used for anti-cancer therapy. However, this method requires a longer computation time than the previous method, so parallel processing should be studied in future research to improve its efficiency regarding computation time.

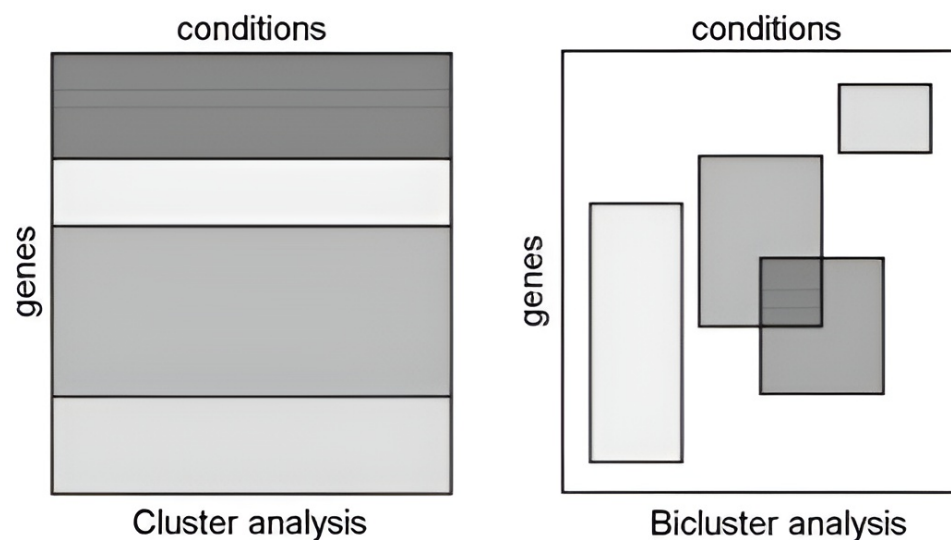
## 2. Materials and Methods

### 2.1. Clustering

Clustering is one of the mostly commonly used data mining techniques and is used in unsupervised learning; this method can analyze data without defining labels. The clustering method is used for grouping observations with highly similar characteristics. The clustering output is the clusters, in which observations in the same cluster have high similarity while observations in a different cluster have low similarity. Clustering can group based on one dimension, that is, based on its rows or columns. K-means clustering is a clustering method commonly used for grouping data based on the smallest Euclidean distance between the observations and the centroid (center of the cluster).

### 2.2. Biclustering

The biclustering method is a derivation of the clustering method that can simultaneously group data based on two dimensions, such as based on both its rows and columns, also called subspace clustering. The output of biclustering is the biclusters (a submatrix of the initial matrix data); a bicluster illustration can be seen in Figure 1. Biclusters also have overlapping characteristics, which means that it is possible to find rows or columns in some biclusters and that it is even possible that some rows or columns have not been categorized into any biclusters [14].



**Figure 1.** Illustration of clustering and biclustering [14].

### 2.3. Triclustering

Triclustering is a subsequent derivation of clustering and biclustering used to simultaneously group three-dimensional data based on its dimensions, observations, attributes, and contexts. This method is also included in subspace clustering, first introduced by Zhao and Zaki in 2005 [4]. Triclustering is used to obtain a set of triclusters; each tricluster has specific criteria that are homogeneous [15]. For example, given three-dimensional data,  $A = (X, Y, Z)$ , that consist of  $n$  observations,  $m$  attributes, and  $p$  context, where  $X = \{x_1, x_2, \dots, x_n\}$ ,  $Y = \{y_1, y_2, \dots, y_m\}$  and  $Z = \{z_1, z_2, \dots, z_p\}$ , the output of a triclustering analysis would be in the form of a tricluster  $B = (I, J, K)$ , which is a submatrix of the three-dimensional data in  $A$ , where  $I \subseteq X$ ,  $J \subseteq Y$  and  $K \subseteq Z$ . A tricluster illustration can be seen in Figure 2.

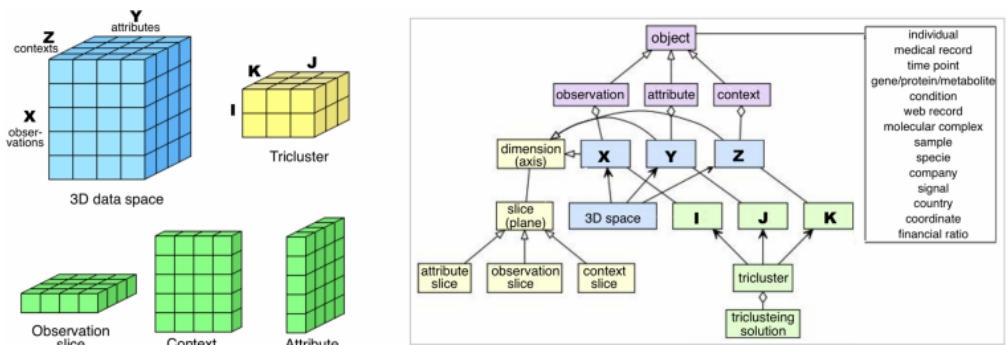


Figure 2. Illustration of triclustering method [15].

### 2.4. Gene Expression Data

Each cell in living things has a role, and the activities of each cell are stored in their genetic code, their deoxyribonucleic acid (DNA). The Central Dogma of Molecular Biology mentions that each cell produces one mRNA per protein, so the gene expression levels for each gene could differ. Microarray technology has also often been used for measuring gene expression value. The output of this technology is in the form of two- or even three-dimensional gene expression data. Generally, three-dimensional gene expression data consist of genes, conditions, and time points. “Time points” were included to observe how a cell develops under each condition over time [16].

### 2.5. Mean Square Residue

A mean square residue (MSR) is used to measure the quality of a tricluster. The concept came from the idea of a perfectly shifting tricluster, which refers to a tricluster with an MSR value equal 0. But, in fact, three-dimensional data, specifically gene expression data always have noise, which means that the MSR value is never equal to 0. Our main focus here is to find a tricluster with a minimal MSR. For example, given a tricluster  $M(I, J, K) = m_{ijk}$ , where  $i \in I, j \in J$  and  $k \in K$ , an MSR value or  $S$  can be obtained by calculating the difference between the actual value  $m_{ijk}$  and the estimate value  $\hat{m}_{ijk}$  using the following equation:

$$\begin{aligned}
 S &= \frac{1}{|I||J||K|} \sum_{i \in I, j \in J, k \in K} r_{ijk}^2 \\
 &= \frac{1}{|I||J||K|} \sum_{i \in I, j \in J, k \in K} (m_{ijk} - m_{iJK} - m_{iJk} - m_{iJk} + 2m_{iJK})^2
 \end{aligned}
 \tag{1}$$

### 2.6. $\delta$ -Trimax Method

The  $\delta$ -Trimax method is a derivation of the biclustering analysis that is able to find a tricluster with a mean square residue (MSR) value below the threshold  $\delta$  (the maximum threshold of the MSR value). This method is included in the greedy approach, in which there is a deletion and addition node algorithm. A node refers to a gene/condition/time point.

Generally,  $\delta$ -Trimax consist of three phases: multiple-node deletion, single-node deletion, and node addition. Multiple-node deletion and single-node deletion produce triclusters with MSR values below  $\delta$ , while node addition is used to optimize the tricluster volume [5,6].

Multiple-node deletion can delete nodes simultaneously suspected of causing the MSR to be below  $\delta$ . This algorithm must be used if the data consist of 50 genes/50 conditions/50 time points.

The  $i$ th  $\in I$  genes are deleted if they satisfy the following inequality:

$$\frac{1}{|J||K|} \sum_{j \in J, k \in K} (m_{ijk} - m_{iJK} - m_{iJk} - m_{iJk} + 2m_{iJK})^2 > \lambda S
 \tag{2}$$

The  $j$ th  $\in J$  experiment conditions are deleted if they satisfy the following inequality:

$$\frac{1}{|I||K|} \sum_{i \in I, k \in K} (m_{ijk} - m_{iJK} - m_{IjK} - m_{IJk} + 2m_{IJK})^2 > \lambda S \tag{3}$$

The  $k$ th  $\in K$  time points are deleted if they satisfy the following inequality:

$$\frac{1}{|I||J|} \sum_{i \in I, j \in J} (m_{ijk} - m_{iJK} - m_{IjK} - m_{IJk} + 2m_{IJK})^2 > \lambda S \tag{4}$$

Single-node deletion can delete nodes one by one if the tricluster produced from the previous algorithm (multiple-node deletion) still has an MSR  $< \delta$ . The gene/condition/time points are deleted if the node has the highest  $\mu$  score.

The  $\mu$  score for the  $i$ th  $\in I$  gene:

$$\mu(i) = \frac{1}{|J||K|} \sum_{j \in J, k \in K} (m_{ijk} - m_{iJK} - m_{IjK} - m_{IJk} + 2m_{IJK})^2 \tag{5}$$

The  $\mu$  score for the  $j$ th  $\in J$  gene:

$$\mu(j) = \frac{1}{|I||K|} \sum_{i \in I, k \in K} (m_{ijk} - m_{iJK} - m_{IjK} - m_{IJk} + 2m_{IJK})^2 \tag{6}$$

The  $\mu$  score for the  $k$ th  $\in K$  gene:

$$\mu(k) = \frac{1}{|I||J|} \sum_{i \in I, j \in J} (m_{ijk} - m_{iJK} - m_{IjK} - m_{IJk} + 2m_{IJK})^2 \tag{7}$$

Node addition can return nodes already deleted in the previous algorithm one by one. The  $i$ th  $\in I$  genes are added if they satisfy the following inequality:

$$\frac{1}{|J||K|} \sum_{j \in J, k \in K} (m_{ijk} - m_{iJK} - m_{IjK} - m_{IJk} + 2m_{IJK})^2 \leq \lambda S \tag{8}$$

The  $j$ th  $\in J$  genes are added if they satisfy the following inequality:

$$\frac{1}{|I||K|} \sum_{i \in I, k \in K} (m_{ijk} - m_{iJK} - m_{IjK} - m_{IJk} + 2m_{IJK})^2 \leq \lambda S \tag{9}$$

The  $k$ th  $\in K$  genes are added if they satisfy the following inequality:

$$\frac{1}{|I||J|} \sum_{i \in I, j \in J} (m_{ijk} - m_{iJK} - m_{IjK} - m_{IJk} + 2m_{IJK})^2 \leq \lambda S \tag{10}$$

### 2.7. Binary Particle Swarm Optimization (Binary PSO)

Particle swarm optimization (PSO) is a type of swarm intelligence method, where the behaviors of bird and fish in searching for food sources within its colony inspired this method. This method is also a meta-heuristic method, which means that it searches for the best solution in its iterations based on global and local searches.

The PSO method is used to find the solution with the optimal fitness function value (the definition of the fitness function is based on the case to be optimized). Instead of other optimization methods, the PSO method is easier to implement [13].

Some of the terms used in PSO method include the following: Swarm (also called population) is a group of particles. Each particle has two characteristics: position and velocity. These two characteristics direct each particle to the best solution. In each iteration, a particle is evaluated using a fitness function. Each particle has a "pbest", which means the particle's position has an optimal fitness function value. In addition, there is also

“gbest”, which means that the particle’s position has an optimal fitness function value within its neighborhood. This algorithm stops if the convergence condition is satisfied or if the maximum iterations are satisfied. The PSO method is illustrated in Figure 3.

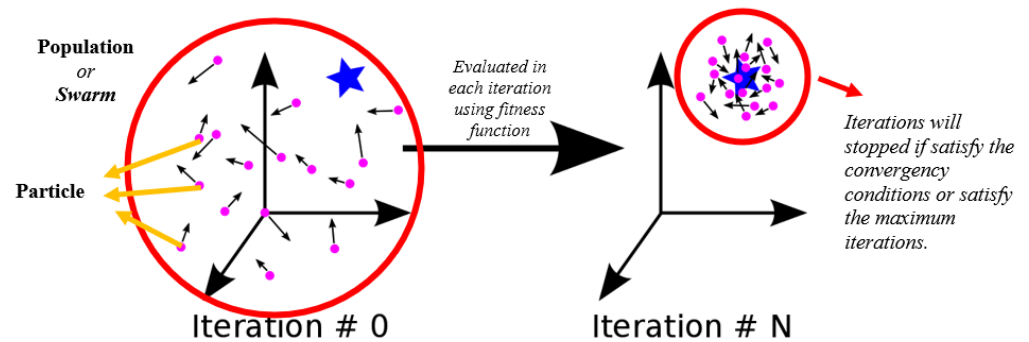


Figure 3. Illustration of the PSO method (source: [www.medium.com](http://www.medium.com) (accessed on 20 June 2021)).

Binary PSO is one of the PSO method’s variations, in which particles are defined as a binary vector with a value of either “1” or “0”, as illustrated in Figure 4.

	$x_{i1}$	$x_{i2}$	$x_{i3}$	$x_{i4}$	$x_{i5}$	$x_{i6}$	...	...	...	$x_{id}$
$X_1$ (Partikel 1)	1	0	0	1	1	0	...	...	...	1
$X_2$ (Partikel 2)	1	0	1	1	0	0	...	...	...	1
$X_3$ (Partikel 3)	0	1	0	1	0	0	...	...	...	1
⋮										
$X_N$ (Partikel N)	1	0	0	1	1	0	...	...	...	1

Figure 4. Illustration of particles in binary PSO.

A particle’s position is represented as a bit with a value of “1”. A particle’s velocity in binary PSO is defined as the probability of the bit having a value of “1”.  $\vec{V}_i = \{v_{i1}, v_{i2}, \dots, v_{id}\}$ ,  $i = 1, 2, \dots, N$  is the  $i$ th particle’s velocity, and the value is restricted by  $\pm v_{max}$ , so  $v_{id} \in [-v_{max}, v_{max}]$ . In each iteration, velocity and position updates are also performed using the following equation: To update the particle’s velocity:

$$v_{id}(t + 1) = \omega v_{id}(t) + c_1 r_1 (p_{id} - x_{id}(t)) + c_2 r_2 (p_{gd} - x_{id}(t)) \tag{11}$$

To obtain the velocity in the form of probability, we used a logistic transformation function with  $s(v_{id}) \in [0, 1]$ .

$$s(v_{id}) = \text{sigmoid}(v_{id}) = \frac{1}{1 + e^{-v_{id}}} \tag{12}$$

To update the particle’s position:

$$x_{id}(t + 1) = \begin{cases} 1, & \text{if rand} < s(v_{id}(t + 1)) \\ 0, & \text{if else} \end{cases} \tag{13}$$

The outputs of binary PSO are “pbest”, for each particle that already has a more optimal fitness function value, and “gbest”, which has the most optimal global fitness function value.

2.8. Triclustering Quality Index (TQI)

The triclustering quality index (TQI) is one of the measurements of a tricluster’s quality. The TQI value can be calculated using the following equation:

$$TQI_i = \frac{MSR_i}{volume_i} \tag{14}$$

where  $MSR_i$  is the mean square residue value from the  $i$ th tricluster and  $volume_i$  can be defined as  $|I_i| \times |J_i| \times |K_i|$ , where  $I_i$  is the number of genes,  $J_i$  is the number of conditions and  $K_i$  is the number of time points in the  $i$ th tricluster. A lower TQI indicates better quality of the tricluster.

3. Proposed Method

This section discusses hybrid  $\delta$ -Trimax particle swarm optimization, combining the node deletion algorithm in  $\delta$ -Trimax, for initializing the population, and binary PSO, for optimizing the triclusters. This method was used to find better-quality triclusters (with minimum MSR values and maximum tricluster volumes). The flowchart of hybrid  $\delta$ -Trimax can be seen in Figure 5.

The initial population was created using the node deletion algorithm in  $\delta$ -Trimax to find the initial triclusters with MSR values  $< \delta$ . The solution obtained from the PSO method depends on the initial population/seed. A homogeneous initial population produces better results than a random initial population [13]. This study only used the node deletion algorithm in  $\delta$ -Trimax to find the homogeneous initial triclusters, as these two algorithm already produced triclusters with an MSR  $< \delta$ .

The initial population that obtained from  $\delta$ -Trimax was then optimized using binary PSO. Triclusters are used as particles and converted to the binary vector (encoded tricluster) with value either “1” or “0”, which “1” represent that genes/conditions/time-points are included in the tricluster, while “0” represent that genes/conditions/time-points are not included in the tricluster. The bit length is the number of genes, conditions, and time points. Illustration can be seen in Table 1.

Table 1. Illustration of encoded tricluster.

$g_1$	$g_2$	$g_3$	...	...	...	$g_{n_g}$	$s_1$	$s_2$	...	...	...	$s_{n_s}$	$t_1$	$t_2$	$t_3$	...	...	$t_{n_t}$
1	1	0	...	...	...	0	1	0	...	...	...	1	0	0	1	...	...	1

The fitness function of binary PSO in this research is defined as the tricluster volume, which is maximized, and can be calculated using the following equation:

$$F(I, J, K) = \begin{cases} |I| \times |J| \times |K|, & \text{if } MSR < \delta \\ 0, & \text{if else} \end{cases} \tag{15}$$

In each iterations, there was an evaluation of the particle using a fitness function value. The particles were moved in each iterations when updating their velocity and position, which were calculated using Equations (11)–(13). This algorithm stops if the maximum iteration is already satisfied or the convergency condition is satisfied. The output of this method is “pbest” of each particle and “gbest” with the optimal global fitness function value.

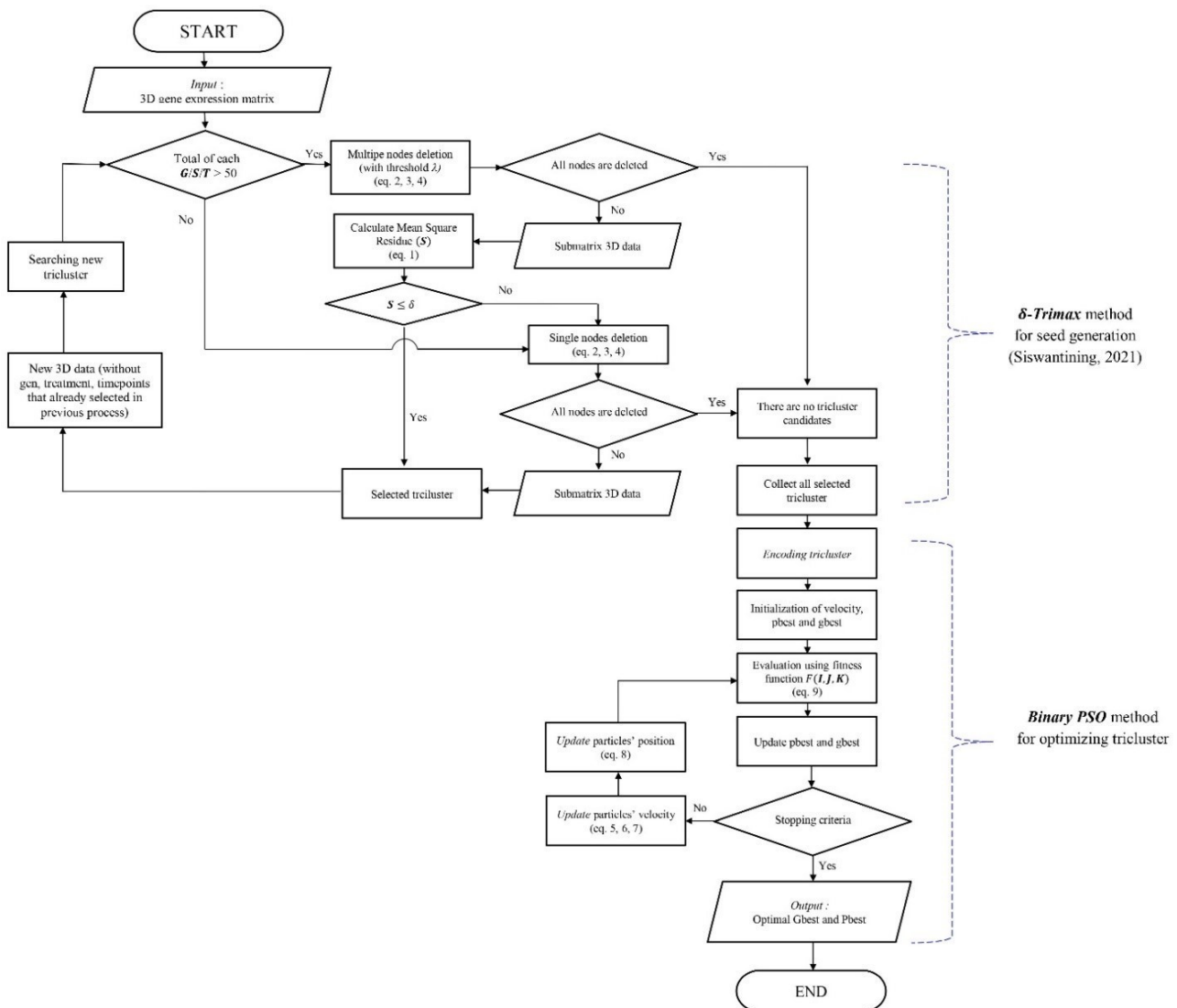


Figure 5. Flowchart of hybrid  $\delta$ -Trimax PSO algorithm [6].

## 4. Result and Discussion

### 4.1. Triclustering Result

Hybrid  $\delta$ -Trimax PSO was implemented on three-dimensional gene expression data with numeric data related to drug treatments using 50 uM Motexafin Gadolinium (MGd) for plateau phase lung cancer cells. These data were downloaded from the NCBI website at the following link: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2189> (accessed on 12 March 2021). The gene expression data consisted of 22,283 genes, which were divided into two different treatments—50 uM MGd and 5% mannitol (control group)—and observed for 4 h, 12 h, and 24 h after the drugs were administered. Each treatment had three observation units. The “time points” are included in this study to determine the short-term response, medium-term response, and long-term response of each drug treatment [17].



Several parameters were used in this triclustering analysis; the parameters used for  $\delta$ -Trimax were obtained based on Bhar et al. [5], while the parameters used in binary PSO were obtained based on Narmadha [13]. The used parameters are listed in Table 2:

**Table 2.** Parameters of hybrid  $\delta$ -Trimax PSO.

Parameters	Value
$\delta$	experimental (0.006; 0.005; and 0.004)
$\lambda$	1.2
$\omega_{max}$	experimental (0.9; 0.8; and 0.7)
$\omega_{min}$	experimental (0.4; 0.3; and 0.2)
$c_1, c_2$	2
$n$ particles	24
$v_{max}$	$\pm 6$
$i_{max}$	50 iterations
neighborhood	experimental (“gbest” and “lbest”)

$\omega_{max}$  and  $\omega_{min}$  in Table 2 are used as inertia weights for controlling the particle’s velocity when it was updated in each iteration using following equation:

$$\omega_i = \omega_{max} - \frac{i \times (\omega_{max} - \omega_{min})}{i_{max}} \tag{16}$$

Using some of the experimental parameters, some simulations were then conducted with each combination of experimental parameters. Each simulation was run three times. The optimal tricluster (gbest) obtained from each simulation was compared based on its average TQI value. A comparison for each simulation can be seen in Table 3.

From Table 3, an optimal tricluster is obtained from the simulation with the parameter combination  $\delta = 0.006$ ;  $\omega_{max} = 0.7$ ;  $\omega_{min} = 0.2$  and the neighborhood type “lbest”, with an average TQI score of  $3.671 \times 10^{-8}$ , which is the smallest TQI. From the best simulations, the characteristics of each tricluster and the subset of genes/conditions/time points in each tricluster can be analyzed. Here, we give the results of the best simulations: particle 23 is gbest with the minimum TQI (Table 4).

**Table 3.** Comparison of the gbest (global best position) results.

$\omega_{max}$	$\omega_{min}$	$\delta = 0.006$		$\delta = 0.005$		$\delta = 0.004$	
		<i>gbest</i>	<i>lbest</i>	<i>gbest</i>	<i>lbest</i>	<i>gbest</i>	<i>lbest</i>
0.9	0.4	$3.8515 \times 10^{-8}$	$3.7867 \times 10^{-8}$	$5.2025 \times 10^{-8}$	$5.1514 \times 10^{-8}$	$6.0009 \times 10^{-8}$	$5.9978 \times 10^{-8}$
0.9	0.3	$3.7570 \times 10^{-8}$	$3.7874 \times 10^{-8}$	$4.8542 \times 10^{-8}$	$5.0695 \times 10^{-8}$	$6.0412 \times 10^{-8}$	$5.7192 \times 10^{-8}$
0.9	0.2	$3.6750 \times 10^{-8}$	$3.7040 \times 10^{-8}$	$5.0139 \times 10^{-8}$	$5.0213 \times 10^{-8}$	$5.6433 \times 10^{-8}$	$5.6874 \times 10^{-8}$
0.8	0.4	$3.8297 \times 10^{-8}$	$3.7931 \times 10^{-8}$	$5.1107 \times 10^{-8}$	$5.1324 \times 10^{-8}$	$5.9724 \times 10^{-8}$	$5.9532 \times 10^{-8}$
0.8	0.3	$3.7504 \times 10^{-8}$	$3.7482 \times 10^{-8}$	$5.0696 \times 10^{-8}$	$5.0553 \times 10^{-8}$	$5.7328 \times 10^{-8}$	$5.8979 \times 10^{-8}$
0.8	0.2	$3.6737 \times 10^{-8}$	$3.6859 \times 10^{-8}$	$4.9522 \times 10^{-8}$	$5.0262 \times 10^{-8}$	$5.7495 \times 10^{-8}$	$5.9847 \times 10^{-8}$
0.7	0.4	$3.7629 \times 10^{-8}$	$3.7999 \times 10^{-8}$	$5.0486 \times 10^{-8}$	$5.1571 \times 10^{-8}$	$6.1041 \times 10^{-8}$	$5.8278 \times 10^{-8}$
0.7	0.3	$3.7644 \times 10^{-8}$	$3.7496 \times 10^{-8}$	$5.0574 \times 10^{-8}$	$5.0554 \times 10^{-8}$	$6.0552 \times 10^{-8}$	$5.9145 \times 10^{-8}$
0.7	0.2	$3.6864 \times 10^{-8}$	$3.6710 \times 10^{-8}$	$5.13752 \times 10^{-8}$	$4.9388 \times 10^{-8}$	$5.9451 \times 10^{-8}$	$5.6180 \times 10^{-8}$

**Table 4.** Pbest and gbest from the best simulation.

Particle	$ I  \times  J  \times  K $	TQI	Experiment Conditions						Time Points		
			$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$t_1$	$t_2$	$t_3$
1	$10,511 \times 4 \times 3$	$4.2872 \times 10^{-8}$	×		×	×		×	×	×	×
2	$10,714 \times 5 \times 2$	$5.4298 \times 10^{-8}$	×	×	×	×	×			×	×
3	$10,534 \times 4 \times 3$	$4.2832 \times 10^{-8}$	×		×	×		×	×	×	×
4	$10,631 \times 5 \times 3$	$3.7478 \times 10^{-8}$	×	×	×	×		×	×	×	×
5	$10,708 \times 3 \times 3$	$5.8190 \times 10^{-8}$	×	×	×				×	×	×
6	$10,276 \times 5 \times 3$	$3.8570 \times 10^{-8}$	×	×	×	×		×	×	×	×
7	$10,382 \times 5 \times 3$	$3.8292 \times 10^{-8}$	×	×	×	×		×	×	×	×
8	$10,572 \times 4 \times 3$	$4.2228 \times 10^{-8}$	×		×	×		×	×	×	×
9	$10,446 \times 4 \times 3$	$4.7691 \times 10^{-8}$	×	×	×			×	×	×	×
10	$10,562 \times 3 \times 3$	$5.9039 \times 10^{-8}$	×	×	×				×	×	×
11	$10,631 \times 4 \times 3$	$4.1879 \times 10^{-8}$	×		×	×		×	×	×	×
12	$10,606 \times 5 \times 2$	$5.6110 \times 10^{-8}$	×		×	×	×	×		×	×
13	$10,437 \times 5 \times 2$	$5.2518 \times 10^{-8}$	×	×	×	×		×	×		×
14	$10,589 \times 5 \times 2$	$5.1170 \times 10^{-8}$	×	×	×	×		×	×		×
15	$10,549 \times 5 \times 3$	$3.7315 \times 10^{-8}$	×	×	×	×		×	×	×	×
16	$10,554 \times 3 \times 3$	$5.0404 \times 10^{-8}$	×		×	×		×	×	×	×
17	$10,442 \times 5 \times 3$	$3.8075 \times 10^{-8}$	×	×	×	×		×	×	×	×
18	$10,667 \times 5 \times 2$	$5.0629 \times 10^{-8}$	×	×	×	×		×	×		×
19	$10,480 \times 3 \times 3$	$5.9310 \times 10^{-8}$	×	×				×	×	×	×
20	$10,513 \times 4 \times 3$	$4.4341 \times 10^{-8}$	×	×		×		×	×	×	×
21	$10,656 \times 4 \times 3$	$4.4152 \times 10^{-8}$	×	×	×	×			×	×	×
22	$10,482 \times 4 \times 3$	$4.7173 \times 10^{-8}$	×	×	×			×	×	×	×
23	$10,826 \times 5 \times 3$	$3.6489 \times 10^{-8}$	×	×	×	×		×	×	×	×
24	$10,315 \times 4 \times 3$	$4.4933 \times 10^{-8}$	×	×		×		×	×	×	×

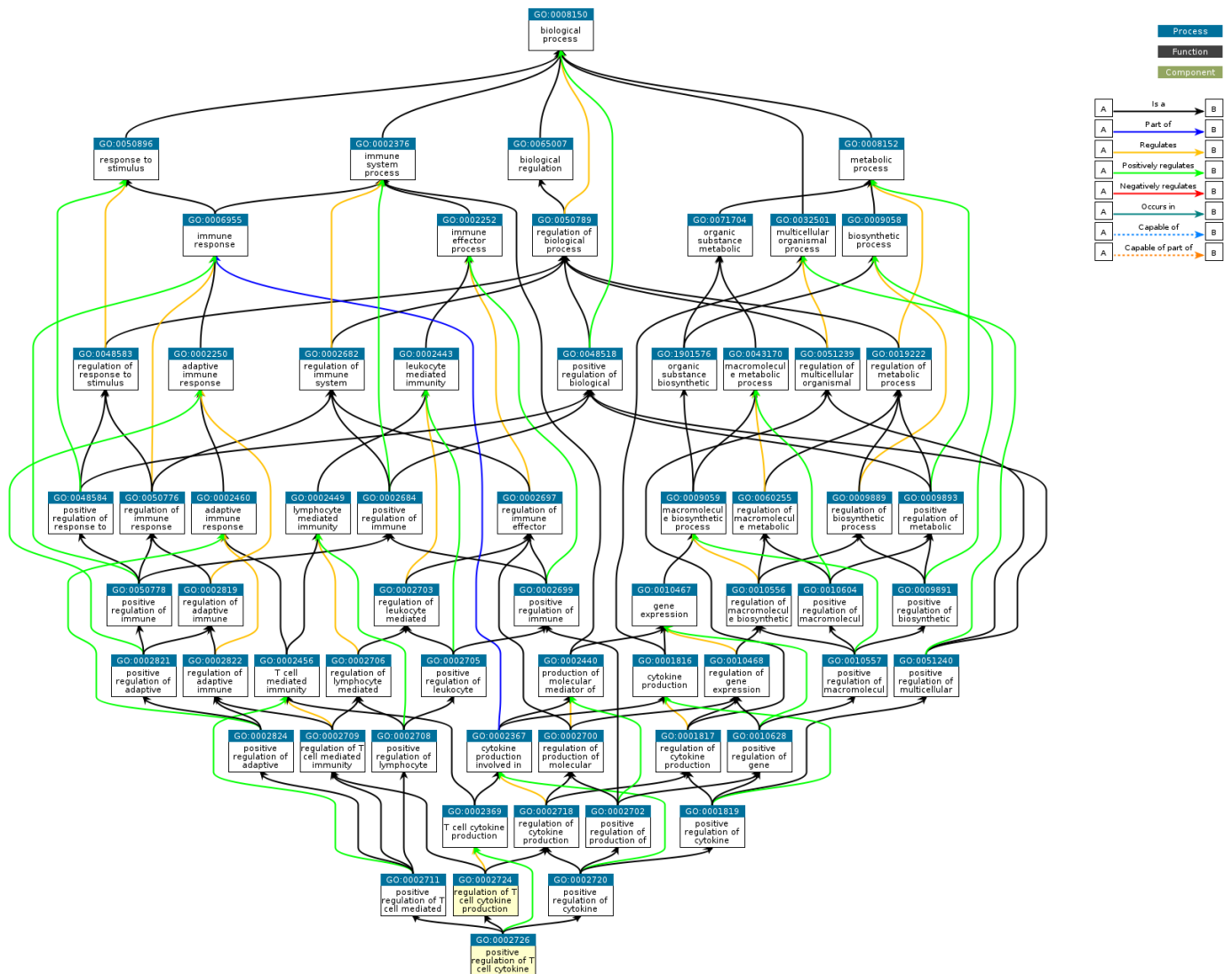
4.2. Gene Ontology Analysis Result

A gene ontology analysis was performed to obtain biological information from the optimal tricluster. A gene ontology analysis was performed on tricluster 10 in Table 4, which has 10,562 genes. The analysis was carried out with the help of the GOnet application available at [www.ebi.ac.uk/QuickGO](http://www.ebi.ac.uk/QuickGO) (accessed on 20 July 2023). The analysis consists of gene ontology for biological processes, cellular components, and molecular components.

From the results of the gene ontology analysis on the biological processes in Figure 6, biological processes associated with lung disease include responses to stimuli, immune system processes, biological regulation, and metabolic processes. A response to a stimulus, such as an irritant or pathogen, is essential to the lung’s immune response to infection or pathological conditions. Research by Barnes [18] linked exaggerated inflammatory responses to stimuli such as exposure to air pollutants with the development of lung diseases, including chronic obstructive pulmonary disease (COPD). Immune system processes are essential in lung disease, especially when responding to pathogens or allergenic agents that enter the respiratory tract. Immune system processes are involved in recognizing, degrading, and regulating lung immune responses associated with lung diseases, such as asthma or interstitial lung disease. Previous studies have identified the role of T cells and adaptive immune responses in the pathogenesis of lung disease, such as the regulation of immune responses involving T cells in obstructive pulmonary disease and asthma [19].

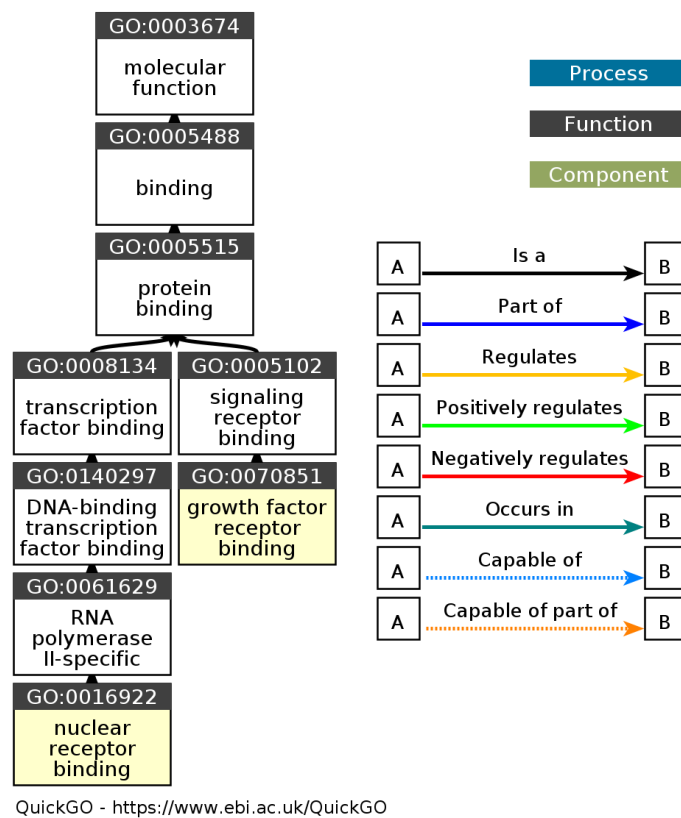
Cellular regulation involves the regulation of activities in the body. In the context of lung disease, biological regulation is related to the processes of adaptation of the lung to stress, regulation of the inflammatory response, and remodeling of lung tissue. Biological regulation is also influenced by the pathogenesis of the lung disease, including the role of growth factors and related signaling pathways in lung tissue repair and change [20]. The metabolic processes involve all cell chemical reactions that produce energy and maintain cellular functions. In the context of lung disease, metabolic processes, including lipid and carbohydrate metabolism, can influence the development of lung diseases, such as COPD

or interstitial lung disease. Research has shown a relationship between lipid metabolism, especially cholesterol, with lung inflammation and the development of obstructive lung disease [21].



**Figure 6.** DAG gene ontology biological process (source: <https://www.ebi.ac.uk/QuickGO/> (accessed on 20 July 2023)).

The results of the gene ontology analysis based on Figure 7 showed that the molecular functions associated with lung disease include molecular binding interactions that play a role in regulation and biological activity. Protein binding is an interaction that plays a vital role in many cellular processes involved in the lung, such as the regulation of signaling pathways and protein–protein interactions that play a role in lung function. Proteins have essential roles in lung disease, one of which is the interaction between the proteins in inflammatory pathways and the immune responses involved in chronic obstructive pulmonary disease (COPD) [22]. Transcription factor binding interactions with DNA also play a role in lung disease. DNA-bound transcription factors can regulate the expression of genes related to lung function, including the regulation of inflammatory pathways and lung tissue remodeling. Research conducted by Barnes [18] identified transcription factors that play a role in the pathogenesis of lung disease, namely transcription factors involved in regulating the pro-inflammatory genes in obstructive lung disease.



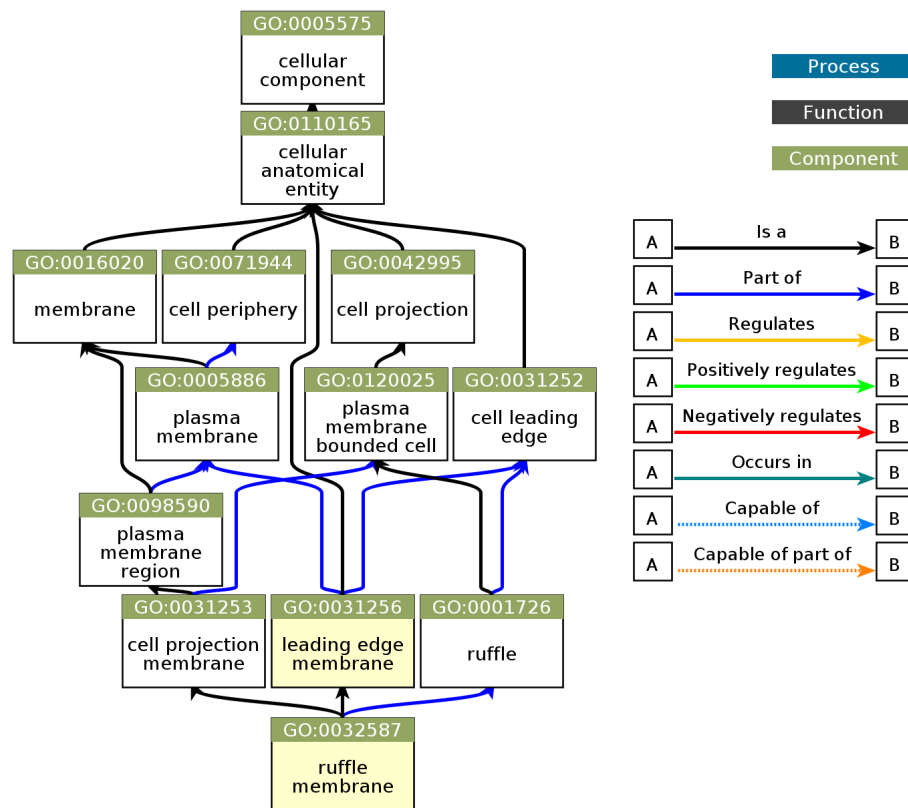
**Figure 7.** DAG gene ontology molecular function (source: <https://www.ebi.ac.uk/QuickGO/> (accessed on 20 July 2023)).

In addition, signal receptor binding also has an important role in lung disease. Signal receptors on the surface of lung cells interact with signaling molecules, such as growth factors, cytokines, or hormones, which can affect cell proliferation, changes in cell phenotype, and inflammatory response in lung disease. Previous studies have linked binding to growth factor receptors such as the epidermal growth factor receptor (EGFR) with the development of lung diseases, including lung cancer [23]. Molecular functions related to RNA polymerase II-specific and nuclear receptor binding also affect the regulation of gene expression and biological pathways related to lung disease.

From the gene ontology analysis of the lung-diseased cells, the cellular components related to lung disease were obtained, including the cellular anatomical structures, membranes, cell parts, cell projections, and plasma membrane components. The DAG gene ontology cellular component can be seen in Figure 8. The cellular anatomical structure includes all the fundamental components in the cell, including the organelles, cytoskeleton structures, and macromolecular complexes. Membranes, particularly the plasma membrane, play roles in maintaining cell integrity, the regulation of molecular transport, and cellular interactions associated with lung disease. Research shows a link between changes in cellular membranes and the development of lung diseases, such as changes in plasma membrane lipid composition in conditions such as chronic obstructive pulmonary disease (COPD) [24].

Cell projections, including cell tip and cellular projections, are essential in lung disease. Cell projection is involved in various processes, such as cell migration, cellular interactions, and changes in cell phenotype. Cell projections play a role in the development and progression of lung disease, including changes in cell projections in lung epithelial cells involved in the pathogenesis of obstructive lung disease [25]. In addition, the plasma membrane component also has an important role in lung disease. Plasma membranes and their components, such as membrane ruffles, are involved in many biological processes in the lung, such as cellular adhesion, cell migration, and inflammatory responses. Previous

research has revealed the role of membrane ruffles in the inflammatory response in lung disease [26].



QuickGO - <https://www.ebi.ac.uk/QuickGO>

**Figure 8.** DAG gene ontology cellular component (source: <https://www.ebi.ac.uk/QuickGO/> (accessed on 20 July 2023)).

The hybrid  $\delta$ -Trimax PSO method has advantages and disadvantages in its application. The advantage is that it can produce a more optimal tricluster than that produced by previous  $\delta$ -Trimax research because it combines the hybrid PSO method and the tricluster that is obtained. However, the combination of these hybrid PSO into  $\delta$ -Trimax can result in increased computation time to obtain the optimal tricluster, so this is something that must be considered when using this method.

### 5. Conclusions

This research was conducted to implement hybrid  $\delta$ -Trimax particle swarm optimization for a triclustering analysis based on the swarm intelligence concept using global and local searches of their particles. Using the fitness function, that is, the tricluster volume, which was maximized, while still maintaining an MSR value below the  $\delta$  threshold, some simulations were carried out to find the best combination of parameters. Based on the average TQI, the optimal tricluster was found from a simulation with  $\delta = 0.006$ ;  $\omega_{max} = 0.7$ ;  $\omega_{min} = 0.2$  and neighborhood type “lbest”.

The tricluster implemented using gene ontology analysis resulted in several processes related to lung disease in the biological processes, molecular functions, and cell components. The biological processes associated with lung disease include response to stimulus, immune system processes, biological regulation, and metabolic processes. The molecular functions related to lung disease include molecular binding interactions that play a role in biological regulation and activity. The cellular components associated with lung disease include cellular anatomical structures, membranes, cell parts, cell projections, and plasma membrane components.

This tricluster could serve as a guideline for further research to develop MGd drugs as an anti-cancer treatment. For further triclustering research, it is advisable to compare the results of several data sets and methods to determine the suitability of a method with a particular type of data.

**Author Contributions:** Conceptualization, T.S.; methodology, M.A.S.I. and N.S.; validation; formal analysis, S.M.S., D.S., S.P. and R.C.I.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the PUTI Q2 Matching Fund (MF) 2022–2023 (Batch 3) from Directorate of Research and Development Universitas Indonesia and The Ministry of Education, Culture, Research, and Technology with contract number NKB-1464/UN2.RST/HKP.05.00/2022.

**Data Availability Statement:** The data are available at the NCBI gene repository: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2189>, accessed on 12 March 2021.

**Acknowledgments:** The authors thank to Agus Salim from Melbourne School of Population and Global Health, The University of Melbourne, who supported this research with his valuable comments, remarks, and suggestions.

**Conflicts of Interest:** The authors declare that there are no conflict of interest.

## References

- Riahi, Y.; Riahi, S. Big data and big data analytics: Concepts, types and technologies. *Int. J. Res. Eng.* **2018**, *5*, 524–528. [[CrossRef](#)]
- Madni, H.A.; Anwar, Z.; Shah, M.A. Data mining techniques and applications—A decade review. In Proceedings of the 2017 23rd International Conference on Automation and Computing (ICAC), Huddersfield, UK, 7–8 September 2017; pp. 1–7.
- SwathyPriyadharsini, P.; Premalatha, K. Particle Swarm Optimization for Triclustering High Dimensional Microarray Gene Expression Data. *Res. J. Pharm. Technol.* **2019**, *12*, 2222–2228. [[CrossRef](#)]
- Zhao, L.; Zaki, M.J. Tricluster: An effective algorithm for mining coherent clusters in 3D microarray data. In Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, Baltimore, MD, USA, 14–16 June 2005; pp. 694–705.
- Bhar, A.; Haubrock, M.; Mukhopadhyay, A.; Maulik, U.; Bandyopadhyay, S.; Wingender, E.  $\delta$ -TRIMAX: Extracting triclusters and analysing coregulation in time series gene expression data. In *Algorithms in Bioinformatics: Proceedings of the 12th International Workshop, WABI 2012, Ljubljana, Slovenia, 10–12 September 2012*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 165–177.
- Siswantining, T.; Saputra, N.; Sarwinda, D.; Al-Ash, H.S. Triclustering discovery using the  $\delta$ -trimax method on microarray gene expression data. *Symmetry* **2021**, *13*, 437. [[CrossRef](#)]
- Daeng, W.A.; Siswantining, T.; Bustamam, A.; Anki, P.  $\delta$ -TRIMAX Method with Silhouette Coefficient on Microarray Gene Expression Data for Early Detection of Heart Failure. In Proceedings of the 2022 5th International Conference on Information and Communications Technology (ICOIACT), Virtually, 24–25 August 2022; pp. 412–416.
- Apriliana, G.D.; Siswantining, T.; Pramana, S.; Anki, P. Evaluation of Order Preserving Triclustering (OPTricluster) in 3 Dimensional Gene Expression Data Analysis Using Gene Ontology. In Proceedings of the 2022 International Conference on Electrical and Information Technology (IEIT), Malang, Indonesia, 15–16 September 2022; pp. 126–131.
- Andika, H.K.; Siswantining, T.; Bustamam, A.; Anki, P. THD-Tricluster Method on Three Dimensional Gene Expression Data of Tuberculosis Patients. In Proceedings of the 2022 6th International Conference on Informatics and Computational Sciences (ICICoS), Virtual, 28–29 September 2022; pp. 48–53.
- Narmadha, N.; Rathipriya, R. Greedy Two Way K-Means Clustering For Optimal Coherent Tricluster. *Int. J. Sci. Technol. Res.* **2019**, *8*, 1916–1921.
- Fahmideh, L.; Kord, H.; Shiri, Y. Importance of microarray technology and its applications. *J. Curr. Res. Sci.* **2016**, *4*, 25.
- Govindarajan, R.; Duraiyan, J.; Kaliyappan, K.; Palanisamy, M. Microarray and its applications. *J. Pharm. Bioallied Sci.* **2012**, *4*, S310.
- Narmadha N, R.R. Gene Ontology Analysis of 3D Microarray Gene Expression Data Using Hybrid PSO Optimization. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *8*, 3890–3896. [[CrossRef](#)]
- Zhao, H.; Wee-Chung Liew, A.; Z Wang, D.; Yan, H. Biclustering analysis for pattern discovery: current techniques, comparative studies and applications. *Curr. Bioinform.* **2012**, *7*, 43–55. [[CrossRef](#)]
- Henriques, R.; Madeira, S.C. Triclustering algorithms for three-dimensional data analysis: a comprehensive survey. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 1–43. [[CrossRef](#)]
- Bar-Joseph, Z. Analyzing time series gene expression data. *Bioinformatics* **2004**, *20*, 2493–2503. [[CrossRef](#)] [[PubMed](#)]
- Magda, D.; Lecane, P.; Miller, R.A.; Lepp, C.; Miles, D.; Mesfin, M.; Biaglow, J.E.; Ho, V.V.; Chawannakul, D.; Nagpal, S.; et al. Motexafin gadolinium disrupts zinc metabolism in human cancer cell lines. *Cancer Res.* **2005**, *65*, 3837–3845. [[CrossRef](#)] [[PubMed](#)]
- Barnes, P.J. Immunology of asthma and chronic obstructive pulmonary disease. *Nat. Rev. Immunol.* **2008**, *8*, 183–192. [[CrossRef](#)] [[PubMed](#)]

19. Lambrecht, B.N.; Hammad, H. The immunology of asthma. *Nat. Immunol.* **2015**, *16*, 45–56. [[CrossRef](#)] [[PubMed](#)]
20. Selman, M.; Pardo, A. Revealing the pathogenic and aging-related mechanisms of the enigmatic idiopathic pulmonary fibrosis. an integral model. *Am. J. Respir. Crit. Care Med.* **2014**, *189*, 1161–1172. [[CrossRef](#)] [[PubMed](#)]
21. Xuan, L.; Han, F.; Gong, L.; Lv, Y.; Wan, Z.; Liu, H.; Zhang, D.; Jia, Y.; Yang, S.; Ren, L.; et al. Association between chronic obstructive pulmonary disease and serum lipid levels: A meta-analysis. *Lipids Health Dis.* **2018**, *17*, 1–8. [[CrossRef](#)] [[PubMed](#)]
22. Vestbo, J.; Hurd, S.S.; Agustí, A.G.; Jones, P.W.; Vogelmeier, C.; Anzueto, A.; Barnes, P.J.; Fabbri, L.M.; Martinez, F.J.; Nishimura, M.; et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am. J. Respir. Crit. Care Med.* **2013**, *187*, 347–365. [[CrossRef](#)] [[PubMed](#)]
23. Bethune, G.; Bethune, D.; Ridgway, N.; Xu, Z. Epidermal growth factor receptor (EGFR) in lung cancer: An overview and update. *J. Thorac. Dis.* **2010**, *2*, 48. [[PubMed](#)]
24. Cong, X.; Hubmayr, R.D.; Li, C.; Zhao, X. Plasma membrane wounding and repair in pulmonary diseases. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **2017**, *312*, L371–L391. [[CrossRef](#)]
25. Caldeira, I.; Fernandes-Silva, H.; Machado-Costa, D.; Correia-Pinto, J.; Moura, R.S. Developmental pathways underlying lung development and congenital lung disorders. *Cells* **2021**, *10*, 2987. [[CrossRef](#)] [[PubMed](#)]
26. Kreider-Letterman, G.; Cooke, M.; Goicoechea, S.M.; Kazanietz, M.G.; Garcia-Mata, R. Quantification of ruffle area and dynamics in live or fixed lung adenocarcinoma cells. *STAR Protoc.* **2022**, *3*, 101437. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.