Gutiérrez-Avilés *et al. BioData Mining* (2018) 11:15
https://doi.org/10.1186/s13040-018-0177-5

BioData Mining

RESEARCH Open Access

TRIQ: a new method to evaluate triclusters

David Gutiérrez-Avilés[1*], Raúl Giráldez[1], Francisco Javier Gil-Cumbreras[1] and Cristina Rubio-Escudero[2]

*Correspondence: dgutavi@upo.es
[1]School of Engineering, Pablo de Olavide University, Seville, Spain
Full list of author information is available at the end of the article

Abstract

Background: Triclustering has shown to be a valuable tool for the analysis of microarray data since its appearance as an improvement of classical clustering and biclustering techniques. The standard for validation of triclustering is based on three different measures: correlation, graphic similarity of the patterns and functional annotations for the genes extracted from the Gene Ontology project (GO).

Results: We propose *TRIQ*, a single evaluation measure that combines the three measures previously described: correlation, graphic validation and functional annotation, providing a single value as result of the validation of a tricluster solution and therefore simplifying the steps inherent to research of comparison and selection of solutions. *TRIQ* has been applied to three datasets already studied and evaluated with single measures based on correlation, graphic similarity and GO terms. Triclusters have been extracted from this three datasets using two different algorithms: TriGen and OPTricluster.

Conclusions: *TRIQ* has successfully provided the same results as a the three single evaluation measures. Furthermore, we have applied *TRIQ* to results from another algorithm, *OPTRicluster*, and we have shown how TRIQ has been a valid tool to compare results from different algorithms in a quantitative straightforward manner. Therefore, it appears as a valid measure to represent and summarize the quality of tricluster solutions. It is also feasible for evaluation of non biological triclusters, due to the parametrization of each component of *TRIQ*.

Keywords: Triclustering, Quality measure, Genetic algorithms, Biological quality, Graphical quality, Correlation

Background

Analysis of data structured in 3D manner is becoming an essential task in fields such as biomedical research, for instance in experiments studying gene expression data taking time into account. There is a lot of interest in this type of longitudinal experiments because they allow an in-depth analysis of molecular processes in which the time evolution is important, for example, cell cycles, development at the molecular level or evolution of diseases [1]. Therefore, the use of specific tools for data analysis in which genes are evaluated under certain conditions considering the time factor becomes necessary. In this sense, triclustering [2] appears as a valuable tool since it allows for the assessment of genes under a subset of the conditions of the experiment and under a subset of time points.

The evaluation of solutions obtained by triclustering algorithms is challenging by the fact that there is no ground truth to describe triclusters present in real 3D data. In

literature, the standard measures to evaluate tricluster solutions are based on three areas as can be seen in the triclustering publications [3–7]. First, correlation measures such as Pearson [8] or Spearman [9]. Second, graphic validation of the patterns extracted based on the graphic representation, i.e., how similar the genes from a tricluster are based on the graphic representation of the genes across conditions and time points. Third, functional annotations extracted from the Gene Ontology project (GO) [10] for the genes in the tricluster.

However, we consider that providing a single evaluation measure capable of combining the information from the three aforementioned sources of validation is a neccesary task. Therefore, in this work we propose *TRIQ*, a validation measure which combines the three previously proposed validation mechanisms (correlation, graphic validation and functional annotation of the genes).

The application of clustering and biclustering techniques to gene expression data has been broadly studied in the literature [11, 12]. Although triclustering is the result from the natural evolution of the clustering and biclustering techniques, is still a very recent concept. However, nowadays, these techniques are arousing a great interest from the scientific community, which has caused a notable increase of the number of researches focused on finding new triclustering approaches. This section is to provide a general overview of triclustering published in literature. We particularly focus on the validation methods applied to assess the quality of the triclusters obtained.

In 2005, Zhao and Zaki [3] introduced the triCluster algorithm to extract patterns in 3D gene expression data. They presented a measure to assess triclusters's quality based on the symmetry property. They validated their triclusters based on their graphical representation and Gene Ontology (GO) results. g-triCluster, an extended and generalized version of Zhao and Zaki's proposal, was published one year later [4]. The authors claimed that the symmetry property is not suitable for all patterns present in biological data and proposed the Spearman rank correlation [9] as a more appropriate tricluster evaluation measure. They also showed validation results based on GO.

An evolutionary computation proposal was made in [13]. The fitness function defined is a multi-objective measure which tries to optimize three conflicting objectives: clusters size, homogeneity and gene-dimension variance of the 3D cluster. The tricluster quality validation was based on GO. LagMiner was introduced in [6] to find time-lagged 3D clusters, what allows to find regulatory relationships among genes. It is based on a novel 3D cluster model called S_2D_3 Cluster. They evaluated their triclusters on homogeneity, regulation, minimum gene number, sample subspace size and time periods length. Their validation was based on graphical representation and GO results. Hu et al. presented an approach focusing on the concept of Low-Variance 3-Cluster [5], which obeys the constraint of a low-variance distribution of cell values. This proposal uses a different functional enrichment tool called CLEAN [14], which uses GO as one of their components. The work in [7] was focused on finding Temporal Dependency Association Rules, which relate patterns of behavior among genes. The rules obtained are used to represent regulated relations among genes. They also validated their triclusters based on their graphical representation and GO results.

Tchagang et al. [15] proposed OPTricluster, a triclustering algorithm which obtains 3D short time series gene expression datasets by applying a statistical methodology. In this case, the authors carried out an in-depth biological validation based on GO, but they tested the robustness of OPTricluster to noise using the Adjusted Rand Index (ARI) [16], which also was used by aforementioned g-tricluster.

In 2013, two new and very interesting approaches were proposed. On the one hand, the $\delta - TRIMAX$ algorithm [17], which applies a variant of the MSR adapted to 3D datasets and yields triclusters that have a MSR score below a threshold δ. This algorithm has a version based on evolutionary multi-objective optimization, named $EMOA - \delta - TRIMAX$ [18], which aims at optimizing the use of $\delta - TRIMAX$ by adding the capabilities of evolutionary algorithms to retrieve overlapping triclusters. On the other hand, OAC-Triclustering was also proposed by Gnatyshak et al. in [19]. In the following years, the authors developed improvements and extensions of this algorithm [20–22].

More recent works have extended the capabilities of the tricluster algorithms by combination of several approaches. Thereby, Liu et al. [23] mixed fuzzy clustering and fuzzy biclustering algorithms in order to expands them to support 3D data and they used the *F*-Measure and Entropy as criteria to evaluate the performance. Also, Kakati et al. [24] combined parallel biclustering and distributed triclustering approaches to obtain improvements on the computational cost. In this work, the authors use a quality measure based on shifting and scaling patterns [25] to optimize the triclusters obtained.

Most of the methods studied base the quality of the triclusters on the graphic representation or on metrics aimed at measuring diverse characteristics of such representation. From a biological point of view, the standard for validation of triclusters quality is based on GO functional annotations.

Methods

This section presents the *TRIQ* (TRIcluster Quality) validation measure [26], a novel method to evaluate the quality of triclusters extracted from gene expression datasets.

From an overall perspective, *TRIQ* takes into account the three principal components of a tricluster, i.e. the genes, experimental conditions and time points, in order to measure its quality from three approaches: the level of biological notoriety of the cluster (biological quality), the graphic quality of the patterns of the genes in the tricluster (graphic quality), and the level of correlation of the genes in the tricluster by means of the Pearson [8] and the Spearman [9] indexes. Therefore, *TRIQ* is composed by a combination of four indexes: *BIOQ* (BIOlogical Quality), *GRQ* (GRaphic Quality), *PEQ* (PEarson Quality) and *SPQ* (SPearman Quality).

In Eq. 1 we define *TRIQ* as the weighted sum of each of the four aforementioned terms. Therefore, four associated weights must be defined: the weight for *BIOQ*, denoted as W_{bio}; the weight for *GRQ*, denoted as W_{gr}; the weight for *PEQ*, denoted as W_{pe}; and the weight for *SPQ*, denoted as W_{sp}.

$$\begin{aligned} TRIQ(TRI) = \frac{1}{W_{bio} + W_{gr} + W_{pe} + W_{sp}} * [W_{bio} * BIOQ(TRI) \\ + W_{gr} * GRQ(TRI) + W_{pe} * PEQ(TRI) + W_{sp} * SPQ(TRI)] \end{aligned} \quad (1)$$

This is a general definition of *TRIQ*. In order to obtain a *TRIQ* index as balanced as possible among the four quality indexes *BIOQ*, *GRQ*, *PEQ*, and *SPQ* we performed an exhaustive testing procedure with well known datasets. Several combinations of values of *BIOQ*, *GRQ*, *PEQ*, and *SPQ* were tested, and in Fig. 1 we show the results obtained.

We see that that the value of *TRIQ* is slightly directly dependent on the weights related to correlation, *PEQ*, and *SPQ*. This is due to the fact that these values rank in the [0-1] interval, being usually high, from 0.7 to 1. The value of *TRIQ* has a higher level of dependence to the graphical quality, *GRQ*, and reverse strong dependence to the biological quality, *BIOQ*, due to the fact that *BIOQ* ranks in low values, usually around 10^{-3} to 10^{-5}. Based on this experiments, we have configured the *TRIQ* measure with the weights showed in Eq. 2 in order to obtain a balanced value of *TRIQ*.

$$W_{bio} = 0.5,\ W_{gr} = 0.4,\ W_{pe} = 0.05,\ W_{sp} = 0.05 \tag{2}$$

Next, we describe in depth each of the terms involved in the *TRIQ* measure.

Correlation measures: PEQ and SPQ

The correlation measures involved in *TRIQ* are Pearson's *PEQ* [8] and Spearman's *SPQ* [9] correlations. They have been chosen since they are the standard correlation measures and they are widely used in literature [4]. The correlation provides a numerical estimation of the dependence among the genes, conditions and times in the tricluster solutions.

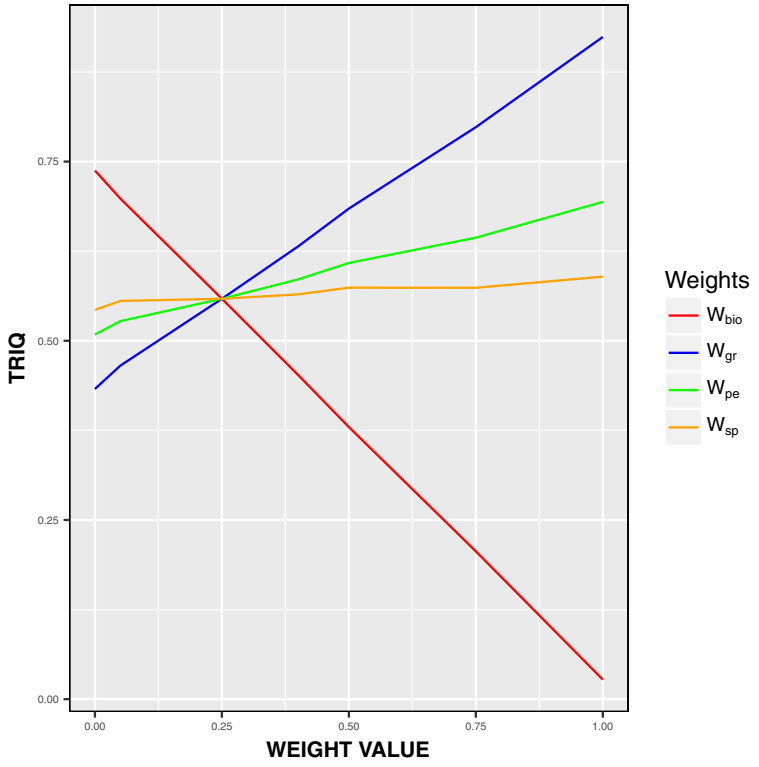

Fig. 1 Representation of *BIOQ*, *GRQ*, *PEQ*, and *SPQ* influence on *TRIQ*

Given a tricluster *TRI*, we compute *PEQ* and *SPQ* by the following mechanism. Given the subset of genes (see Eq. 3a), conditions (see Eq. 3b) and time stamps (see Eq. 3c), we obtain a value of expression for each combination gene, condition and time. For instance, for a tricluster consisting of four genes, two conditions and three time points, we have twenty four expression values. We then compute the Pearson correlation for each pair of values, and compute *PEQ* as the average of the absolute values to avoid negative and positive correlations canceling each other (see Eq. 4). Furthermore, for this measure we do not care if the correlation is positive or negative between values, we only want to know the level of correlation. The *SPQ* value is the equivalent using the Spearman correlation (see Eq. 5).

$$TRI_G = < g_0, g_1, \ldots, g_G > \tag{3a}$$

$$TRI_C = < c_0, c_1, \ldots, c_C > \tag{3b}$$

$$TRI_T = < t_0, t_1, \ldots, t_T > \tag{3c}$$

$$PEQ(TRI) = \frac{\sum_{i=0,j=0}^{\#exp} \left| Pearson_{i \neq j} \left(exp_i, exp_j \right) \right|}{\#pairs\ of\ exp} \tag{4}$$

$$SPQ(TRI) = \frac{\sum_{i=0,j=0}^{\#exp} \left| Spearman_{i \neq j} \left(exp_i, exp_j \right) \right|}{\#pairs\ of\ exp} \tag{5}$$

with *exp* representing the expressions in each tricluster *TRI*.

Graphical validation: GRQ

The *GRQ* member of Eq. 1 measures the graphical quality of the tricluster. This graphical quality of a tricluster is a quantitative representation of a qualitative measure: how homogeneous the members of the tricluster are. This method is widely used in literature for visual validation of the results by means of graphically representing the triclusters on their three components: genes, conditions and time points [3, 6, 7].

The *GRQ* index is described in Eq. 6. This measure is defined based on the normalization of the angle value given by *MSL*. The Multi SLope (*MSL*) evaluation function was defined in [27] and, given a tricluster *TRI*, provides a numerical value of the similarity among the angles of the slopes formed by each profile shaped by the genes, conditions, and times of the tricluster.

$$GRQ(TRI) = 1 - \frac{MSL(TRI)}{2\pi} \tag{6}$$

The *MSL* measure considers the three graphical views of a tricluster, also defined in [27]: TRI_{gct}, TRI_{gtc}, and TRI_{tgc}. These three terms are generally defined as TRI_{xop}, with the expression levels of the tricluster represented in the Y axis, x represents the tricluster component in the X axis (genes or time points), o represents the lines plotted in the graph (genes, conditions or time lines) and p the type of facets or panels represented (time points or conditions). We can observe an example of the TRI_{tgc} view of a tricluster with the genes g_1, g_4, g_7 and g_{10}, the experimental conditions c_2, c_5 and c_8 and the time points t_0, t_2, t_{11} in Fig. 2 and see how each line or gene forms a set of angles (two for this particular example) defined by each time point in the X axis for every panel or experimental condition. Thus, *MSL* measures the differences among the angles formed by every series traced on each of the three graphic representations taking into account TRI_{gct}, TRI_{gtc}, and TRI_{tgc}. A near to zero value of *MSL* implies a better graphical quality of a tricluster

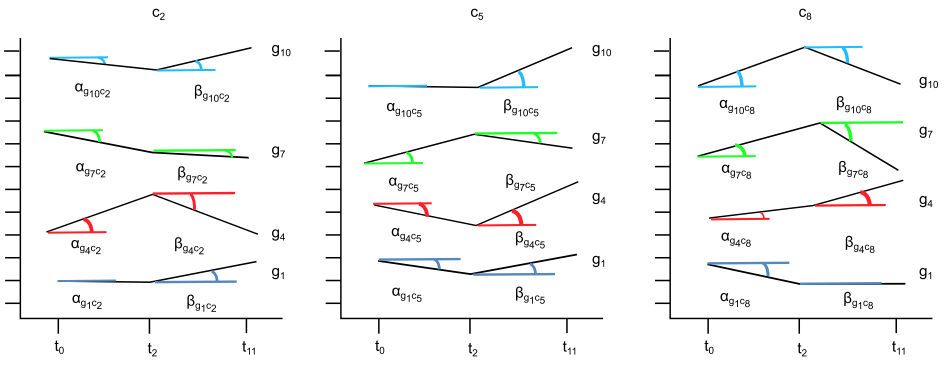

Fig. 2 Representation of how the *MSL* measure is calculated. This figure shows three graphics containing four lines each one with a representation of their slopes

therefore, according to *GRQ* formulation in Eq. 6, a tricluster is graphically better the smaller the value of *MSL*.

Biological validation: BIOQ

The *BIOQ* member of Eq. 1 measures the biological quality of the tricluster. Specifically, *BIOQ* uses the genes (TRI_G) of the input tricluster *TRI* to compute this index. As you can see in Eq. 7, the biological quality of a tricluster *TRI* is defined as the biological significance, SIG_{bio}, of the set of genes TRI_G divided by the S_{max} value.

$$BIOQ(TRI) = \frac{SIG_{bio}\,(TRI_G)}{S_{max}} \tag{7}$$

The SIG_{bio} and S_{max} elements of the *BIOQ* index have been designed in order to represent, by means of a quantitative score, the value of the Gene Ontology analysis of the genes that compose the measured tricluster.

The Gene Ontology Project (GO) [10] is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases, besides identifying the annotated terms, performs the statistical analysis for the over-representation of those terms, also providing a statistical significance *p*-value. However, it is also important to take into account how deep in the ontology the terms are annotated, with the deeper terms being more specific than the superficial ones [28]. The SIG_{bio} and S_{max} elements are calculated based on the GO analysis that identifies, for a set of genes in a tricluster, the terms listed in each of the three available ontologies: biological processes, cellular components, and molecular functions. This GO analysis is performed with the software Ontologizer [29].

The computation of SIG_{bio} consists on counting how many terms of the annotated genes of the tricluster in the GO analysis are in a particular intervals of *p*-value. Table 1 represents the ah-hoc designed system of intervals of *p*-value and scoring system. The intervals and the scoring system are defined in Eq. 8 where for a given level, $Inter_l$ is defined by a weight value w_l for the level, and by the lower and upper bounds (inf_l and sup_l, respectively), being an open-closed *p*-values interval (Eq. 8a). The set of existing *LV* consists of all levels with Inf_l smaller or equal to a minimum *p*-value, *th*. For each interval of each level $Inter_l$, the weight value w_l is defined in Eq. 8c; Inf_l is defined in Eq. 8d, and sup_l is defined in Eq. 8e.

Table 1 Biological significance intervals

Level (l)	Weight (w_l)	Interval ($inter_l$)
41	401	(0.0E-00,1.0E-40]
40	391	(1.0E-40,1.0E-39]
39	381	(1.0E-39,1.0E-38]
38	371	(1.0E-38,1.0E-37]
37	361	(1.0E-37,1.0E-36]
36	351	(1.0E-36,1.0E-35]
35	341	(1.0E-35,1.0E-34]
34	331	(1.0E-34,1.0E-33]
33	321	(1.0E-33,1.0E-32]
32	311	(1.0E-32,1.0E-31]
31	301	(1.0E-31,1.0E-30]
30	291	(1.0E-30,1.0E-29]
29	281	(1.0E-29,1.0E-28]
28	271	(1.0E-28,1.0E-27]
27	261	(1.0E-27,1.0E-26]
26	251	(1.0E-26,1.0E-25]
25	241	(1.0E-25,1.0E-24]
24	231	(1.0E-24,1.0E-23]
23	221	(1.0E-23,1.0E-22]
22	211	(1.0E-22,1.0E-21]
21	201	(1.0E-21,1.0E-20]
20	191	(1.0E-20,1.0E-19]
19	181	(1.0E-19,1.0E-18]
18	171	(1.0E-18,1.0E-17]
17	161	(1.0E-17,1.0E-16]
16	151	(1.0E-16,1.0E-15]
15	141	(1.0E-15,1.0E-14]
14	131	(1.0E-14,1.0E-13]
13	121	(1.0E-13,1.0E-12]
12	111	(1.0E-12,1.0E-11]
11	101	(1.0E-11,1.0E-10]
10	91	(1.0E-10,1.0E-09]
9	81	(1.0E-09,1.0E-08]
8	71	(1.0E-08,1.0E-07]
7	61	(1.0E-07,1.0E-06]
6	51	(1.0E-06,1.0E-05]
5	41	(1.0E-05,1.0E-04]
4	31	(1.0E-04,1.0E-03]
3	21	(1.0E-03,1.0E-02]
2	11	(1.0E-02,1.0E-01]
1	1	(1.0E-01,1.0E-00]

$$inter_l = \langle w_l,\ (inf_l, sup_l] \rangle \tag{8a}$$

$$LV = \forall\ l \in \mathbb{N} : inf_l \leq th \tag{8b}$$

$$w_l = [(l-1) * d] + 1 \tag{8c}$$

$$inf_l = \frac{s}{b^l} \tag{8d}$$

$$sup_l = \frac{s}{b^{(l-1)}} \tag{8e}$$

This definition is made in order to establish a general interval system dependent on the parameters described above. For our purpose, we have settled these parameters as shown in Eq. 9; this configuration produces the intervals of Table 1, furthermore, it describes all the biological significance intervals for the configuration detailed in Eq. 9. For each row, weight (wl) and range ($inter_l$) for each level (L) sorted in ascending order are shown. Each interval provides a set of p-values where their significance is directly related to the corresponding level, that is, a p-value is better the higher the level to which it belongs, and a p-value is better the closer to zero it is.

$$th = 1.0 \times 10^{-40} \tag{9a}$$
$$d = 10.0 \tag{9b}$$
$$b = 10.0 \tag{9c}$$
$$s = 1.0 \tag{9d}$$
$$LV = \{1, \ldots, 41\} \tag{9e}$$

Taking into account each level l and each predefined interval $inter_l$, the biological significance for the genes of the measured tricluster is defined in Eq. 10a as the addition of all scores for each level l from the LV level set Eq. 9e. The score function S for a level l (Eq. 10b) is defined by the multiplication of the concentration of terms for this level $C(l)$, defined in Eq. 10c as the number of terms of the level l divided by the total number of terms, by the weight of the level, and by the level plus a bonus function f_{bonus}, defined in Eq. 10d as the sum of the level plus a bonus value V_{bonus} if the current level is the maximum level of LV or zero in any another case.

$$SIG_{bio}(TRI_G) = \sum_{l \in LV} S(l) \tag{10a}$$
$$S(l) = C(l) * w_l * l + f_{bonus}(l) \tag{10b}$$
$$C(l) = \frac{\#terms(l)}{\#total\ terms} \tag{10c}$$
$$f_{bonus}(l) = if\ \ (l\ equal\ to\ l_{max})\ \ then\ \ l + V_{bonus}\ \ else\ 0 \tag{10d}$$

Again, this definition is made in order to establish a general system of SIG_{bio}. For our purpose and as a result of an exhaustive testing, the V_{bonus} parameter has been settled to 0; this fact produces S_{max} as the maximum achievable score for the interval configuration as you can see in Eq. (11), that has been used to the SIG_{bio} normalization in Eq. 7.

$$S_{max} = \left(C_{max} * w_{l_{41}} * l_{41}\right) + f_{bonus}(41) = (1 * 401 * 41) + (41 + 0) = 16482 \tag{11}$$

Results

In this section, we present how *TRIQ* works in an experimental environment. To reach this goal, we have used the TriGen algorithm [2] and the OPTricluster algorithm [15] in order to analyze the datasets, find triclusters and measure them with *TRIQ*.

TriGen is based on an heuristic, genetic algorithm, and its performance greatly depends on the fitness function used to find the triclusters. There are three fitness functions available in *TriGen*: Mean Squared Residue 3D (MSR_{3D}) [30], Least Squared Lines (LSL) [31] and Multi SLope Measure (MSL) [27]. *OPTricluster* identifies triclusters of genes with

expression levels having the same direction across the time point experiments in subsets of samples taking into consideration the sequential nature of the time-series.

The three datasets analyzed that involve genes and experimental conditions examined under certain time points are:

- $D_{elu_{3D}}$: The yeast cell cycle *(Saccharomyces Cerevisiae)* [32], in particular, the elutriation experiment.
- $D_{GDS4510_{3D}}$: The GDS4510 dataset from an experiment with mice (Mus Musculus) [33].
- $D_{GSD4472_{3D}}$: The GDS4472 dataset from an experiments with humans (Homo Sapiens) [34].

The first dataset is available at the Stanford University website. The last two datasets have been retrieved from Gene Expression Omnibus [35], a repository of high throughput gene expression data.

For each dataset, we have performed four algorithm executions: *TriGen* with MSR_{3D} (hereon MSR_{3D}), *TriGen* with *LSL* (hereon *LSL*), *TriGen* with *MSL* and (hereon *MSL*) and *OPTricluster* (hereon *OPT*).

For each algorithm execution and dataset, we have yielded 10 triclusters and the *TRIQ* measure has been used to evaluate their quality. We have found 10 triclusters for each execution in order to have a high number of solutions where *TRIQ* can show its suitability. In the case of MSR_{3D}, *LSL*, and *MSL* executions the number of triclusters has been chosen as one of the *TriGen* algorithm parameters and for *OPT* executions, the tricluster have been randomly selected from the wide collection of triclusters yielded.

Summarizing, we present three experimental batches (*Yeast Elutriation Dataset*, *Mouse GDS4510 Dataset* and, *Human GDS4472 Dataset*) with four experiments each one: MSR_{3D}, *LSL*, *MSL* and *OPT*.

Yeast elutriation dataset

This batch corresponds to the yeast *(Saccharomyces Cerevisiae)* cell cycle problem [32]. The yeast cell cycle analysis project's goal is to identify all genes whose mRNA levels are regulated by the cell cycle. The resources used are public and available in http://genome-www.stanford.edu/cellcycle/. There, we can find information relative to gene expression values obtained from different experiments using microarrays.

For our purpose, we have created a dataset $D_{elu_{3D}}$ from the elutriation experiment with 7744 genes, 13 experimental conditions, and 14 time points. Experimental conditions correspond to different statistical measures of the Cy3 and Cy5 channels while time points represent different moments of taking measures from 0 to 390 min.

$D_{elu_{3D}}$ has been used as the input of the *TriGen* and the *OPTtricluster* algorithm in four experiments: MSR_{3D}, *LSL*, *MSL* and, *OPT*.

Elutriation MSR_{3D} experiment

We can verify in Table 2 how TRI_9 has the best values of *BIOQ*, *PEQ* and *SPQ* whereas TRI_{10} has the best value of *GRQ*. The *GRQ*, *PEQ* and *SPQ* values are stabilized from TRI_2 to TRI_8 until $TRI_9 - TRI_{10}$ when these values reach the maximum. Regarding *BIOQ* values, these vary around 0.0012. Furthermore, *TRIQ* values are stable for all solutions except $TRI_9 - TRI_{10}$ due to the genetic algorithms nature. In conclusion, TRI_9 is the best solution since it has the best value of *TRIQ*, closely followed by TRI_{10}.

Table 2 MSR_{3D} Elutriation solution table

SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
TRI_1	0.289957861	0.001180518	0.627911696	0.400860192	0.363198285
TRI_2	0.283154367	0.001118227	0.610190268	0.397890126	0.372492792
TRI_3	0.292658244	0.001217404	0.632360796	0.39778358	0.384320901
TRI_4	0.283891286	0.001085482	0.614807027	0.38713593	0.36137875
TRI_5	0.282839356	0.001224203	0.613862367	0.379462124	0.354184014
TRI_6	0.290639052	0.001129778	0.625267377	0.4293412	0.370003051
TRI_7	0.259777538	0.001208157	0.553613841	0.382072259	0.372486191
TRI_8	0.281909708	0.001215203	0.606726347	0.407953984	0.36427737
TRI_9	0.453932884	0.001330144	0.896650615	0.986952953	0.905198358
TRI_{10}	0.45152166	0.001148045	0.934659815	0.776480244	0.765193987

Elutriation LSL experiment

In Table 3 you can see how TRI_3 has the best value of *BIOQ*, TRI_2 has the best value of *GRQ*, TRI_6 has the best value of *PEQ* and, TRI_1 has the best value of *SPQ*. In general, the *GRQ*, *PEQ* and *SPQ* values vary around an average value from TRI_1 until TRI_8. Then, these values decrease in $TRI_9 - TRI_{10}$ solutions due to the fact that the algorithm reached a local minimum in this two solutions; the *BIOQ* values fluctuate around 0.0012 value reaching a maximum in TRI_3 and a minimum in TRI_4. The values of *TRIQ* reach the maximum values at the first two solutions, then remain stable and finally fall in local minimum in the last two solutions. In conclusion, TRI_1 is the best solution since it has the best value of *TRIQ*.

Elutriation MSL experiment

We can observe in Table 4 how TRI_2 has the best value of *BIOQ*, *PEQ* and *SPQ* whereas TRI_1 has the best value of *GRQ*. The *GRQ*, *PEQ* and *SPQ* have a stable fluctuation throughout the solutions whilst *BIOQ* varying around the central value 0.0011. The *TRIQ* values reach their maximum value at TRI_2, the minimum at TRI_3 and the rest are stabilized. In conclusion, TRI_2 is the best solution since it has the best value of *TRIQ*.

Elutriation OPT experiment

We can verify in Table 5 how all triclusters have the same value of *BIOQ* since all triclusters grouped the same collection of genes. Regarding *GRQ* index, the triclusters have values between 0.70 and 0.86 with the exception of TRI_1, TRI_9 and, TRI_8 being TRI_4 the solution with better *GRQ*. The *PEQ* and *SPQ* indexes have fluctuating values being TRI_7

Table 3 *LSL* Elutriation solution table

SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
TRI_1	0.444841672	0.001147115	0.925741144	0.737449684	0.741983455
TRI_2	0.444050729	0.001217804	0.927628308	0.725178526	0.722631553
TRI_3	0.434940552	0.001327826	0.912385527	0.697309668	0.689138899
TRI_4	0.431591352	0.001071675	0.905144571	0.692097513	0.6878562
TRI_5	0.433960732	0.001125858	0.913689264	0.683063155	0.675378795
TRI_6	0.440497687	0.001192667	0.916680329	0.743684691	0.720899755
TRI_7	0.437721769	0.001143537	0.916956452	0.702066665	0.705281726
TRI_8	0.441054484	0.001233014	0.919127603	0.730818495	0.724920229
TRI_9	0.41970611	0.001200657	0.894690897	0.629273489	0.595314967
TRI_{10}	0.399331119	0.001102605	0.848695823	0.597009139	0.589020606

Table 4 *MSL* Elutriation solution table

SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
TRI_1	0.492819589	0.001051563	0.999519361	0.929642164	0.9200791
TRI_2	0.493539244	0.001240807	0.997800501	0.930758605	0.945214193
TRI_3	0.476117422	0.001118134	0.991760508	0.78527775	0.791805282
TRI_4	0.478990452	0.001166044	0.991627468	0.813400974	0.821727882
TRI_5	0.480938627	0.001090473	0.995151057	0.820019348	0.826640002
TRI_6	0.475974935	0.001085123	0.992527638	0.779644523	0.788781847
TRI_7	0.478754345	0.00100258	0.994551592	0.806892319	0.801756043
TRI_8	0.478200414	0.001199622	0.993176848	0.804634565	0.801962727
TRI_9	0.477639873	0.001147773	0.991707562	0.805881226	0.801778004
TRI_{10}	0.475505918	0.001132268	0.989937077	0.788265502	0.791033557

the tricluster with the better *PEQ* and *SPQ*. In conclusion, TRI_7 is the best solution since it has the best value of *TRIQ*.

Elutriation summary

We can see in Fig. 3 how the solutions are distributed regarding *BIOQ* and *GRQ* for each experiment. We observe that all points are concentrated in a *BIOQ* interval of $[0.000728, 0.0013]$ for each experiment meanwhile the *MSL* experiment stands out because all its solutions have a *GRQ* near to 1. Regarding the *PEQ* and *SPQ* solutions distribution, we can see in Fig. 4 how the majority of the solutions are concentrated around the point $PEQ = 0.325, SPQ = 0.325$ in the MSR_{3D} experiment, all solutions are concentrated in $[0.50, 0.75]$ interval for *PEQ* and *SPQ* in the *LSL* experiment, all solutions are concentrated in $[0.75, 1.00]$ interval for *PEQ* and *SPQ* in the *MSL* experiment and, all solutions are concentrated in $[0.30, 0.70]$ interval for *PEQ* and *SPQ* in the *OPT* experiment.

The global *TRIQ*-based ranking of solutions is showed in Table 6; we can see how the solutions of the *MSL* experiment are placed on the first positions followed by two outstanding solutions of the MSR_{3D} experiment, all solutions of the *LSL* experiment, all solutions of the *OPT* experiment and, the remaining solutions of the MSR_{3D} experiment.

The *MSL* experiment has the best average values of *TRIQ* and the lowest standard deviation of *TRIQ* as seen in Table 7. This fact is reflected in Fig. 5 wherein the *MSL* point is located on the bottom-right side of the graph which implies that the *MSL* experiment has

Table 5 *OPT* Elutriation solution table

SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
TRI_1	0.25439082	0.000728	0.55556687	0.32575013	0.31025512
TRI_2	0.31786223	0.000728	0.7005279	0.36494172	0.38080349
TRI_3	0.38238284	0.000728	0.84763736	0.40215787	0.45712372
TRI_4	0.39914764	0.000728	0.86882797	0.49884203	0.52621082
TRI_5	0.39749144	0.000728	0.86565058	0.50040614	0.51694181
TRI_6	0.40017866	0.000728	0.86455717	0.53452807	0.5453116
TRI_7	0.40707391	0.000728	0.84656685	0.66128956	0.70037758
TRI_8	0.25896921	0.000728	0.56897207	0.31626703	0.30406432
TRI_9	0.25904229	0.000728	0.56900655	0.31749708	0.30402
TRI_{10}	0.32249076	0.000728	0.72222718	0.33095502	0.33376653

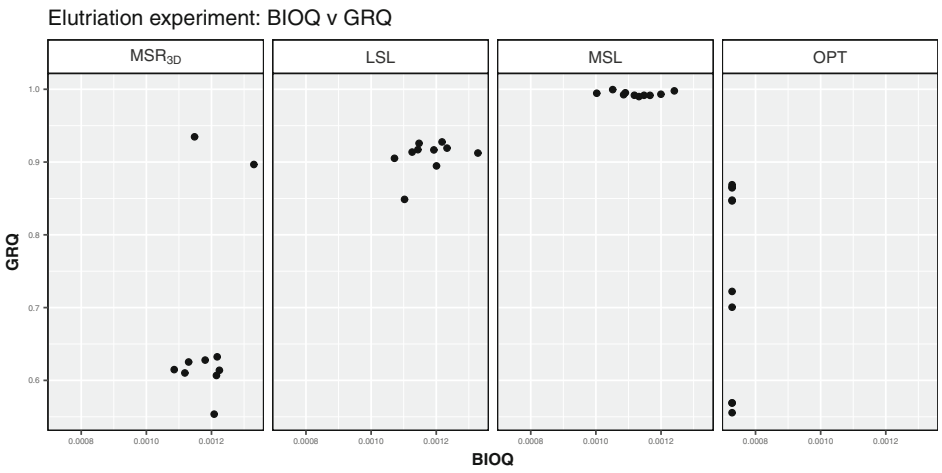

Fig. 3 *BIOQ* vs *GRQ* dispersion graph for each Elutriation solution of each experiment

the highest values of *TRIQ* and a sparsely dispersed distribution, thus this is a high-quality experiment.

The most valuable solution of all experiments is the tricluster TRI_2 of the *MSL* experiment. We can see in Fig. 6 its three graphic views showing that its high value of *GRQ* is reflected in the patterns depicted. Furthermore, in Table 8 we observe terms with moderately low *p*-value as *fermentation*, *vesicle fusion to plasma membrane* and *exocytosis*. *Fermentation* is a biological process that is part of the process called *energy derivation by oxidation of organic compounds* and, in turn, belongs to *the generation of precursor metabolites and energy* process and *the oxidation-reduction process*; *Vesicle fusion to plasma membrane* is a biological process that is part of the *exocytosis proccess*; the first term is a process of *cellular component organization* whereas the second is an *establishment of localization process*.

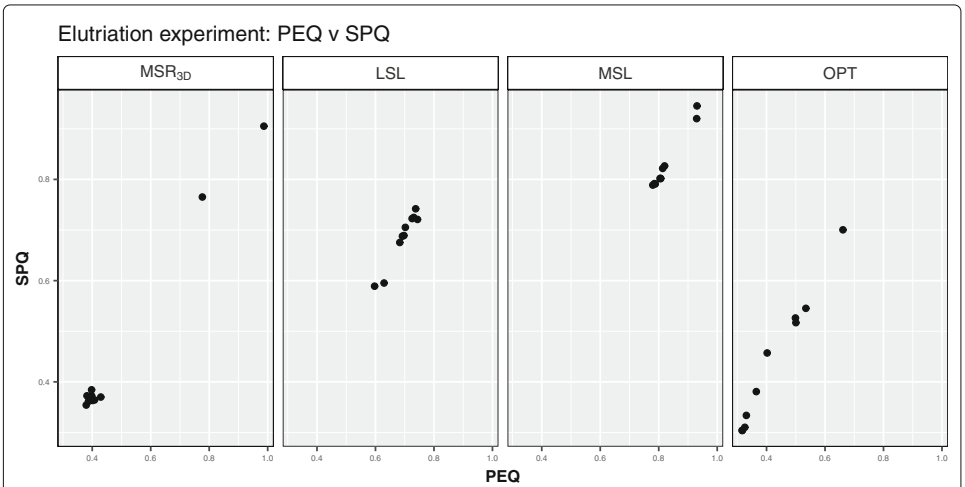

Fig. 4 *PEQ* vs *SPQ* dispersion graph for each Elutriation solution of each experiment

Table 6 Elutriation ranking table

EXPERIMENT	SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
MSL	TRI_2	0.493539244	0.001240807	0.997800501	0.930758605	0.945214193
MSL	TRI_1	0.492819589	0.001051563	0.999519361	0.929642164	0.9200791
MSL	TRI_5	0.480938627	0.001090473	0.995151057	0.820019348	0.826640002
MSL	TRI_4	0.478990452	0.001166044	0.991627468	0.813400974	0.821727882
MSL	TRI_7	0.478754345	0.00100258	0.994551592	0.806892319	0.801756043
MSL	TRI_8	0.478200414	0.001199622	0.993176848	0.804634565	0.801962727
MSL	TRI_9	0.477639873	0.001147773	0.991707562	0.805881226	0.801778004
MSL	TRI_3	0.476117422	0.001118134	0.991760508	0.78527775	0.791805282
MSL	TRI_6	0.475974935	0.001085123	0.992527638	0.779644523	0.788781847
MSL	TRI_{10}	0.475505918	0.001132268	0.989937077	0.788265502	0.791033557
MSR_{3D}	TRI_9	0.453932884	0.001330144	0.896650615	0.986952953	0.905198358
MSR_{3D}	TRI_{10}	0.45152166	0.001148045	0.934659815	0.776480244	0.765193987
LSL	TRI_1	0.444841672	0.001147115	0.925741144	0.737449684	0.741983455
LSL	TRI_2	0.444050729	0.001217804	0.927628308	0.725178526	0.722631553
LSL	TRI_8	0.441054484	0.001233014	0.919127603	0.730818495	0.724920229
LSL	TRI_6	0.440497687	0.001192667	0.916680329	0.743684691	0.720899755
LSL	TRI_7	0.437721769	0.001143537	0.916956452	0.702066665	0.705281726
LSL	TRI_3	0.434940552	0.001327826	0.912385527	0.697309668	0.689138899
LSL	TRI_5	0.433960732	0.001125858	0.913689264	0.683063155	0.675378795
LSL	TRI_4	0.431591352	0.001071675	0.905144571	0.692097513	0.6878562
LSL	TRI_9	0.41970611	0.001200657	0.894690897	0.629273489	0.595314967
OPT	TRI_7	0.40707391	0.000728	0.84656685	0.66128956	0.70037758
OPT	TRI_6	0.40017866	0.000728	0.86455717	0.53452807	0.5453116
LSL	TRI_{10}	0.399331119	0.001102605	0.848695823	0.597009139	0.589020606
OPT	TRI_4	0.39914764	0.000728	0.86882797	0.49884203	0.52621082
OPT	TRI_5	0.39749144	0.000728	0.86565058	0.50040614	0.51694181
OPT	TRI_3	0.38238284	0.000728	0.84763736	0.40215787	0.45712372
OPT	TRI_{10}	0.32249076	0.000728	0.72222718	0.33095502	0.33376653
OPT	TRI_2	0.31786223	0.000728	0.7005279	0.36494172	0.38080349
MSR_{3D}	TRI_3	0.292658244	0.001217404	0.632360796	0.39778358	0.384320901
MSR_{3D}	TRI_6	0.290639052	0.001129778	0.625267377	0.4293412	0.370003051
MSR_{3D}	TRI_1	0.289957861	0.001180518	0.627911696	0.400860192	0.363198285
MSR_{3D}	TRI_4	0.283891286	0.001085482	0.614807027	0.38713593	0.36137875
MSR_{3D}	TRI_2	0.283154367	0.001118227	0.610190268	0.397890126	0.372492792
MSR_{3D}	TRI_5	0.282839356	0.001224203	0.613862367	0.379462124	0.354184014
MSR_{3D}	TRI_8	0.281909708	0.001215203	0.606726347	0.407953984	0.36427737
MSR_{3D}	TRI_7	0.259777538	0.001208157	0.553613841	0.382072259	0.372486191
OPT	TRI_9	0.25904229	0.000728	0.56900655	0.31749708	0.30402
OPT	TRI_8	0.25896921	0.000728	0.56897207	0.31626703	0.30406432
OPT	TRI_1	0.25439082	0.000728	0.55556687	0.32575013	0.31025512

Table 7 Elutriation summary table

EXPERIMENT	BEST SOLUTION	BEST TRIQ	MEAN	STDEV
MSR_{3D}	TRI_9	0.453932884	0.317028196	0.072095449
LSL	TRI_1	0.444841672	0.432769621	0.013833102
MSL	TRI_2	0.493539244	0.480848082	0.006701521
OPT	TRI_7	0.40707391	0.33990298	0.064949576

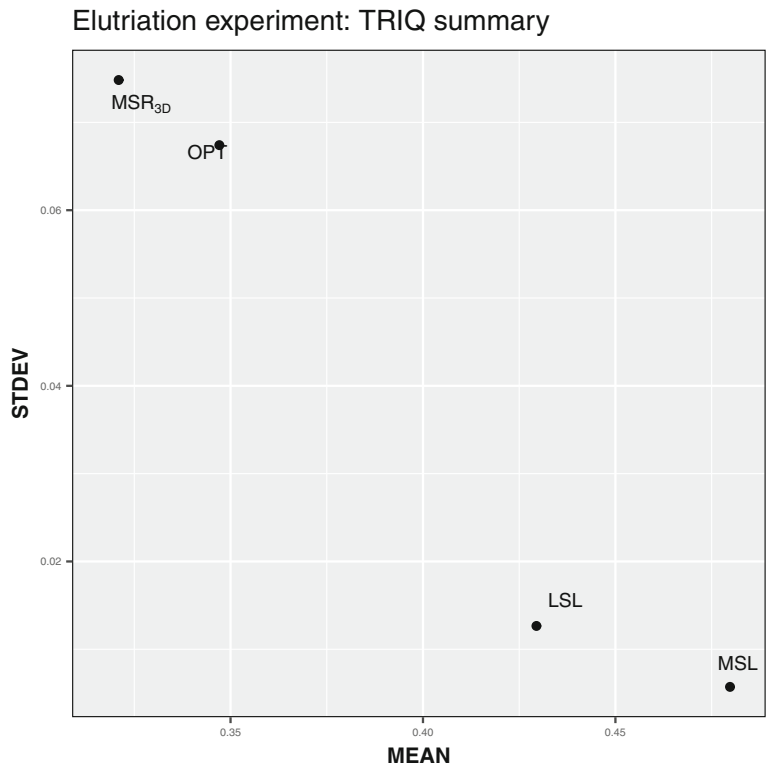

Fig. 5 *MEAN* vs *STDEV* dispersion graph for each Elutriation experiment

Mouse GDS4510 dataset

This batch corresponds to the mouse GDS4510 dataset. This dataset was obtained from GEO [35] with accession code GDS4510 and title *rd1 model of retinal degeneration: time course* [33]. In this experiment, the degeneration of retinal cells in different individuals of home mouse *(Mus musculus)* is analyzed over 4 days just after birth, specifically on days 2, 4, 6 and 8.

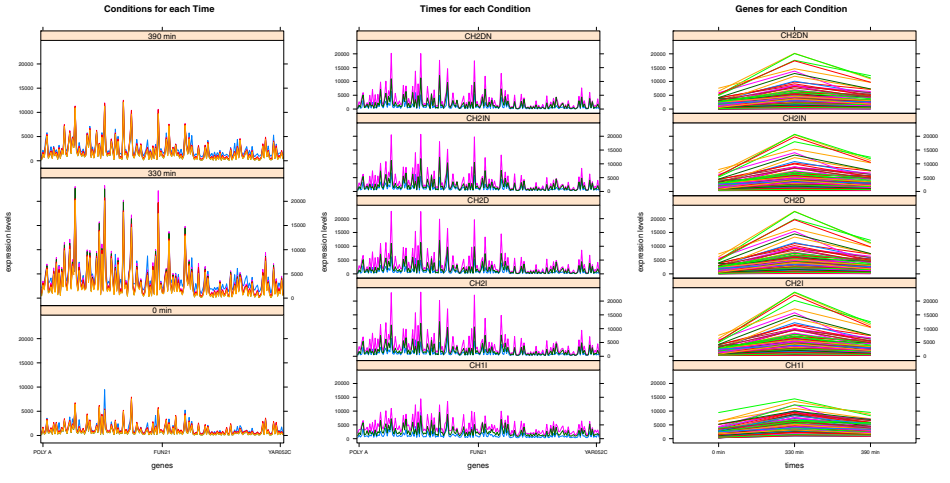

Fig. 6 TRI_2 graphic views of the Elutriation *MSL* experiment

Table 8 TRI_2 GO table of the *MSL* Elutriation experiment

TERM ID	TERM	*P*-VALUE
GO:0006113	Fermentation	7.39E-04
GO:0099500	Vesicle fusion to plasma membrane	0.001183063
GO:0006887	Exocytosis	0.001183063
GO:0140029	Exocytic process	0.001183063
GO:0045026	Plasma membrane fusion	0.00141327
GO:0000145	Exocyst	0.001794132
GO:0048193	Golgi vesicle transport	0.002271213
GO:0061025	Membrane fusion	0.002444417
GO:0051039	Positive regulation of transcription involved in meiotic cell cycle	0.002483587
GO:0051436	Negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	0.002483587
GO:0051439	Regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	0.002483587
GO:1904667	Negative regulation of ubiquitin protein ligase activity	0.002483587
GO:0032940	Secretion by cell	0.00251212
GO:0046903	Secretion	0.00251212
GO:0051049	Regulation of transport	0.002574368
GO:0061024	Membrane organization	0.002785422
GO:0051321	Meiotic cell cycle	0.003307558
GO:1903046	Meiotic cell cycle process	0.003307558
GO:0140013	Meiotic nuclear division	0.003307558
GO:0044275	Cellular carbohydrate catabolic process	0.004058262

For our purpose, we have created a dataset $D_{GDS4510_{3D}}$ composed of 22690 genes, 8 experimental conditions (one for each individual involved in the biological experiment) and 4 time points.

$D_{GDS4510_{3D}}$ has been used as the input of the *TriGen* and the *OPTtricluster* algorithm in four experiments: MSR_{3D}, *LSL*, *MSL* and, *OPT*.

GDS4510 MSR_{3D} experiment

We can verify in Table 9 how TRI_7 has the best value of *BIOQ*, *GRQ*, *PEQ*, *SPQ*. The *GRQ*, *PEQ* and *SPQ* indexes vary uniformly among all the solutions. *BIOQ* has a peak of TRI_7 which has the maximum value. The *TRIQ* values oscillate between 0.385 and 0.4 with the exception of TRI_7, therefore this is the best solution since it has the best value of *TRIQ*.

GDS4510 LSL experiment

In Table 10 we can see how TRI_1 has the best value of *BIOQ* and *GRQ* meanwhile TRI_2 has the best values of *PEQ* and *SPQ*. The *GRQ*, *PEQ* and *SPQ* values vary uniformly around

Table 9 MSR_{3D} GDS4510 solution table

SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
TRI_1	0.399937853	0.001348086	0.870211819	0.516112583	0.507469069
TRI_2	0.397972383	0.001177971	0.866535835	0.511042941	0.504338338
TRI_3	0.391066411	0.001255371	0.849235889	0.508613518	0.506273874
TRI_4	0.397028323	0.0014405	0.863853884	0.512208068	0.503122322
TRI_5	0.388644055	0.001187588	0.842929885	0.511734835	0.505831309
TRI_6	0.392316477	0.00190466	0.850869722	0.513791033	0.506534134
TRI_7	0.40677296	0.004479209	0.882477468	0.520330064	0.510517317
TRI_8	0.3851186	0.001240227	0.834606686	0.508323861	0.504792392
TRI_9	0.390891083	0.001281294	0.848296937	0.510926324	0.507706903
TRI_{10}	0.390730352	0.001137925	0.8484396	0.50930819	0.506402803

Table 10 *LSL* GDS4510 solution table

SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
TRI_1	0.435171938	0.005902591	0.902662935	0.6949723	0.728137064
TRI_2	0.427168871	0.002676716	0.885320349	0.700788434	0.733259027
TRI_3	0.422560787	0.002909652	0.887599773	0.648196371	0.673124663
TRI_4	0.42987221	0.004218243	0.901813346	0.657798589	0.68295641
TRI_5	0.416008121	0.002006917	0.869706286	0.658784335	0.683658641
TRI_6	0.41490654	0.001678815	0.866767983	0.661174595	0.686024185
TRI_7	0.417068507	0.001262531	0.87574877	0.648950843	0.673803828
TRI_8	0.40861261	0.001179271	0.854772399	0.649739272	0.672541022
TRI_9	0.42454573	0.001683951	0.890635412	0.663209255	0.68578253
TRI_{10}	0.417718487	0.002071672	0.874308872	0.657210052	0.681971994

a central value among the triclusters whereas *BIOQ* has peak values in TRI_1 and TRI_4. The *TRIQ* values oscillates between 0.40 and 0.43 being TRI_1, TRI_4 and TRI_9 the most outstanding solutions. We can conclude that TRI_1 is the best solution since it has the best value of *TRIQ*.

GDS4510 MSL experiment

For this experiment, we can observe in Table 11 how TRI_1 has the best value of *BIOQ* and *GRQ* meanwhile TRI_2 has the best value of *PEQ* and TRI_8 has the best value of *SPQ*. The *PEQ* and *SPQ* indexes of all solutions vary uniformly around 0.5 whereas all the *GRQ* values are close to 0.9. The *BIOQ* values oscillate between 0.0012 and 0.0019 reaching its higher value in the TRI_1 solution. The *TRIQ* values are in the $[0.42, 0.44]$ interval, therefore we can conclude that they are good results for this experiment. The highest value of *TRIQ* is reached by TRI_1, hence it is the best solution for this experiment.

GDS4510 OPT experiment

In Table 12 we can see how TRI_2 has the best value of *BIOQ*, TRI_4 has the best value of *GRQ* and, TRI_9 and TRI_1 have the best value of *PEQ* and *SPQ* respectively. The *BIOQ* values vary among $[0.0012, 0.0016]$ interval with the exception of TRI_2 and TRI_3 whilst the *GRQ* values vary uniformly around the 0.80 value excepting TRI_4. The *PEQ* and *SPQ* values oscillate among the $[0.5, 0.8]$ interval.The highest value of *TRIQ* is reached by TRI_4, thus it is the best solution for this experiment.

Table 11 *MSL* GDS4510 solution table

SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
TRI_1	0.446289279	0.003624207	0.990551544	0.496833632	0.468297522
TRI_2	0.430638622	0.001399471	0.945717127	0.515568434	0.51747227
TRI_3	0.429698209	0.00149303	0.943951098	0.506740977	0.520684131
TRI_4	0.425844616	0.001388422	0.935696236	0.506485147	0.510953062
TRI_5	0.431185402	0.001224915	0.948344121	0.507510722	0.517195194
TRI_6	0.422692807	0.001367523	0.927464145	0.507112049	0.513355693
TRI_7	0.429129078	0.001401202	0.944156645	0.501640545	0.513675839
TRI_8	0.436192976	0.001999141	0.958438573	0.512285251	0.524074273
TRI_9	0.433173322	0.001604555	0.95182792	0.510885656	0.521911883
TRI_{10}	0.422409162	0.001390319	0.928018397	0.501351072	0.508781791

Table 12 *OPT* GDS4510 solution table

SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
TRI_1	0.42009257	0.00134246	0.83749772	0.8574642	0.8309809
TRI_2	0.42005895	0.00356033	0.85404927	0.82019206	0.71298946
TRI_3	0.42901805	0.00256279	0.87133294	0.85498273	0.72908679
TRI_4	0.44490172	0.00136003	0.93478154	0.77222529	0.63395639
TRI_5	0.37648925	0.00128696	0.81223451	0.51226675	0.50677264
TRI_6	0.37500834	0.00122195	0.80916885	0.50865504	0.50594149
TRI_7	0.37783613	0.00125442	0.81768005	0.50153608	0.50120198
TRI_8	0.37545313	0.00144891	0.80990327	0.50841987	0.50692736
TRI_9	0.43860855	0.00167827	0.89471647	0.86719834	0.73045828
TRI_{10}	0.37115418	0.00120689	0.80002563	0.50727999	0.50352975

GDS4510 summary

We can see how the solutions are distributed regarding *BIOQ* and *GRQ* in Fig. 7; we observe that all points of all experiments are concentrated in a *BIOQ* interval of $[0.0011, 0.0059]$. Regarding the *GRQ* values, the MSR_{3D} and *LSL* experiments have all the solutions in the $[0.83, 0.90]$ interval, the *MSL* experiment has all the solutions in the $[0.92, 0.99]$ interval and, the *OPT* experiment has all the solutions in the $[0.80, 0.95]$ interval. Regarding the *PEQ* and *SPQ* distribution we can see in Fig. 8 how the majority of solutions are concentrated around the point $PEQ = 0.5, SPQ = 0.5$ in the MSR_{3D} and *MSL* experiments, meanwhile the solutions of *LSL* experiment are concentrated in the interval $[0.625, 0.75]$ for *PEQ* and *SPQ* values and, the *OPT* experiment has his solutions dispersed in two groups: one group around the $PEQ = 0.5, SPQ = 0.5$ point and the other in an interval of $[0.60, 0.83]$ for both *PEQ* and *SPQ* values.

A global *TRIQ*-based ranking of solutions is shown in Table 13. The *MSL*, *LSL* and a part of *OPT* solutions are placed alternatively on the first positions and the MSR_{3D} and the remaining of *OPT* solutions are in the last positions.

We can see in Table 14 how the GDS4510 *MSL* experiment has the best value of the mean of *TRIQ* and the four experiments have low values of standard deviation having the MSR_{3D} experiment the lowest value but very close to the *MSL* one. This fact implies

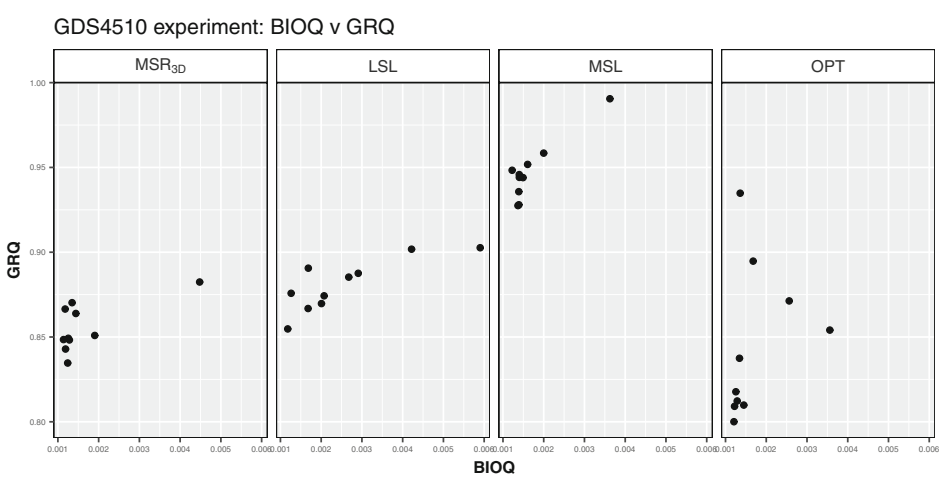

Fig. 7 *BIOQ* vs *GRQ* dispersion graph for each GDS4510 solution of each experiment

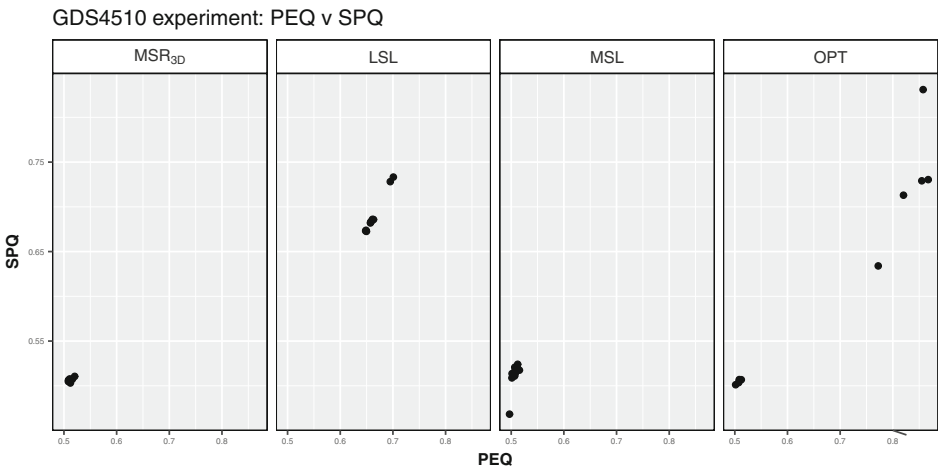

Fig. 8 *PEQ* vs *SPQ* dispersion graph for each GDS4510 solution of each experiment

that the four experiments have a low sparse distribution and solutions with high quality. We can see in Fig. 9 how the MSR_{3D}, *LSL* and, *MSL* points are located on the bottom side of the graph meanwhile the *OPT* point is located in a high level of the standard deviation axis; on the other hand, *LSL* and, *MSL* points are located on the right side of the graph meanwhile the MSR_{3D} and *OPT* points are located in a left level of the average axis. Hence, in terms of standard deviation and average, we can conclude that *MSL* is the best experiment.

The most valuable solution of all experiments is the tricluster TRI_1 of the *MSL* experiment. We can see in Fig. 10 how this solution depicts very uniform patterns consistent with the *GRQ* value. Also, we can see in Table 15 that this solution has Gene Ontology terms with low *p*-value such as *sensory perception of chemical stimulus*, *olfactory receptor activity* or *detection of chemical stimulus involved in sensory perception of smell*. The term *olfactory receptor activity* is a molecular function that combining with an odorant and transmitting the signal from one side of the membrane to the other to initiate a change in cell activity in response to detection of smell; this function is part of the biological process *detection of chemical stimulus involved in sensory perception of smell* that is the series of events involved in the perception of smell in which an olfactory chemical stimulus is received and converted into a molecular signal. Finally, that process is framed in a more general biological process called *sensory perception of chemical stimulus* that is the series of events required for an organism to receive a sensory chemical stimulus, convert it to a molecular signal, and recognize and characterize the signal.

Human GDS4472 dataset

The dataset, corresponding to this batch, has been obtained from GEO [35] under code GDS4472 titled *Transcription factor oncogene OTX2 silencing effect on D425 medulloblastoma cell line: time course* [34]. In this experiment, the effect of doxycycline on medulloblastoma cancerous cells at six times after induction (0, 8, 16, 24, 48 and 96 h) had been studied.

Table 13 GDS4510 ranking table

EXPERIMENT	SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
MSL	TRI_1	0.446289279	0.003624207	0.990551544	0.496833632	0.468297522
OPT	TRI_4	0.44490172	0.00136003	0.93478154	0.77222529	0.63395639
OPT	TRI_9	0.43860855	0.00167827	0.89471647	0.86719834	0.73045828
MSL	TRI_8	0.436192976	0.001999141	0.958438573	0.512285251	0.524074273
LSL	TRI_1	0.435171938	0.005902591	0.902662935	0.6949723	0.728137064
MSL	TRI_9	0.433173322	0.001604555	0.95182792	0.510885656	0.521911883
MSL	TRI_5	0.431185402	0.001224915	0.948344121	0.507510722	0.517195194
MSL	TRI_2	0.430638622	0.001399471	0.945717127	0.515568434	0.51747227
LSL	TRI_4	0.42987221	0.004218243	0.901813346	0.657798589	0.68295641
MSL	TRI_3	0.429698209	0.00149303	0.943951098	0.506740977	0.520684131
MSL	TRI_7	0.429129078	0.001401202	0.944156645	0.501640545	0.513675839
OPT	TRI_3	0.42901805	0.00256279	0.87133294	0.85498273	0.72908679
LSL	TRI_2	0.427168871	0.002676716	0.885320349	0.700788434	0.733259027
MSL	TRI_4	0.425844616	0.001388422	0.935696236	0.506485147	0.510953062
LSL	TRI_9	0.42454573	0.001683951	0.890635412	0.663209255	0.68578253
MSL	TRI_6	0.422692807	0.001367523	0.927464145	0.507112049	0.513355693
LSL	TRI_3	0.422560787	0.002909652	0.887599773	0.648196371	0.673124663
MSL	TRI_{10}	0.422409162	0.001390319	0.928018397	0.501351072	0.508781791
OPT	TRI_1	0.42009257	0.00134246	0.83749772	0.8574642	0.8309809
OPT	TRI_2	0.42005895	0.00356033	0.85404927	0.82019206	0.71298946
LSL	TRI_{10}	0.417718487	0.002071672	0.874308872	0.657210052	0.681971994
LSL	TRI_7	0.417068507	0.001262531	0.87574877	0.648950843	0.673803828
LSL	TRI_5	0.416008121	0.002006917	0.869706286	0.658784335	0.683658641
LSL	TRI_6	0.41490654	0.001678815	0.866767983	0.661174595	0.686024185
LSL	TRI_8	0.40861261	0.001179271	0.854772399	0.649739272	0.672541022
MSR_{3D}	TRI_7	0.40677296	0.004479209	0.882477468	0.520330064	0.510517317
MSR_{3D}	TRI_1	0.399937853	0.001348086	0.870211819	0.516112583	0.507469069
MSR_{3D}	TRI_2	0.397972383	0.001177971	0.866535835	0.511042941	0.504338338
MSR_{3D}	TRI_4	0.397028323	0.0014405	0.863853884	0.512208068	0.503122322
MSR_{3D}	TRI_6	0.392316477	0.00190466	0.850869722	0.513791033	0.506534134
MSR_{3D}	TRI_3	0.391066411	0.001255371	0.849235889	0.508613518	0.506273874
MSR_{3D}	TRI_9	0.390891083	0.001281294	0.848296937	0.510926324	0.507706903
MSR_{3D}	TRI_{10}	0.390730352	0.001137925	0.8484396	0.50930819	0.506402803
MSR_{3D}	TRI_5	0.388644055	0.001187588	0.842929885	0.511734835	0.505831309
MSR_{3D}	TRI_8	0.3851186	0.001240227	0.834606686	0.508323861	0.504792392
OPT	TRI_7	0.37783613	0.00125442	0.81768005	0.50153608	0.50120198
OPT	TRI_5	0.37648925	0.00128696	0.81223451	0.51226675	0.50677264
OPT	TRI_8	0.37545313	0.00144891	0.80990327	0.50841987	0.50692736
OPT	TRI_6	0.37500834	0.00122195	0.80916885	0.50865504	0.50594149
OPT	TRI_{10}	0.37115418	0.00120689	0.80002563	0.50727999	0.50352975

Table 14 GDS4510 summary table

EXPERIMENT	BEST SOLUTION	BEST TRIQ	MEAN	STDEV
MSR_{3D}	TRI_7	0.40677296	0.39404785	0.006348192
LSL	TRI_1	0.435171938	0.42136338	0.007979308
MSL	TRI_1	0.446289279	0.430725347	0.006987671
OPT	TRI_4	0.44490172	0.402862087	0.030140772

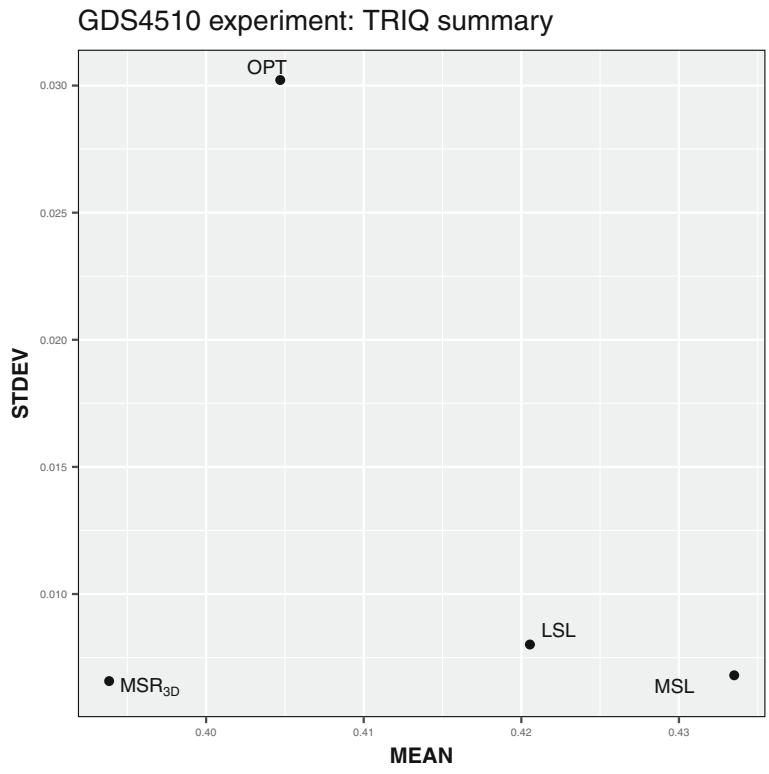

Fig. 9 *MEAN* vs *STDEV* dispersion graph for each GDS4510 experiment

Our input dataset $D_{GSD4472_{3D}}$ is composed of 54675 genes, 4 conditions (one for each individual involved) and 6 time points (one per hour) and has been used as the input of the *TriGen* and the *OPTtricluster* algorithm in four experiments: MSR_{3D}, *LSL*, *MSL* and, *OPT*.

GDS4472 MSR_{3D} experiment

For this experiment, TRI_4 has the best value of *BIOQ*, TRI_6 has the best value of *PEQ*, TRI_3 has the best value of *SPQ* and TRI_5 has the best value of *GRQ* as you can see Table 16.

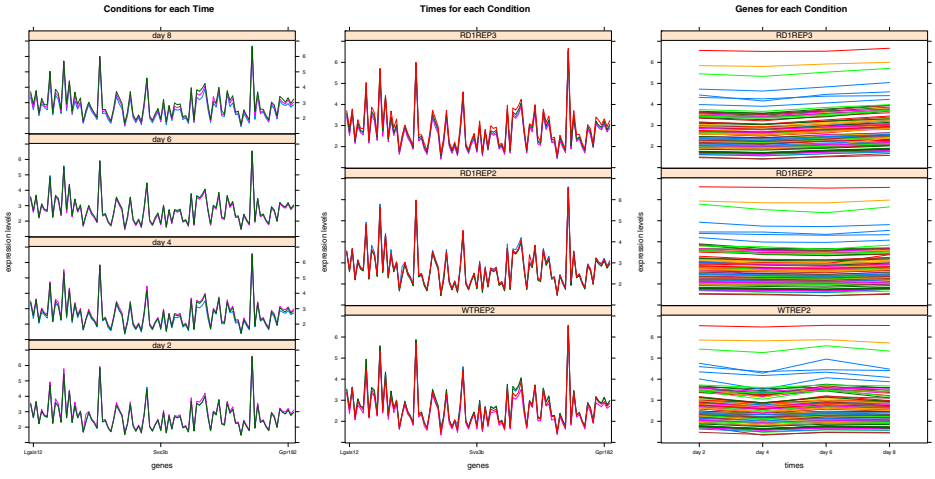

Fig. 10 TRI_1 graphic views of the GDS4510 *MSL* experiment

Table 15 TRI_1 GO table of the *MSL* GDS4510 experiment

TERM ID	TERM	*P*-VALUE
GO:0007606	Sensory perception of chemical stimulus	1.68E-25
GO:0004984	Olfactory receptor activity	6.56E-19
GO:0050911	Detection of chemical stimulus involved in sensory perception of smell	6.56E-19
GO:0050907	Detection of chemical stimulus involved in sensory perception	2.97E-18
GO:0004930	G-protein coupled receptor activity	4.68E-17
GO:0007186	G-protein coupled receptor signaling pathway	4.68E-17
GO:0007608	Sensory perception of smell	6.93E-16
GO:0009593	Detection of chemical stimulus	1.16E-15
GO:0007600	Sensory perception	5.28E-15
GO:0050906	Detection of stimulus involved in sensory perception	8.32E-14
GO:0004872	Receptor activity	9.34E-14
GO:0060089	Molecular transducer activity	6.27E-13
GO:0004888	Transmembrane signaling receptor activity	8.08E-13
GO:0050877	Nervous system process	1.07E-12
GO:0099600	Transmembrane receptor activity	2.01E-12
GO:0038023	Signaling receptor activity	1.43E-11
GO:0004871	Signal transducer activity	2.69E-11
GO:0051606	Detection of stimulus	1.64E-10
GO:0003008	System process	1.09E-09
GO:0005549	Odorant binding	1.85E-08

The *PEQ* and *SPQ* values of the solutions oscillate around 0.64 and the *GRQ* values vary between 0.76 and 0.64; the *BIOQ* index oscillates around 0.0014 reaching two peaks at TRI_4 and TRI_8. In general, the *TRIQ* value of solutions are in $[0.32, 0.37]$ having TRI_3 and TRI_7 as outstanding ones and TRI_5 as the best solution in this experiment.

GDS4472 LSL experiment

We can verify in Table 17 how TRI_1 has the best values of *BIOQ*, *GRQ*, *PEQ* and *SPQ*. In general, the *GRQ*, *PEQ* and *SPQ* indexes of the solutions depicts homogeneous values with the exception of TRI_1 where they reach their maximum; regarding *BIOQ* values, those reach three peaks at TRI_1, TRI_4 and TRI_{10}. The *TRIQ* values vary between 0.39 and 0.44 being TRI_1 the best solution of this experiment.

GDS4472 MSL experiment

In Table 18 we can see how TRI_9 has the best values of *BIOQ* and *GRQ* while TRI_7 has the best value of *PEQ* and TRI_{10} has the best value of *SPQ*. The *PEQ* values of the solutions

Table 16 MSR_{3D} GDS4472 solution table

SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
TRI_1	0.339109333	0.001444791	0.696219908	0.596698979	0.601280513
TRI_2	0.321761941	0.001591523	0.645157719	0.633303534	0.624758294
TRI_3	0.363970471	0.001440455	0.742093089	0.650828401	0.677431755
TRI_4	0.343765956	0.001732664	0.69802844	0.623399523	0.650365438
TRI_5	0.370128492	0.001337649	0.761586904	0.637388072	0.659110049
TRI_6	0.360725206	0.001406179	0.730735981	0.688724917	0.665829566
TRI_7	0.366252916	0.001263468	0.750692098	0.655100071	0.651786783
TRI_8	0.351001074	0.00159526	0.709109493	0.674238924	0.656954002
TRI_9	0.327754495	0.001401494	0.664697595	0.606214508	0.617279679
TRI_{10}	0.360821995	0.001434449	0.743345617	0.631027541	0.624302919

Table 17 *LSL* GDS4472 solution table

SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
TRI_1	0.447346181	0.027287612	0.923852614	0.69377633	0.589450252
TRI_2	0.392576223	0.004031468	0.862302229	0.468881448	0.443910489
TRI_3	0.409737004	0.002294049	0.886097803	0.570674741	0.512342415
TRI_4	0.421749212	0.00967779	0.895313993	0.60768017	0.568014215
TRI_5	0.402016193	0.002568856	0.869691073	0.55912722	0.497979503
TRI_6	0.394065329	0.00467901	0.864596579	0.474398477	0.443345363
TRI_7	0.39497655	0.005644179	0.865341762	0.477395239	0.442959875
TRI_8	0.397055929	0.005748916	0.868777413	0.482543181	0.450866935
TRI_9	0.40510461	0.007434596	0.881069048	0.514114336	0.465079524
TRI_{10}	0.411954946	0.019662416	0.89416486	0.458207225	0.430948663

vary in the $[0.43, 0.46]$ interval and the *SEQ* values are in the $[0.40, 0.44]$ interval while all solutions have high *GRQ* values close to 0.90; the *BIOQ* values have three peaks at TRI_5, TRI_7 and TRI_9. Regarding *TRIQ* values, they vary in $[0.40, 0.42]$ interval being TRI_1, TRI_5 and TRI_7 the outstanding solutions and being TRI_9 the best solution.

GDS4472 OPT experiment

For this experiment, TRI_5 has the best value of *BIOQ*, TRI_{10} has the best value of *GRQ* and *SPQ* and, TRI_8 has the best value of *PEQ* as you can see Table 19. The *BIOQ* index oscillates around 0.0015 reaching three peaks at TRI_5, TRI_9 and, TRI_{10}. The *GRQ* index vary in the $[0.6, 07]$ interval reaching an outstanding value in the TRI_{10} solution. Regarding the *PEQ* values they vary in a interval of $[0.42, 0.86]$ and the *SPQ* values in the $[0.34, 0.76]$ interval. The *TRIQ* values vary between 0.28 and 0.44 being TRI_{10} the best solution of this experiment.

GDS4472 summary

We can observe in Fig. 11 how the solutions of the four experiments are in a *BIOQ* interval of $[0.0012, 0.0272]$ meanwhile the *GRQ* values of the solutions of MSR_{3D} are in the $[0.6451, 0.7615]$ interval, the solutions of *LSL* are in the $[0.8623, 0.8953]$ interval, the solutions of *MSL* are in the $[0.8964, 0.9238]$ interval and, the solutions of *OPT* are in the $[0.6, 0.7]$ interval with an outstanding point near to $GRQ = 0.92$. Regarding the *PEQ* and *SPQ* solutions distribution we can see in Fig. 12 how the *PEQ* and *SPQ* of MSR_{3D} are concentrated in the $[0.50, 0.75]$ interval, the values *PEQ* and *SPQ* of *LSL* are in the $[0.325, 0.75]$ interval, the values *PEQ* and *SPQ* of *MSL* are in the $[0.325, 0.50]$

Table 18 *MSL* GDS4472 solution table

SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
TRI_1	0.413005918	0.008623332	0.909739803	0.463874665	0.432091958
TRI_2	0.406682712	0.005351847	0.901242812	0.449739986	0.420453301
TRI_3	0.404078935	0.004069221	0.896447319	0.445616204	0.423691724
TRI_4	0.409123273	0.004869646	0.9053715	0.456215467	0.43458153
TRI_5	0.410786658	0.011209144	0.903088937	0.453954127	0.424976095
TRI_6	0.404207143	0.004999521	0.896798986	0.44398627	0.415769491
TRI_7	0.411937377	0.012628523	0.901459314	0.468134175	0.432653616
TRI_8	0.405644251	0.0030952	0.902252364	0.445061054	0.418853066
TRI_9	0.42006885	0.025664213	0.912118818	0.439476488	0.408307841
TRI_{10}	0.41078403	0.006450477	0.90556104	0.465916366	0.440771152

Table 19 *OPT* GDS4472 solution table

SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
TRI_1	0.361091084	0.001165855	0.728642443	0.841096481	0.539927104
TRI_2	0.302530473	0.001445096	0.649190316	0.42227834	0.420357627
TRI_3	0.298417139	0.001567996	0.639083858	0.421848063	0.418143898
TRI_4	0.290997577	0.0013925	0.620125263	0.423388521	0.421635907
TRI_5	0.353327655	0.00233497	0.7175687	0.832938328	0.469715461
TRI_6	0.298612766	0.001430159	0.640127397	0.421427176	0.415507376
TRI_7	0.282392223	0.0018726	0.610369316	0.397933947	0.348229987
TRI_8	0.35196608	0.00159536	0.707996999	0.865220464	0.49417155
TRI_9	0.328919371	0.001916523	0.649746035	0.838138155	0.523115758
TRI_{10}	0.446233789	0.002289266	0.924944835	0.740548556	0.761675883

interval and, the values *PEQ* and *SPQ* of *OPT* are dispersed in three groups: the first in the $[0.42, 0.45]$ interval for *PEQ* and *SPQ*, the second in the $[0.70, 0.85]$ interval for *PEQ* and the $[0.46, 0.54]$ interval for *SPQ* and the third, that is a single point, in $PEQ = 0.74$, $SPQ = 0.76$.

We can see the global *TRIQ*-based ranking of solutions in Table 20; the *MSL* solutions, one *OPT* solution and, the *LSL* solutions are placed alternatively on the first positions and the MSR_{3D} and the remaining of *OPT* solutions are on the last positions.

We can see in Table 21 how the *MSL* experiment has the best value of the average and standard deviation of *TRIQ*, however, the *LSL* experiment has the best tricluster closely followed by the *OPT* experiment. In Fig. 13 we can see how the *MSL* is placed in the bottom-right position being the best experiment in terms of standard deviation and average.

The most valuable solution of all experiments is the tricluster TRI_1 of the *LSL* experiment. This solution depicts very uniform patterns since has a very high *GRQ* value, we can check this fact in Fig. 14. Also, we can see in Table 22 that this solution has Gene Ontology terms with very low *p*-value such as *SRP-dependent cotranslational protein targeting to membrane, nuclear-transcribed mRNA catabolic process, nonsense-mediated decay* or *ribonucleoprotein complex*.

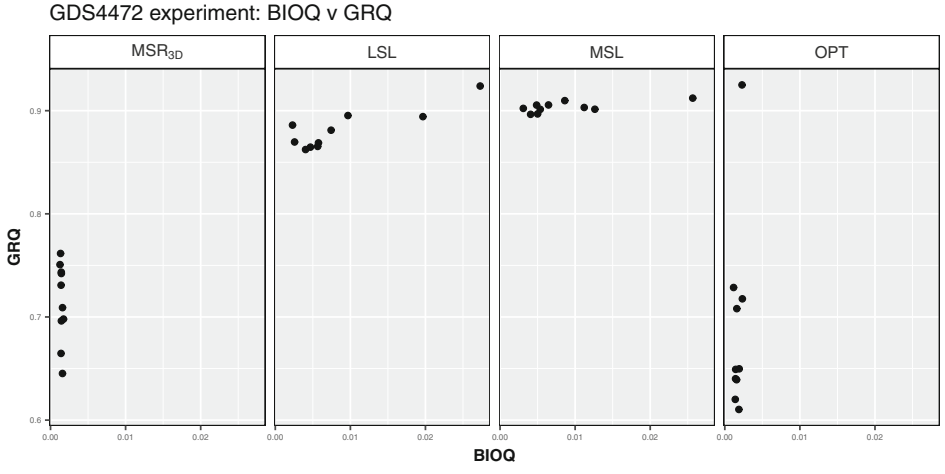

Fig. 11 *BIOQ* vs *GRQ* dispersion graph for each GDS4472 solution of each experiment

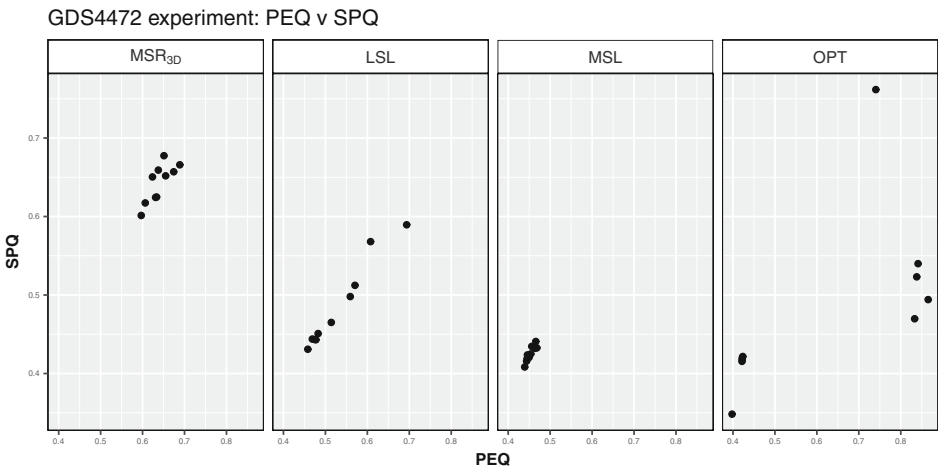

Fig. 12 *PEQ* vs *SPQ* dispersion graph for each GDS4472 solution of each experiment

The *SRP-dependent cotranslational protein targeting to membrane* process is described as the targeting of proteins to a membrane that occurs during translation and is dependent upon two key components, the signal-recognition particle (SRP) and the SRP receptor. SRP is a cytosolic particle that transiently binds to the endoplasmic reticulum (ER) signal sequence in a nascent protein, to the large ribosomal unit, and to the SRP receptor in the ER membrane; it is a protein targeting process that occurs in the intracellular component and is part of the cellular protein localization process. The *nuclear-transcribed mRNA catabolic process, nonsense-mediated decay* is a biological process that describes the nonsense-mediated decay pathway for nuclear-transcribed mRNAs degrades mRNAs in which an amino-acid codon has changed to a nonsense codon; this prevents the translation of such mRNAs into truncated, and potentially harmful, proteins; it is a negative regulation of gene expression process that negatively regulates the macromolecule metabolic process. Finally the *ribonucleoprotein complex* is a cellular component that is defined as a macromolecular complex containing both protein and RNA molecules.

Conclusions and discussion

Although triclustering has emerged as an essential task to study 3D datasets, there is no consensus on how to evaluate tricluster solutions obtained from each data set. Different authors validate their triclusters on different measures, with correlation, graphic validation and Gene Ontology terms being the most common ones. In this work we have presented a tricluster validation measure, *TRIQ*, a single evaluation measure that combines the information from the three aforementioned sources of validation.

We have applied *TRIQ* to three different datasets: the yeast cell cycle (*Saccharomyces Cerevisiae*), in particular the elutriation experiment, an experiment with mice (*Mus Musculus*) called GDS4510 and data from an experiments with humans (*Homo Sapiens*) called GDS4472.

We have shown that *TRIQ* has successfully resumed the three validation measures (correlation, graphic validation and Gene Ontology terms) yielding the same validation results

Table 20 GDS4472 ranking table

EXPERIMENT	SOLUTION	TRIQ	BIOQ	GRQ	PEQ	SPQ
LSL	TRI_1	0.447346181	0.027287612	0.923852614	0.69377633	0.589450252
OPT	TRI_{10}	0.446233789	0.002289266	0.924944835	0.740548556	0.761675883
LSL	TRI_4	0.421749212	0.00967779	0.895313993	0.60768017	0.568014215
MSL	TRI_9	0.42006885	0.025664213	0.912118818	0.439476488	0.408307841
MSL	TRI_1	0.413005918	0.008623332	0.909739803	0.463874665	0.432091958
LSL	TRI_{10}	0.411954946	0.019662416	0.89416486	0.458207225	0.430948663
MSL	TRI_7	0.411937377	0.012628523	0.901459314	0.468134175	0.432653616
MSL	TRI_5	0.410786658	0.011209144	0.903088937	0.453954127	0.424976095
MSL	TRI_{10}	0.41078403	0.006450477	0.90556104	0.465916366	0.440771152
LSL	TRI_3	0.409737004	0.002294049	0.886097803	0.570674741	0.512342415
MSL	TRI_4	0.409123273	0.004869646	0.9053715	0.456215467	0.43458153
MSL	TRI_2	0.406682712	0.005351847	0.901242812	0.449739986	0.420453301
MSL	TRI_8	0.405644251	0.0030952	0.902252364	0.445061054	0.418853066
LSL	TRI_9	0.40510461	0.007434596	0.881069048	0.514114336	0.465079524
MSL	TRI_6	0.404207143	0.004999521	0.896798986	0.44398627	0.415769491
MSL	TRI_3	0.404078935	0.004069221	0.896447319	0.445616204	0.423691724
LSL	TRI_5	0.402016193	0.002568856	0.869691073	0.55912722	0.497979503
LSL	TRI_8	0.397055929	0.005748916	0.868777413	0.482543181	0.450866935
LSL	TRI_7	0.39497655	0.005644179	0.865341762	0.477395239	0.442959875
LSL	TRI_6	0.394065329	0.00467901	0.864596579	0.474398477	0.443345363
LSL	TRI_2	0.392576223	0.004031468	0.862302229	0.468881448	0.443910489
MSR_{3D}	TRI_5	0.370128492	0.001337649	0.761586904	0.637388072	0.659110049
MSR_{3D}	TRI_7	0.366252916	0.001263468	0.750692098	0.655100071	0.651786783
MSR_{3D}	TRI_3	0.363970471	0.001440455	0.742093089	0.650828401	0.677431755
OPT	TRI_1	0.361091084	0.001165855	0.728642443	0.841096481	0.539927104
MSR_{3D}	TRI_{10}	0.360821995	0.001434449	0.743345617	0.631027541	0.624302919
MSR_{3D}	TRI_6	0.360725206	0.001406179	0.730735981	0.688724917	0.665829566
OPT	TRI_5	0.353327655	0.00233497	0.7175687	0.832938328	0.469715461
OPT	TRI_8	0.35196608	0.00159536	0.707996999	0.865220464	0.49417155
MSR_{3D}	TRI_8	0.351001074	0.00159526	0.709109493	0.674238924	0.656954002
MSR_{3D}	TRI_4	0.343765956	0.001732664	0.69802844	0.623399523	0.650365438
MSR_{3D}	TRI_1	0.339109333	0.001444791	0.696219908	0.596698979	0.601280513
OPT	TRI_9	0.328919371	0.001916523	0.649746035	0.838138155	0.523115758
MSR_{3D}	TRI_9	0.327754495	0.001401494	0.664697595	0.606214508	0.617279679
MSR_{3D}	TRI_2	0.321761941	0.001591523	0.645157719	0.633303534	0.624758294
OPT	TRI_2	0.302530473	0.001445096	0.649190316	0.42227834	0.420357627
OPT	TRI_6	0.298612766	0.001430159	0.640127397	0.421427176	0.415507376
OPT	TRI_3	0.298417139	0.001567996	0.639083858	0.421848063	0.418143898
OPT	TRI_4	0.290997577	0.0013925	0.620125263	0.423388521	0.421635907
OPT	TRI_7	0.282392223	0.0018726	0.610369316	0.397933947	0.348229987

Table 21 GDS4472 summary table

EXPERIMENT	BEST SOLUTION	BEST TRIQ	MEAN	STDEV
MSR_{3D}	TRI_5	0.370128492	0.350529188	0.016814529
LSL	TRI_1	0.447346181	0.407658218	0.016734175
MSL	TRI_9	0.42006885	0.409631915	0.004869533
OPT	TRI_{10}	0.446233789	0.331448816	0.049451114

Fig. 13 *MEAN* vs *STDEV* dispersion graph for each GDS4472 experiment

as in [27] where each of the components of *TRIQ* (*BIOQ*, *GRQ*, *PEQ*, and *SPQ*) where applied separately. In that publication we presented the *MSL* measure, comparing it to MSR_{3D} and *LSL*, with the same datasets used in this article. We concluded that *MSL* was the best fitness function. In this publication, we have seen how *MSL* has obtained the best general results, with high values of *TRIQ* and low standard deviation for all solutions presented. Therefore, we can conclude that *TRIQ* has been successful in representing

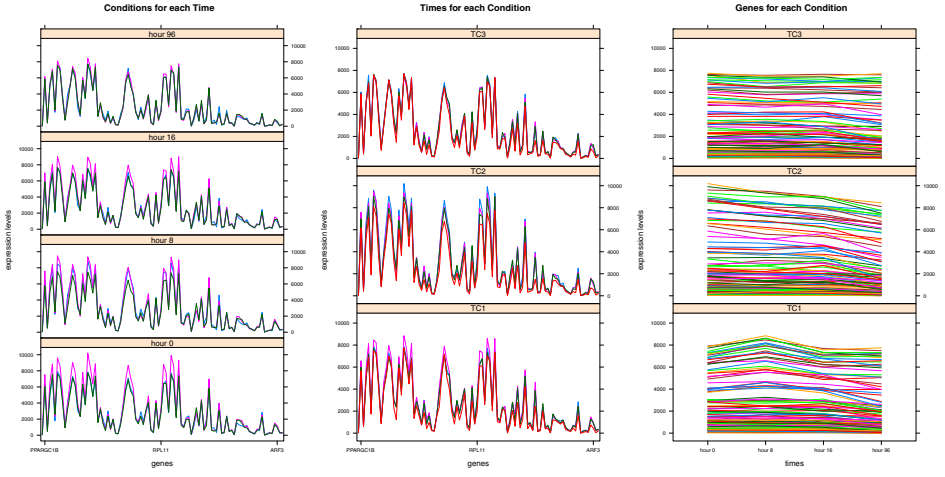

Fig. 14 TRI_1 graphic views of the GDS4472 *LSL* experiment

Table 22 TRI_1 GO table of the *LSL* GDS4472 experiment

TERM ID	TERM	*P*-VALUE
GO:1990904	Ribonucleoprotein complex	1.52E-41
GO:0030529	Intracellular ribonucleoprotein complex	1.52E-41
GO:0044403	Symbiosis, encompassing mutualism through parasitism	6.40E-40
GO:0044419	Interspecies interaction between organisms	1.52E-39
GO:0016032	Viral process	2.20E-39
GO:0045047	Protein targeting to ER	2.91E-38
GO:0006613	Cotranslational protein targeting to membrane	3.78E-38
GO:0072599	Establishment of protein localization to endoplasmic reticulum	8.16E-38
GO:0006614	SRP-dependent cotranslational protein targeting to membrane	8.33E-37
GO:0070972	Protein localization to endoplasmic reticulum	6.26E-36
GO:0005840	Ribosome	6.47E-36
GO:0022626	Cytosolic ribosome	2.56E-35
GO:0019080	Viral gene expression	1.03E-34
GO:0043624	Cellular protein complex disassembly	1.10E-34
GO:0022618	Ribonucleoprotein complex assembly	1.54E-34
GO:0071826	Ribonucleoprotein complex subunit organization	1.96E-34
GO:0000184	Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	4.39E-34
GO:0044391	Ribosomal subunit	5.80E-34
GO:0001677	Formation of translation initiation ternary complex	6.45E-34
GO:0006412	Translation	6.45E-34

and summarizing the individual values provided by *BIOQ*, *GRQ*, *PEQ*, and *SPQ*. Furthermore, we have applied *TRIQ* to results from another algorithm, *OPTRicluster*, and we have shown how TRIQ has been a valid tool to compare results from different algorithms in a quantitative straightforward manner.

For the case of triclustering being applied to not biologically related fields as in [36], *TRIQ* can also cope with the analysis of the tricluster solutions thanks to the weighting system (see "Methods" section), which allows for each term to be included or removed in the final measure.

Acknowledgements
The authors thank financial support by the Spanish Ministry of Science and Technology project TIN2014-55894-C2-1-R and Junta de Andalucía's project P12-TIC-7528.

Funding
Spanish Ministry of Science and Technology project TIN2014-55894-C2-1-R and Junta de Andalucía's project P12-TIC-7528.

Availability of data and materials
TriGen and *TRIQ* application resources (TrLab Application): https://github.com/davgutavi/trlab-application/releases. Yest Cell Cycle resources: http://genome-www.stanford.edu/cellcycle/. Mouse GDS4510 resources: https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS4510. Human GDS4472 resources: https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS4472.

Authors' contributions
Conceived and designed the experiments: DGA, CRE. Analyzed the data: DGA. Wrote the first draft of the manuscript: DGA, RGR, FJGC, CRE. Contributed to the writing of the manuscript: DGA, RGR, FJGC, CRE. Agree with manuscript results and conclusions: DGA, RGR, FJGC, CRE. Jointly developed the structure and arguments for the paper: DGA, RGR, FJGC, CRE. All authors reviewed and approved of the final manuscript.

Consent for publication
All data used in this paper has been obtained from public repositories therefore the consent for publication is not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]School of Engineering, Pablo de Olavide University, Seville, Spain. [2]Department of computer Science, University of Seville, Seville, Spain.

Received: 29 November 2017 Accepted: 26 July 2018
Published online: 06 August 2018

References

1. Bar-Joseph Z. Analyzing time series gene expression data. Bioinformatics. 2004;20(16):2493–503.
2. Gutiérrez-Avilés D, Rubio-Escudero C, Martínez-Álvarez F, Riquelme JC. TriGen: A genetic algorithm to mine triclusters in temporal gene expression data. Neurocomputing. 2014;132(0):42–53.
3. Zhao L, Zaki MJ. 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 - SIGMOD '05. New York: ACM Press; 2005. p. 694.
4. Jiang H, Zhou S, Guan J, Zheng Y. gTRICLUSTER : A More General and Effective 3D Clustering Algorithm for Gene-Sample-Time Microarray Data. In: BioDM; 2006. p. 48–59.
5. Hu Z, Bhatnagar R. Algorithm for discovering low-variance 3-clusters from real-valued datasets. Sydney: IEEE International Conference on Data Mining; 2010, pp. 236–45.
6. Xu X, Lu Y, Tan K-L, Tung AKH. Finding Time-Lagged 3D Clusters. In: 2009 IEEE 25th International Conference on Data Engineering; 2009. p. 445–56.
7. Liu Y, Lee C, Chen W, Shin JW, Hsu H, Tseng VS. A novel method for mining temporally dependent association rules in three-dimensional microarray datasets. Tainan: 2010 International Computer Symposium (ICS2010); 2010, pp. 759–64.
8. Pearson K, Filon LNG. Mathematical contributions to the Theory of Evolution. IV. On the Probable Errors of Frequency Constants and on the Influence of Random Selection on Variation and Correlation. Proc R Soc Lond (1854-1905). 1897;62(1):173–6.
9. Spearman C. Correlation calculated from faulty data. Br J Psychol, 1904-1920. 1910;3(3):271–95.
10. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25(1):25–9.
11. Oyelade J, Isewon I, Oladipupo F, Aromolaran O, Uwoghiren E, Ameh F, Achas M, Adebiyi E. Clustering algorithms: Their application to gene expression data. Bioinforma Biol Insights. 2016;10:38316.
12. Pontes B, Giráldez R, Aguilar-Ruiz JS. Biclustering on expression data. J Biomed Inform. 2015;57(C):163–80.
13. Liu J, Li Z, Hu X, Chen Y. Multi-objective evolutionary algorithm for mining 3D clusters in gene-sample-time microarray data. In: 2008 IEEE International Conference on Granular Computing; 2008. p. 442–7.
14. Freudenberg JM, Joshi VK, Hu Z, Medvedovic M. Clean: Clustering enrichment analysis. BMC Bioinforma. 2009;10(1):234.
15. Tchagang AB, Phan S, Famili F, Shearer H, Fobert P, Huang Y, Zou J, Huang D, Cutler A, Liu Z, Pan Y. Mining biological information from 3d short time-series gene expression data: the optricluster algorithm. BMC Bioinforma. 2012;1:54.
16. Yeung KY, Ruzzo WL. Principal component analysis for clustering gene expression data. Bioinformatics. 2001;17(9): 763–74.
17. Bhar A, Haubrock M, Mukhopadhyay A, Maulik U, Bandyopadhyay S, Wingender E. Coexpression and coregulation analysis of time-series gene expression data in estrogen-induced breast cancer cell. Algorithm Mol Biol. 2013;8(1):9.
18. Bhar A, Haubrock M, Mukhopadhyay A, Wingender E. Multiobjective triclustering of time-series transcriptome data reveals key genes of biological processes. BMC Bioinforma. 2015;16(1):200.
19. Gnatyshak D, Ignatov DI, Kuznetsov SO. From triadic FCA to triclustering: Experimental comparison of some triclustering algorithms. In: Proceedings of the Tenth International Conference on Concept Lattices and Their Applications; 2013. p. 249–60. La Rochelle, France, October 15-18, 2013.
20. Gnatyshak DV. Greedy modifications of oac-triclustering algorithm. Procedia Computer Science. 2014;31(Supplement C):1116–23. 2nd International Conference on Information Technology and Quantitative Management, ITQM.
21. Gnatyshak DV. A single-pass triclustering algorithm. Autom Doc Math Linguist. 2015;49(1):27–41.
22. Egurnov D, Ignatov DI, Nguifo EM. On containment of triclusters collections generated by quantified box operators. In: Foundations of Intelligent Systems - 23rd International Symposium, ISMIS 2017, Warsaw, Poland, June 26-29, 2017, Proceedings; 2017. p. 573–79.
23. Liu Y, Yang T, Fu L. A partitioning based algorithm to fuzzy tricluster. Math Probl Eng. 2015;2015:10. Article ID 235790.
24. Kakati T, Ahmed HA, Bhattacharyya DK, Kalita JK. A fast gene expression analysis using parallel biclustering and distributed triclustering approach. In: Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies. ICTCS '16. New York: ACM; 2016. p. 122–11226.
25. Ahmed HA, Mahanta P, Bhattacharyya DK, Kalita JK. Shifting-and-scaling correlation based biclustering algorithm. IEEE/ACM Trans Comput Biol Bioinforma. 2014;11(6):1239–52.
26. Gutiérrez-Avilés D, Rubio-Escudero C. Triq: A comprehensive evaluation measure for triclustering algorithms. In: Martínez-Álvarez F, Troncoso A, Quintián H, Corchado E, editors. Hybrid Artificial Intelligent Systems. Cham: Springer; 2016. p. 673–84.
27. Gutiérrez-Avilés D, Rubio-Escudero C. MSL: A measure to evaluate Three-dimensional patterns in gene expression data. Evol Bioinforma. 2015;11:. EBO.S25822.

28. Romero-Zaliz RC, Rubio-Escudero C, Cobb JP, Herrera F, Cordón O, Zwir I. A Multiobjective Evolutionary Conceptual Clustering Methodology for Gene Annotation Within Structural Databases : A Case of Study on the Gene Ontology Database. IEEE Trans Evol Comput. 2008;12(6):679–701.
29. Bauer S, Grossmann S, Vingron M, Robinson PN. Ontologizer 2.0–a multifunctional tool for GO term enrichment analysis and data exploration,. Bioinforma (Oxford, England). 2008;24(14):1650–1.
30. Gutiérrez-Avilés D, Rubio-Escudero C. Mining 3D Patterns from Gene Expression Temporal Data: A New Tricluster Evaluation Measure. Sci World J. 2014;2014:1–16.
31. Gutiérrez-Avilés D, Rubio-escudero C. LSL : A new measure to evaluate triclusters. In: IEEE International Conference on Bioinformatics and Biomedicine; 2014. p. 30–7.
32. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B. Comprehensive Identification of Cell Cycle-regulated Genes of the Yeast Saccharomyces cerevisiae by Microarray Hybridization. Mol Biol Cell. 1998;9(12):3273–97.
33. Dickison VM, Richmond AM, Abu Irqeba A, Martak JG, Hoge SCE, Brooks MJ, Othman MI, Khanna R, Mears AJ, Chowdhury AY, Swaroop A, Ogilvie JM. A role for prenylated rab acceptor 1 in vertebrate photoreceptor development. BMC Neurosci. 2012;13:152.
34. Bunt J, Hasselt NE, Zwijnenburg Da, Hamdi M, Koster J, Versteeg R, Kool M. OTX2 directly activates cell cycle genes and inhibits differentiation in medulloblastoma cells. Int J Cancer Journal international du cancer. 2012;131(2):21–32.
35. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013;41(Database issue):991–5.
36. Martínez-Álvarez F, Gutiérrez-Avilés D, Morales-Esteban A, Reyes J, Amaro-Mellado J, Rubio-Escudero C. A novel method for seismogenic zoning based on triclustering: application to the Iberian Peninsula. Entropy. 2015;17(12):5000–21.