Computers in Biology and Medicine 155 (2023) 106653

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

GENECI: A novel evolutionary machine learning consensus-based approach for the inference of gene regulatory networks

Adrián Segura-Ortiz [a], José García-Nieto [a,b,*], José F. Aldana-Montes [a,b], Ismael Navas-Delgado [a,b]

[a] Dept. de Lenguajes y Ciencias de la Computación, ITIS Software, Universidad de Málaga, Málaga, 29071, Spain
[b] Biomedical Research Institute of Málaga (IBIMA), Universidad de Málaga, Málaga, Spain

ARTICLE INFO

Keywords:
Gene regulatory networks
Differential expression
Machine learning
Evolutionary algorithms

ABSTRACT

Gene regulatory networks define the interactions between DNA products and other substances in cells. Increasing knowledge of these networks improves the level of detail with which the processes that trigger different diseases are described and fosters the development of new therapeutic targets. These networks are usually represented by graphs, and the primary sources for their correct construction are usually time series from differential expression data. The inference of networks from this data type has been approached differently in the literature. Mostly, computational learning techniques have been implemented, which have finally shown some specialization in specific datasets. For this reason, the need arises to create new and more robust strategies for reaching a consensus based on previous results to gain a particular capacity for generalization. This paper presents GENECI (GEne NEtwork Consensus Inference), an evolutionary machine learning approach that acts as an organizer for constructing ensembles to process the results of the main inference techniques reported in the literature and to optimize the consensus network derived from them, according to their confidence levels and topological characteristics. After its design, the proposal was confronted with datasets collected from academic benchmarks (DREAM challenges and IRMA network) to quantify its accuracy. Subsequently, it was applied to a real-world biological network of melanoma patients whose results could be contrasted with medical research collected in the literature. Finally, it has been proved that its ability to optimize the consensus of several networks leads to outstanding robustness and accuracy, gaining a certain generalization capacity after facing the inference of multiple datasets. The source code is hosted in a public repository at GitHub under MIT license: https://github.com/AdrianSeguraOrtiz/GENECI. Moreover, to facilitate its installation and use, the software associated with this implementation has been encapsulated in a python package available at PyPI: https://pypi.org/project/geneci/.

1. Introduction

Gene expression is the process of reading information encoded in the genome to produce the set of proteins necessary for the development and functioning of a living organism [1]. The sequence of bases provided by DNA ultimately contains information about which amino acids and in what order they must be joined to produce various cellular proteins. DNA is divided into genes, defined as fragments of nucleotide sequences responsible for storing information necessary to manufacture specific polypeptide chains, i.e., a given set of proteins.

Genes are selectively transcribed into RNA in each cell, so although the genome is common to all cells, the transcriptome is a consequence of their biological functions and, therefore, of the cellular tissue to which it belongs [2]. This process of activation and deactivation of genes that makes cell specialization possible is known as gene expression regulation [3] and is usually explained by interactions between gene products and transcription factors.

However, it should be taken into account that this whole process is efficient thanks to the presence of a certain balance, so the alteration of this balance can be the trigger of diseases and functional problems in the organism [4,5]. Therefore, the study of the regulation of gene expression is of vital importance for modern medicine and future work related to its personalization [6,7]. Furthermore, expanding knowledge regarding this topic and having the ability to discover new interactions between genes allows for improving the level of detail with which the processes that trigger different diseases are described and fosters the development of new therapeutic targets [8,9]. However, despite the wealth of experimental data currently generated, there is very little

* Corresponding author at: Dept. de Lenguajes y Ciencias de la Computación, ITIS Software, Universidad de Málaga, Málaga, 29071, Spain.
E-mail address: jnieto@uma.es (J. García-Nieto).

https://doi.org/10.1016/j.compbiomed.2023.106653
Received 16 November 2022; Received in revised form 9 January 2023; Accepted 8 February 2023
Available online 14 February 2023

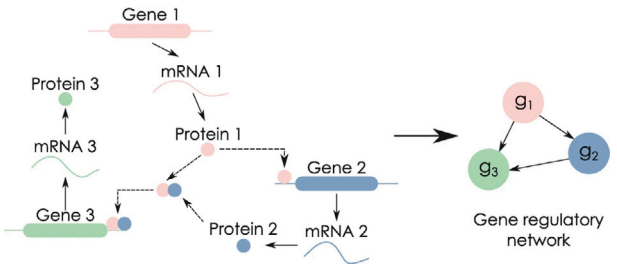

Fig. 1. Representation of gene regulatory networks [11].

specific knowledge about this regulation in different tissues and cell types.

Gene Regulatory Networks (GRNs) are collections of DNA segments whose products interact with each other and other substances in the cell [10]. The connections between the network elements and the directionality and intensity of their interactions determine the expression levels of the genes that make up the system within the cell. Therefore, understanding and analyzing these networks is a fundamental aspect of extending the existing knowledge on the regulation of gene expression in different organisms.

As shown in Fig. 1, these networks are usually represented by graphs. The primary sources for their correct construction are usually time series from differential expression data obtained by biological procedures and machinery [12]. After processing and cleaning of the data using the relevant bioinformatics procedures [13–15], these data reflect the expression levels manifested by different genes over a specific period under different experimental conditions.

The exhaustive analysis of these expression levels makes it possible to deduce the different relationships that exist between the genes that make up the network (see Fig. 2), making it possible to infer and subsequently represent them [16]. However, the techniques designed so far to infer gene regulatory networks and which apply this approach [17–26] seem to show a certain level of specialization on datasets with specific characteristics. That is, depending on the gene network to be inferred, the same technique can provide more or less competitive results [27], and in many cases, the reason for this is unknown.

Given that, in most cases, it is unknown which technique may be most appropriate for a real-world dataset whose results are unknown, numerous proposals have subsequently appeared. These proposals work on the consensus of several techniques to provide more general and robust proposals that guarantee a certain quality in the results. [28–31]. This paper presents the design of an evolutionary machine learning algorithm called GENECI, which acts as an organizer for the construction of ensembles, considering the results of the main inference techniques reported in the literature (ARACNE [17], C3NET [18], BC3NET [19], CLR [20], GENIE3 [21], KBOOST [22], MRNET [23], MRNETB [24], PCIT [25] and TIGRESS [26]) and optimizing the consensus network derived from them, according to their confidence levels and topological characteristics.

The results obtained by GENECI are compared with those provided by the individual techniques for several datasets. First, datasets were taken from known solutions in academic benchmarks (Gold Standards), such as the networks from the DREAM challenges [32] (in particular, editions 3, 4 and 5) and the IRMA yeast network [33] will be used. In this case, to quantify the accuracy obtained by the different techniques, the Area Under the ROC curve (AUROC) and Area Under the Precision–Recall curve (AUPR) metrics will be calculated concerning the gold standards. Finally, this proposal will be confronted with a real non-simulated biological network of patients with melanoma [34]. In the absence of a known reference for this case, the interactions inferred by GENECI will be validated by searching for adequate support in the literature. In addition to obtaining competitive results, robustness

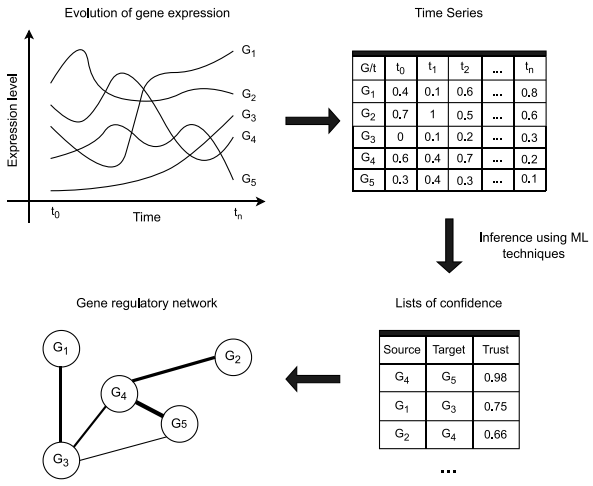

Fig. 2. Standard gene network inference procedure.

and generalization capacity have been achieved so that the algorithm presents promising results for data from different sources and with different characteristics.

This article is organized into six sections. In Section 2, the state-of-the-art of the main efforts undertaken so far in the task of inferring GRNs from differential expression data is presented. Both the individual inference techniques with the most significant impact in the literature and the first approaches aimed at an efficient consensus method are discussed. In Section 3, the algorithmic proposal and all the software development carried out in this work are shown. Subsequently, the experimentation and the choice of the main parameters of the evolutionary algorithm will be discussed in Section 4. Section 5 shows the results and discussion of this work with the accuracy values obtained by GENECI for each of the studied networks in comparison with the individual inference techniques. Finally, in Section 6, the conclusions and possible future lines considered feasible for future versions of the approached strategy are presented.

2. State of the art

The challenge of accurately inferring gene regulatory networks has encouraged the emergence of a wide range of specialized techniques in this area. A representative set of these individual techniques are conceived as candidate input of the evolutionary machine learning algorithm and therefore defines the initial quality of the optimization process. Therefore, GENECI has integrated those tools frequently appearing in the literature and showing competent results within this domain. Among them are: ARACNE (Algorithm for the Reconstruction of Accurate Cellular NEtworks) [17], C3NET (Conservative Causal Core NETwork) [18], BC3NET (Bagging C3NET) [19], CLR (Context Likelihood or Relatedness network) [20], GENIE3 (GEne Network Inference with Ensemble of trees) [21], KBOOST (kernel PCA regression and gradient boosting to reconstruct gene regulatory networks) [22], MRNET (Minimum Redundancy NETworks) [23], MRNETB (Minimum Redundancy NETworks using Backward elimination) [24], PCIT (Partial Correlation coefficient with Information Theory) [25] and TIGRESS (Trustful Inference of Gene REgulation with Stability Selection) [26]. All techniques aim to infer gene regulatory networks through the analysis of time series that record the evolution of the expression levels of different genes in a limited time period. All of them try to solve the same problem and are applied to the same datasets, which we will also use in the GENECI entry. In order to outline an idea of how each of these techniques works, the main characteristics of each of them are shown in Table 1.

Table 1
Main characteristics of each of the inference techniques.

Technique	Main calculation	Filtering/Procedure	Outstanding properties	Quantity
ARACNE	Mutual information coefficient	Statistical threshold based on the method of relevance networks and subsequent application of the data processing inequality property (DPI)	Elimination of indirect relationships	++
C3NET	Mutual information coefficient	Only the most significant interaction of each gene is finally selected	Prioritizing the presence of false negatives over the usual number of false positives	+
BC3NET	Mutual information coefficient	Attempts to relax the constraints imposed on C3NET filtering by generating several versions of the input data with the use of a nonparametric bootstrap and the individual application of the C3NET algorithm	Consensus perspective that aims to refine the confidence values	++
CLR	Mutual information coefficient	Filtering based on the calculation of the statistical probability of each value of mutual information within the context of its network	Elimination of false positives and correction of errors caused by inadequate or unequal sampling	+++
GENIE3	Decomposition into regression subproblems	Higher confidence to those interactions where the expression profile of the factor gene takes a high coefficient in the mathematical formula constructed for the target gene	Handling of combinatorial and nonlinear interactions, produces targeted GRNs but recovery of a regulator of a gene has been shown to degrade as the number of regulators of the gene increases	+++++
KBOOST	Decomposition into regression subproblems	Filtering based on comparison of different models and estimation of the probability of one gene regulating another using Bayesian Model Averaging (BMA)	Fast execution and high false positive detection	+++++
MRNET	Mutual information coefficient and redundancy value	Filtering based on the contemplation of both values by applying the forward feature selection method of Maximum Relevance Minimum Redundancy (MRMR)	High detection of indirect connections due to the discarding of those with high redundancy during the filtering process. However, due to the forward feature selection method the algorithm is strongly conditioned by the first selected variables	+++
MRNETB	Mutual information coefficient and redundancy value	Attempts to address the limitations of MRNET by replacing the forward feature selection method with a backward elimination procedure combined with sequential replacement	Elimination of indirect relationships by studying redundancy	+++
PCIT	Partial correlation coefficients combined with an information theory approach	For each trio of genes, the algorithm calculates the three first-order partial correlation coefficients and then applies the data processing inequality theorem (DPI) of information theory to obtain a local threshold that determines candidate filtering	It usually rules out bidirectionality among factors, opting for a single direction of regulation	++
TIGRESS	Decomposition into regression subproblems	use of the LARS method during feature selection, stabilized by a procedure that iteratively executes the above method on randomly perturbed data	High computational cost and stochastic method	++++

Following the DREAM5 challenge [32], all the proposals put forward by the participants were collected. The potential of their joint application was studied in [35]. Specifically, comparisons were made between the accuracies obtained by the proposed techniques individually versus different possible combinations of the techniques. The results showed that if the appropriate techniques were chosen, the consensus gene networks were undoubted of higher quality than those obtained by the individual techniques.

This tendency towards consensus has been further refined by applying different strategies. For example, in [28] several assembly methods based on normalization and analysis of the topological features present in the networks inferred by the different individual techniques are proposed. On the other hand, in [29] they study the consensus of various inference tools by using graph mining. Specifically, their work analyses the networks resulting from each technique and performs intelligent pattern detection to provide some reliability for export to the final network.

Finally, some papers in the literature present consensus network constructions by applying evolutionary algorithms that aim to find the optimal combination of individual techniques. Although none of these works presents an objective function similar to the proposal presented in this article, two significant cases are mentioned below. While in [30] the participation of the techniques in the different ensembles (binary values) is simply discussed, in [31] the assignment of weights to the different inference techniques is studied, constructing a weighted voting system that results in a final gene network.

The objective is to find the combination that maximizes the accuracy of the resulting network; however, this optimization process is, in both cases, guided by the network to be inferred (gold standard) associated with the input data, which this work is intended to avoid. On the one hand, in [30] each individual is evaluated by directly equating the fitness value to the AUROC obtained by the consensus network. On the other hand, in [31] the proposed system uses labeled information to positively evaluate the assignment of higher weights to tools with more links correctly predicted. Although these strategies achieve competitive accuracy values for the data presented, their fitness functions may present certain limitations concerning the discovery of new interactions and their application on real biological networks for which their outcome is unknown.

This article presents an alternative strategy that aims to overcome these limitations. For this purpose, the idea of optimization towards consensus has been taken up but establishing the pursuit of appropriate objectives is entirely different from the gold standards.

3. Algorithmic proposal

GENECI (GEne NEtwork Consensus Inference) takes up the idea of weight assignment seen in [31], but tries to improve the optimization process, choosing an approach similar to the one presented in [36] although adapted for GRNs. This article is helpful to know in practice the concept of evolutionary machine learning [37] despite having a purpose outside network inference. This work aims to solve these problems through computational learning by using its results to optimize the

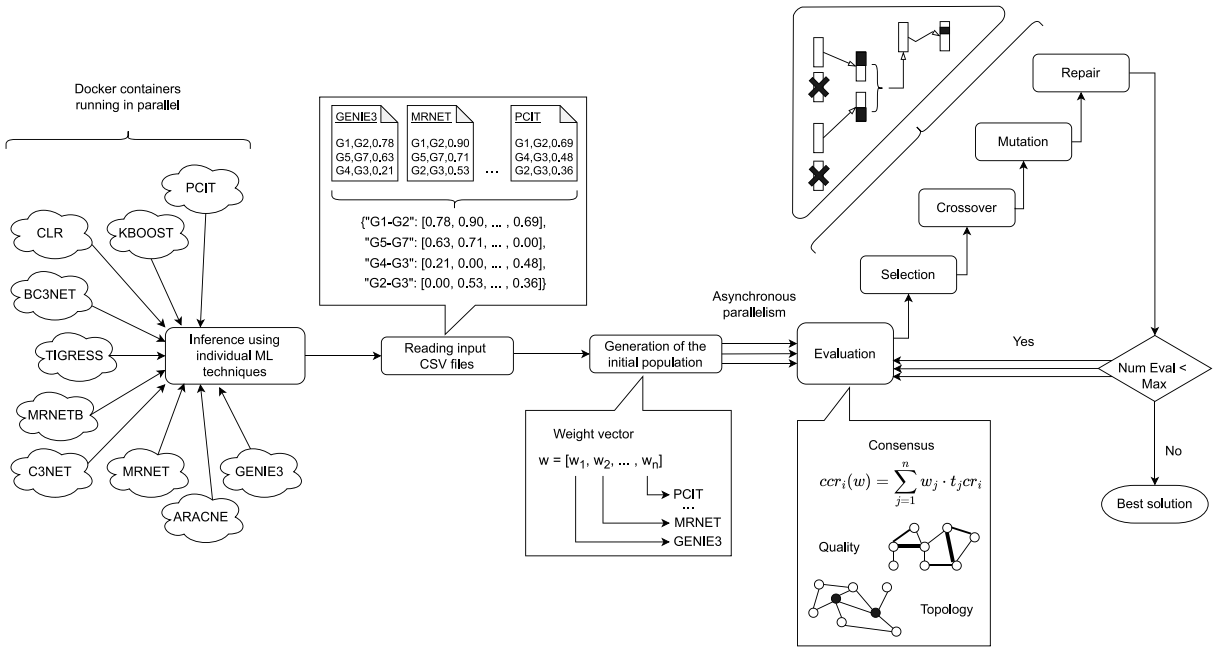

Fig. 3. Architecture and workflow covered in GENECI. First, the execution of multiple individual inference techniques in parallel is enabled by encapsulating their implementations in Docker containers. After that, their results are normalized and collected by the evolutionary algorithm in order to optimize weight vectors that assign a value to each technique. The weight vectors are iteratively subjected to evaluation (depending on the quality and topology of the consensus networks they represent), selection, crossover, mutation and finally an additional repair step to keep the sum of values at unity.

quality of a solution derived from them, thanks to the application of an evolutionary algorithm.

It has been decided to implement a genetic algorithm within the evolutionary branch. Firstly, the flexibility of its operators allows for searching for implementations that adapt correctly to the problem. Secondly, as it is an algorithm widely used in the literature, it has numerous execution environments that facilitate the design of our algorithm and the establishment of a solid and efficient construction base.

The genetic algorithm has been implemented using the jMetal framework [38] and takes as input a set of files with confidence lists related to interactions between genes. These lists can come from complementary runs on the Python package built or from external runs produced on other techniques. Although both options exist, it is recommended to run these techniques on the GENECI environment as it guarantees uniformity in the output format (tables made up of Source, Target and Trust columns) as well as the standardization of their confidence levels between 0 and 1, which is crucial for consensus building.

It is important to mention that all the functionalities incorporated in GENECI (including the execution of the evolutionary algorithm itself) have been encapsulated in Docker images. The main reason for choosing this design is that GENECI can enjoy the conveniences and advantages of the Python programming language (such as the construction of the package in PyPI or the hierarchical design of commands using the Typer library) while tolerating the implementation of each functionality in the most suitable programming language for this purpose, i.e. the most efficient or the one with the most specific and suitable libraries.

In the GENECI optimization process, the presence of the gold standard is avoided, and the consensus networks are evaluated according to their confidence levels and topological characteristics. Specifically, the designed objective function examines measures such as the confidence levels of the different links, the adequacy of the weights assigned in the ensemble to produce these values, the number of hubs[1] present in the consensus network and their similarity to a scale-free distribution.[2]

Algorithm 1 Main code of the generalized EA

Input files: List of input csv files with trusted lists.
Output consensusList: Final consensus list.

```
inferredNetworks = ReadAll(files)
population = GeneratePop(len(inferredNetworks))
while numEvaluations < max do
    fitness = Evaluation(population)
    numEvaluations += len(population)
    selectedPopulation = Selection(fitness)
    crossPopulation = Crossover(selectedPopulation)
    mutPopulation = Mutation(crossPopulation)
    repPopulation = Repair(mutPopulation)
    population = repPopulation
fitness = Evaluation(population)
bestIndividual = GetBest(fitness)
consensusList = MakeConsensus(bestIndividual)
return consensusList
```

Fig. 3 shows a schematic diagram where each stage addressed in GENECI is contemplated. As usual, the algorithm includes the main stages of evaluation, selection, crossover and mutation, to which an additional repair stage is added because of the chosen representation. In addition, it can be seen how the stopping criterion imposed in this case is determined by a maximum number of evaluations specified as the input parameter. In more detail, its implementation pseudocode is presented in Algorithm 1.

[1] Nodes with a statistically significant number of links concerning the rest

[2] That which is characterized by having a small number of high-degree nodes, concerning the rest that usually has a low number of links. This distribution is the opposite of the random distribution, where all nodes have a fairly similar number of connections.

Table 2
Example of input to the evaluation process. This table presents a gene network of 4 interactions (column 1) that has been inferred using three different individual techniques (columns 2, 3 and 4), the results of which are intended to be consensual. In this case, an individual is evaluated, proposing the following vector of weights: (0.5, 0.3, 0.2). The first significant value to be calculated is the consensus confidence (column 6), which consists of a simple weighted sum where the weight of each technique is multiplied by the level of individual confidence reported by the technique for the interaction in question. Second, a vector (column 7) is constructed, storing in each position the mean between the weight of the technique and the distance normalized to the median of the confidence levels of all techniques (column 5). This approach allows the calculation of the second significant value, the distance. This value consists of the difference between the maximum and the minimum of the vector constructed above, and the fitness function will try to minimize it.

Interaction	Tec 1	Tec 2	Tec 3	Median	Ind [0.5, 0.3, 0.2]		
					Consensus confidence	Vector	Distance
G1–G2	0.78	0.9	0.69	0.78	0.78 * 0.5 + 0.9 * 0.3 + 0.69 * 0.2 = 0.798	[(0+0.5)/2, (1+0.3)/2, (0.75+0.2)/2]	0.4
G5–G6	0.63	0.71	–	0.63	0.63 * 0.5 + 0.71 * 0.3 + 0 * 0.2 = 0.528	[(0+0.5)/2, (0.13+0.3)/2, (1+0.2)/2]	0.39
G4–G3	0.21	–	0.48	0.21	0.21 * 0.5 + 0 * 0.3 + 0.48 * 0.2 = 0.201	[(0+0.5)/2, (0.78+0.3)/2, (1+0.2)/2]	0.35
G2–G3	–	0.53	0.36	0.36	0 * 0.5 + 0.53 * 0.3 + 0.36 * 0.2 = 0.231	[(1+0.5)/2, (0.47+0.3)/2, (0+0.2)/2]	0.65

To speed up the execution of the algorithm, an asynchronous parallelization has been implemented to evaluate, cross and mutate several individuals simultaneously. This approach considerably reduces the execution time in those machines with the appropriate performance despite increasing the amount of RAM consumed during the algorithm's progress.

The following sections describe the aspects and stages of the designed evolutionary algorithm, concluding with a compilation of all the necessary input parameters.

3.1. Solution representation

The individuals of the population are represented by vectors of weights, where each position refers to a certain reconstruction of the network (indirectly a machine learning technique) and its value means the power or weight that this list has to vote for the consensus. Therefore, given a list of techniques $t = \{t_1, t_2, \ldots, t_n\}$ where each of them contains a list of relations $r_t = \{tr_1, tr_2, \ldots, tr_m\}$ with their respective confidence values $cr_t = \{tcr_1, tcr_2, \ldots, tcr_m\}$, an individual with a vector of weights $w = \{w_1, w_2, \ldots, w_n\}$ implies that the confidence value of the relation r_i in the consensus list (ccr_i) is given by:

$$ccr_i(w) = \sum_{j=1}^{n} w_j \cdot t_j cr_i$$

where $t_j cr_i$ is the confidence value of the relationship r_i in the list produced by the technique t_j. If the relationship r_i is not contemplated in that list, it will be understood that the technique has not detected that interaction, and therefore, its confidence level for t_j is 0. A more detailed example of this calculation can be seen in the column *Consensus confidence* of Table 2, where the construction of a 6-gene consensus network is simulated for a set of three techniques and a given individual.

The representation of individuals (solutions) as weight vectors implies the need to constantly keep the sum of their values at one. During the generation of offspring, crossover and mutation operators can alter this property. To solve this problem, it has been decided to add an additional operator in the optimization process to recover this feature of the solutions.

Two different repairers were designed. The first one, named *StandardizationRepairer*, performs a simple standardization where the values are rescaled, so their sum is 1. The second repairer, named as *GreedyRepairer*, performs a greedy repair where after choosing a position of the vector at random, it keeps its values until the sum exceeds unity (in which case it sets the following values to 0) or reaches the last position and adds the number needed to sum to 1. Finally, it was shown that the greedy repairer obtains worse results as a consequence of distorting the weights of the individuals excessively, bringing the optimization process closer to a random search. The repairer in charge of rescaling the values favors the algorithm's performance by maintaining the proportions of the values.

3.2. Evaluation

At each iteration, all the individuals in the population are evaluated to know the quality of their proposals and to make the selection step possible. The process begins with the translation of the vector of weights to the evaluated concept, the consensus network derived from the voting system seen above. However, for evaluation purposes, more than the confidence level alone is needed as a reference of reliability since a relationship between genes may obtain a fairly high value but has originated from an inadequate distribution of weights. For example, distributions that give too much weight to a particular technique, to a very small subset of them, or to techniques whose confidence values are not supported by any others.

For this reason, the evaluation process is responsible for calculating a vector per interaction and the confidence value of each relationship after consensus. The average distance between the median confidence of all the techniques (scaled between 0 and 1) and the weight assigned to the evaluated individual are stored for each technique. After that, the distance between that vector's maximum and minimum values is calculated and stored, together with the previously calculated confidence.

This distance will try to be minimized in the first term of the fitness function, so the aim is to establish a compensation system between the distance to the median and the weight assigned by the individual. In this way, as all the values of the calculated vector have a similar value (and therefore, the distance between the maximum and the minimum is minimized), the techniques whose proposal is different from the rest (greater distance to the median) will be penalized by being given a lower weight. On the contrary, the techniques with a proposal quite close to the rest (smaller distance to the median) will be compensated by assigning a high weight.

Therefore, as can be observed in Table 2, having given a higher weight to the first technique is beneficial for the first three interactions. This is because the first technique is the closest to the median for these cases, which provides some reliability to its proposal and consequently rewards the individual for having assigned it a higher weight. However, for the last interaction (G2–G3), it can be seen that the calculated distance has grown because the technique whose value is more reliable (closer to the median) for this case is the third one, to which the individual has assigned the lowest weight.

3.2.1. Objective function

An aggregate objective function with two weighted terms has been designed: Quality and Topology.

The **Quality** term aims to encourage the emergence of solutions with high confidence levels that also come from consistent weight distributions that assign greater importance to those techniques whose values have high support concerning the rest. Its purpose is to establish a certain contrast between good and bad links so that the links finally reported are of high reliability. In this term, a quality value is assigned to each interaction considered in the problem. For this assignment, two significant values are considered, the consensus confidence

Algorithm 2 First term of the fitness function: Quality

Input c: consensus list with confidence and distance values.
Output $quality$: value of the first term of the fitness function.

1: $distConf = []$
2: **for** i **in** $len(c)$ **do**
3: $\quad distConf[i] = \left(conf_i + (1 - dist_i)\right)/2$
4: $mean = \frac{1}{len(c)} \cdot \sum_{i=1}^{len(c)} distConf[i]$
5: $distConfSum = 0$
6: $cnt = 0$
7: **for** i **in** $len(c)$ **do**
8: $\quad$ **if** $distConf[i] > \ mean$ **then**
9: $\quad\quad distConfSum \mathrel{+}= distConf[i]$
10: $\quad\quad cnt \mathrel{+}= 1$
11: $numPosLinks = N_{genes}^2$
12: $q_1 = |cnt$ - $0.1 \cdot numPosLinks| \ / \ (0.9 \cdot numPosLinks)$
13: $q_2 = 1 - (distConfSum/cnt)$
14: $quality = 0.25 \cdot q_1 + 0.75 \cdot q_2$
15: **return** $quality$

Algorithm 3 Second term of the fitness function: Topology

Input b: binary network originated after cut-off.
Output $topology$: value of the second term of the fitness function.

1: $degree = []$
2: **for** i **in** N_{genes} **do**
3: $\quad$ **for** j **in** N_{genes} **do**
4: $\quad\quad degree[i] = n[i][j]$
5: $mean = \frac{1}{N_{genes}} \cdot \sum_{i=1}^{N_{genes}} degree[i]$
6: $hubsDegreeSum = 0$
7: $hubs = 0$
8: **for** i **in** N_{genes} **do**
9: $\quad$ **if** $degree[i] > \ mean$ **then**
10: $\quad\quad hubsDegreeSum \mathrel{+}= degree[i]$
11: $\quad\quad hubs \mathrel{+}= 1$
12: $t_1 = |hubs$ - $0.1 \cdot N_{genes}| \ / \ (0.9 \cdot N_{genes})$
13: **if** $hubs > \ 0$ **then**
14: $\quad t_2 = 1 - (hubsDegreeSum/hubs) \ / \ \left(N_{genes} - 1\right)$
15: **else**
16: $\quad t_2 = 1$
17: $topology = \left(t_1 + t_2\right)/2$
18: **return** $topology$

level and another previously introduced value called "distance". The first considerable value (consensus confidence) is calculated through a weighted sum. The weight assigned by the individual being evaluated and the individual confidence level reported by that technique for the interaction in question is multiplied for each technique.

On the other hand, the "distance" value, as explained with the example shown in Table 2, is the difference between the maximum and the minimum of a vector that stores for each technique the mean between its weight and the distance between its individual confidence value and the median of the set of techniques. Finally, the quality value associated with interaction will be the mean between its consensus confidence level and the unit subtracted by the distance. In other words, an interaction will have a good quality when it has a high confidence level and a small distance value. A small distance value means that the maximum and minimum of the calculated vector are close values and that, therefore, the techniques with a confidence value far from the median have been assigned a lower weight than the rest. Likewise, it means that the techniques with a confidence level close to the median (smaller distance) have been compensated with a higher weight. Finally, those interactions whose quality exceeds the mean are chosen, and an attempt is made to maximize their quality while approximating their quantity to 10% of the total number of interactions. This threshold is set up because minimizing the number of good links would result in a fuzzy network. The aim is to establish a clear contrast that allows us to report truly reliable interactions.

Its implementation can be found in Algorithm 2. For each link in the consensus list, the mean between its confidence value and the unit subtracted by the distance mentioned above is calculated and stored in the $distConf$ vector in the pseudocode (lines 1 to 3). The mean of the previous vector is then calculated (line 4), and those whose value manages to exceed it are set as good links. On the one hand, the sum of the values of all these good links is carried out. On the other hand, their quantity is stored (lines 5 to 10). Since the algorithm is oriented to minimization, the result of this term is lower when the quality of the consensus network is higher.

First, to optimize the number of these good links, q_1 is calculated (lines 11 and 12 in Algorithm 2). Specifically, this variable will reduce its value when the number of good links is closer to 10% of the total number of possible links in the network. Second, for these links to have the highest possible quality, the sum of their averages between confidence and unity minus distance should be maximized. For this purpose, in q_2 the mean of the $distConf$ of these good links is calculated and adapted to the minimization objective by subtracting its value from unity (line 13 in Algorithm 2).

The fact that in q_2 only the quality of good links is reported allows GENECI to establish the balance mentioned above between distance to the median and weight by focusing exclusively on the most relevant interactions in the network. This allows to eliminate possible noise and contradictions that lower intensity relationships could cause and also not constantly penalize more selective and strict inference techniques such as C3NET.

Finally, the result of the term *Quality* is the value of q_1 multiplied by 0.25 plus that of q_2 multiplied by 0.75 (line 14 in Algorithm 2). This means that this function gives more importance to good links' quality than quantity.

The second term, called **Topology**, is more oriented towards improving the structure of the consensus network. To this end, it intends to positively evaluate those proposals that present networks with a scale-free distribution (as real biological networks usually are). Mathematically, it tries to increase the degree (number of links) of those genes with a high potential to be considered hubs. At the same time, it is intended that the number of genes that meet this condition should be relatively low since this is usually observed in real genetic networks. The goal is to promote the approximation of the network to a scale-free configuration and to move away from a random structure.

However, the input of this second term is not the same as the first one. In order to correctly study the network topology, it should be decided which links are finally labeled as definitive and which are not reliable enough to do so. This decision is made by applying a certain cut-off criterion. Three different criteria were designed for this issue:

- The first one called *MaxNumLinksBestConfCriteria* takes as input the number of links k to be obtained in the definitive network. After knowing that input, it simply sorts from highest to lowest all interactions based on their confidence value and returns the top k.
- In the second one called *MinConfidenceCriteria*, the minimum confidence required to report a link is entered as the input value. Therefore, all interactions are filtered and only those that have exceeded the threshold are returned.
- The third one identified as *MinConfDistCriteria* has a similar operation to the previous criterion, except that in this case the threshold refers to the average between the confidence and the unit subtracted by the distance of the links.

After putting the three criteria to the test, it was found that the one that considers both metrics is the most effective.

As shown in Algorithm 3, the code starts by calculating the average of the degrees of all the nodes in the network (lines 1 to 5). After that, it selects as hubs those genes with a degree higher than the average to quantify them and store the sum of their degrees (lines 6 to 11).

Similarly to the previous term, with t_1 an optimization of the number of hubs is carried out trying to approximate its value to 10% of the total number of genes (line 12 in Algorithm 3). This quantity is the one that has been considered appropriate to bring the node degree distribution closer to a scale-free distribution, where a small number of nodes concentrate most of the network connections. On the other hand, with t_2, the sum of the degrees of all the hubs is maximized. For this purpose, the average of these degrees is calculated. Then, it is normalized by dividing it by the maximum achievable degree and subtracted from the unit to adapt the variable to the minimization objective (lines 13 to 16 in Algorithm 3).

Finally, the value of the returned term is the average of these two metrics (line 17 in Algorithm 3). Therefore, the fitness value assigned to an individual is given by the formula:

$$Fitness(Ind) = w_Q \cdot Quality(Ind) + w_T \cdot Topology(Ind)$$

where w_Q and w_T are the weights assigned to the *Quality* and *Topology* terms respectively, which are given as input parameters.

3.3. Selection

Selection is carried out using the classical binary tournament in which individuals are randomly grouped in pairs and pitted against each other, so only those with better scores are selected. Pairwise matching is performed as often as necessary to cover the indicated population size. In other words, the same individual can be tested on more than one occasion and even selected for the next phase. However, in the next generation, it will certainly be crossed with other individuals, and its offspring will subsequently be subjected to different mutations.

3.4. Crossover

The crossover operation simulates a reproduction process between individuals, where their respective genetic materials are crossed to procreate offspring. This operation occurs with a fairly high probability modifiable in the jMetal framework. How this genetic material is crossed is what leads to multiple possible operators. Depending on the characteristics of the problem, the type of crossover chosen will have better or worse results. However, the best way to check the choice of a good operator is by testing and comparison.

The jMetal framework offers a wide range of crossover operators. These include *SBXCrossover* (Simulated Binary Crossover) [39], *BLXAlphaCrossover* (Blend Alpha Crossover) [40], *DifferentialEvolutionCrossover* [41], *NPointCrossover*, *NullCrossover* and *WholeArithmeticCrossover*. Finally, the *SBXCrossover* operator was chosen after several scores tests. Firstly, this operator was the one that reported the best results concerning the others. Secondly, its choice was the most coherent if considered a linear expression between the two weight vectors. It tends to maintain the feasibility of the solution and reduce the distortion cost produced by the repairer.

3.5. Mutation

After crossing the individuals of the previous generation, the offspring are subjected to a mutation process to incorporate new genetic material into the population. Otherwise, the resolution of the problem would be completely limited by the genetic content of the initial population, which reduces the search procedure and conditions the solution to the initial decisions of the algorithm. This operation occurs with a fairly low probability, again modifiable from the jMetal framework. As with the crossover stage, jMetal integrates a wide variety of operators to cover this phase of the evolutionary algorithm. Specifically, the mutation operators available are the following: *PolynomialMutation*, *CDGMutation*, *LinkedPolynomialMutation*, *GroupedPolynomialMutation*, *GroupedAndLinkedPolynomialMutation*, *SimpleRandomMutation*, *UniformMutation*, *NonUniformMutation* and *NullMutation*. Finally, after verifying the good results generated in combination with *SBXCrossover*, the *PolynomialMutation* was chosen. In this case, the modification of the mutated values is compensated by the rest of the vector weights by repairing individuals.

Table 3
GENECI input parameters.

Parameter	Description
–confidence-list	CSV file paths with trusted lists.
–gene-names	Path to the TXT file with the name of the genes separated by comma and without space. If not specified, only genes specified in the confidence lists will be considered.
–crossover	Crossover operator.
–crossover-probability	Crossover probability.
–mutation	Mutation operator.
–mutation-probability	Mutation probability.
–repairer	Repairer to keep the sum of weights equal to 1.
–population-size	Population size.
–num-evaluations	Number of evaluations.
–cut-off-criteria	Cut-off criteria for network binarization.
–cut-off-value	Numeric value associated with the selected criterion.
–Q-weight	Weight associated with term *Quality*.
–T-weight	Weight associated with term *Topology*.
–threads	Number of threads to be used during parallelization. By default, the maximum number of threads available in the system is used.
–graphics	Graphical representation of the evolution of the fitness value.
–output-dir	Path to the output folder.

3.6. Output

After completing the number of evaluations set in the input parameter, GENECI selects the best vector of weights found during the execution and produces an output consisting of 5 files:

- List of optimized interactions with their respective consensus confidence values.
- Binary network resulting from applying the selected cut-off criterion to the previous list.
- Weights assigned to the different techniques in the final solution.
- A plain text file with the evolution of the fitness values.
- Optionally, a pdf file with a graphical representation of this evolution.

4. Experimentation

GENECI has a fairly large number of parameters, which are shown in Table 3 along with their respective descriptions. Before elaborating on the real experimentation of this work, it was necessary to carry out a parameterization exercise to guarantee the optimal performance of the algorithm.

4.1. Parameter settings

Similarly, before parameter refinement, it was necessary to ensure a certain consistency between the fitness values of the individuals and

their accuracy in the prediction of gene networks. That is to say, a good fitness value should translate into a good quality index in the subsequent network prediction. To guide the evolutionary algorithm to some extent, several networks (mainly from the DREAM challenges) were tested with different values of the parameters associated with the weights of the fitness function terms. Finally, the combination with the best accuracy results (regardless of their fitness values) was chosen. This combination was 0.75 for the first term and 0.25 for the second.

Once this was done, the rest of the parameters were tested to optimize the fitness values. Since testing all combinations of values was completely unfeasible, an incremental procedure was carried out to try to progressively fix the values of the parameters, starting with the analysis of the most fundamental ones and ending with those of lesser importance.

Parameters related to crossover and mutation probabilities were set before. It is already known in the literature that an adequately constructed evolutionary algorithm should respond well to a high crossover probability and a low mutation probability [42,43]. However, testing their values ensures the consistency of the implementation. Although the crossover probability seemed to be clearly fixed at 0.9, the mutation probability varied depending on the number of consensual lists. After reviewing the literature, it could be seen how the recommended mutation probability for these cases is the maximum between 0.01 and $1/n$ [36,44], where n is the length of the vector, which in this case is the number of lists provided in the input. Therefore, since having a number less than 0.01 would mean trying to agree on more than 100 lists (which is infeasible), $1/n$ was set as the optimal value associated with the mutation probability.

The next parameters to be optimized were those related to the repairer and cut-off criterion. As mentioned in their respective sections, *StandardizationRepairer* and *MinConfDist* were finally established as the optimal values for these parameters. In this case, since both parameters are quite specific to our problem, instead of doing point executions, we proceeded to perform a systematic test on all datasets. For reasons of length, these results can be found in the supplementary material in the repository. The winners were easily predictable since the alternatives lacked adequate meaning. First, concerning the repairers, it was to be expected that the greedy ones would offer worse results. There was some randomness in its operation, and it did not fully maintain the essence of the vector it was intended to repair. On the other hand, in the cut-off criteria, the justification discussed in previous sections regarding the reliability based on the confidence values together with the distance between weight means and distance concerning the median confidence allowed to expect that the criterion contemplating both metrics would indeed be the most effective.

Finally, the parameters of population size and the number of evaluations remain to be analyzed. It is evident that the higher the value given to them, the higher the quality of the result obtained, but the more execution time they consume. To find a certain balance, several combinations were tested to find the one that would ensure the algorithm's convergence at a suitable stage under a reasonable number of iterations. It should be mentioned that, regardless of the size of the input network, the number of techniques applied in the experimentation is the same, so the vector of weights to be optimized is always of the same size. This means that, although the evaluation is slower for large networks, the possible combinations of weights assigned to the lists cover the same search space as for the rest of the networks. Finally, a population size of 100 individuals and a total of 50,000 evaluations were established for the experimentation addressed in this work.

4.2. Experimental procedure

After setting all the GENECI parameters, an experimental procedure was constructed to demonstrate the validity of the work carried out and the benefits of the proposed strategy. The first part of this study takes data from academic benchmarks to quantify the accuracy of GENECI. First, data from some of the DREAM challenges [32] (specifically editions 3, 4 and 5) are used, as they have been extensively studied in the literature [26,45–49] and provide specific evaluation scripts that allow us to compare accuracy values with other research articles. And secondly, the IRMA 5-gene network [33] is considered, whose gold standard allows us to evaluate the quality of the results as a binary classification problem. Finally, GENECI is confronted with a real-world biological network of melanoma patients [34] whose interactions are validated by specific literature searches. For each dataset, the process starts by inferring their corresponding gene regulatory networks using all the individual techniques integrated into the proposal.

For the benchmark data, the predictive capacity of the results provided by the individual techniques is evaluated to conduct a subsequent comparison exercise concerning the quality of the GENECI consensus networks. Specifically, metrics AUROC (Area Under the ROC curve) and AUPR (Area Under the Precision–Recall curve) are calculated. The area under the receiver operating characteristic curve (AUROC) is a single scalar value that quantifies the overall performance of a binary classifier. Its value is bounded by the interval [0.5–1.0], where the minimum value represents the performance of a random classifier, and the maximum value is associated with a perfect classifier. Secondly, the area under the precision–recall curve (AUCPR) is a model performance metric that has been recognized as useful for classification performance assessment for unbalanced binary responses in bioinformatics [50]. This is the case for predicting interactions between genes that form the GRNs, as the number of truly interconnected genes is small compared to all the possible connections. Its value increases the better the classifier is evaluated.

For the DREAM challenges, a subcommand integrated into the package is used to call the evaluation scripts presented in the respective challenges. Another generic subcommand is used for the IRMA data that treats the case as a binary classification problem. Subsequently, a total of 25 independent runs are elaborated by the evolutionary algorithm (except for the last DREAM5 network due to its size), and the quality of the prediction made for each is validated. Finally, a comparison is made between the median AUROC and AUPR values of these 25 runs and the individual values of the different techniques.

The GENECI result is studied for real-world data by validating the presence of the main interactions reported in the literature from a biomedical point of view.

5. Results and discussion

This section presents the results provided by GENECI for the experimental procedure described previously. A section is dedicated to each dataset, illustrating the precision values obtained in each case and graphical representations that allow visualizing network topologies, distribution of fitness values, weights assigned by GENECI and a series of comparisons between the results of different inference techniques.

5.1. Benchmarks

5.1.1. DREAM3

The DREAM3 challenge on "in silico" network inference [51] sought to discover the ability to exist technology to infer gene networks of various sizes and connection densities. The data provided in the challenge were based on subnetworks associated with two organisms: E. coli (*Escherichia coli*) and yeast (*Saccharomyces cerevisiae*). Specifically, data were generated using continuous differential equations representing reasonable approximations for their gene expression regulatory functions. These values were subsequently altered by introducing a small amount of Gaussian noise to simulate measurement error. Five networks (Ecoli1, Ecoli2, Yeast1, Yeast2, Yeast3) were contemplated for each subchallenge, which divided the challenge by network sizes.

Table 4
Accuracy values for DREAM3 and size 10 networks. AUPR and AUROC values are provided for each technique and problem, highlighting in bold the results obtained by GENECI and the best obtained by any of the individual techniques.

Técnica	D3_10_Ecoli1		D3_10_Ecoli2		D3_10_Yeast1		D3_10_Yeast2		D3_10_Yeast3	
	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR
ARACNE	0.561	0.1529	0.524	0.1852	0.6269	0.1546	**0.6031**	0.417	0.5043	0.2463
BC3NET	0.6116	0.2146	0.4804	0.2158	0.5606	0.2359	0.5914	0.3459	0.4846	0.228
C3NET	0.5754	0.1599	0.5218	0.1832	0.5719	0.1368	0.6003	0.414	0.5167	0.2538
CLR	0.5719	0.1477	0.4542	0.1599	0.5788	0.1724	0.5782	0.3783	0.4893	0.2619
GENIE3_ET	**0.6157**	0.1762	0.6524	0.2172	0.525	0.1212	0.5858	0.3417	**0.5809**	0.2855
GENIE3_GBM	0.5673	0.139	**0.7458**	0.2848	0.4863	0.1341	0.5643	**0.4183**	0.5381	0.2695
GENIE3_RF	0.5938	0.1598	0.6818	0.2369	0.5	0.1064	0.5545	0.345	0.5689	0.3072
KBOOST	0.5949	0.1739	0.648	0.2392	0.3438	0.0928	0.5415	0.3246	0.3168	0.1786
MRNETB	0.5472	0.141	0.4631	0.1685	0.5487	0.1285	0.5754	0.4043	0.4485	0.223
MRNET	0.5155	0.1469	0.5116	0.1785	0.5206	0.1264	0.5788	0.4122	0.4402	0.2318
PCIT	0.5455	**0.2987**	0.4862	0.1626	0.5631	0.1694	0.412	0.2336	0.5675	**0.3501**
TIGRESS	0.481	0.1174	0.6569	**0.4301**	**0.6763**	**0.242**	0.4892	0.2491	0.4305	0.2013
Median GENECI	**0.5627**	**0.1707**	**0.6089**	**0.2468**	**0.5175**	**0.1311**	**0.5982**	**0.3645**	**0.5127**	**0.2711**
Best GENECI	**0.5685**	**0.1791**	**0.6196**	**0.2523**	**0.5275**	**0.1369**	**0.6025**	**0.3956**	**0.5287**	**0.3245**

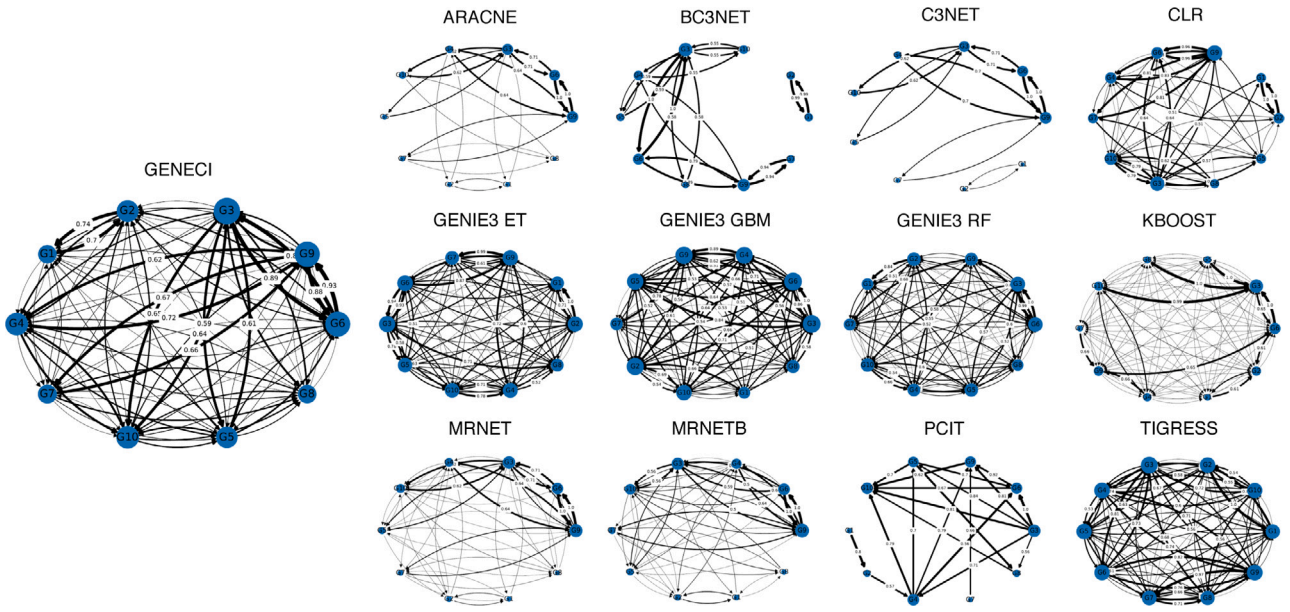

Fig. 4. For the first 10-gene yeast network of the DREAM3 challenge, the gene networks inferred by the individual techniques and the consensus gene network computed in the run whose AUROC corresponds to the median exposed in Table 4 are illustrated. Graphs attempt to represent gene regulatory networks by setting up genes in the form of nodes and interactions through links. In addition, it can be seen that the directionality and confidence of these interactions are represented in these networks.

Specifically, the sub-challenges were of 10, 50 and 100 nodes, for which 4, 23 and 46 different trajectories were provided, respectively.

Currently, the DREAM3 challenge has become a standard benchmark for evaluating the reconstruction of GRNs from expression data. The scripts provided in the challenge have established a unified evaluation mechanism that facilitates the comparison of quality between the different proposals in the literature to some extent.

Size 10

First, the results for the 10-node networks of the DREAM3 challenge are shown. Table 4 shows the AUROC and AUPR values for each of the individual techniques and the median of these same metrics for the 25 independent runs of the evolutionary algorithm. In addition, for comparative purposes, the best value obtained by the inference techniques and the one achieved by GENECI are highlighted in bold for each column. In this case, the values of the accuracy metrics reported for the consensus networks are in a competitive range but without standing out from the rest. The exception can be observed for the AUROC obtained for the second yeast network, where the distance between the best result of the techniques and that of GENECI is relatively small. However, this case is considered insignificant, considering its isolated character and the low precision quality reported by the individual techniques.

The explanation of these results is that the second term of the fitness function is practically frozen for cases of such a small size. When faced with 10-node networks, the evolutionary algorithm does not seem to have enough margin to carry out the optimization part aimed at improving the consensus network topology. However, this is not considered a problem if one remembers that the goal of GENECI is to optimize the consensus of real-world gene networks, where the number of transcription factors is much larger than that contained in this subchallenge.

Fig. 4 shows the graphs associated with each of the networks inferred by the individual techniques and the consensus network whose AUROC corresponds to the median of the runs. These graphs show the directionality of the interactions (direction of the arrows), the degree of the genes (size of the nodes) and the intensity of the relationships (thickness of the links and numerical specification for the highest values). In addition to appreciating the small size of these networks, it can be seen how the techniques that have obtained the best results in Table 4 (TIGRESS and those derived from GENIE3) present random networks that are far from the scale-free configuration that usually

Table 5
Accuracy values for DREAM3 and size 50 networks. In this table, a gene network is contemplated for each pair of columns, where in each row the AUPR and AUROC values are provided for each inference technique.

Técnica	D3_50_Ecoli1		D3_50_Ecoli2		D3_50_Yeast1		D3_50_Yeast2		D3_50_Yeast3	
	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR
ARACNE	0.4893	0.0241	0.539	0.0424	0.537	0.0505	0.5252	0.0761	0.5283	0.0887
BC3NET	0.5015	0.0263	0.5271	0.0372	0.5455	0.0429	0.5144	0.0715	0.5245	0.084
C3NET	0.5083	0.0265	0.5209	0.0391	0.5363	0.0508	0.5198	0.0764	0.5244	0.0868
CLR	0.5831	0.0334	0.627	0.0606	0.5401	0.0504	0.5292	0.0814	0.5491	0.1014
GENIE3_ET	0.5338	0.0294	0.6373	0.0783	0.5463	0.0514	0.5709	0.0872	0.5752	0.0965
GENIE3_GBM	0.4951	0.0299	0.6269	**0.087**	**0.5704**	**0.0778**	0.571	**0.0906**	0.5775	**0.1058**
GENIE3_RF	0.5563	0.0338	0.6318	0.0811	0.5665	0.07	**0.5892**	0.0895	0.5739	0.1004
KBOOST	0.5459	0.0277	0.528	0.0379	0.4628	0.0347	0.4659	0.0668	0.5168	0.0738
MRNETB	**0.6046**	0.0363	**0.6401**	0.057	0.5467	0.0557	0.5357	0.0755	0.5507	0.0992
MRNET	0.5803	0.0329	0.6237	0.0557	0.5409	0.0507	0.5383	0.0786	0.5474	0.0983
PCIT	0.5765	**0.0403**	0.5953	0.0677	0.5636	0.0588	0.499	0.0644	0.539	0.0843
TIGRESS	0.5794	0.0289	0.3846	0.0246	0.5591	0.0335	0.5876	0.084	**0.5807**	0.0983
Median GENECI	**0.5998**	**0.0348**	**0.6461**	**0.0713**	**0.5517**	**0.0657**	**0.5688**	**0.0839**	**0.574**	**0.1001**
Best GENECI	**0.6027**	**0.0349**	**0.6475**	**0.0719**	**0.5553**	**0.0662**	**0.5694**	**0.0846**	**0.5821**	**0.1014**

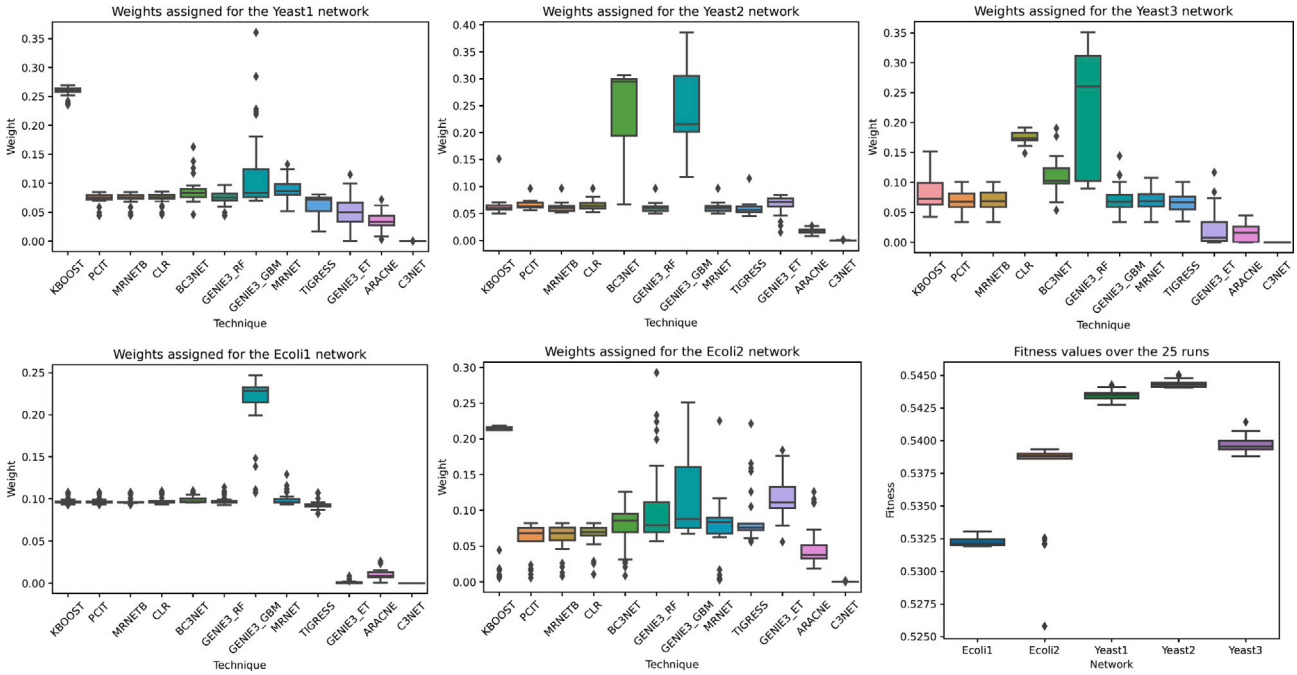

Fig. 5. Boxplots of the fitness values and weights over the 25 independent runs. The first 5 graphs represent the distribution of the weights assigned by GENECI across all the runs for each of the techniques. Finally, the sixth figure shows the distribution of the fitness values obtained in the different runs performed.

appears in real-world gene networks of larger size. This means that the GENECI target is far from the topological characteristics of the gold standard in these cases, which again explains the limited results shown.

Size 50

This section shows the results obtained for the 50-node networks of the DREAM3 challenge. In Table 5, it can be seen that the increase in the size of the networks to be inferred has led to better results than in the previous section. This is because a size of 50 nodes already gives a certain margin to the evolutionary algorithm to optimize the topological characteristics of the consensus networks. In this case, it can be observed that the medians of the AUROC and AUPR values of GENECI are quite close to the best results of the individual techniques. In addition, it is worth mentioning that the maxima of these metrics are selected individually so that, in many cases, the method that provides the best result concerning AUROC is not the same as the best about AUPR. Therefore, the fact that GENECI is able to approach the maximums of both metrics (or even surpass them), is in many cases an improvement over any of the individual techniques.

Fig. 5 shows boxplots representing the weights assigned to the different techniques by the final solutions of the 25 independent runs. An additional diagram concerning the fitness values of these solutions is depicted at the bottom right, where the variation of the achieved values can be observed for each network.

Regarding the assignment of weights, it can be seen that in most of the networks, the proposed solutions throughout the 25 runs are quite similar. However, there is the exception of the Ecoli2 network, which in addition to showing outliers in the fitness boxplot also presents a slightly more random distribution of weights. This may be because this network presents more local minima that hinder the algorithm's progress. Most runs seem to have stalled at one of them, as there are sporadic runs with better results. Consequently, both the distributions of good solutions and those related to premature convergences coexist in the boxplots of the weights, which explains the variability shown in the graph.

From a different perspective, specific techniques in certain networks do not seem to converge to a given weight. Two different situations

Table 6
Accuracy values for DREAM3 and size 100 networks. The AUPR and AUROC values are presented in two clearly distinguishable bands. The first band shows the precision values for the individual inference techniques, while the second band shows the values obtained by GENECI after the consensus of the techniques, distinguishing between the median of the runs and the best result obtained from them.

Técnica	D3_100_Ecoli1		D3_100_Ecoli2		D3_100_Yeast1		D3_100_Yeast2		D3_100_Yeast3	
	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR
ARACNE	0.5512	0.0238	0.5323	0.0187	0.553	0.0332	0.5216	0.053	0.5126	0.064
BC3NET	0.5442	0.0175	0.5286	0.0175	0.523	0.0191	0.5148	0.0435	0.5063	0.0569
C3NET	0.5413	0.0233	0.5189	0.017	0.5232	0.0263	0.5165	0.0507	0.5108	0.0638
CLR	0.659	0.0311	0.5909	0.0268	0.5553	0.0472	0.521	0.0557	**0.5284**	0.0693
GENIE3_ET	0.6596	0.0338	0.5822	0.0374	0.6235	0.0496	0.5343	0.0528	0.5176	0.0672
GENIE3_GBM	0.6176	0.0363	0.5729	0.0455	**0.6515**	**0.0629**	**0.5566**	**0.0621**	0.5274	0.0713
GENIE3_RF	**0.6673**	**0.042**	0.6001	**0.0506**	0.6465	0.0557	0.5548	0.0602	0.5269	**0.0719**
KBOOST	0.4975	0.0153	0.506	0.0132	0.4934	0.0193	0.4692	0.0389	0.4721	0.0516
MRNETB	0.6422	0.0332	**0.6045**	0.0235	0.5478	0.0357	0.5116	0.0482	0.5246	0.0658
MRNET	0.6352	0.032	0.6002	0.0229	0.5505	0.032	0.513	0.05	0.5284	0.0671
PCIT	0.5972	0.0236	0.5879	0.0242	0.5133	0.0269	0.4909	0.0432	0.5084	0.0564
TIGRESS	0.6257	0.0178	0.557	0.0258	0.5051	0.0163	0.5251	0.0412	0.4863	0.0504
Median GENECI	**0.6813**	**0.0368**	**0.6093**	**0.0347**	**0.5869**	**0.0433**	**0.5305**	**0.0577**	**0.5291**	**0.0693**
Best GENECI	**0.6918**	**0.0373**	**0.6115**	**0.0351**	**0.5905**	**0.0436**	**0.5336**	**0.0583**	**0.5299**	**0.0694**

can be seen in the illustrated graphs. First, in the case of Yeast2 graph, BC3NET and GENIE3_GBM obtain quite different weights depending on the execution. The fact that the fitness values remain constant is a sign that their variability is not related to the algorithm's convergence. Moreover, since all other techniques remain constant, it can be deduced that the weight increase in one is usually reflected in a decrease in the weight of the other, although being this balance completely indifferent with regard to the fitness value of the solution. The only case in which this is possible is when the techniques infer quite similar networks, and therefore the granting of greater weight to one or the other is practically indiscernible for the consensual network.

Second, in the case of GENIE3_RF for the Yeast3 network, there is a quite similar situation to the previous one, except that on this occasion, the increase or decrease in the weight of this technique is uniformly assumed by the rest. This may be a consequence of the fact that the confidence values of this list are quite close to the median of the remaining ones, and therefore, voting during consensus is somewhat redundant.

Size 100

Finally, the experimentation on the DREAM3 challenge networks is concluded by presenting the results associated with the 100-node networks. Similar to the previous section, Table 6 shows how the results provided by GENECI are practically at the same level as those obtained by the best individual techniques.

In fact, it is observed that the increase in size continues to bring benefits to the results. While in the 50-node table, only in one case GENECI came to exceed the maximum AUROC of the individual techniques, in the 100-node table, this occurs for 3 of the 5 networks. This means that the optimization performed on the topological characteristics of the network becomes more meaningful the larger the size of the network to be inferred.

Fig. 6 shows the Ecoli1 network agreed upon by GENECI in the run corresponding to the median of the AUROC values. As can be seen, the network is clearly scale-free, as the degrees of the different nodes are distributed non-uniformly. The interactive 3D representation generated using the Python package built in this proposal is very useful for network analysis, allowing rotations, overlapping techniques, zooming, the query of confidence values, etc. The fitness curves for the 25 runs performed on this network are shown on the right. In them, one can visualize how a few runs seem to have stagnated at a local minimum, which is also appreciable in the violin plot located at the top right. After several tests, it has been shown that GENECI tends to converge well before 50,000 evaluations, so reducing this value would represent some gain concerning the execution time without harming the quality of the results.

Finally, it is worth mentioning that the AUROC and AUPR values for some of the individual techniques had already been calculated previously in the literature. Specifically, some accuracy values for the DREAM3 challenge networks can be observed in [46,48], where the resemblance of the results to those obtained in this work provides some reliability in the study addressed. It is striking to note the low quality of accuracy that is achieved today in the task of inferring GRNs, and it is for this reason that it remains a significant area of research.

5.1.2. DREAM4

The DREAM4 challenges were a new edition of the challenges already exploited in DREAM3 by incorporating new datasets. Similar to what was seen in the previous edition, the in silico network inference challenge was classified in two parts according to the size of the networks. In this case, they were established in network sizes of 10 and 100 nodes, and as for DREAM3, there were 5 networks in each group. For the 10-node networks, expression levels of 21 time points and 5 replicates were provided, and for the 100-node networks, another 21 time points, but in this case 10 replicates.

Networks vary in their topology, but all of them try to mimic real organisms such as Escherichia coli or Saccharomyces cerevisiae. The data attempt to simulate their known dynamic properties, which are simulated using different initial conditions and kinetic parameters. Stochastic differential equations were used to generate the expression data for each topology, followed by the addition of noise proportional to the gene expression level (as seen in the real microarray datasets). In addition, four sets of observations are available for each network: time series, wild type, knock-out and knockdown.

This edition can be considered the most studied of all the DREAM challenges. The proportion of evaluation scripts and the background imposed in the previous editions prompted several participants to test their inference proposals. As in the previous section, the presentation of the results will be divided into sub-challenges, detailing first the results obtained for the 10-node networks and then for the 100-node networks.

Size 10

Table 7 shows the AUROC and AUPR values for each of the individual techniques and the median of these metrics for the 25 independent runs of the proposal. Indeed, as in the previous challenge, the nodes' scarcity and the networks' small size lead to moderate results for GENECI. This is visible in networks 1, 2 and 5, as well as in the AUPR of network 3. However, the exception of network 4 stands out, where GENECI achieved competitive results. In this case, unlike the exception seen in DREAM3 for networks of the same size, the AUROC and AUPR values provided by the individual techniques are neither homogeneous nor of low quality, so this time it is considered a merit on the part of the algorithm.

To analyze this sporadic behavior, Fig. 7 shows for Net-4 10-gene the graphs related to the different individual techniques and

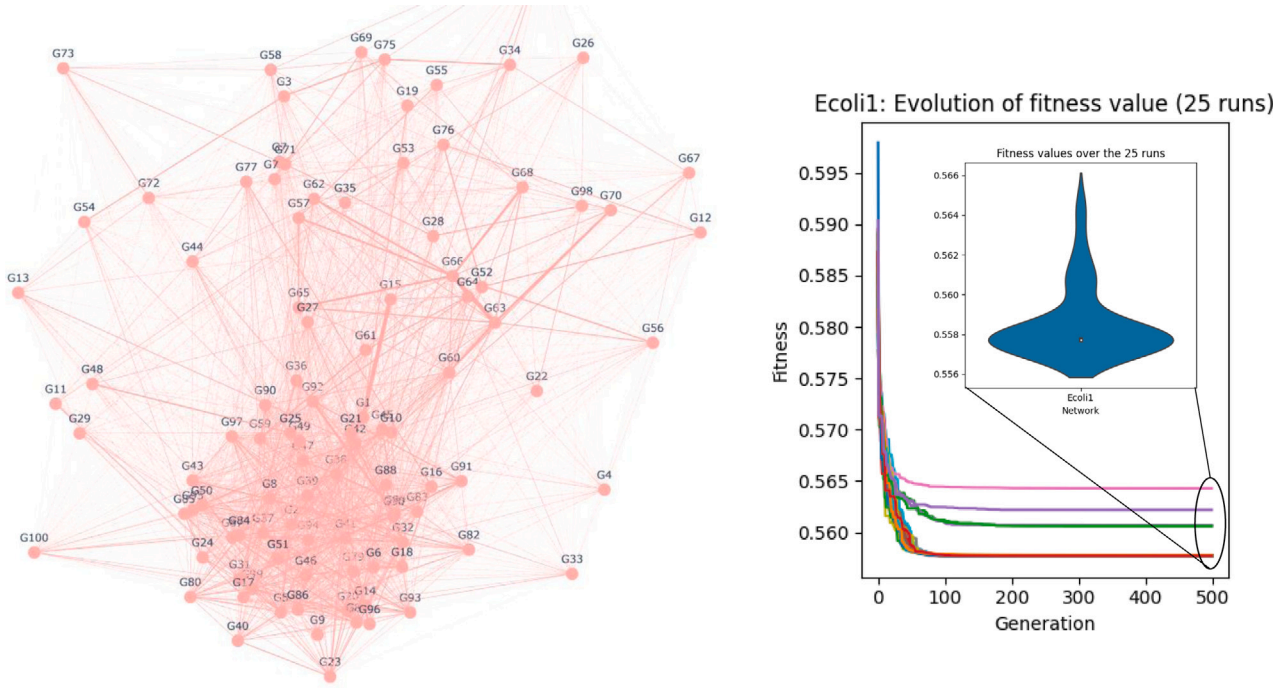

Fig. 6. For the first 100-gene Ecoli network of the DREAM3 challenge, the consensus gene network calculated in the run whose AUROC corresponds to the median illustrated in Table 6 is plotted on the left. On the right, the evolution of the fitness values obtained during the 25 runs and a violin plot representing the distribution of their corresponding final values.

Table 7
Accuracy values for DREAM4 and size 10 networks. For each gene regulatory network included in this dataset (columns), the AUPR and AUROC values are shown after comparing the networks inferred by the different techniques (rows) with the respective gold standards.

Técnica	D4_10_1		D4_10_2		D4_10_3		D4_10_4		D4_10_5	
	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR
ARACNE	0.6236	0.331	0.4489	0.171	0.5618	0.2771	0.6693	0.3033	0.6688	0.3168
BC3NET	0.7236	0.4237	0.4954	0.1841	0.5671	0.1936	0.6349	0.2785	0.6613	0.2166
C3NET	0.6636	0.354	0.4865	0.1823	0.5262	0.2625	0.6344	0.2866	0.6966	0.3497
CLR	0.6507	0.3483	0.4861	0.1881	0.5947	0.2693	0.6893	0.2694	0.6912	0.3648
GENIE3_ET	**0.8631**	**0.4533**	0.614	0.2401	0.6533	0.2437	0.7073	0.2686	**0.8259**	0.4167
GENIE3_GBM	0.664	0.2638	0.5971	0.2296	**0.696**	0.3017	0.6883	0.2658	0.6613	0.3193
GENIE3_RF	0.8284	0.4441	**0.6326**	0.2546	0.6898	**0.3486**	0.6883	0.3097	0.8024	**0.4379**
KBOOST	0.5858	0.2324	0.603	0.2368	0.576	0.2533	**0.7383**	**0.3412**	0.6667	0.2393
MRNETB	0.6867	0.3614	0.4907	0.1838	0.6302	0.2936	0.6718	0.317	0.6704	0.34
MRNET	0.6493	0.3439	0.5046	0.1876	0.5422	0.2744	0.7118	0.3173	0.672	0.3416
PCIT	0.5884	0.3262	0.5819	**0.2948**	0.5649	0.2084	0.5395	0.2562	0.5577	0.2579
TIGRESS	0.5973	0.3257	0.614	0.2191	0.4871	0.1674	0.4915	0.15	0.4017	0.1268
Median GENECI	**0.7689**	**0.4371**	**0.5887**	**0.2709**	**0.6933**	**0.274**	**0.7493**	**0.3445**	**0.7906**	**0.4248**
Best GENECI	**0.7733**	**0.4393**	**0.5938**	**0.2724**	**0.6987**	**0.2777**	**0.7572**	**0.3561**	**0.8077**	**0.4638**

the GENECI consensus gene network corresponding to the median. In the graphs, despite the small size of the network, a certain scale-free distribution is shown, where the most interconnected node also has the most intense relationships. This means that when GENECI rewards individuals that increase the degree of the only existing hub (1/10 = 10% of genes), it is actually bringing the consensus network closer to the gold standard. This behavior in the rest of the networks is unsatisfactory because there is no single hub in the network since, as explained in the section on DREAM3, networks of this size tend to have a random configuration.

Size 100

Again, incorporating a larger number of nodes favors the optimization of the topological characteristics of the consensus network. Table 8 shows the results for the 100-node networks (DREAM4). It should be remembered that, as in the rest of the sections, obtaining good precision values by GENECI implies incorporating a reliable method that guarantees outstanding results for networks of various densities and characteristics (discarding the excessively small ones of 10 nodes). This implies that when it is desired to infer a gene regulatory network whose structure is unknown, the application of GENECI provides a reliable and effective resolution method. This is not feasible with the individual techniques since, as shown in the tables, they tend to provide good results only for specific subsets of problems, reporting less satisfactory results for the rest.

Unlike the other cases, on this occasion, GENECI seems to overcome the individual inference techniques through the AUPR values. This occurs in 3 of the 5 exposed networks, highlighting notably the consensus addressed on network number 5 of this subchallenge.

Fig. 8 shows the consensus gene network corresponding to the median calculated by GENECI. Again, it can be observed how the distribution of node degrees is not uniform, with a large group being mostly interconnected in contrast to the rest of the network. On the right are the fitness curves, which in this case, reflect the correct convergence of all the executions and the end in a fixed and concrete fitness value.

Finally, it is worth mentioning that, as with DREAM3, previous works have tested different techniques for inferring networks from this

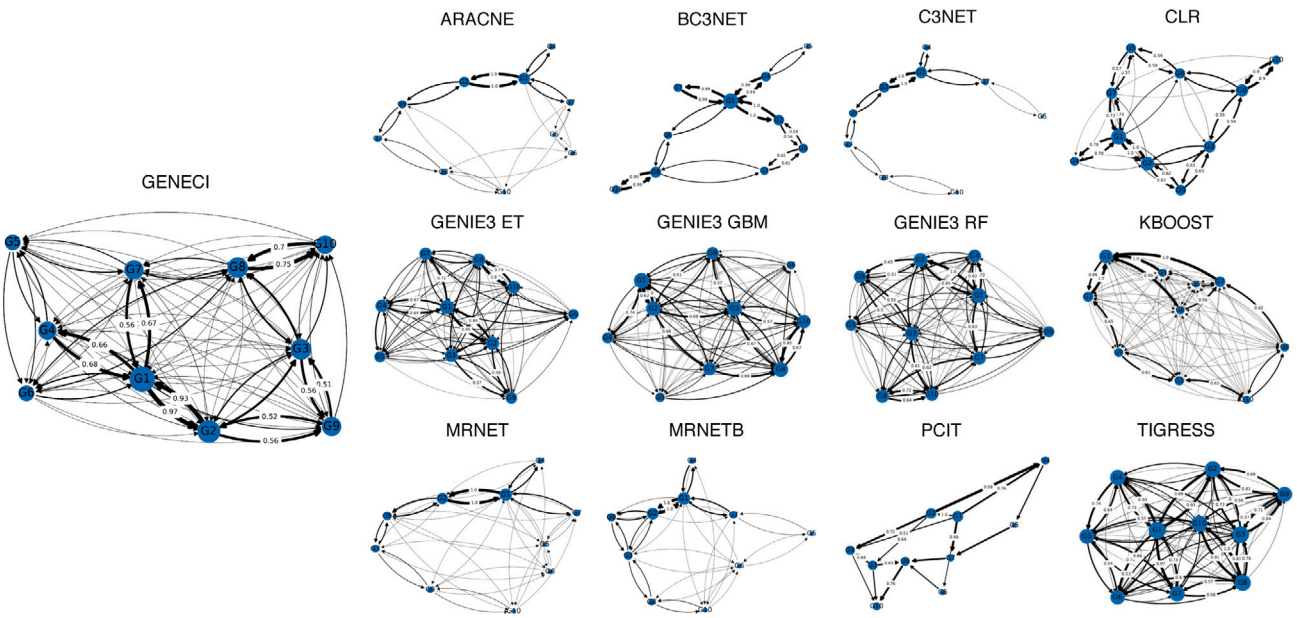

Fig. 7. For Net-4 10-gene of the DREAM4 challenge, the gene networks inferred by the individual techniques and the consensus gene network computed in the run whose AUROC corresponds to the median exposed in Table 7 are illustrated. In these graphs we can see how each gene corresponds to a node and each edge to a specific gene interaction. The directionalities are expressed by arrows and the confidence values by the thickness of the links, even specifying their value when this is highly significant.

Table 8
Accuracy values for DREAM4 and size 100 networks. This table shows the results of the evaluation scripts run on each of the individual (first band) and consensus (second band) results for each problem network (columns). The best value of the individual techniques and the values of the consensus networks per geneci (best and median of all runs) are shown in bold.

Técnica	D4_100_1		D4_100_2		D4_100_3		D4_100_4		D4_100_5	
	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR
ARACNE	0.5568	0.0317	0.5412	0.0453	0.559	0.0673	0.5525	0.0419	0.5826	0.0594
BC3NET	0.5629	0.0334	0.5312	0.0315	0.5905	0.0617	0.5537	0.06	0.5939	0.0565
C3NET	0.5345	0.029	0.518	0.038	0.5522	0.0664	0.54	0.0383	0.5622	0.0575
CLR	0.6963	0.048	0.6291	0.0578	0.7079	0.1036	0.6654	0.0621	0.6768	0.0764
GENIE3_ET	**0.7733**	**0.0756**	0.6741	0.0526	0.7289	0.1143	0.7046	0.0683	0.7555	0.0842
GENIE3_GBM	0.7608	0.0567	**0.6929**	0.0624	0.719	0.0983	0.7046	0.0634	**0.7707**	0.0815
GENIE3_RF	0.756	0.062	0.6873	0.0633	**0.7411**	**0.1182**	**0.7195**	0.0698	0.7694	0.082
KBOOST	0.6135	0.0461	0.5234	0.0431	0.5764	0.054	0.5247	0.0367	0.5171	0.0414
MRNETB	0.6848	0.047	0.6334	**0.0639**	0.7169	0.1076	0.6667	0.0604	0.6798	0.0739
MRNET	0.6771	0.0446	0.6322	0.0583	0.7124	0.1022	0.6622	0.0568	0.6786	0.081
PCIT	0.6172	0.0614	0.5649	0.0486	0.6149	0.1017	0.5988	**0.0772**	0.6339	**0.0894**
TIGRESS	0.6581	0.0261	0.5595	0.0617	0.6476	0.0319	0.6281	0.0402	0.6509	0.0312
Median GENECI	**0.7613**	**0.0674**	**0.6631**	**0.0685**	**0.7316**	**0.1241**	**0.7078**	**0.0742**	**0.7368**	**0.1143**
Best GENECI	**0.7623**	**0.0676**	**0.665**	**0.0691**	**0.7396**	**0.1265**	**0.7097**	**0.0748**	**0.7377**	**0.1144**

challenge. Results and quality metrics related to DREAM4 networks can be seen in [26,45–49], where again the kinship between these values and those calculated in this work bring some confidence to this study.

5.1.3. DREAM5

In the next edition of the DREAM challenges, the scientific community was invited to infer genome-scale transcriptional regulatory networks from gene expression data derived from an in silico benchmark (Net 1), the human pathogen S.aureus (Net 2), the prokaryotic model organism E. coli (Net 3) and finally the eukaryotic model organism S.cerevisiae (Net 4). These nets had a size of 1643, 2810, 4511 and 5950 genes, respectively, which gave them greater veracity from a biological point of view. However, during the evolution of the challenge, the network related to the human pathogen, i.e., the second network of the challenge, was eventually discarded from the evaluation process. It is clear that since the interactions between the organisms involved were not fully known, the gold standards were somewhat incomplete. Although this phenomenon has been present in many of the benchmarks collected in the literature, the case of the S. aureus network stood out for its few interactions available with experimental support, invalidating the evaluation process too much. For this reason, like many of the works presented in the literature, only networks 1, 3 and 4 of this challenge will be worked with.

Due to its large size, 15 runs have been carried out for the last network instead of 25. Therefore, in this case, the values reported for Net 4 in Table 9 refer to the AUROC and AUPR values concerning those 15 runs (denoted by *). Due to the slowness of the single TIGRESS technique, it has been finally discarded from the optimization process for this challenge. In the table, it can be observed that the results are again favorable. There is an exception for the AUROC obtained in the third network of the challenge. However, it is worth mentioning that the distance observed concerning the best individual technique is due to its accuracy being a clear outlier for the rest of the values. These results finally demonstrate that GENECI responds satisfactorily to the input of gene expression data from large networks with biological support.

For individual inference techniques, similar results can be observed in [22,26], where the absence of the second gene network during the evaluation process is indeed appreciated.

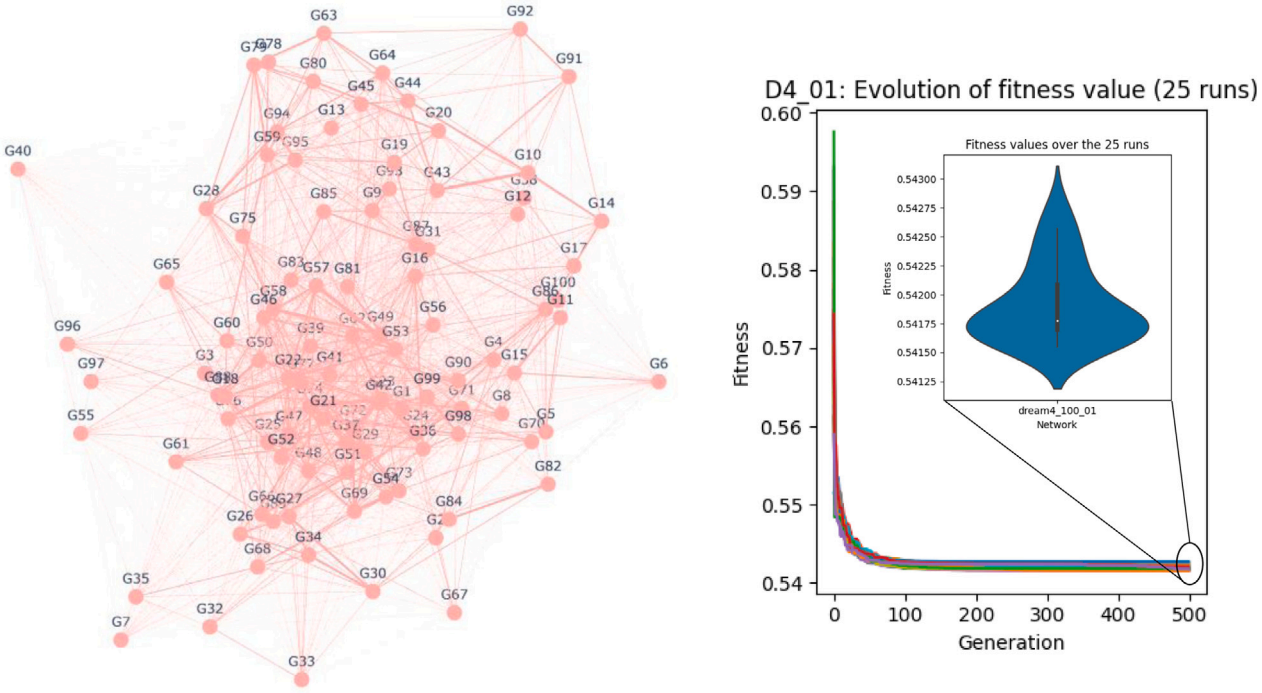

Fig. 8. For the first 100-gene network of the DREAM4 challenge, the consensus gene network calculated in the run whose AUROC corresponds to the median illustrated in Table 8 is plotted on the left. On the right, the evolution of the fitness values obtained during the 25 runs and a violin plot representing the distribution of their corresponding final values.

Table 9
Accuracy values for DREAM5 networks. AUPR and AUROC values are provided for each technique and problem, highlighting in bold the results obtained by GENECI and the best obtained by any of the individual techniques.

Técnica	D5_1		D5_3		D5_4	
	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR
ARACNE	0.5538	0.1011	0.5131	0.0296	0.5005	0.0176
BC3NET	0.5673	0.0966	0.5149	0.0257	0.5006	0.0176
C3NET	0.5393	0.0865	0.5071	0.0243	0.5005	0.0176
CLR	0.7402	0.2234	0.5892	0.0602	0.5211	0.0212
GENIE3_ET	**0.8149**	0.2502	**0.6632**	**0.0879**	0.543	0.0222
GENIE3_GBM	0.7952	**0.3042**	0.6108	0.0682	0.5318	0.0208
GENIE3_RF	0.8135	0.2802	0.6535	0.0844	**0.5494**	**0.0223**
KBOOST	0.4679	0.0634	0.5572	0.0445	0.5037	0.0182
MRNETB	0.7421	0.2	0.5948	0.0697	0.52	0.0195
MRNET	0.7404	0.2054	0.5943	0.0557	0.521	0.0196
PCIT	0.6761	0.1712	0.5751	0.0621	0.5173	0.0194
Median GENECI	**0.8007**	**0.2788**	**0.6211**	**0.0751**	**0.5316***	**0.0214***
Best GENECI	**0.8024**	**0.2801**	**0.6264**	**0.0762**	**0.5317***	**0.0214***

Table 10
Accuracy values for IRMA networks. In this table, a gene network is contemplated for each pair of columns, where in each row the AUPR and AUROC values are provided for each inference technique.

Técnica	IRMA_switch-off		IRMA_switch-on	
	AUROC	AUPR	AUROC	AUPR
ARACNE	0.6667	0.6815	0.6667	0.6815
BC3NET	0.5833	0.5679	0.5833	0.5679
C3NET	0.6667	0.6815	0.6667	0.6815
CLR	0.6111	0.5339	0.7222	0.609
GENIE3_ET	0.6667	0.6815	**0.8611**	**0.7865**
GENIE3_GBM	0.5	0.4	0.75	0.7759
GENIE3_RF	0.6944	0.6261	0.75	0.7759
KBOOST	**0.7778**	**0.7099**	0.6111	0.5339
MRNETB	0.6667	0.6815	0.6667	0.6815
MRNET	0.6667	0.6815	0.6667	0.6815
PCIT	0.5	0.4	0.5	0.4
TIGRESS	**0.7778**	0.6	0.7778	0.6
Median GENECI	**0.8611**	**0.7865**	**0.8889**	**0.75**
Best GENECI	**0.8611**	**0.7865**	**0.8889**	**0.75**

5.1.4. IRMA

The In vivo Reverse-engineering and Modeling Assessment (IRMA) network was created to evaluate the performance of different gene network reconstruction methods. Specifically, quantitative RT-PCR on the yeast Saccharomyces cerevisiae measured expression levels at different time points. The network has 5 genes (CBF1, GAL4, SWI5, GAL80 and ASH1) and 6 regulatory interactions, giving rise to "switch on" and "switch off" versions by culturing cells in galactose or glucose, respectively.

The synthetic network includes several regulatory interactions, capturing the behavior of large eukaryotic gene networks but on a smaller scale. The network was designed to be negligibly affected by endogenous genes and to respond to galactose, which triggers the transcription of its genes. This network is deceptively simple, yet it is actually quite articulate in its interconnections. In fact, they include regulatory chains, single-entry motifs, and multiple feedback loops generated by the combination of transcriptional activators and repressors.

Table 10 shows the AUROC and AUPR results for each of the individual techniques and the median of the 25 independent runs of GENECI. It should be noted that the evaluation process carried out on these networks differs from the previous ones. In this case, IRMA does not offer specific evaluation scripts, so the confidence lists reported by the different techniques have been binarized to perform this task. Subsequently, the generic evaluation subcommand has been used, facing the binary classification problem.

Finally, it can be seen that after applying the same evaluation criteria to all the networks, GENECI is again shown to perform well (see Table 10). It outperforms the AUROC and AUPR maxima of the individual techniques for most cases, with a remarkably significant difference in the case of the "switch off" network. The graphs reported by the different inference techniques and the consensus network constructed by GENECI are shown in Fig. 9. It can be seen that the network is

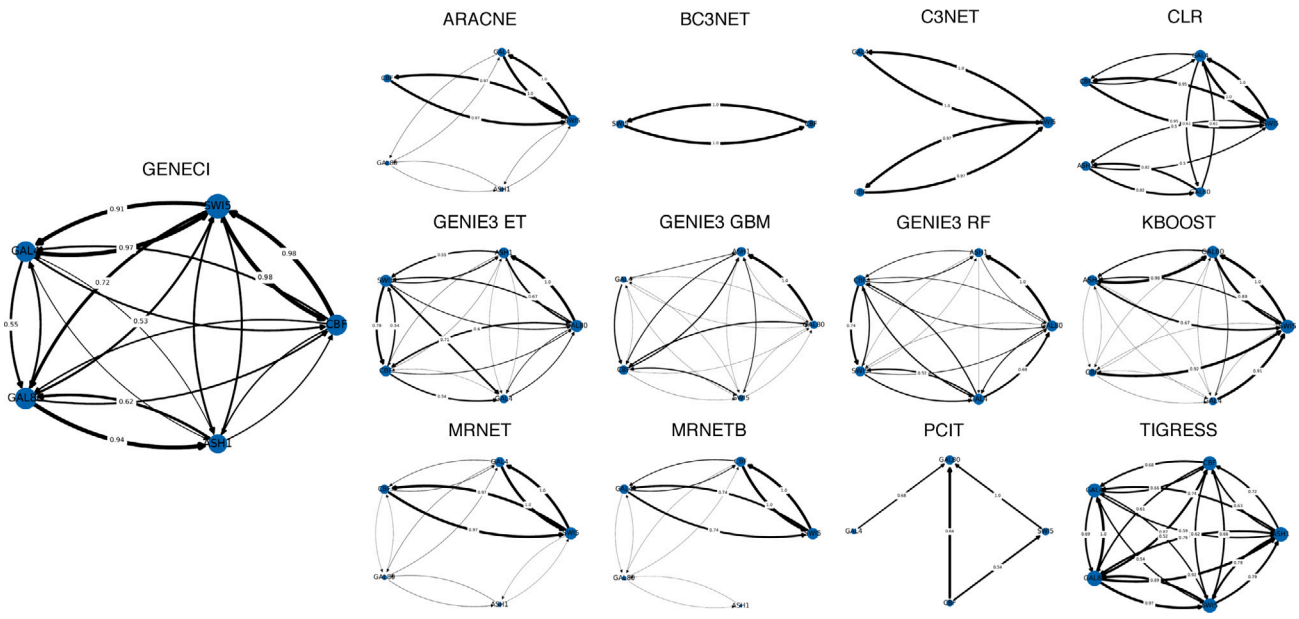

Fig. 9. For the "switch on" IRMA network, the gene networks inferred by the individual techniques and the consensus gene network calculated in the run whose AUROC corresponds to the median exposed in Table 10 are illustrated. The graphs shown in this figure try to represent the different inferred networks, arranging the genes in the same layout in order to facilitate visual comparison between them.

really small, with a clear disagreement in the set of proposals when establishing the existing interactions.

The IRMA network has been frequently used in the literature to test various techniques for network reconstruction [22,46–49]. However, the lack of a rigorous official criterion for measuring the accuracy of the different tools has given the scientific community a certain margin to choose the procedure that best suits its proposal in each case. For this reason, the results reported in the literature on the levels of quality differ considerably between articles, making it difficult to compare them in the first instance.

5.1.5. Statistical significance

In this section, a statistical analysis is addressed that allows a rigorous comparison of the precision obtained by each of the individual inference techniques, as well as the consensus of these techniques developed by GENECI, from a global point of view that considers all the results presented so far (except for the MELANOMA dataset).

According to Friedman's statistical ranking and Holm's non-parametric tests [52] performed for both AUROC and AUPR values (see Table 11 and 12 respectively), the best GENECI result is the one that obtains the first position in both cases (thus acting as a control denoted with *).

After this, it can be seen that the median of GENECI and the techniques derived from GENIE3 are also in good positions, and no statistical difference in their performance can be assured. The rest of the techniques show statistically lower performances since, in their case, the null hypothesis of Holm's test is rejected.

GENIE3 obtained such good results during our experimentation due to the main use of the DREAM challenges as a data source. This algorithm shows a clear specialization of the time series exposed in these challenges, as it won several times. This is part of what is discussed in the manuscript about the unidentified specialization of the different inference techniques, and that causes that in the absence of a gold standard, it is not possible to know a priori which is the best tool to infer the problem network. This is what GENECI tries to solve, i.e., it does not try to outperform all the individual techniques in their domains of specialization (since, by the No Free Lunch theorem itself, this would be impossible) but aims to obtain quality results (competitive concerning the best technique) for a wide range of problems, gaining generalization capacity and relieving the researcher of the need to choose and rely on an individual technique.

On the other hand, it is worth mentioning that GENECI, as well as boosting the weight of the most promising techniques for each dataset,

Table 11
Friedman mean rank with Holm's adjusted p values (0.05) for AUROC.

AUROC		
Algorithm	$Friedman's Rank$	$Holm's Adj-p$
***Best GENECI**	**3.100e+00**	–
GENIE3_RF	3.833e+00	0.742e+00
GENIE3_ET	4.067e+00	0.742e+00
Median GENECI	4.367e+00	0.723e+00
GENIE3_GBM	5.617e+00	0.079e+00
MRNETB	7.417e+00	3.215e−04
CLR	7.467e+00	3.170e−04
MRNET	8.083e+00	2.769e−05
TIGRESS	9.117e+00	2.034e−07
PCIT	1.002e+01	1.518e−09
ARACNE	1.002e+01	1.518e−09
BC3NET	1.033e+01	2.344e−10
C3NET	1.052e+01	7.896e−11
KBOOST	1.105e+01	2.386e−12

Table 12
Friedman mean rank with Holm's adjusted p values (0.05) for AUPR.

AUPR		
Algorithm	$Friedman's Rank$	$Holm's Adj-p$
***Best GENECI**	**2.867e+00**	–
GENIE3_RF	3.783e+00	0.412e+00
Median GENECI	4.233e+00	0.412e+00
GENIE3_GBM	5.200e+00	0.092e+00
GENIE3_ET	5.433e+00	0.070e+00
CLR	7.333e+00	1.772e−04
MRNETB	7.517e+00	1.002e−04
MRNET	8.117e+00	8.194e−06
PCIT	8.350e+00	3.074e−06
ARACNE	9.250e+00	3.082e−08
C3NET	9.750e+00	1.857e−09
BC3NET	1.077e+01	2.853e−12
KBOOST	1.117e+01	1.846e−13
TIGRESS	1.123e+01	1.233e−13

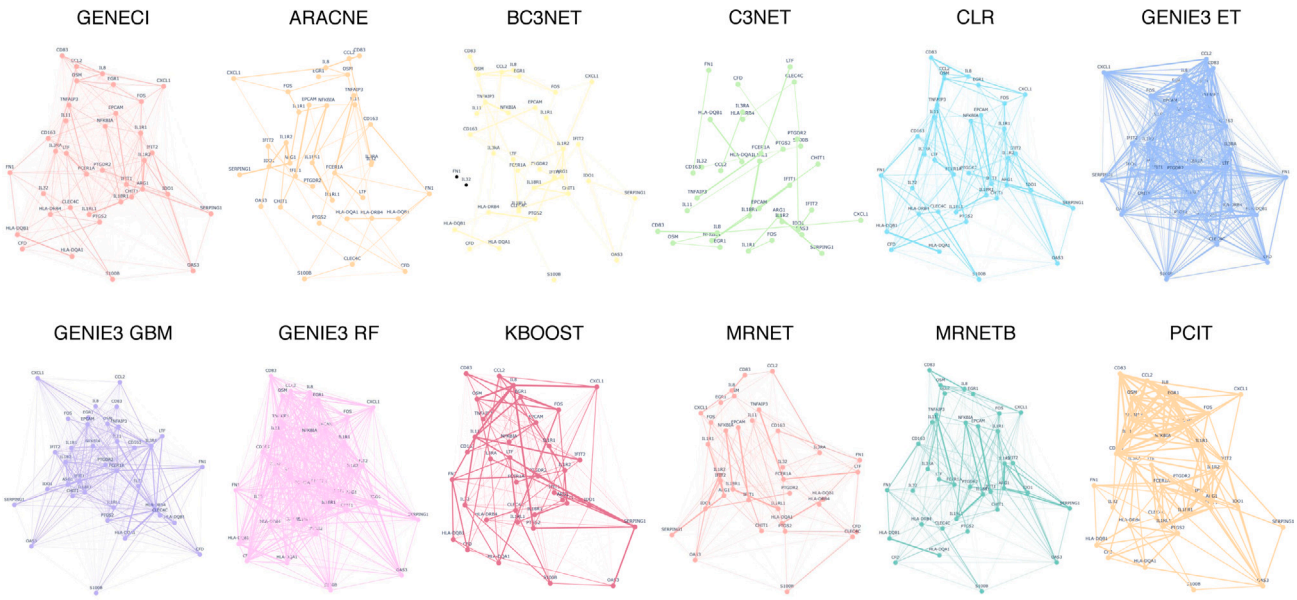

Fig. 10. For Melanoma expression data, gene networks inferred by multiple individual techniques as well as the consensus one produced by GENECI are represented. Graphs attempt to represent gene regulatory networks by setting up genes through nodes and interactions through links. In this case, as they are captures of interactive representations, the directionality of the interactions is not visible to the naked eye. However, the equal arrangement of nodes allows the topology of the different networks to be easily compared.

has shown that it can quickly silence those that are not. The presence of noise from the less competitive techniques has not been detected, and GENECI has not hesitated to assign low weights to them in those cases where it has been considered pertinent.

All this assures that the GENECI proposal obtains a highly competitive performance and ensures robustness for its use in the inference of real-world datasets for which no previous solutions are known.

5.2. Real-world: MELANOMA

Finally, GENECI has also experimented on non-simulated gene expression data. Specifically, real-world clinical data from melanoma patients are used. These include gene expression levels obtained from NanoString,[3] i.e., from the platform comprising the immunological profiling panel. This panel was treated and subjected to specific filtering techniques, resulting in the elimination of 35 genes showing less stable results.

Since, in this case, there is no gold standard to validate the accuracy of the consensus network, a review of the current literature will proceed to manually check the existence of the most relevant inferred relationships in studies related to melanoma and immunology. Fig. 10 shows the results of the different individual techniques and the GENECI consensus gene network at the top left. The three relationships that obtained the highest level of confidence after consensus are (1) IL1R2-ARG1, (2) IL18R1-IL1RL1 and (3) HLA-DQA1-HLA-DQB1.

Regarding the first interaction, several studies relate both genes within the context of cancer [53,54]. Specifically, in [55] it is concluded that the amebiasis pathway could be involved in melanoma metastasis through these genes. Regarding the second connection, despite appearing in the literature as highly associated with allergic pathologies such as asthma or dermatitis [56–58], in [59] both genes are related to repressors involved in epithelial cancers. Finally, for the third interaction, in addition to numerous articles that relate the HLA antigen family to this type of cancer [60,61], in some of them, this relationship becomes the protagonist, and the central axis of the study [62].

6. Conclusions

GENECI has demonstrated that its ability to optimize the consensus of several inference techniques leads to high-quality and accurate results. Moreover, its application on several datasets has allowed it to be defined as an effective general methodology that provides good results for networks of diverse characteristics, thus overcoming the limitations imposed by some individual techniques due to their specialization in specific regulatory networks.

This novel ensemble approach organized by an optimization algorithm is proposed in this study for GRNs inference. Therefore, it is intended to bring to the state-of-the-art novel and proven efficient techniques encapsulated in a ready-to-use software package.

The construction of this procedure is a practice with clear indications of lasting over time since implementing new inference techniques that improve the results of the current ones would not mean that this work would be outdated. The higher the quality of the confidence lists provided to GENECI, the higher the quality of the consensus network built. Therefore, the emergence of new techniques could be beneficial and worthy of study for incorporation into the ensemble produced by the evolutionary algorithm.

Moreover, the integration of all the necessary software in a single tool implemented in a language as well known as Python facilitates its subsequent use by external users, who have easy access and installation procedure thanks to the hosting of the code in a repository in GitHub[4] and its availability in PyPI.[5] The complementary functionalities incorporated into the tool (inference using known individual techniques, application of cut-off criteria, generic evaluation, static and interactive representation of networks, etc.) and its easy execution using a simple hierarchy of commands encourage the installation and use of this tool. Its flexibility to incorporate other expression data or confidence lists from other techniques eliminates any limitation and allows the user to elaborate on new experiments beyond the one presented in this work.

After developing this study and gaining a more detailed understanding of the gene network inference problem, a series of possible modifications that could further improve the results have emerged.

[3] https://www.nanostring.com/

[4] https://github.com/AdrianSeguraOrtiz/GENECI

[5] https://pypi.org/project/geneci/

First, the design of new fitness functions with some biological context or that explores in greater detail the topological term of the current one could further align the optimization process with the accuracy obtained during inference. From a different perspective, introducing a multi-objective approach that deals with existing conflicts between several functions are a territory to be explored whose application could provide significant benefits. Finally, integrating new inference techniques or even allowing the input of labeled data could be a great idea to bring more knowledge to the consensus process, improving its trajectory and results.

Regarding future experiments, based on previous collaboration with Dr. Miguel Berciano [34,63], the intention is to work on the clinical analysis of the network of melanoma patients to discover new biomarkers for this problem. In addition, to expand the scope of this proposal in the field of real biological networks, new and more current datasets will be used to cover other clinical pathologies.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been partially funded by grant (funded by MCIN/AEI/10.13039/501100011033/) PID2020-112540RB-C41, AETHER-UMA, Spain (A smart data holistic approach for context-aware data analytics: semantics and context exploitation) and Andalusian PAIDI program, Spain with grant P18-RT-2799. Funding for open access charge: Universidad de Málaga, Spain/CBUA. Adrián Segura-Ortiz is supported by Grant FPU21/03837 (Spanish Ministry of Science, Innovation and Universities, Spain)

References

[1] G. Orphanides, D. Reinberg, A unified theory of gene expression, Cell 108 (2002) 439–451.
[2] M. Pertea, The human transcriptome: An unfinished story, Genes 3 (2012) 344–360.
[3] Overview: Eukaryotic gene regulation (article) | Khan Academy, 2016.
[4] C. Barbosa, I. Peixeiro, L. Romão, Gene expression regulation by upstream open reading frames and human disease, PLoS Genetics 9 (2013) e1003529.
[5] P.H.L. Krijger, W.D. Laat, Regulation of disease-associated gene expression in the 3D genome, Nat. Rev. Mol. Cell Biol. 2016 17:12 17 (2016) 771–782.
[6] M.G.V.D. Wijst, D.H.D. Vries, H. Brugge, H.J. Westra, L. Franke, An integrative approach for building personalized gene regulatory networks for precision medicine, Genome Med. 2018 10:1 10 (2018) 1–15.
[7] A.N. Burska, K. Roget, M. Blits, L.S. Gomez, F.V.D. Loo, L.D. Hazelwood, C.L. Verweij, A. Rowe, G.N. Goulielmos, L.G.V. Baarsen, F. Ponchel, Gene expression analysis in RA: towards personalized medicine, Pharmacogenomics J. 2014 14:2 14 (2014) 93–106.
[8] S. Romagnoli, E. Fasoli, V. Vaira, M. Falleni, C. Pellegrini, A. Catania, M. Roncalli, A. Marchetti, L. Santambrogio, G. Coggi, S. Bosari, Identification of potential therapeutic targets in malignant mesothelioma using cell-cycle gene expression analysis, Am. J. Pathol. 174 (2009) 762–770.
[9] X. Yang, H. Han, D.D. DeCarvalho, F.D. Lay, P.A. Jones, G. Liang, Gene body methylation can alter gene expression and is a therapeutic target in cancer, Cancer Cell 26 (2014) 577–590.
[10] E. Davidson, M. Levin, Gene regulatory networks, Proc. Natl. Acad. Sci. 102 (14) (2005) 4935.
[11] V.A. Huynh-Thu, G. Sanguinetti, Gene regulatory network inference: An introductory survey, in: Gene Regulatory Networks, Springer, 2019, pp. 1–23.
[12] C. Soneson, M. Delorenzi, A comparison of methods for differential expression analysis of RNA-seq data, BMC Bioinformatics 14 (2013) 1–18.
[13] Q. Yang, Y. Wang, Y. Zhang, F. Li, W. Xia, Y. Zhou, Y. Qiu, H. Li, F. Zhu, NOREVA: Enhanced normalization and evaluation of time-course and multi-class metabolomic data, Nucleic Acids Res. 48 (2020) W436–W448.
[14] J. Fu, Y. Zhang, Y. Wang, H. Zhang, J. Liu, J. Tang, Q. Yang, H. Sun, W. Qiu, Y. Ma, Z. Li, M. Zheng, F. Zhu, Optimization of metabolomic data processing using NOREVA, Nat. Protoc. 17 (2022) 129–151.
[15] Q. Yang, B. Li, J. Tang, X. Cui, Y. Wang, X. Li, J. Hu, Y. Chen, W. Xue, Y. Lou, Y. Qiu, F. Zhu, Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data, Brief. Bioinform. 21 (2020) 1058–1068.
[16] M. Hecker, S. Lambeck, S. Toepfer, E. van Someren, R. Guthke, Gene regulatory network inference: Data integration in dynamic models—A review, Biosystems 96 (2009) 86–103.
[17] A.A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R.D. Favera, A. Califano, ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context, BMC Bioinformatics 7 (2006) 1–15.
[18] G. Altay, F. Emmert-Streib, Inferring the conservative causal core of gene regulatory networks, BMC Syst. Biol. 4 (2010) 1–13.
[19] R. de Matos Simoes, F. Emmert-Streib, Bagging statistical network inference from large-scale gene expression data, PLoS One 7 (2012) e33624.
[20] J.J. Faith, B. Hayete, J.T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J.J. Collins, T.S. Gardner, Large-scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles, PLoS Biol. 5 (2007) 0054–0066.
[21] V.A. Huynh-Thu, A. Irrthum, L. Wehenkel, P. Geurts, Inferring regulatory networks from expression data using tree-based methods, PLoS One 5 (2010) e12776.
[22] L.F. Iglesias-Martinez, B.D. Kegel, W. Kolch, KBoost: A new method to infer gene regulatory networks from gene expression data, Sci. Rep. 2021 11:1 11 (2021) 1–13.
[23] P.E. Meyer, K. Kontos, F. Lafitte, G. Bontempi, Information-theoretic inference of large transcriptional regulatory networks, EURASIP J. Bioinform. Syst. Biol. 2007 (2007).
[24] P. Meyer, D. Marbach, S. Roy, M. Kellis, Information-theoretic inference of gene networks using backward elimination, in: BioComp, Citeseer, 2010, pp. 700–705.
[25] A. Reverter, E.K. Chan, Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks, Bioinformatics (Oxford, England) 24 (2008) 2491–2497.
[26] A.C. Haury, F. Mordelet, P. Vera-Licona, J.P. Vert, TIGRESS: Trustful inference of gene regulation using stability selection, BMC Syst. Biol. 6 (2012) 1–17.
[27] P. Bellot Pujalte, Study of Gene Regulatory Networks Inference Methods from Gene Expression Data, Universitat Politècnica de Catalunya, 2017.
[28] H. Khojasteh, A. Khanteymoori, M.H. Olyaee, EnGRNT: Inference of gene regulatory networks using ensemble methods and topological feature extraction, Inform. Med. Unlocked 27 (2021) 100773.
[29] H. Jiang, T. Turki, S. Zhang, J.T. Wang, Reverse engineering gene regulatory networks using graph mining, in: International Conference on Machine Learning and Data Mining in Pattern Recognition, Springer, 2018, pp. 335–349.
[30] S. Peignier, B. Sorin, F. Calevro, Ensemble learning based gene regulatory network inference, in: 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence, ICTAI, 2021, pp. 113–120.
[31] C. Fujii, H. Kuwahara, G. Yu, L. Guo, X. Gao, Learning gene regulatory networks from gene expression data using weighted consensus, Neurocomputing 220 (2017) 23–33.
[32] P. Meyer, J. Saez-Rodriguez, Advances in systems biology modeling: 10 years of crowdsourcing DREAM challenges, Cell Syst. 12 (6) (2021) 636–653.
[33] I. Cantone, L. Marucci, F. Iorio, M.A. Ricci, V. Belcastro, M. Bansal, S. Santini, M. di Bernardo, D. di Bernardo, M.P. Cosma, A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches, Cell 137 (1) (2009) 172–181.
[34] I. Navas-Delgado, J. García-Nieto, E. López-Camacho, M. Rybinski, R. Lavado, M.Á.B. Guerrero, J.F. Aldana-Montes, VIGLA-M: Visual gene expression data analytics, BMC Bioinformatics 20 (4) (2019) 1–11.
[35] D. Marbach, J.C. Costello, R. Küffner, N.M. Vega, R.J. Prill, D.M. Camacho, et al., Wisdom of crowds for robust gene network inference, Nat. Methods 2012 9:8 9 (2012) 796–804.
[36] M.N. Haque, N. Noman, R. Berretta, P. Moscato, Heterogeneous ensemble combination search using genetic algorithm for class imbalanced data classification, PLoS One 11 (1) (2016) e0146116.
[37] H. Al-Sahaf, Y. Bi, Q. Chen, A. Lensen, Y. Mei, Y. Sun, B. Tran, B. Xue, M. Zhang, A survey on evolutionary machine learning, J. R. Soc. New Zealand 49 (2) (2019) 205–228.
[38] A.J. Nebro, J.J. Durillo, M. Vergne, Redesigning the jmetal multi-objective optimization framework, in: Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation, ACM, New York, NY, USA, 2015.
[39] K. Deb, R.B. Agrawal, et al., Simulated binary crossover for continuous search space, Complex Syst. 9 (2) (1995) 115–148.
[40] L.J. Eshelman, J.D. Schaffer, Real-coded genetic algorithms and interval-schemata, in: L.D. WHITLEY (Ed.), Foundations of Genetic Algorithms, Vol. 2, Elsevier, 1993, pp. 187–202.
[41] R. Storn, K. Price, Differential evolution–A simple and efficient heuristic for global optimization over continuous spaces, J. Global Optim. 11 (4) (1997) 341–359.
[42] M. Srinivas, L. Patnaik, Adaptive probabilities of crossover and mutation in genetic algorithms, IEEE Trans. Syst. Man Cybern. 24 (4) (1994) 656–667.

[43] H. Chiroma, S. Abdulkareem, A. Abubakar, A. Zeki, A.Y. Gital, M.J. Usman, Correlation study of genetic algorithm operators: Crossover and mutation probabilities, in: Proceedings of the International Symposium on Mathematical Sciences and Computing Research, 2013, pp. 6–7.
[44] E. Cox, Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration, Elsevier, 2005.
[45] R. Aghdam, M. Ganjali, X. Zhang, C. Eslahchi, CN: A consensus algorithm for inferring gene regulatory networks using the SORDER algorithm and conditional mutual information test, Mol. Biosyst. 11 (3) (2015) 942–949.
[46] S. Hurtado, J. García-Nieto, I. Navas-Delgado, A.J. Nebro, J.F. Aldana-Montes, Reconstruction of gene regulatory networks with multi-objective particle swarm optimisers, Appl. Intell. 51 (2021) 1972–1991.
[47] J. Pirgazi, A.R. Khanteymoori, A robust gene regulatory network inference method base on Kalman filter and linear regression, PLoS One 13 (7) (2018) e0200094.
[48] J. García-Nieto, A.J. Nebro, J.F. Aldana-Montes, Inference of gene regulatory networks with multi-objective cellular genetic algorithm, Comput. Biol. Chem. 80 (2019) 409–418.
[49] L.F. Iglesias-Martinez, W. Kolch, T. Santra, BGRMI: A method for inferring gene regulatory networks from time-course gene expression data and its application in breast cancer research, Sci. Rep. 6 (1) (2016) 1–12.
[50] T. Saito, M. Rehmsmeier, The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets, PLoS One 10 (3) (2015) e0118432.
[51] R.J. Prill, D. Marbach, J. Saez-Rodriguez, P.K. Sorger, L.G. Alexopoulos, X. Xue, N.D. Clarke, G. Altan-Bonnet, G. Stolovitzky, Towards a rigorous assessment of systems biology models: The DREAM3 challenges, PLoS One 5 (2) (2010) e9202.
[52] J. Derrac, S. García, D. Molina, F. Herrera, A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, Swarm Evol. Comput. 1 (1) (2011) 3–18.
[53] P. Andersson, Y. Yang, K. Hosaka, Y. Zhang, C. Fischer, H. Braun, S. Liu, G. Yu, S. Liu, R. Beyaert, et al., Molecular mechanisms of IL-33–mediated stromal interactions in cancer metastasis, JCI Insight 3 (20) (2018).
[54] S. Ahmad, P. Singh, A. Sharma, S. Arora, N. Shriwash, A.H. Rahmani, S.A. Almatroodi, K. Manda, R. Dohare, M.A. Syed, Transcriptome meta-analysis deciphers a dysregulation in immune response-associated gene signatures during sepsis, Genes 10 (12) (2019).
[55] R. Xie, B. Li, L. Jia, Y. Li, Identification of core genes and pathways in melanoma metastasis via bioinformatics analysis, Int. J. Mol. Sci. 23 (2) (2022) 794.
[56] N.E. Reijmerink, D.S. Postma, M. Bruinenberg, I.M. Nolte, D.A. Meyers, E.R. Bleecker, G.H. Koppelman, Association of IL1RL1, IL18R1, and IL18RAP gene cluster polymorphisms with asthma and atopy, J. Allergy Clin. Immunol. 122 (3) (2008) 651–654.
[57] O.E. Savenije, J.M.M. John, R. Granell, M. Kerkhof, F.N. Dijk, J.C. de Jongste, H.A. Smit, B. Brunekreef, D.S. Postma, K. Van Steen, et al., Association of IL33–IL-1 receptor–like 1 (IL1rl1) pathway polymorphisms with wheezing phenotypes and asthma in childhood, J. Allergy Clin. Immunol. 134 (1) (2014) 170–177.
[58] T. Hirota, A. Takahashi, M. Kubo, T. Tsunoda, K. Tomita, M. Sakashita, T. Yamada, S. Fujieda, S. Tanaka, S. Doi, et al., Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population, Nature Genet. 44 (11) (2012) 1222–1226.
[59] D. Ellinghaus, H. Baurecht, J. Esparza-Gordillo, E. Rodríguez, A. Matanovic, I. Marenholz, N. Hübner, H. Schaarschmidt, N. Novak, S. Michel, et al., High-density genotyping study identifies four new susceptibility loci for atopic dermatitis, Nature Genet. 45 (7) (2013) 808–812.
[60] P. Larrieu, L.-H. Ouisse, Y. Guilloux, F. Jotereau, J.-F. Fonteneau, A HLA-DQ5 restricted melan-A/MART-1 epitope presented by melanoma tumor cells to CD4+ T lymphocytes, Cancer Immunol. Immunotherapy 56 (10) (2007) 1565–1575.
[61] B. Huang, W. Han, Z.-F. Sheng, G.-L. Shen, Identification of immune-related biomarkers associated with tumorigenesis and prognosis in cutaneous melanoma patients, Cancer Cell Int. 20 (1) (2020) 1–15.
[62] X. Wang, F. Almazan, Y.G. Montoyo-Pujol, A. Martin-Casares, A. Martin, T. Cabrera, M.A. López-Nevot, HLA-DRB116: 01 and HLA-DQB105: 02 Alleles influence the susceptibility and progression of cutaneous malignant melanoma, J. Oncology 2021 (2021).
[63] S. Hurtado, H. Nematzadeh, J. García-Nieto, M.-Á. Berciano-Guerrero, I. Navas-Delgado, On the use of explainable artificial intelligence for the differential diagnosis of pigmented skin lesions, in: I. Rojas, O. Valenzuela, F. Rojas, L.J. Herrera, F. Ortuño (Eds.), Bioinformatics and Biomedical Engineering, Springer International Publishing, Cham, 2022, pp. 319–329.