Computers in Biology and Medicine 179 (2024) 108850

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Multi-objective context-guided consensus of a massive array of techniques 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] Department 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 network
Inference
Expression data
Multi-objective evolutionary algorithm
Consensus

ABSTRACT

Background and Objective: Gene Regulatory Network (GRN) inference is a fundamental task in biology and medicine, as it enables a deeper understanding of the intricate mechanisms of gene expression present in organisms. This bioinformatics problem has been addressed in the literature through multiple computational approaches. Techniques developed for inferring from expression data have employed Bayesian networks, ordinary differential equations (ODEs), machine learning, information theory measures and neural networks, among others. The diversity of implementations and their respective customization have led to the emergence of many tools and multiple specialized domains derived from them, understood as subsets of networks with specific characteristics that are challenging to detect a priori. This specialization has introduced significant uncertainty when choosing the most appropriate technique for a particular dataset. This proposal, named MO-GENECI, builds upon the basic idea of the previous proposal GENECI and optimizes consensus among different inference techniques, through a carefully refined multi-objective evolutionary algorithm guided by various objective functions, linked to the biological context at hand.

Methods: MO-GENECI has been tested on an extensive and diverse academic benchmark of 106 gene regulatory networks from multiple sources and sizes. The evaluation of MO-GENECI compared its performance to individual techniques using key metrics (AUROC and AUPR) for gene regulatory network inference. Friedman's statistical ranking provided an ordered classification, followed by non-parametric Holm tests to determine statistical significance.

Results: MO-GENECI's Pareto front approximation facilitates easy selection of an appropriate solution based on generic input data characteristics. The best solution consistently emerged as the winner in all statistical tests, and in many cases, the median precision solution showed no statistically significant difference compared to the winner.

Conclusions: MO-GENECI has not only demonstrated achieving more accurate results than individual techniques, but has also overcome the uncertainty associated with the initial choice due to its flexibility and adaptability. It is shown intelligently to select the most suitable techniques for each case. The source code is hosted in a public repository at GitHub under MIT license: https://github.com/AdrianSeguraOrtiz/MO-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

Understanding the mechanisms that control gene expression is fundamental in biomedical research, as they play a crucial role in a wide variety of cellular processes and diseases. A gene regulatory network is a model that represents the interactions between genes and transcription factors, providing valuable insights into functional relationships and information flows within a cell or tissue [1].

The inference of gene regulatory networks from expression data has become an essential tool for discovering and comprehending the complexity of these networks. By analyzing expression data obtained from real biological experiments, it is possible to identify gene interactions and hence, to construct models that capture network dynamics [2–4].

However, this inference process is not without its challenges. Multiple techniques and approaches have been developed for inferring gene regulatory networks, each with its strengths and limitations. These

* Corresponding author.
E-mail address: adrianseor.99@uma.es (A. Segura-Ortiz).

https://doi.org/10.1016/j.compbiomed.2024.108850
Received 8 February 2024; Received in revised form 3 July 2024; Accepted 3 July 2024
Available online 15 July 2024
0010-4825/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

techniques have evolved to address different aspects of regulatory inference, such as indirect relationships [5,6], redundancy elimination [7,8], or false positive detection [9,10]. Each technique has a specific domain of expertise and can provide accurate results in its target area [11,12].

Another important consideration in the inference of gene regulatory networks is the high variability and inherent noise in expression data. Biological experiments can be subject to several sources of variability, such as genetic variation among individuals, experimental noise, and environmental conditions. These factors can introduce biases and make it challenging to identify underlying genetic interactions accurately [13].

Both problems reinforce the need for a consensus algorithm for gene regulatory network inference. On the one hand, researchers often face the challenge of determining which technique is most appropriate for their expression data set, as it is difficult to know the specific domain to which it belongs. Choosing the wrong technique can lead to suboptimal or even incorrect results. A consensus algorithm removes the burden of selecting a technique from the researcher, allowing them to rely on a set of techniques instead of a single one. On the other hand, it has already been demonstrated that by combining the results of different methods, the analysis's robustness is improved [14–17], and the influence of noise and data variability in expression data is reduced [18]. Therefore, constructing an intelligent consensus algorithm beyond the simple combination of results is a field of study that can offer outstanding contributions.

In this article, MO-GENECI (Multi-Objective GEne NEtwork Consensus Inference) is introduced, presenting a novel approach with a multi-objective focus to guide the optimization of consensus inference from an extensive pool of inference techniques within the context of gene regulatory networks. It builds upon the groundwork laid out in GENECI [19] and enhances its implementation through:

- Implementation of a multi-objective approach involving new fitness functions. The individual aggregate terms of GENECI have been thoroughly analyzed, a deeper exploration has been conducted within the context of real biological networks, and a more rigorous procedure has been designed to enhance decision-making during the implementation of the functions (comparison of versions). First, a new *Quality* function has been implemented and improved through exhaustive comparisons of modifications. Second, the *Topology* function has been renamed as *Degree distribution* and refined. Thus, the *Topology* now dispenses with the binarization of the consensus network and has incorporated a goodness-of-fit test to a Pareto distribution directly applied to the confidence levels of interactions. Additionally, a third objective named *Motifs* has been implemented from scratch, based on detecting frequently occurring patterns in real gene regulation networks.
- New crossover and mutation operators specific to the addressed problem have been implemented, ensuring the feasibility of solutions after execution to eliminate the search space distortion caused by the use of repair operators.
- Replacement of the genetic algorithm using a more suitable and robust multi-objective model based on NSGA-II [20].
- The number of available techniques for the initial inference of gene regulation networks has been expanded. In the initial version, ten techniques were considered, whereas now, a total of 26 techniques are used. Furthermore, to minimize the execution time of this phase as much as possible, not only has the parallel execution of containers been maintained, but a function has been implemented to automatically and intelligently distribute the available cores, allocating more resources to computationally expensive parallelizable techniques.
- A parameter setting and analysis procedure has been designed using a larger dataset, thanks to the addition of new known gene networks and the integration of the SysGenSIM simulator [21]. This collection forms an experimental benchmark comprising 106 gene regulation networks sourced from diverse origins, ensuring comprehensive coverage across all specialization domains in this study.

MO-GENECI has been tested on 106 gene regulatory networks, including networks of several characteristics and sizes (up to 2000 genes). This provides an ideal experimental scenario to determine the scope of this proposal and its capacity to optimize consensus under different situations. The results have determined that MO-GENECI provides more accurate networks than individual inference techniques and that its ability to exploit the best of each technique in each case is beneficial across the diversity of the dataset.

Therefore, applying MO-GENECI in the inference of gene regulatory networks from expression data offers an efficient and reliable solution to overcome the uncertainty associated with choosing the appropriate technique. Through this approach, researchers enhance their ability to obtain accurate and relevant results in the biomedical context. This has significant implications in this field, where a more precise understanding of gene regulatory networks can lead to discoveries and advancements in the understanding and treatment of diseases [22–25].

This article is organized into five sections. In Section 2, the efforts made so far in the inference of gene regulatory networks from expression data are discussed, along with proposals from the literature that design consensus mechanisms for this problem. Section 3 presents the algorithmic proposal of this work, detailing the chosen multi-objective model, the implemented operators, and the encoding of several fitness functions. Section 4 describes the experimentation undertaken in this study, including the description of the network set to be inferred, the parameter setting procedure, and the execution strategy. In Section 5, the results of the previous experimentation are presented, and the accuracy values obtained by this proposal are discussed in comparison to other inference techniques. Finally, Section 6 explains the conclusions of this work and its potential implications in the field of biomedicine.

2. State of the art

2.1. First approaches using artificial intelligence

The inference of gene regulatory networks (GRNs) from expression data is crucial in computational biology. Over the years, several computational approaches have been developed to tackle this challenging problem. These approaches leverage different algorithms and techniques to infer the regulatory relationships between genes from expression data.

One common approach is based on the use of probabilistic graphical models, such as Bayesian networks [26]. Bayesian networks provide a probabilistic framework to model the dependencies between genes and infer the causal relationships in a GRN. These models have been widely used due to their ability to handle uncertainty and incorporate prior knowledge. Another approach involves using ordinary differential equations (ODEs) to model gene expression dynamics and infer the regulatory interactions [27–29]. ODE-based methods capture the temporal behavior of gene expression and can reveal the causal relationships between genes.

In recent years, machine learning techniques, particularly artificial neural networks, have gained popularity in GRN inference [30–33]. Neural networks can capture complex patterns in gene expression data and learn the underlying regulatory relationships. These methods often use recurrent neural networks (RNNs) to model the temporal dependencies in gene expression time-series data. Additionally, integrative approaches that combine multiple types of omics data, such as gene expression, DNA methylation, and protein-protein interactions, have been proposed to improve the accuracy of GRN inference [34].

Furthermore, some methods leverage the concept of causality to infer GRNs. For example, the Granger causality method uses time series data to identify causal relationships between genes [35]. This approach has been successfully applied to infer GRNs in several biological systems. Information theory measures, such as mutual information, have also been explored to infer GRNs [36]. These methods quantify the statistical dependencies between genes and can reveal the regulatory interactions.

In relation to this work, it is worth mentioning that multi-objective evolutionary algorithms have previously been used to optimize some of these models. Due to the mathematical complexity of ordinary differential equations, these algorithms have frequently been employed to identify the parameters of the S-System, a mathematical model based on ODEs used to describe the dynamics of biological systems, especially in the inference of GRNs [28,29]. This, in a way, validates the use of such algorithms in this field when the objectives to be met are highly complex.

2.2. Single techique approaches

These different approaches to the inference of GRNs have led to the emergence of a wide range of computational techniques. Each is based on a particular approach or a combination of several, adding modifications and strategies that distinguish them from the others.

The following are some of them, the most prominent in the literature due to their accuracy and popularity[1]:

- ARACNE (Algorithm for the Reconstruction of Accurate Cellular NEtworks) [5]
- C3NET (Conservative Causal Core NETwork) [9]
- BC3NET (Bagging C3NET) [37]
- CLR (Context Likelihood or Relatedness network) [10]
- GENIE3 (GEne Network Inference with Ensemble of trees) [2]
- GRNBOOST2 (Gene Regulatory Network inference using gradient BOOSTing) [38]
- KBOOST (kernel PCA regression and gradient boosting to reconstruct gene regulatory networks) [39]
- MRNET (Minimum Redundancy NETworks) [6]
- MRNETB (Minimum Redundancy NETworks using Backward elimination) [7]
- PCIT (Partial Correlation coefficient with Information Theory) [8]
- TIGRESS (Trustful Inference of Gene REgulation with Stability Selection) [40]
- CMI2NI (Conditional Mutual Inclusivity principle-based Network Inference) [41]
- GRNVBEM (Gene Regulatory Network inference using Variational Bayesian Expectation-Maximization algorithm) [42]
- INFERELATOR (regression and variable selection to identify transcriptional influences on genes) [43]
- JUMP3 (based on a formal on/off model of gene expression but uses a non-parametric procedure based on decision trees, called “jump trees”, to reconstruct the GRN topology) [3]
- LEAP (Lag-based Expression Association for Pseudotime-series) [44]
- LOCPCACMI (Local Path Consistency Algorithm based on Conditional Mutual Information) [45]
- MEOMI (Mixed Entropy Optimizing context-related likelihood Mutual Information) [46]
- NARROMI (Noise And Redundancy reduction technology by combining Recursive Optimization and Mutual Information) [47]
- NONLINEARODES (NON-LINEAR Ordinary Differential EquationS) [48]
- PCACMI (Path Consistency Algorithm based on Conditional Mutual Information) [4]
- PIDC (Partial Information Decomposition and context) [49]
- PLSNET (PLS-based gene NETwork inference method) [50]
- PUC (Proportional Unique Contribution) [49]
- RSNET (Redundancy Silencing and Network Enhancement Technique) [51]

Furthermore, it is important to note the recent emergence of individual inference techniques with innovative approaches and ideas [52–55], underscoring the ongoing evolution and experimentation in the field. This highlights that inferring gene regulatory networks from expression data continues to be a bioinformatics problem of significant interest. These studies reflect the dynamic and exploratory nature of the current research landscape, where new methods are continually evolving and maturing to address this complex challenge.

2.3. Consensus based approaches

In the realm of gene regulatory network inference, the pursuit of consensus among the results of multiple techniques has been a prominent trend. The DREAM challenge [57] marked a significant turning point, as demonstrated in [11], by showcasing that combining outputs from several techniques yielded more accurate solutions compared to individual methods. This discovery spurred the exploration of diverse avenues in achieving consensus, such as topological feature analysis [14], graph mining [15], and evolutionary algorithms similar to the approach proposed [16,17,58].

However, it is emphasized that in the case of the latter algorithms, the optimization process is guided by labeled data from the problem at hand. In other words, the presence of the gold standard is required, making it impossible to infer networks that lack prior information.

Subsequent advancements in the field have unveiled novel strategies aiming at consensus among inference techniques. However, they still do not provide a robust methodology adapted to the realm of biological networks. On the one hand, EnGRaiN [59] approaches consensus from a purely mathematical perspective that does not consider the biological context of the problem. Furthermore, like previous cases, it employs a supervised procedure that again requires previously known interactions from the gold standard, limiting its applicability and conditioning the inference to a subset of interactions. On the other hand, AGRN [60] considers a fairly limited number of techniques, it has been tested on a small data set of a few networks and the proposal is again supervised. Nevertheless, the development and publication of both approaches demonstrate that the belief in the potential of a consensus solution from multiple techniques persists.

The challenge of overcoming the limitations arising from linking consensus techniques to labeled data was addressed in GENECI [19], a previous proposal that introduced the consensus of several techniques, considering the biological nature of the problem.

The current proposal, known as MO-GENECI, builds upon the original concept of GENECI, aiming to enhance it through a multi-objective approach that incorporates numerous innovations mentioned earlier. By embracing the multi-objective approach and considering a broader range of biological features, MO-GENECI strives to bridge further the gap between the inference of GRNs and biological reality, distancing the results from the purely mathematical realm.

Table 1 summarizes the main previously mentioned proposals for achieving consensus among different GRN inference techniques. As can be seen, most of these proposals require labeled data, which does not align with the objectives of this work. Among the remaining proposals, firstly, graph mining involves a high computational cost, limiting its testing to networks with a maximum of 30 nodes. Secondly, GENECI aims to reach consensus on a subset of current MO-GENECI techniques, pursuing only a preliminary subset of the current objectives.

[1] Since describing each of these techniques in this document would be overly lengthy, it has been decided to add them to the supplementary material in the main project repository: https://github.com/AdrianSeguraOrtiz/MO-GENECI/tree/main#integrated-techniques.

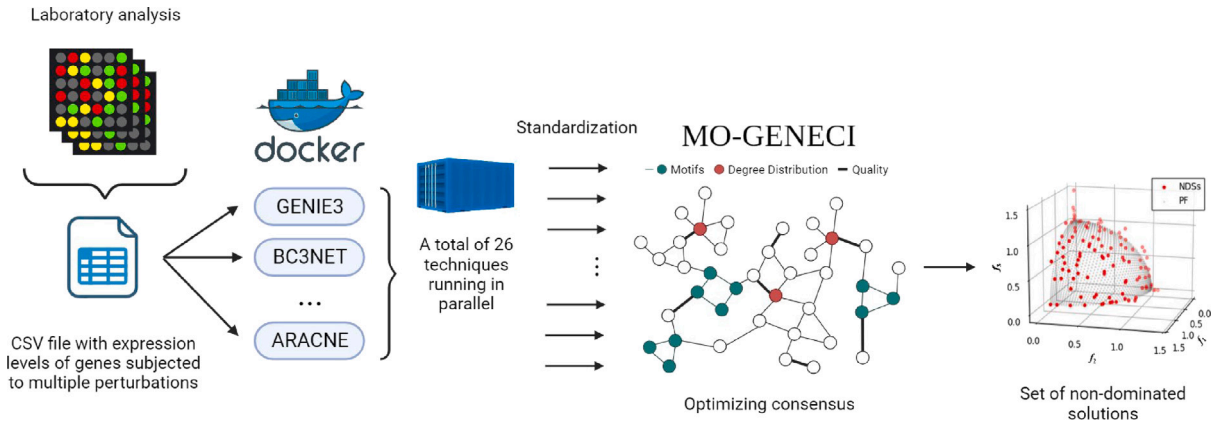


Fig. 1. Workflow Implemented in MO-GENECI. After providing the input expression data, the inference techniques are executed in parallel thanks to their encapsulation in Docker containers. Once the results of all techniques are obtained, the lists containing confidence values for each interaction are handed over to the multi-objective evolutionary algorithm. This algorithm generates an initial random population (weight vectors) and undergoes the iterative process of evaluation (*Quality*, *Degree distribution*, *Motifs*), selection, crossover, and mutation until the maximum number of iterations is reached. Upon completion of the multi-objective evolutionary algorithm, the result consists of the set of non-dominated solutions from the last generation [56].

Table 1
Summary of proposals for addressing the consensus of individual methods in the inference of gene regulatory networks. For each tool, it is provided the principle that its consensus strategy is intended to follow, whether the model training is supervised or unsupervised (requires the use of labeled data or gold standard), number of individual techniques involved in the process, number of individual techniques against which the results are compared and finally the number of problem instances that have been used to validate the proposal in the experimentation of the corresponding research article.

Tool	Based on	Supervision	Consensus	Comparative	Benchmark
EnGRNT [14]	Topology	Supervised	1 × n times	3	21
Graph Mining [15]	Common patterns	Unsupervised	3	3	6
Gen. Alg. (Binary) [16]	AUROC (GS)	Supervised	18	18	4
Gen. Alg. (Weight) [17,58]	True Positive (GS)	Supervised	9	9	20
EnGRaiN [59]	True Positive/Negative (GS)	Supervised	15	15	15
AGRN [60]	SHAP importance scores	Supervised	3	5	7
GENECI [19]	Coherence and Hubs	Unsupervised	12	12	30
MO-GENECI	Coherence, Topology and Motifs	Unsupervised	26	26	106

3. Algorithmic proposal

MO-GENECI represents a groundbreaking approach that combines multiple state-of-the-art inference techniques, harnessing their collective power to optimize the confidence levels associated with each interaction. This optimization is accomplished through a cutting-edge multi-objective evolutionary algorithm, carefully designed with customized operators and fitness functions directly aligned with the specific biological context under investigation.

The main workflow is outlined in detail in Fig. 1. This graphical representation illustrates the procedure to be followed, starting from the initial expression data analysis to the exportation of consensus networks. The Python package developed (available on PyPi[2]) includes complementary functions that cover all these steps, including individual GRN inference, which can be performed using 26 different inference techniques within the same working environment. However, the key and truly innovative phase is the "optimizing consensus" stage, where the specifically designed multi-objective evolutionary algorithm comes into play.

In this sense, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) [20] provides the structural support for MO-GENECI. This skeleton, which defines the main stages of the algorithm, is customized through the choice of a specific representation of a weighted voting system, the incorporation of carefully designed crossover and mutation operators, and the design of several fitness functions tailored to address this bioinformatic problem from the biological context to which it belongs. Fig. 2 shows a flow chart with the main phases of the designed algorithm. In the following sections, the logic implemented in each of these phases will be explained in greater detail.

NSGA-II is an evolutionary algorithm designed to solve multi-objective optimization problems. It is based on the principles of genetic algorithms and is characterized by three main features that are quite beneficial for the problem addressed:

- **Elitism**: The best solutions from the current population are carried over to the next generation. This ensures that the quality of solutions does not degrade over time.
- **Diversity Preservation**: NSGA-II uses a mechanism known as "crowding distance" to maintain diversity in the population. The crowding distance measures how close an individual solution is to its neighbors. Solutions with a larger crowding distance are preferred as they are less crowded.
- **Non-Dominated Sorting**: This is a method of ranking solutions based on their dominance. Solutions are sorted into different "fronts" based on the number of solutions that dominate them. The first front contains non-dominated solutions, the second front contains solutions dominated by the first front, and so on.

MO-GENECI follows the general outline of an NSGA-II, as depicted in Algorithm 1. In this pseudocode, the primary phases of the evolutionary algorithm are outlined. It is worth noting that some of these phases, such as the representation, fitness functions, and crossover and mutation operators, have been designed from scratch for the specific case under consideration in this study. Therefore, they will be further elaborated upon in subsequent sections to provide a comprehensive understanding of the approach.

Initially, a random population is created (line 1 in Algorithm 1), which is then evaluated using several fitness functions designed in this study (line 2 in Algorithm 1). Following this, an iterative process begins

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

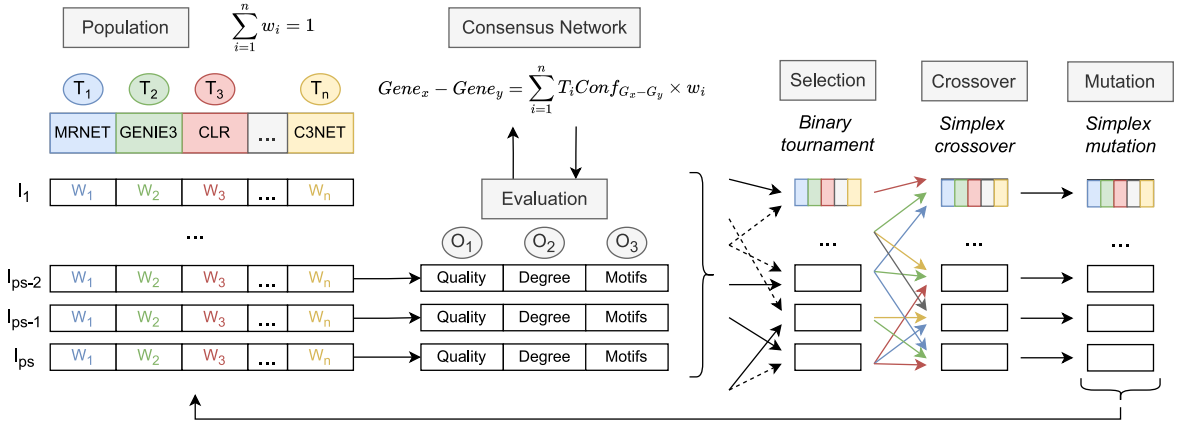


Fig. 2. Flowchart of the multi-objective evolutionary algorithm developed in this proposal. The algorithm starts with the generation of an initial population, whose representation is in the form of a vector of weights (simplex). The individuals are evaluated for each of the fitness functions, previously applying the conversion of the individual to its corresponding consensus network. This is followed by a selection process guided by a binary tournament. Finally, the individuals are subjected to the crossover and mutation operators selected for the representation of this problem, giving rise to the next generation of individuals.

Algorithm 1 MO-GENECI Algorithm.

Input Num of generations T, Population size P, Num of objectives M, Fitness functions $F : f_1, f_2, \ldots, f_M$, Crossover operator x, Mutation operator m

Output Pareto-optimal front PF

1: $P \leftarrow$ generate_random_population(P)
2: $E \leftarrow$ evaluate_population(P, F)
3: $t \leftarrow 1$
4: **while** $t < T$ **do**
5: $ranks \leftarrow$ rank_population(E)
6: $crowd\_dist \leftarrow$ crowding_distance($E, ranks$)
7: $selected \leftarrow$ select_population($P, ranks, crowd\_dist$)
8: $offspring \leftarrow$ crossover($selected, x$) ▷ Section 3.2
9: $offspring \leftarrow$ mutate($offspring, m$) ▷ Section 3.3
10: $P \leftarrow$ replace_population($P, offspring$)
11: $E \leftarrow$ evaluate_population(P, F) ▷ Section 3.4
12: $t \leftarrow t + 1$
13: $PF \leftarrow$ get_pareto_front(E)
14: **return** PF

and continues until the maximum number of generations is reached (lines 3, 4, and 12 in Algorithm 1). During this process, solutions in the population are classified based on non-dominance and assigned a rank and crowding distance (lines 5 and 6 in Algorithm 1). The crowding distance is used to maintain diversity, especially when splitting a front. The MO-GENECI selection process consists of two steps: first, individuals are selected based on their rank, and if ranks are equal, crowding distance is used. This approach is known as crowding distance tournament selection (line 7 in Algorithm 1).

After selection, specific algorithmic operators for adapted crossover and mutation are applied to generate offspring (lines 8 and 9 in Algorithm 1). Subsequently, parent and offspring populations are merged (line 10 in Algorithm 1), forming the next generation. This process repeats in a loop or, in the case of the final generation, the result becomes the Pareto front obtained from the algorithm (line 14 in Algorithm 1).

As the algorithmic skeleton of MO-GENECI, NSGA-II has been found to be effective in finding a better spread of solutions and better convergence near the true Pareto-optimal front compared to other multi-objective evolutionary algorithms [20]. Its advantages include its ability to solve multi-objective optimization problems, its elitism (which increases the convergence speed), and its parameter-less sharing approach.

However, despite all these advantages, other options were considered and ultimately discarded. Firstly, MOEA/D [61] does not perform well with non-normalized fitness functions, which does not fit in this case considering the *Motifs* fitness function explained below. Secondly, OMOPSO [62] has its own mutation operator and does not maintain solution feasibility. Thirdly, GDE3 [63] is designed to use differential evolution crossover, which also pushes individuals out of the search space. Finally, while compatible, SMPSO [64] does not take advantage of the specific problem-based design described below due to its lack of a crossover phase.

3.1. Solution representation

For the representation of individuals, weight vectors are employed, where each position specifies the weight assigned to a particular technique in the voting system. Therefore, the sum of all positions in the vector must always be equal to 1. This greatly influenced the implementation of the crossover and mutation operators in the evolutionary algorithm, which are deeply detailed in Sections 3.2 and 3.3, respectively.

3.2. Crossover

As mentioned earlier, it has been decided to implement a crossover operator that generates new solutions within feasible regions in the search space. Therefore, the chosen crossover operator is the *Simplex Crossover* [65], whose Java implementation has been adapted from the MOEA framework [66].

The *Simplex Crossover* is a multi-parent recombination operator proposed for real-coded genetic algorithms. This operator generates offspring by uniformly sampling values from the simplex formed by m ($2 \leq m \leq len\_vector + 1$) parental vectors. Experimental results with commonly used benchmark functions in evolutionary algorithm studies [65] have shown that *Simplex Crossover* performs well on functions with multimodality and/or epistasis with a moderate number of parents: 3 parents for low-dimensional functions or 4 parents for high-dimensional functions. Both values have been considered in the parameter setting procedure to ensure the proper choice of this variable.

Algorithm 2 Main code of Simplex Mutation operator.

Input Individual from population *ind*, Mutation probability *mutProb*, Mutation strength *mutStr*
Output Mutated individual *ind*

```
1: if randomDouble(0, 1) < mutProb then
2:     team1Size = randomInt(1, size(ind) - 1)
3:     team1 = randomSubset(ind, team1Size)
4:     team2Size = randomInt(1, size(ind) - team1Size)
5:     team2 = randomSubset(ind - team1, team2Size)
6:     amount = sum(team1) * mutStr
7:     mutateVariables(ind, norm(team1), -amount)
8:     mutateVariables(ind, norm(team2), +amount)
9: return ind
```

3.3. Mutation

Similar to the crossover phase, in the mutation phase, it was necessary to use an operator that allows the maintenance of solution feasibility after its execution. However, in this case, instead of adopting an operator from the literature, a new one has been designed specifically for this problem. This operator is named the *Simplex Mutation*, and it is based on applying a negative perturbation to a subset of the solution in such a way that its subsequent positive adjustment, rather than being spread across the entire vector, is directed towards another specific subset.

The *Simplex Mutation* operator requires the specification of two parameters: mutation probability and mutation strength. The former is commonly used in such operators and determines the probability with which an individual in the population undergoes the mutation process. The latter refers to the magnitude of the perturbation applied to the vector, also ranging between 0 and 1.

The pseudocode for its implementation is shown in Algorithm 2. The mutation procedure begins by extracting two subsets from the solution vector (lines 2 to 5 in Algorithm 2). The first subset will undergo the negative perturbation, while the second one will dampen this impact by positively reinforcing its values. Both the size of the groups and the positions in the vector that compose them are random. The only constraint is that a position cannot belong to both groups simultaneously.

Then, the values of the positions belonging to the first group are summed up. This sum, multiplied by the second input parameter (mutation strength), determines the value to be subtracted from the first subset (line 6 in Algorithm 2). This value is not uniformly distributed across the subset; instead, each member has a percentage of this value subtracted based on the normalization of the subset. Therefore, each member of the group has the calculated value subtracted, multiplied by its normalized figure within the subset (line 7 in Algorithm 2). After applying this negative perturbation, the amount subtracted from the first group is added to the second group following the same distribution procedure (line 8 in Algorithm 2).

An example is shown in Fig. 3 to clarify this implementation. It offers a possible mutation of the individual (0.1, 0.05, 0.2, 0.1, 0.15, 0.1, 0.05, 0.15, 0.1), which, given its length, would belong to a run that tries to agree on 9 concrete inference techniques. The first step is forming groups whose sizes and members are entirely random. Team 1 (in red) will suffer a loss in their values, while Team 2 (in green) will increase their numbers to compensate for the previous loss. In this example, the mutation strength value is set to 0.2. The total amount to subtract in team 1 is the sum of its members multiplied by this factor, in this case: $0.2 * 0.4 = 0.08$. This amount is divided to subtract each member's share based on the team's normalization. Since the second member is half of the team's total weight, half of the total amount will be subtracted: $0.08 * 0.5 = 0.04$. After performing this operation on each member, the reverse process is applied to the members of team 2. That is, each member's portion of this amount is added to each member based on their normalization.

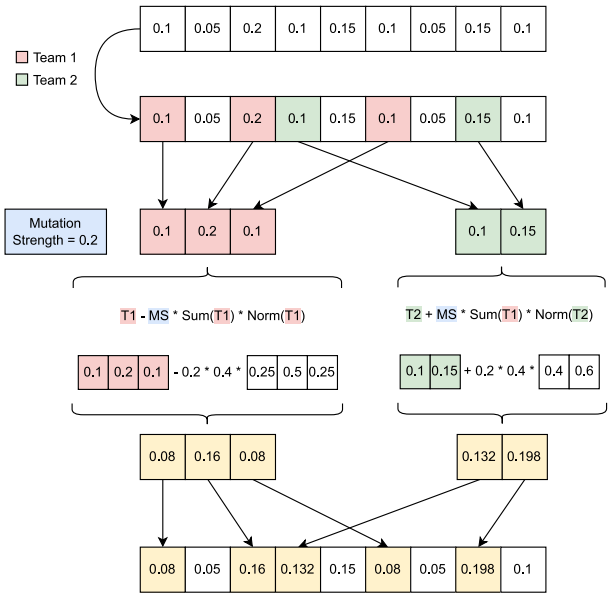


Fig. 3. Example of mutation of an individual belonging to an execution trying to agree on 9 inference techniques. In red are highlighted the members of team 1 (intended to reduce their values), in green the members of team 2 (intended to increase their values), in blue the value of the mutation strength factor (set in this case to 0.2) and finally the values finally modified in yellow.

Algorithm 3 First fitness function: Quality.

Input Consensus list with confidence and distance values c
Output Value of the fitness function *result*

```
1: quality = []
2: for i in len(c) do
3:     quality[i] = (conf_i + (1 − dist_i)) /2
4: qualitySum = 0
5: cnt = 0
6: for i in len(c) do
7:     if quality[i] > mean(quality) then
8:         qualitySum += quality[i]
9:         cnt += 1
10: result = 1 − (qualitySum/cnt)
11: return result
```

3.4. Evaluation

The evaluation process begins by transforming each individual into its corresponding consensus network proposal. In other words, for each interaction in each individual, the sum of the products between the confidence level reported by a technique and the weight assigned to it by the individual for the voting system is calculated. Afterwards, the individual and its corresponding consensus network are provided to each objective function for evaluation. MO-GENECI follows a 3-objective strategy, combining: *Quality*, *Degree distribution* and *Motifs*.

3.4.1. Objective 1: Quality

The *Quality* function aims to encourage the emergence of solutions whose consensus networks have a subgroup of interactions distinguished from the rest with high confidence levels that, in turn, originate from consistent weight distributions. In other words, it assigns greater importance to individual techniques whose reported values exhibit higher concordance than the remaining ones. The pseudocode for its implementation can be seen in Algorithm 3.

This function assigns a quality value to each interaction considered in the problem based on the solution being evaluated (lines 1 to 3 in

Algorithm 3). Two critical values are considered for each interaction to make this assignment. The first one is the consensus confidence level, which is obtained by summing the products of the confidence level given by each technique and its corresponding weight in the vector. The second value is called "distance" and is calculated as the difference between a specific vector's maximum and minimum values for the interaction. For each technique, this vector stores the average between its weight and, the difference between the confidence value assigned to the interaction and the median of the confidence levels of all techniques for that same case.

Finally, the quality value assigned to an interaction is obtained by calculating the average between its consensus confidence level and the result of subtracting one from the distance value. In other words, an interaction will be considered of good quality when its confidence level is high, and its distance value is small. A small distance value indicates that the maximum and minimum of the calculated vector are close to each other, which implies that lower weights have been assigned to techniques with confidence values far from the median compared to other techniques. It also indicates that techniques with confidence levels close to the median (i.e., with a smaller distance) have been compensated with high weights.

Finally, for the evaluated individual, the *Quality* function selects those interactions whose quality surpasses the mean (lines 4 to 9 in Algorithm 3) and returns the unit subtracted by the mean of this subset (lines 10 and 11 in Algorithm 3). Other options were tested to arrive at this definitive version of the function, including:

1. Changing the way the vector from which the distance originates is calculated by replacing the median with the mean.
2. In the calculation of the final quality, interactions above the cutoff criterion or all interactions are selected instead of those above the mean.
3. An additional step that influenced the result by evaluating the number of interactions that entered the previous subset. This step was present in the initial proposal under the term "contrast".

A temporary configuration was executed to make a decision regarding the definitive implementation of this objective in MO-GENECI. This involved replacing NSGA-II with a genetic algorithm while retaining the designed representation and operators. The only change was in the evaluation process, now considered a single fitness function. In each execution, this fitness function would be the specific version of *Quality* from which its individual accuracy was desired. For each version, the algorithm was run on all networks considered in this study with a size of less than 1000 genes (see Section 4.1). Subsequently, the accuracy of the results for each implementation was evaluated by comparing, for each network, the gold standard with the consensus network derived from the obtained solution.

This experimentation yielded an AUROC and AUPR value for each network and version. On the one hand, AUROC (Area Under the Receiver Operating Characteristic Curve) is a metric used to assess the discriminative ability of a classification model. It represents the model's ability to distinguish between positive and negative classes. On the other hand, AUPR (Area Under the Precision-Recall Curve) is another metric especially useful when classes are imbalanced, which is common in the inference of gene regulatory networks. It focuses on the true positive rate (recall) and the model's precision. Both metrics have been widely used to report on the efficacy of gene regulatory network inference techniques [2,3,38–40,46,48]

Finally, a statistical ranking of Friedman was computed using non-parametric Holm tests for both metrics [67]. Friedman's statistical ranking compares the performance of various proposals across multiple datasets. It calculates a test statistic based on the differences between the average ranks of the models across datasets. After calculating Friedman's statistical ranking, Holm tests are used to determine whether the winning version has significant differences in its performance compared to the rest. Holm tests control the Type I error by adjusting the p-values for each comparison, helping to prevent incorrect conclusions.

Table 2
Friedman mean rank with Holm's adjusted p values (0.05) for AUPR.

	AUPR	
Quality version	$Friedman's Rank$	$Holm's Adj - p$
***Median Average**	**3.2093**	–
Mean Average	3.38953	0.62946
Mean Cut-off	4.23256	0.01231
Median Cut-off	4.44186	0.0029
Median	5.02907	5.534e−06
Mean Average Contrast	5.02907	5.534e−06
Mean	5.15698	1.109e−06
Median Average Contrast	5.51163	4.982e−09

Table 3
Friedman mean rank with Holm's adjusted p values (0.05) for AUROC.

	AUROC	
Quality version	$Friedman's Rank$	$Holm's Adj - p$
***Median Average**	**2.85465**	–
Mean Average	3.36628	0.170793
Mean Cut-off	4.61047	5.19242e−06
Median Cut-off	4.70349	2.23266e−06
Median	4.8314	4.84291e−07
Mean	5.02326	3.20887e−08
Mean Average Contrast	5.07558	1.65298e−08
Median Average Contrast	5.53488	5.0559e−12

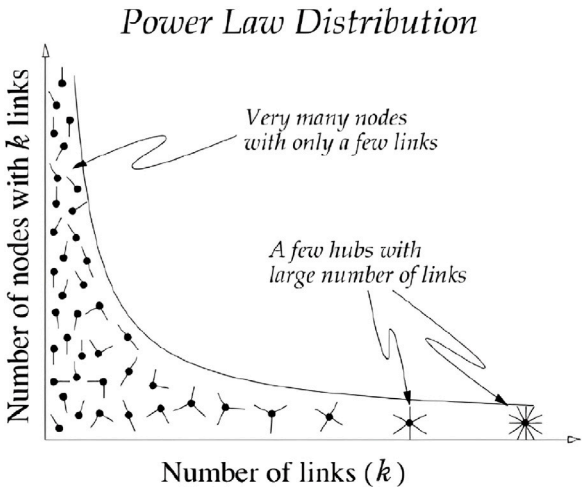


Fig. 4. Visualization of a degree distribution following a power-law [70].

The result can be seen in Tables 2 and 3. After observing the results, it was decided to choose the *Median Average* version as the definitive implementation of the *Quality* fitness function.

3.4.2. Objective 2: Degree distribution

The objective of *Degree distribution* is designed to favor solutions that lead to consensus networks with a distribution following a power-law. This characteristic of biological networks has been asserted in the literature [68,69], and even the most straightforward implementation supporting this idea yielded excellent results in large networks [19].

In a network with a degree distribution following a power-law, the frequency of nodes with a degree k is inversely related to k raised to a certain power. This leads to most nodes having few links while a few highly connected nodes, called "hubs", are present in the network. This idea can be visually appreciated with greater clarity in Fig. 4.

Algorithm 4 Second fitness function: Degree Distribution.

Input Consensus list with confidence values c
Output Value of the fitness function $result$
1: $degreeMap = \{String : Float\}$
2: **for** i **in** $len(c)$ **do**
3: $\quad weight = c[i][\text{“weight”}]$
4: $\quad degreeMap[c[i][\text{“source”}]] \mathrel{+}= weight$
5: $\quad degreeMap[c[i][\text{“target”}]] \mathrel{+}= weight$
6: $degreeArray = mapToSortedArray(degreeMap)$
7: $result = goodnessFitParetoTest(degreeArray)$
8: **return** $result$

Table 4
Friedman mean rank with Wilcoxon p values (0.05) for AUPR and AUROC.

Top. version	AUPR		AUROC	
	Friedman Rank	*Wilcox p*	*Friedman Rank*	*Wilcox p*
***Weighted**	**1.4186**	–	**1.3256**	–
Binarized	1.5814	0.0935	1.6744	0.0015

Given that this concept is quite broad, an attempt was made to find a more precise distribution that fits within this framework and allows for a more refined procedure, such as a goodness-of-fit test. This is the case with the Pareto distribution, a continuous probability distribution that fits within the scale-free distribution and follows a power-law.

In the context of gene regulatory networks, this distribution aligns with the idea that most genes have relatively low connections. In contrast, a few genes have a very high number of interactions [68]. This organization of the gene regulatory network has significant implications for network dynamics and robustness. Highly connected hubs play a crucial role in signal propagation and information integration within the network [71,72].

The implementation of this fitness function is reflected in Algorithm 4 and analyzes the consensus network calculated from the individual. For each node, it calculates its degree as the decimal sum of confidence levels for interactions, where the gene appears as a source or target (lines 1 to 5 in Algorithm 4). After storing the degree of all genes in a vector (line 6 in Algorithm 4), it is provided to the goodness-of-fit test for a Pareto distribution (line 7 in Algorithm 4, implemented based on [73]). It then returns a value related to the probability that the network follows this distribution.

Like with *Quality*, several versions of this function were considered. The first version (weighted), which has been described and is the definitive one, and a second version (binarized) that, instead of using decimal confidence levels, applied a cutoff criterion to binarize the network. In this case, since there are only two versions to compare, a Wilcoxon test has been applied. The results obtained are presented in Table 4, and despite not being able to claim a statistically significant difference between the two versions for the AUPR metric, the first one from the ranking was ultimately chosen.

3.4.3. Objective 3: Motifs

In the field of systems biology, it is known that biological networks commonly exhibit a series of patterns known as motifs. These patterns are typically specific configurations of interactions between molecules, such as proteins, genes, or metabolites, that repeat in different parts of the network. Motifs are considered basic structural and functional units of biological networks, and their study helps understand principles of organization, modularity, and dynamics in complex systems. The detection and analysis of motifs provide valuable information about molecular interactions and signal propagation in a network, contributing to the understanding of the function and regulation of biological systems [74–77].

The importance and well-established existence of these motifs in biological networks motivated the design of this third fitness function.

Algorithm 5 Third fitness function: Motifs.

Input Consensus list with confidence values c, List of motifs ids to detect $motifs$
Output Value of the fitness function $result$
1: $binaryMatrix = getBinaryMatrix(c)$
2: $key = deepHashCode(binaryMatrix)$
3: **if** key **in** $keys(cache)$ **then**
4: $\quad result = cache[key]$
5: **else**
6: $\quad g = getDirectedJGraph(binaryMatrix)$
7: $\quad result = 0.0$
8: $\quad$ **for** m **in** $motifs$ **do**
9: $\quad\quad result \mathrel{-}= count(m, g)$
10: **return** $result$

The idea is to encourage the emergence of solutions whose consensus networks have a high density of motifs that have already been confirmed in previous studies as common in biological networks and gene regulatory networks.

The implementation of this fitness function is represented in the pseudocode presented in Algorithm 5. First of all, it is necessary to perform a prior binarization exercise (line 1 in Algorithm 5) to detect these motifs in the consensus networks generated by the algorithm's solutions. In other words, to convert the consensually agreed confidence levels of interactions into a definitive statement of their existence or absence in the network. To do this, an appropriate cutoff criterion must be designed to establish as definitive those interactions that are more reliable.

The implemented cutoff criterion selects the top $x\%$ interactions in the network, i.e., those with the highest consensual confidence levels. Assigning this $x\%$ threshold does not need to be done through a rigorous parametric exercise, because the number of motifs found will be closely tied to how permissive the cutoff criterion is. In this regard, the more relationships are ultimately reported in the network, the greater the number of motifs detected. This is why the decision was made to choose the top 40% of interactions as the cutoff. It is a sufficiently high quantity to form a functional gene network, but not too large to prevent a substantial distinction between the consensus networks generated by different solutions.

However, it is evident that by integrating the binarization step into this fitness function, there will be cases where multiple solutions lead to the same binary network. To optimize the computational cost of the algorithm, a cache was created to avoid repeating the motif detection exercise in binary networks that were previously evaluated during the algorithm's execution (lines 2 to 5 in Algorithm 5).

Finally, the binary matrix is converted to a directed graph from the JGrapht library [78] to facilitate the detection of different motifs in the consensus binary network (line 6 in Algorithm 5). Since JMetal is oriented towards goal minimization, the count of each motif will be subtracted to a cumulative variable that will represent the final score of this function (lines 7 to 9 in Algorithm 5).

In this case, unlike the two previous functions, instead of considering various versions, different motifs were analyzed to see which of them could better approximate the algorithm's solutions to real gene networks. Although 8 specific motifs were rigorously compared in the end, it should be noted that more manual testing was done earlier, which included other patterns. This testing led to a final list of candidates, represented in Fig. 5, and described below:

- **Regulatory Route**: It is a specific sequence of interactions between nodes in a directed network or graph. This pattern arises as a result of the regulation of biological processes or any other phenomenon where information flows sequentially and orderly through a series of components. In this pattern, an initial node

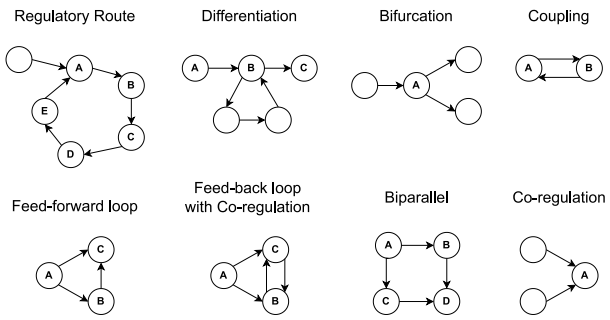


Fig. 5. Motifs considered in the study of the third objective function.

has a single successor, and each descendant node in the path has a single predecessor. Additionally, it is required that the last node in the path has a back edge to the initial node. The method counts both, the number of regulatory paths present in the graph and their size.

- **Differentiation**: It is a pattern in a directed graph where nodes represent a cell differentiation process. This motif is fundamental in developing and maintaining tissues and organs in multicellular organisms, as it allows for the precise regulation of cell differentiation and prevents overproduction or underproduction of differentiated cells. A node is considered part of a differentiation pathway if it can be reached from an initial node without forming cycles. The method counts the number of nodes participating in the graph's differentiation pathways.
- **Bifurcation**: It represents a branching point in the graph, where a node has at least two successors and a single predecessor, i.e., the node branches into multiple pathways. Branch points in gene regulatory networks allow cells to respond to a wide range of signals and stimuli. When a cell receives a signal, it can activate or deactivate specific genes, leading to various outcomes.
- **Coupling**: It represents a coupling relationship between two nodes in the graph, with an edge connecting them in both directions of regulation. The method counts the number of coupled node pairs present in the graph. In gene regulatory networks, coupling can contribute to stability and homeostasis. This is supported in [79] where this motif is considered as a self-regulation of length 2. Coupled genes can maintain balanced expression and avoid excessive fluctuations with reciprocal interaction.
- **Feed-forward loop**: It involves the interaction of three main elements: a regulatory transcription factor (TF1), an intermediate transcription factor (TF2), and a target gene (G). In this motif, TF1 directly regulates the expression of the target gene G while activating the expression of the intermediate transcription factor TF2, which also modulates the expression of G, creating a cascade of regulation. This motif contributes stability and robustness to gene regulatory networks. The presence of TF2 acts as an additional regulator that dampens fluctuations in the TF1 signal, preventing excessively rapid or inappropriate responses to changes in the environment.
- **Feedback Loop with CoRegulation**: This motif is similar to the previous one, but it incorporates an additional interaction between the target gene G and the intermediate transcription factor TF2 (which, in this case, could be considered as another gene) in the opposite direction to the existing one. Feedback allows for the adjustment and maintenance of gene expression levels within certain limits, while co-regulation ensures that genes in the circuit are activated or repressed simultaneously and coordinately, enhancing the precision and efficiency of the biological response.

Table 5
Friedman mean rank with Holm's adjusted p values (0.05) for AUPR.

AUPR		
Motifs version	$Friedman's Rank$	$Holm's Adj - p$
***Regulatory route**	**3.00581**	–
Differentiation	3.02326	0.96276
Bifurcation	3.38953	0.60861
Coupling	3.79651	0.10284
Feed-forward loop	5.38953	7.020e−10
Feed-back loop co-regulation	5.48256	1.674e−10
Biparallel	5.55233	5.571e−11
Co-regulation	6.36047	1.885e−18

Table 6
Friedman mean rank with Holm's adjusted p values (0.05) for AUROC.

AUROC		
Motifs version	$Friedman's Rank$	$Holm's Adj - p$
***Bifurcation**	**3.32558**	–
Regulatory route	3.48837	1.02661
Differentiation	3.56977	1.02661
Coupling	4.3314	0.02127
Feed-forward loop	4.91279	8.587e−05
Biparallel	5.27326	9.238e−07
Co-regulation	5.5407	1.818e−08
Feed-back loop co-regulation	5.55814	1.594e−08

- **Biparallel**: In this motif, situations are sought where a transcription factor indirectly regulates the expression of a gene through multiple pathways, i.e., indirect interactions that individually lack significance, but together signify a close relationship between both elements.
- **Co-regulation**: It occurs when a gene is simultaneously regulated by more than one transcription factor. This motif is common in gene regulatory networks because a gene often has multiple regulatory genes or regulatory pathways that converge on it. Gene expression is a highly coordinated process that requires the appropriate activation or repression of genes at different times and in different tissues. The presence of multiple transcription factors acting in concert on a gene provides the possibility of finer and more adaptable regulation.

In the first instance, a fitness function was designed to count each motif (considering additional aspects such as size in specific cases mentioned earlier). To assess the accuracy of each fitness function individually, the same temporal configuration used for comparing versions seen in the previous objectives was employed again.

The results are shown in Tables 5 and 6. In this case, instead of selecting a specific option and limiting the optimization of this objective to a single motif, it was considered more appropriate to build a function that considers the joint detection of the best motifs. In this case, the top 4 from the ranking have been chosen, as no statistically significant difference can be claimed between them for both metrics.

4. Experimentation

In this section, precise information will be presented regarding the study conducted to evaluate the quality of MO-GENECI compared to the set of individual techniques. The data collection process is described in Section 4.1, individual inference techniques are presented in 4.2, parameter setting procedure is addressed in 4.3, and finally, the experimental procedure is outlined in 4.4.

4.1. Data collection

The first step in the study is to define the set of networks aimed to be inferred. Although this proposal carries out the consensus of different

Table 7
Summary of the academic benchmark collected for the experimentation of this study. It is composed of a total of 106 inference problems that try to cover as much diversity as possible in order to obtain strong conclusions. Each network is subjected to each of the perturbations available for its case, obtaining a different set of data that make up an instance. Legend: KO (Knock-Out), KD (Knock-Down), and OE (Over-Expression).

Source	Networks	Sizes	Simulator	Disturbance	Instances
DREAM3 [89]	15	10, 50 and 100	DREAM team	–	15
DREAM4 [57]	10	10 and 100	DREAM team	–	10
IRMA [80]	1	5	Cell culture: RT-PCR	Switch on/off	2
TFLink [81]	4	12, 75, 163 and 371	SysGenSIM	Mixed	4
RegulonDB [82]	1	2234	SysGenSIM	Mixed	1
RegNetwork [83]	2	983 and 1033	SysGenSIM	Mixed	2
BioGRID [84]	32	6 - 1505	SysGenSIM	Mixed	32
GRNdb [85]	11	320 - 1598	SysGenSIM	Mixed	11
From scratch	4	20, 50, 100 and 200	SysGenSIM (EIPO Modular)	KO, KD and OE	12
From scratch	4	20, 50, 100 and 200	SysGenSIM (Scale Free)	KO, KD and OE	12
GRNdata [90]	2	300 and 1000	SynTReN	–	2
GRNdata [90]	1	1000	Rogers	–	1
GRNdata [90]	2	1565 and 2000	GeneNetWeaver	–	2

techniques without considering or knowing the true interactions of the networks, it is necessary to use datasets (in the training phase) that have a gold standard to perform accuracy comparisons later.

It has been decided to collect gene regulation networks from several sources to cover all possible domains of specialization. On the one hand, there are widely used benchmark networks in this research area to compare and measure the performance of different proposals. Among them, it is worthy to consider the DREAM challenges [57] (specifically, editions 3 and 4) and the yeast IRMA network [80]. On the other hand, there are databases and resources that construct gene regulation networks by gathering experimental data, bibliographic content, inference tested by multiple algorithms, etc. Examples of these sources include TFLink [81], RegulonDB [82], RegNetwork [83], BioGrid [84], and GRNdb [85]. However, none of them provide precise expression data measured in the laboratory and cover the entire published network. This is why it was decided to integrate a gene expression data simulator into the MO-GENECI software package that generates valid and coherent data to address the inference exercise.

In particular, SysGenSIM [21] is a software package that simulates systems genetics experiments in model organisms. This simulator allows the user to input the reference network structure in the form of an edge list and generate expression data using its nonlinear dynamic model. Among the perturbations available for simulation are: *knock-out* (transcription rate of deleted genes is multiplied by 0), *knock-down* (transcription rate of downregulated genes is reduced by multiplication to values smaller than 1), *over-expression* (transcription rate of overexpressed genes is increased by multiplication to values greater than 1), and *mixed perturbations* (a combination of the three previous types).

In this study, gene regulation networks have been collected from previously mentioned databases. Different gene expression levels have been generated by simulating a series of mixed perturbations on the regulatory system. In addition to benchmark-type networks and those collected in biological databases, SysGenSIM has been used to construct gene networks from scratch with standard topology models typical of gene regulation networks. Specifically, networks of 20, 50, 100, and 200 genes with both, scale-free and modular topology, have been created applying knock-out, knock-down, and over-expression perturbations.

Moreover, to further increase the diversity of the dataset used in the experimentation, gene networks have been collected from other well-known simulators, such as SynTReN [86], Rogers [87], and GeneNet Weaver [88].

All of this constitutes an academic benchmark of 106 problem instances. The origin of each of these is summarized in Table 7 and discussed in more detail in the following sections. The aim is to have a wide range of test cases extracted from multiple diverse sources to ensure proper coverage of all specialization domains.

4.1.1. DREAM

The DREAM challenges [57] are a series of scientific competitions in which participants work as teams to develop innovative methods and algorithms that can solve specific problems in biology and medicine. Participants compete to find the best solutions, and often, an open-source approach is used to share the results and the developed code.

These challenges have addressed various problems in bioinformatics, such as drug response prediction, protein structure prediction, identification of genetic markers associated with diseases, and many others. These challenges have brought together researchers from around the world and have made significant contributions to advancing bioinformatics and understanding molecular biology.

- **DREAM3**: The DREAM3 challenge in "in silico" network inference [89] aimed to assess the technology's capacity to infer gene networks with varying sizes and levels of connectivity. The challenge data focused on subnetworks associated with two organisms: *Escherichia coli* (E. coli) and *Saccharomyces cerevisiae* (yeast). Specifically, these datasets were created using continuous differential equations, which reasonably approximated the regulatory functions governing gene expression. A small amount of Gaussian noise was added to these values to introduce a level of measurement error. Each subchallenge was divided based on network sizes, with five networks (Ecoli1, Ecoli2, Yeast1, Yeast2, Yeast3) considered for each size category. These sub-challenges encompassed networks of 10, 50, and 100 nodes, providing 4, 23, and 46 different trajectories, respectively. The DREAM3 challenge has evolved into a standard benchmark for assessing the reconstruction of gene regulatory networks from expression data. The scripts included in the challenge have established a unified evaluation framework that, to some extent, simplifies the comparison of the quality of different approaches in the literature.
- **DREAM4**: The DREAM4 challenges introduced new elements to the challenges that were previously featured in DREAM3. These enhancements included the incorporation of new datasets. Similar to the format of the previous edition, the in silico network inference challenge was divided into two parts based on network size. Specifically, networks were categorized into two groups: those with 10 nodes and those with 100 nodes. Each group consisted of 5 networks, mirroring the setup in DREAM3. For the 10-node networks, participants were provided with expression data comprising 21 time points and 5 replicates. Meanwhile, the 100-node networks offered another 21 time points but with 10 replicates instead. These networks exhibited varying topologies, all aimed at emulating real organisms like *Escherichia coli* or *Saccharomyces cerevisiae*. The data aimed to replicate the dynamic properties of these organisms, incorporating different initial conditions and

kinetic parameters. To generate the expression data for each network, stochastic differential equations were employed, followed by the addition of noise proportional to the gene expression level, a feature reminiscent of real microarray datasets. Furthermore, each network was associated with four sets of observations: time series, wild type, knock-out, and knock-down. It is worth noting that this edition of the challenge stands out as one of the most extensively explored in the DREAM series. The increased number of evaluation scripts and background information provided in comparison to previous editions encouraged numerous participants to put their inference methods to the test.

4.1.2. IRMA

The In vivo Reverse-engineering and Modeling Assessment (IRMA) network [80] was established for the purpose of assessing the effectiveness of various methods for reconstructing gene networks. To achieve this, quantitative RT-PCR was employed to measure the expression levels of the yeast *Saccharomyces cerevisiae* at various time points. This network comprises 5 genes (CBF1, GAL4, SWI5, GAL80, and ASH1) and encompasses 6 regulatory interactions. These interactions lead to the creation of both "switch on" and "switch off" versions of the network, achieved by cultivating cells in either galactose or glucose conditions, respectively.

This synthetic network encompasses several regulatory interactions, capturing the behavior of extensive eukaryotic gene networks, but on a smaller scale. The network was intentionally designed to minimize its susceptibility to the influence of endogenous genes while being responsive to galactose, which serves as a trigger for the transcription of its genes. Despite its seemingly uncomplicated nature, this network is surprisingly intricate in its connections. It includes regulatory chains, single-entry motifs, and multiple feedback loops generated through the interplay of transcriptional activators and repressors.

4.1.3. Simulated datasets based on real-world networks

With the aim of expanding the diversity of the networks considered in the study, gene regulation networks (experimentally tested in the literature) have been extracted from biological databases, and expression data have been simulated from these networks to evaluate the accuracy with which different techniques, including this proposal, infer them. These datasets are:

- **TFLink**: TFLink [81] provides precise information about transcription factors, including details about target gene interactions, nucleotide sequences, and genomic locations of the binding sites for these factors. Data is collected for the human organism and six model organisms: mouse (*Mus musculus*), rat (*Rattus norvegicus*), zebrafish (*Danio rerio*), fruit fly (*Drosophila melanogaster*), nematode (*Caenorhabditis elegans*), and yeast (*Saccharomyces cerevisiae*). In TFLink, the data is accompanied by its respective sources, such as databases, experimental methods, and publications. To develop TFLink, several databases were considered, and ten resources were selected for integration: DoRothEA [91], GTRD [92], HTRIdb [93], JASPAR [94], ORegAnno [95], REDfly [96], ReMap [97], TRED [98], TRRUST [99], and Yeastract [100]. For this study, it has been decided to use the networks that TFLink labels as "small-scale" for the organisms *Caenorhabditis elegans*, *Drosophila melanogaster*, *Rattus norvegicus*, and *Saccharomyces cerevisiae*. After a filtering process based on the number of detection methods, the 4 networks have sizes of 75, 163, 12, and 371 genes, respectively.
- **RegulonDB**: RegulonDB [82] is a comprehensive and up-to-date database that focuses on the gene regulation of *Escherichia coli K-12*. It contains a wide range of information, from regulatory elements and transcription factors to gene interactions and experimental conditions. The database has been created from manually curated scientific publications data, high-throughput datasets and computational predictions. After downloading the file labeled as "TF-gene interactions" from the dataset with experimental evidence, a filtering process was performed in which all interactions with "weak" confidence or uncertain regulation signs were removed. This resulted in one network with a total of 2234 genes.
- **RegNetwork**: RegNetwork [83] is a comprehensive database that hosts a wide range of transcriptional and post-transcriptional regulatory relationships for humans and mice. It includes five types of interactions: TF-TF, TF-gene, TF-miRNA, miRNA-TF, and miRNA-gene. This valuable resource integrates curated regulations from various databases, along with inferred regulations based on transcription factor binding sites (TFBSs). Additionally, RegNetwork leverages conserved knowledge of TFBSs to establish potential regulatory relationships between regulators and their targets, further enhancing its utility. In this case, both networks (human and mouse) have been downloaded without any filtering or modification. Therefore, it will be an attempt to infer the gene regulation network for humans, consisting of 983 genes, and the mouse network with a total of 1033 genes.
- **BioGRID**: BioGRID [84], officially known as the Biological General Repository for Interaction Datasets, is a publicly accessible repository dedicated to cataloging and sharing genetic and protein interaction information originating from both model organisms and humans. This valuable resource, boasts an extensive collection of over 1,740,000 interactions. These interactions have been meticulously curated from a combination of high-throughput datasets and individual-focused research studies, all sourced from a staggering pool of 70,000+ publications within the primary literature. Within the BioGRID database, comprehensive coverage is upheld for specific organisms like budding yeast (*S. cerevisiae*), fission yeast (*S. pombe*), and thale cress (*A. thaliana*). Additionally, ongoing efforts are dedicated to expanding curation efforts to encompass various *metazoan* species. The primary focus of these curation initiatives is to shed light on specific aspects of biology, facilitating the exploration of conserved networks and pathways that hold relevance for human health. In this study, the following gene networks have been incorporated: *Human papillomavirus 5*, *Human papillomavirus 6b*, *Bacillus subtilis 168*, *Bos taurus*, *Macaca mulatta*, *Middle-East Respiratory Syndrome-related Coronavirus*, *Canis familiaris*, *Chlamydomonas reinhardtii*, *Chlorocebus sabaeus*, *Neurospora crassa OR74A*, *Cricetulus griseus*, *Danio rerio*, *Oryctolagus cuniculus*, *Oryza sativa Japonica*, *Emericella nidulans FGSC A4*, *Plasmodium falciparum 3D7*, *Gallus gallus*, *Glycine max*, *Simian Immunodeficiency Virus*, *Human Herpesvirus 1*, *Simian Virus 40*, *Human Herpesvirus 4*, *Human Herpesvirus 5*, *Streptococcus pneumoniae ATCCBAA255*, *Strongylocentrotus purpuratus*, *Sus scrofa*, *Human Herpesvirus 8*, *Vaccinia Virus*, *Human Immunodeficiency Virus 2*, *Xenopus laevis*, *Human papillomavirus 16* and *Zea mays*. These 32 networks cover a range of sizes between 6 and 1505 genes, bringing a wide diversity to the experimental dataset of this study.
- **GRNdb**: GRNdb [85] stands out as a user-friendly, open-access database designed to provide a seamless platform for exploring and visualizing anticipated regulatory networks. These networks are established through the interactions between transcription factors (TFs) and their downstream target genes, known as regulons. These connections are inferred from extensive RNA-seq data and existing TF-target relationships across various conditions in both humans and mice. It is worth noting that all regulatory insights within GRNdb are derived from omics data analysis rather than direct experimental validation. This resource empowers users to effortlessly search, navigate, and retrieve TF-target pairings and associated motifs under diverse conditions, whether at the single-cell or bulk level. Additionally, researchers can simultaneously delve into gene expression profiles and investigate the correlation between gene expression levels and patient

survival for a wide range of TCGA cancer types. In this study, the following gene networks have been incorporated: *Fetal-Brain*, *Fetal-Thymus*, *Adult-Pancreas*, *Adult-Muscle*, *Adult-Adipose*, *Adult-Ascending-Colon*, *Adult-Lung*, *Adult-Liver*, *Fetal-Calvaria*, *Adult-Epityphlon* and *Adult-Rectum*. These 11 networks are up to 1598 genes in size.

4.1.4. Simulated data from scratch

In addition to the expression data simulated from real gene regulation networks, other simulations have been conducted from scratch to follow degree distributions frequently observed in the literature.

- **EIPO Modular distribution**: The Exponential In-degree and Power law Out-degree Modular distribution [101] is a network model inspired by real gene networks. It combines two key features: an exponential in-degree distribution and an out-degree distribution that follows a power law. Additionally, it introduces modularity by creating multiple "modules" (densely connected groups of genes) that are interconnected by reorganizing connections with a specific probability. This modular structure and degree distributions capture the properties observed in biological networks, allowing for a more realistic model of gene interactions. Using this distribution, networks of 20, 50, 100, and 200 genes have been constructed, simulating knock-out, knock-down, and over-expression perturbations for each size. This results in 12 instances of simulated expression data.
- **Scale Free distribution**: The scale-free distribution refers to a pattern of node degree distribution in a network that follows a power law. This means that the probability of a node having a specific degree (i.e., the number of connections a node has) decreases according to a power law, rather than following a typical normal or Gaussian probability distribution. This distribution has also been observed in biological networks, which is why the SysGenSIM simulator includes it in its implementation. The network sizes and perturbations applied are identical to the previous distribution, summing up to 12 datasets.
- **Other simulators**: To add even more diversity to the dataset considered in this study, networks were collected from other simulators beyond SysGenSIM. The R package GRNdata [90] was used for this purpose, and networks from other well-known simulators were retrieved. From SynTReN [86], two networks of 300 and 1000 nodes were collected. From Rogers [87], a single network of 1000 genes was obtained. And from GeneNetWeaver [88], two networks of 1565 and 2000 nodes were retrieved.

4.2. Inference techniques

As commented in the state of the art, a well-grounded set of 26 inference techniques is used for the proposed consensus strategy conducted by MO-GENECI. However, due to the limitations and high computational costs of some of them, certain techniques have been eliminated for certain network size ranges. In this sense, all techniques are employed for networks with fewer than 25 genes. For those with a size between 25 and 110 genes, all are used except for JUMP3. For networks with sizes between 110 and 250 genes, the following techniques are removed from the list: JUMP3, TIGRESS, CMI2NI, LOCPCACMI, GRNVBEM, and NONLINEARODES. For networks larger than 250 and smaller than 2000, the following are also discarded: PCACMI, PLSNET, INFERELATOR, GENIE3_RF, GENIE3_ET, GRNBOOST2, and MEOMI. Lastly, for networks larger than 2000 genes, PUC and PIDC are also eliminated.

Regarding computational resources, it has been decided to parallelize the execution of all techniques, including an intelligent allocation of cores, to provide more resources to the techniques that need them most to synchronize their completion times. Therefore, the techniques have been divided into four groups based on their resource priority, from highest to lowest. In the first group, the most resource-intensive techniques are JUMP3, LOCPCACMI, NONLINEARODES, GRNVBEM, and CMI2NI. The second group has TIGRESS, PCACMI, PLSNET, INFERELATOR, GENIE3_RF, GRNBOOST2, and GENIE3_ET. In the third group, only KBOOST and LEAP are included. In the last group, the rest of the techniques either have minimal computational cost or their implementations prevent parallelization, so they would not use the allocated cores.

4.3. Parameter settings

To enhance the quality of MO-GENECI, a parameter setting procedure has been carried out to find the optimal combination of values for the following parameters: population size (100, 200, or 300), crossover probability (0.7, 0.8, or 0.9), number of parents (3 or 4), mutation probability (0.05, 0.1, or 0.2), and mutation strength (0.1, 0.2, or 0.3). In this exercise, it was decided to include all gene networks from the benchmark with a size of fewer than 500 genes, consisting of 82 networks. Five independent runs were performed for each parameter combination and gene regulatory network, each consisting of a total of 100,000 evaluations. The termination criterion used was PercLinksWithBestConf, with a threshold set at 0.4 (as previously mentioned during the explanation of the third fitness function).

Once all the results were obtained, a reference Pareto front was constructed for each problem by selecting the best solutions from each combination. Subsequently, the reference front was compared to each independent result, and the following metrics were calculated: Epsilon (EP) [102], Generational Distance (GD) [103], PISAHypervolume (HV) [104], Inverted Generational Distance (IGD) [103], Inverted Generational Distance Plus (IGD+) [105], and Spread (SP) [20]. With these results, a statistical ranking using Friedman's test with non-parametric Holm tests was applied for each metric to determine which combination of values performed better.

The total of 162 parameter combinations with 5 independent runs for each of the 82 gene networks considered amounts to a total of 66,420 executions. The extent of this is the reason why it was decided to present the results of these executions and the subsequent statistical ranking of each metric in the documentation section of the main repository as supplementary material.[3] Nevertheless, Table 8 provides a summary of the results.

After reviewing the results, it can be observed that two candidate configurations stand out: PS200-CP0.7-MP0.2-NP3-MS0.3 and PS300-CP0.9-MP0.05-NP4-MS0.1. While the first configuration leads the ranking for the metrics Epsilon (EP) and PISAHypervolume (HV), the second one performs well in Inverted Generational Distance (IGD) and Inverted Generational Distance Plus (IGD+). Despite this information, which may make the decision seem complex, it is important to note that the first candidate ranks near the bottom for the rest of the metrics, while the second one falls in more intermediate positions. For this reason, it has been decided to select the second configuration as the optimal choice for MO-GENECI.

4.4. Experimental procedure

After collecting all the gene networks to be inferred, selecting the individual techniques to use for each of them, and setting the parameters, an experimental procedure has been designed to assess the actual performance of the proposal concerning a large set of techniques in the state of the art.

Firstly, the corresponding techniques were executed individually in parallel for each gene expression dataset. This provides a proposed network for each technique and instance. Secondly, to consolidate the

[3] https://github.com/AdrianSeguraOrtiz/MO-GENECI/tree/main/docs/parameterization

Table 8
The first 5 configurations from the statistical Friedman ranking for each metric, indicating significance compared to the winner based on non-parametric Holm tests (R = Rejected and A = Accepted). The configurations are presented using the following abbreviations: PS = Population Size, CP = Crossover Probability, NP = Number of Parents, MP = Mutation Probability y MS = Mutation Strength.

Ranking	EP	GD	HV
1	PS200-CP0.7-MP0.2-NP3-MS0.3	PS300-CP0.8-MP0.2-NP3-MS0.3	PS200-CP0.7-MP0.2-NP3-MS0.3
2	PS100-CP0.7-MP0.2-NP3-MS0.3 (R)	PS300-CP0.9-MP0.2-NP3-MS0.3 (R)	PS100-CP0.7-MP0.2-NP3-MS0.3 (R)
3	PS200-CP0.8-MP0.2-NP3-MS0.3 (R)	PS300-CP0.7-MP0.2-NP3-MS0.3 (R)	PS100-CP0.8-MP0.2-NP3-MS0.3 (R)
4	PS100-CP0.8-MP0.2-NP3-MS0.3 (R)	PS300-CP0.7-MP0.1-NP3-MS0.3 (R)	PS200-CP0.8-MP0.2-NP3-MS0.3 (R)
5	PS300-CP0.7-MP0.2-NP3-MS0.3 (R)	PS300-CP0.9-MP0.1-NP3-MS0.3 (R)	PS300-CP0.7-MP0.2-NP3-MS0.3 (R)

Ranking	IGD	IGD+	SP
1	PS300-CP0.9-MP0.05-NP4-MS0.1	PS300-CP0.7-MP0.05-NP4-MS0.1	PS100-CP0.9-MP0.2-NP4-MS0.2
2	PS300-CP0.8-MP0.05-NP4-MS0.1 (A)	PS300-CP0.8-MP0.05-NP4-MS0.1 (A)	PS100-CP0.7-MP0.05-NP4-MS0.2 (R)
3	PS300-CP0.7-MP0.05-NP4-MS0.1 (A)	PS300-CP0.9-MP0.05-NP4-MS0.1 (A)	PS100-CP0.9-MP0.05-NP4-MS0.2 (R)
4	PS200-CP0.9-MP0.05-NP4-MS0.1 (R)	PS200-CP0.9-MP0.05-NP4-MS0.1 (R)	PS100-CP0.7-MP0.05-NP4-MS0.1 (R)
5	PS200-CP0.8-MP0.05-NP4-MS0.1 (R)	PS200-CP0.8-MP0.05-NP4-MS0.1 (R)	PS100-CP0.7-MP0.1-NP4-MS0.2 (R)

networks proposed by these techniques, for each problem, MO-GENECI was executed with the previously fixed parameters, setting a maximum of 250,000 evaluations. This provides a Pareto front with multiple solutions for each problem.

Thirdly, the networks proposed by the individual techniques and all those derived from MO-GENECI (one consensus network for each solution from the front) were compared to the gold standard in the test phase. This yields an AUROC and AUPR value for each network proposal.

Fourthly, since choosing a single solution from the Pareto front is left to the domain expert, it was decided to extract two statistically significant solutions for subsequent comparison with the individual techniques. The first one was called BEST_MO-GENECI, referring to the solution from the front whose consensus network achieves the best AUPR and AUROC values. The second one was named MEDIAN_MO-GENECI, representing the solution whose accuracy is the median of the front. In this way, the goal is to demonstrate the potential of MO-GENECI when choosing the appropriate solution (BEST_MO-GENECI) and the good quality it can provide even when this choice is made without prior knowledge (MEDIAN_MO-GENECI). It is worth mentioning that during the subsequent results discussion (specifically in the explanation of Fig. 8), a clear trend will be shown that undoubtedly facilitates the selection of a good solution from the front based on basic network characteristics such as size.

Finally, a systematic comparison procedure was initiated after collecting the precision levels for each gene network from the individual techniques and the two solutions extracted from MO-GENECI. The different proposals were compared for each size range using the standard Friedman statistical ranking and Holm's non-parametric tests. The results are presented in the following section of this manuscript.

Furthermore, once the efficiency of this approach was validated, it was decided to run MO-GENECI on a real-world dataset from melanoma patients. This same set was also employed in GENECI and this allows us to observe whether the performance improvements translate into new discoveries that did not emerge after running the preliminary algorithm.

5. Results and discussion

This section analyses the results obtained by MO-GENECI for the total benchmark of 106 networks considered in this study. A chronological order based on the workflow described earlier is followed to facilitate the comprehension of this analysis. In addition to analyzing the accuracy of MO-GENECI, a couple of subsections are included to discuss the computational cost of the algorithm and the results of MO-GENECI when run on a real-world data set.

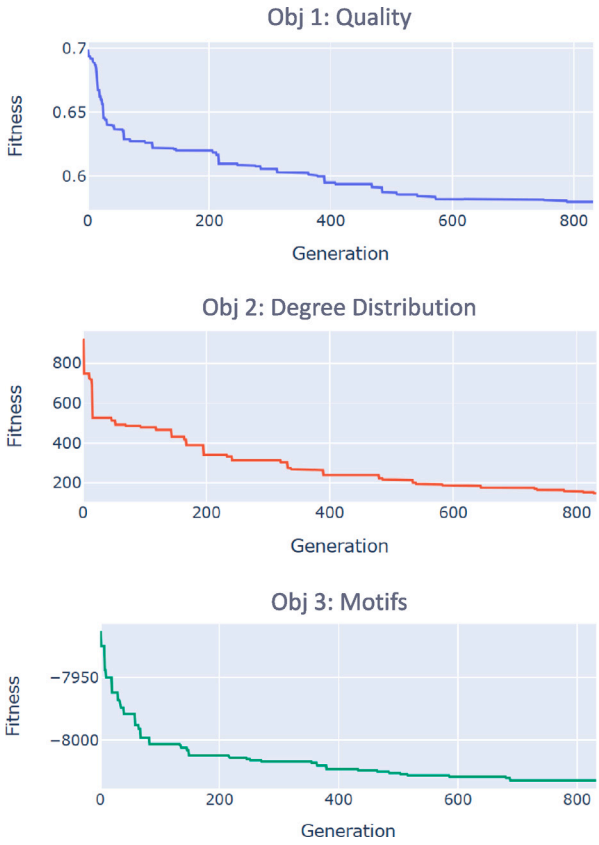
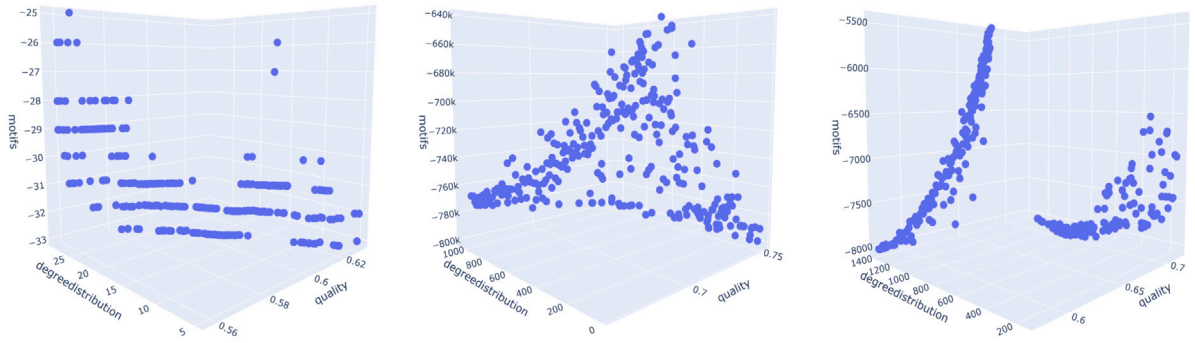


Fig. 6. Evolution of fitness functions for the 200-node network constructed from scratch with a scale-free degree distribution and subjected to overexpression perturbation.

5.1. MO-GENECI internal behavior

In MO-GENECI, the three objective functions are optimized simultaneously over the iterations. Despite not pursuing the individual optimization of each one, it proves quite useful to graph the best individual result for each function in each generation. This allows for the verification of the algorithm's learning procedure, the proper selection of individuals, and the ability of the crossover and mutation operators to achieve uniform optimization of the three objectives, while maintaining the right balance between exploration and exploitation.

In Fig. 6, a typical execution of MO-GENECI for the 200-node network, constructed from scratch with a scale-free degree distribution



(a) 13-gene network from rattus norvegicus organism obtained from TFLink subjected to mixed perturbation.

(b) Simulated 2000-gene network by GeneNetWeaver.

(c) 200-node network constructed from scratch with a scale-free degree distribution and subjected to over-expression perturbation.

Fig. 7. Set of non-dominated solutions in Pareto front approximation from the final population.

and subjected to over-expression perturbation, is depicted. It can be observed that the three objective functions indeed show improvement in their results as the evolutionary algorithm progresses. The slight stepwise behavior in the plots is attributed to the network's size. In small networks, this stepping is more prominent, particularly in functions related to network topology, while in large networks comprising thousands of genes, this stepping smoothens and ultimately disappears.

The number of generations is approximately 800, which aligns with the chosen population size of 300 individuals and a maximum of 250,000 evaluations per run. This configuration was established after verifying that the proposed approach is able to converge at this number of generations, for the three objectives, as shown in Fig. 6, so no significant improvements were detected after this stopping condition (hence, to save on computational effort).

Concerning the visible range of values for each function, it aligns with the previously specified implementation. The *quality* values are normalized between 0 and 1, *degreedistribution* values fall within a range between 0 and ∞ due to the goodness-of-fit test, and *motif* values range between 0 and $-\infty$ due to negative motif counting.

After executing MO-GENECI, the output is obtained as the set of non-dominated solutions from the last generation (in the form of Pareto front approximation). These individuals are represented in interactive 3D plots that facilitate visualization of the obtained Pareto front for each problem. Fig. 7 displays three specific cases of interest, allowing for the derivation of conclusions from this approach.

Firstly, in Fig. 7(a), the obtained Pareto front for a relatively small network consisting of 13 genes can be observed. The most notable aspect of this case relates to what was mentioned previously regarding the stepping behavior. In such small networks, it can be observed that the *motifs* objective function has limited room for optimization, with a relatively narrow and discretized range of values. Therefore, the representation of the front appears as a set of curves parallel at different heights.

Nevertheless, in Fig. 7(b), the opposite scenario is shown: a large network with 2000 genes where all objective functions have significant room for optimization. Consequently, the solutions are more evenly distributed throughout the plot. As expected, three vertices stand out, corresponding to the extreme cases of each objective function, where one objective is maximized at the expense of receiving poor scores in the other two functions.

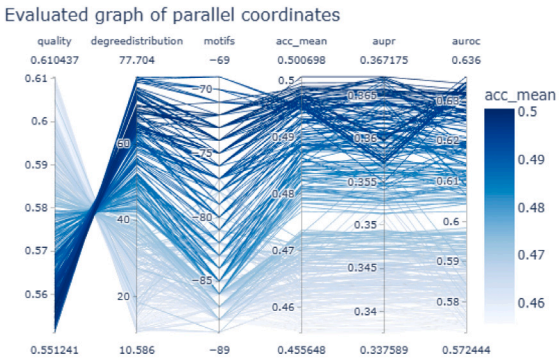
However, there are exceptions, as shown in Fig. 7(c), where a network's characteristics break the opposition between the functions. In this case, it is impossible to optimize *degreedistribution* until *quality* presents extremely unfavorable values (curve with a shallower slope), and vice versa, *quality* cannot be optimized until *degreedistribution* values move away from the minimum (curve with a steeper slope).

From a different perspective, MO-GENECI's output also provides a representation of parallel coordinates for the set of non-dominated solutions from the last population. However, it is important to remember that work has been carried out with gold standards in this case in preparation for subsequent comparison between techniques. That is why the decision was made to expand the parallel coordinates initially presented by MO-GENECI with the accuracy values for each individual concerning the known network. Additionally, to improve visualization, apart from adding the columns AUROC and AUPR, a new column representing the average of the two previous ones has been created. This allows individuals to be color-coded with varying intensity based on their accuracy. To obtain this expanded representation, the command *evaluate dream/generic-prediction dream/generic-pareto-front* is executed, which takes a front of solutions and the known network as input.

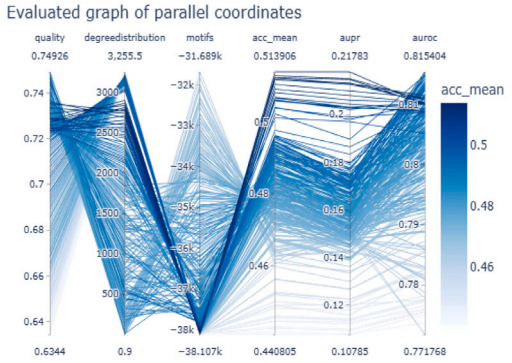
These newly evaluated parallel coordinates serve a dual purpose. First, they allow for an analysis of the intensity and shape of the trade-offs between different objectives. Second, they help identify certain patterns regarding optimization levels in the graphs that lead to high-quality results. In Fig. 8, three significant cases have been represented in this sense.

In Fig. 8(a), the "evaluated" parallel coordinates for a 20-gene network are represented, from which, four observations can be made. Firstly, the stepping behavior mentioned in the previous plots for small networks is still present in this representation. As can be seen, the *motifs* column has a limited number of values, grouping individuals at quite specific heights. Secondly, the opposition between different objectives is evident, with the differences between *quality* and *degreedistribution* being particularly noticeable in this case, showing a complete contrast. Thirdly, by analyzing the graphics, the coherence of this approach can be verified. Solutions with similar optimization profiles result in networks with similar levels of accuracy. Lastly, due to the network's size, the objective function that contributes the most accuracy to the solutions is *quality*. This is because neither *degreedistribution* nor *motifs* make sense in such small networks. There is no room for the existence of hubs or the appearance of complex motifs. Therefore, improving solution accuracy relies solely on giving more weight to the most reliable individual techniques than the others.

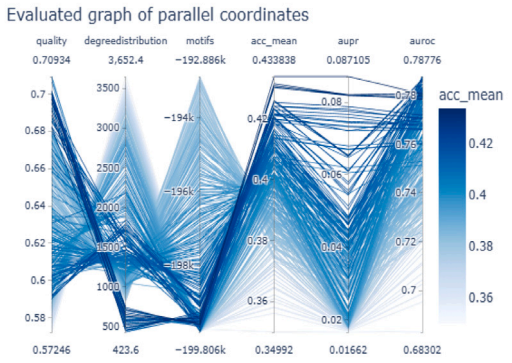
The second significant case is a medium-small network with 436 genes. In Fig. 8(b), it can be observed that for this case, *quality* loses the prominence mentioned for networks with few genes, *degreedistribution* still lacks significant room to contribute accuracy to the solutions. Finally, *motifs* begins to play a key role in the optimization process. In fact, it is evident how this objective begins to show greater opposition to the other ones, which was not as visible in the previous example. Once again, the coherence of this approach can be checked, with a clear gradient of color in all represented objectives.



(a) 20-node network constructed from scratch with a scale-free degree distribution and subjected to a knock-down perturbation.



(b) 436-gene network from the gallus gallus organism obtained from BioGrid subjected to a mixed perturbation.



(c) Simulated 2000-gene network by SynTReN.

Fig. 8. Extended parallel coordinate plots with accuracy metrics.

Finally, in Fig. 8(c), the case of a large network with 2000 genes is shown. The increase in size compared to the previous case is the reason why *degreedistribution* gains importance in improving solution accuracy. When networks have more realistic sizes, both *degreedistribution* and *motifs* serve a more coherent purpose and, therefore, play a more significant role in the effectiveness of this approach.

It should be noted that the last and first examples are opposite cases, and the difference in their sizes has a similar effect on the optimization profiles for higher accuracy. This is why, when explaining BEST_MO-GENECI and MEDIAN_MO-GENECI, reference was made to this section to clarify that choosing a good solution is not a complicated task considering the analysis just presented. For smaller networks, it is logical to focus on the reliability of the techniques rather than topological characteristics. In contrast, for networks of realistic size, the concordance between techniques diminishes, and more biological aspects like degree distribution or motifs become more relevant.

5.2. Experimental comparisons

The next step in this study is the comparison of MO-GENECI with all the individual inference techniques. After evaluating all the networks generated by these techniques and extracting the aforementioned BEST_MO-GENECI and MEDIAN_MO-GENECI solutions, AUROC and AUPR values were obtained for each network and inference method. Since presenting these values for all 106 networks would be extensive, a series of visual plots are provided to summarize and facilitate understanding of the results. However, detailed data for both metrics can be found in the main project repository as supplementary material.[4]

In Fig. 9, the comparison of accuracy values for a subset of networks with a size of up to 25 genes is presented. It can be observed that the main objective of this proposal is fully achieved: to provide solutions that, without attempting to outperform individual techniques in their respective domains of expertise, have high accuracy for any type of network to be inferred. The existence of these specializations is also clearly exposed in the plot. Not only is there diversity in the winning individual techniques (e.g. CMI2NI, GENIE3_ET, ARACNE, JUMP3, PCACMI, etc.), but there are significant differences in their rankings depending on the case. For instance, JUMP3 is the technique with the highest AUPR value for the D4_10_3 network, yet it ranks last for the AUROC metric in the TFL-RN-M network. PUC is the AUROC winner for D3_10_E1, but it holds a rather unfavorable position in the AUROC of its neighboring BG-SV40-M. Another example is ARACNE, which seems not to excel in any network except for the exceptional case of the TFL-RN-M network. MO-GENECI surpasses these individual specializations and attains excellent rankings across all networks thanks to its adaptability and flexibility.

Furthermore, it should be noted that even when the best solution from the front is not chosen appropriately, the effectiveness of MO-GENECI remains relatively high. As can be seen, the values of MEDIAN_MO-GENECI are pretty close to those of BEST_MO-GENECI (for example, in D3_10_E1, D4_10_3 or FS-SF20-O).

It is also worth mentioning that when running MO-GENECI with all inference techniques, any information that the researcher may have about individual inference techniques is ignored. While it is true that domains of specialization are difficult to delineate, there is still some minimum knowledge that could help researchers eliminate certain techniques for their specific case. This would filter out some noise in the consensus and further improve the results obtained.

Additionally, this analysis is not only helpful in comparing the proposal of this work with other individual techniques but also serves as a systematic and robust comparison among them. This study can contribute to the state of the art by providing knowledge about the performance of individual techniques and their effectiveness in this field. This can be explored in more detail in Section 5.3.

Fig. 10 shows the AUPR and AUROC values obtained by the inference methods for a subset of networks with sizes ranging from 25 to 110 genes. The same information as in Fig. 9 is presented for different networks and using a different visual representation.

In this case, the fact that some techniques perform well for certain networks and worse for others can be observed at the intersections

[4] https://github.com/AdrianSeguraOrtiz/MO-GENECI/tree/main/docs/experimentation

Fig. 9. AUROC and AUPR values for networks up to 25 genes. Each column represents a gene network, and each row represents a different metric. The upper row shows AUROC values, while the lower row displays AUPR values. Each cell contains two sets of bars stacked from lowest to highest height. In the first set of bars, the values obtained by BEST_MO-GENECI and MEDIAN_MO-GENECI are represented, while the second set contains the values from the other individual techniques. The number of individual techniques surpassed by BEST_MO-GENECI is indicated on the first set of bars, which can be visualized thanks to the dashed horizontal line that sets the threshold. The name of the winning individual technique that achieved the best results for the given network and metric is displayed on the second set of bars.

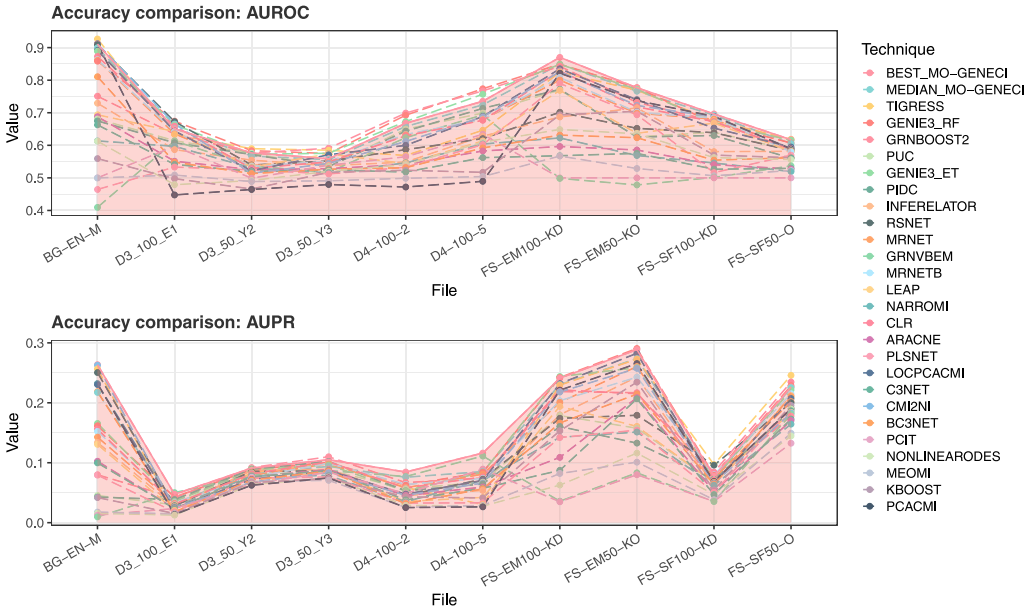


Fig. 10. AUROC and AUPR values for networks between 25 and 110 genes in size. Each plot displays the results for a specific metric. The upper plot shows the AUROC values, while the lower one shows the AUPR values. In both plots, the vertical axis represents the values of the corresponding metric, and the horizontal axis represents the different networks considered in the figure. Each technique is assigned a color and represented by a series of points connected by dashed lines. To facilitate the comparison of BEST_MO-GENECI with the other techniques, a continuous representation of its curve and shading over its area have been used.

between the dashed lines. The most noticeable case due to its color is PCACMI, as its low accuracy for networks derived from the DREAM challenges and its high quality for the remaining networks in the plot are clearly visible.

Once again, MO-GENECI gains generalization capacity and achieves good results for all cases, overshadowing the outcomes of individual techniques. However, as mentioned earlier, this proposal is eventually surpassed by techniques specialized in the domain of expertise in which the inferred gene network falls.

Fig. 11 presents the accuracy values obtained by different techniques for a subset of networks with sizes between 110 and 250 genes. The intersections between the lines again highlight the specialization of techniques.

In this sense, it is worth noting the case of the simulated BioGrid network "*Human papillomavirus 5*" subjected to a mixed perturbation (BG-HPV5-M). In this network, techniques like MEOMI, which show unfavorable results for the other networks, improve their quality. In contrast, other clearly outstanding techniques for the rest of the networks significantly reduce their accuracy. This is a clear example of the risk that a researcher runs when choosing a specific technique based on the results shown in other gene regulation networks. This is where the utility of MO-GENECI becomes evident, as it has remained immune to

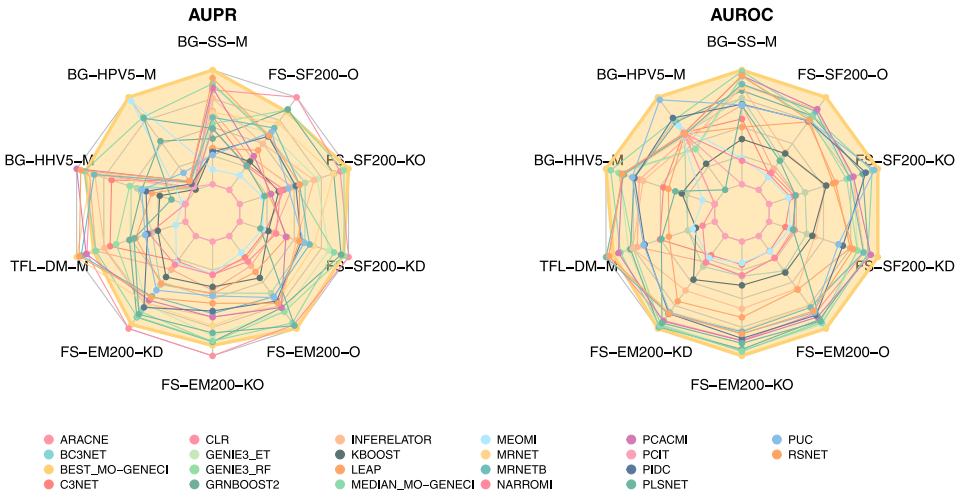


Fig. 11. AUROC and AUPR values for networks between 110 and 250 genes in size. In this case, radar charts have been used, which clearly represent the specializations of different techniques. Once again, shading has been applied to BEST_MO-GENECI (yellow), highlighting its ability to cover all domains of expertise.

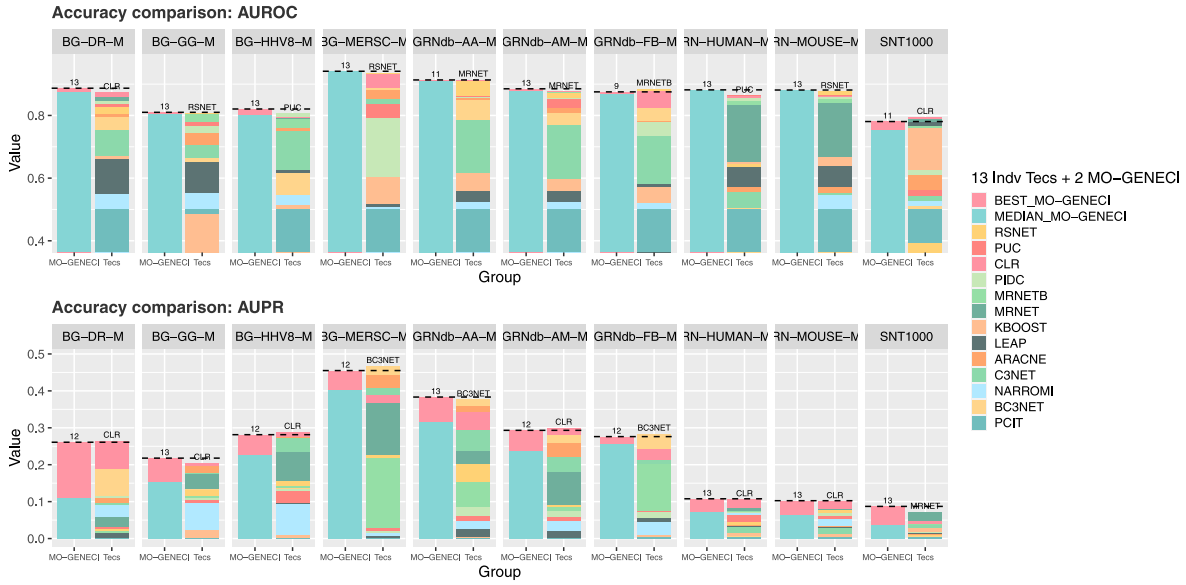


Fig. 12. AUROC and AUPR values for networks between 250 and 2000 genes in size. La explicación de esta representación es la misma que en la Figura 9.

this change in techniques and has been able to ultimately redistribute its weights to achieve a high-quality result.

Additionally, it can be observed that PCIT achieves poor accuracy for all networks in this representation, yet it was declared the winner in Fig. 9 for the AUPR metric in the D3_10_E1 network.

The last representation is shown in Fig. 12, for a subset of networks with sizes between 250 and 2000 genes. It is evident how the accuracy of MO-GENECI and the number of surpassed individual techniques increase with the size of the networks. Furthermore, it can be observed that the distances between BEST_MO-GENECI and MEDIAN_MO-GENECI are further reduced in these cases. Once again, all the observations made so far about the capabilities of MO-GENECI can be clearly confirmed. This proposal has demonstrated versatility for different domains of expertise and shows its potential to be more than competitive in networks of all sizes. The specialization by individual techniques continues to be present, as seen in the previous comparisons.

Finally, two networks with sizes exceeding 2000 genes are left to analyze. Due to the small number of networks in this category, it has been decided to create Table 9 displaying the metric values explicitly instead of representing them graphically. It highlights in bold the best value achieved by individual inference techniques and the one obtained by BEST_MO-GENECI for each network and metric.

For the simulated 2000-gene network by GeneNetWeaver, it can be seen that BEST_MO-GENECI achieves a competitive AUPR value, while for the AUROC metric, it clearly outperforms the rest of the techniques (also MEDIAN_MO-GENECI). In the case of the RegulonDB network, it is quite similar to the previous one. However, this time, for the AUPR metric, the CLR technique presents a clear outlier unsupported by the rest of the techniques, which is difficult for MO-GENECI to detect. It is a clear example of a technique that surpasses the consensus due to its domain of expertise. Nevertheless, MO-GENECI's values are clearly competitive compared to the other techniques.

It is also worth mentioning that during this section, MO-GENECI has been compared to the best AUROC and AUPR values for each network, regardless of whether these values come from the same technique or not. In other words, for a specific network, when MO-GENECI performance is just below one technique in the AUPR metric, and below another in AUROC, it does not mean it is below both. The two metrics should be taken into account, and the fact that a technique excels in one of them but performs poorly in another is counterproductive. However,

Table 9
AUROC and AUPR values for networks of more than 2,000 genes.

Network	gnw2000		regulonDB	
technique	AUPR	AUROC	AUPR	AUROC
ARACNE	0.3835	0.6990	0.0154	0.5739
BC3NET	0.2090	0.6342	0.0393	0.5975
C3NET	0.3183	0.6620	0.0131	0.5621
CLR	0.1782	0.8150	**0.0960**	0.9635
KBOOST	**0.3959**	0.7997	0.0064	0.6072
LEAP	0.0505	0.7052	0.0140	0.8228
MRNETB	0.0733	0.8205	0.0198	0.9591
MRNET	0.2715	**0.8270**	0.0264	0.9609
NARROMI	0.0196	0.5490	0.0341	0.6144
PCIT	0.0026	0.5000	0.0008	0.5000
RSNET	0.0564	0.8031	0.0300	**0.9664**
MEDIAN_MO-GENECI	0.1590	0.8711	0.0207	0.9681
BEST_MO-GENECI	**0.3486**	**0.8750**	**0.0353**	**0.9751**

MO-GENECI has demonstrated stability for both cases with consistently high values.

5.3. Statistical significance

This section presents a statistical analysis with the aim of rigorously comparing the performance of individual inference techniques, as well as BEST_MO-GENECI and MEDIAN_MO-GENECI. This analysis takes a global perspective, considering AUROC and AUPR metrics for all the networks in the dataset of this study.

Specifically, the Friedman statistical ranking with non-parametric Holm tests has been calculated for each subset of networks organized by sizes. This is because there are techniques that could not be executed for certain sizes, so they can only be compared with the other techniques that have been executed within their subset.

There are a total of 5 groups determined by the thresholds of 25, 110, 250, and 2000 genes. However, in the last group, this statistical study has been omitted as it consists of only two networks, which is an insufficient quantity to draw reliable conclusions. This results in a total of 8 tables, that is, two for each subset of networks, one for comparing AUPR and another for AUROC.

The first subset consists of networks with up to 25 genes. Table 10 shows the results for the AUPR metric, and Table 11 presents the results for AUROC. In this set of networks, most of the techniques participate, including the 26 individual techniques and the two extractions from MO-GENECI. It is observed that BEST_MO-GENECI is the best-ranked resulting technique for both metrics.

It is worth noting the CMI2NI technique, which is above the significance threshold in both cases. However, as seen later in the next subset regarding large-size networks, it acquires a rather unfavorable position and is surpassed by other inference techniques. In fact, CMI2NI becomes computationally infeasible for networks with more than 110 genes.

The next subset consists of networks with 25 to 110 nodes. The results for the AUPR and AUROC metrics are shown in Table 12 and Table 13, respectively. In these tables, once again, BEST_MO-GENECI leads these two statistical rankings.

In this case, it is worth highlighting the accuracy of techniques derived from the original GENIE3. That is the proposal that applies Random Forest regression (GENIE3_RF), the one that uses ExtraTrees regression (ET), and the one that employs the Stochastic Gradient Boosting Machine (GRNBOOST2). Although only GENIE3_RF is above the significance threshold in both cases, it can be seen that the other two also obtain good positions for the AUPR metric.

This was also observed in the previous proposal [19], so that GENIE3 has a clear domain of specialization in the DREAM networks, and this subset of sizes contains a significantly higher concentration of networks from these challenges. GENIE3, despite obtaining good positions in the other tables, does not stand out as the best individual technique in the ranking in any other case, and in particular, for large sizes is much more computationally expensive than other techniques.

The third subset consists of networks with a size ranging from 110 to 250 genes. Table 14 displays the results for the AUPR metric, while

Table 10
Friedman mean rank with Holm's adjusted p values (0.05) for the AUPR metric measured across all techniques in networks with 0 to 25 genes.

AUPR 0–25 genes		
Technique	$Friedman's Rank$	$Holm's Adj - p$
***BEST_MO-GENECI**	**4.94444**	–
CMI2NI	8.0926	1.59677e−01
MEDIAN_MO-GENECI	9.2778	1.05849e−01
INFERELATOR	10.8889	2.37812e−02
RSNET	11.1852	2.124633-02
GENIE3_ET	11.4444	2.10120e−02
LOCPCACMI	11.4815	2.10120e−02
PCACMI	11.8704	1.38443e−02
GENIE3_RF	12.1296	1.06427e−02
LEAP	13.3148	1.66457e−03
CLR	13.8519	6.93208e−04
MRNET	14.2222	3.75372e−04
MRNETB	14.5370	2.19631e−04
TIGRESS	14.8889	1.15967e−04
BC3NET	15.0185	9.52642e−05
ARACNE	15.1667	7.46093e−05
KBOOST	15.2963	6.02883e−05
GRNBOOST2	15.7407	2.41250e−05
PUC	16.0185	1.36096e−05
C3NET	16.1852	9.77589e−06
PLSNET	16.3333	7.27584e−06
PIDC	17.0926	1.20942e−06
JUMP3	17.4259	5.44535e−07
NONLINEARODES	18.3333	5.12226e−08
MEOMI	19.2963	3.48254e−09
NARROMI	20.0185	4.15433e−10
GRNVBEM	20.0370	4.08150e−10
PCIT	21.9074	9.56885e−13

Table 11
Friedman mean rank with Holm's adjusted p values (0.05) for the AUROC metric measured across all techniques in networks with 0 to 25 genes.

AUROC 0–25 genes		
Technique	$Friedman's Rank$	$Holm's Adj - p$
***BEST_MO-GENECI**	**4.57407**	–
MEDIAN_MO-GENECI	8.0185	1.23925e−01
GENIE3_ET	9.3519	6.56767e−02
CMI2NI	9.8519	5.52124e−02
GENIE3_RF	10.2778	4.33829e−02
INFERELATOR	11.1111	1.75100e−02
RSNET	12.2037	3.92833e−03
LEAP	12.8519	1.52490e−03
PCACMI	13.0926	1.13488e−03
CLR	13.5556	5.42568e−04
LOCPCACMI	13.7222	4.38619e−04
GRNBOOST2	14.0000	2.80650e−04
TIGRESS	14.2407	1.89164e−04
MRNETB	14.3148	1.76315e−04
KBOOST	14.4630	1.40115e−04
PLSNET	14.7222	8.73115e−05
MRNET	15.0926	4.19886e−05
NONLINEARODES	16.3148	2.70204e−06
ARACNE	16.3333	2.70204e−06
JUMP3	16.6481	1.31626e−06
BC3NET	16.7593	1.04978e−06
PIDC	17.1111	4.50595e−07
PUC	17.1667	4.08972e−07
C3NET	17.7037	1.03604e−07
MEOMI	18.8148	4.81678e−09
NARROMI	20.4630	3.18802e−11
GRNVBEM	20.5926	2.17753e−11
PCIT	22.6481	1.85180e−14

Table 12
Friedman mean rank with Holm's adjusted p values (0.05) for the AUPR metric measured across all techniques in networks with 25 to 110 genes.

AUPR 25–110 genes		
Technique	$Friedman's Rank$	$Holm's Adj-p$
***BEST_MO-GENECI**	**3.55882**	–
GENIE3_RF	6.0882	1.88868e−01
CLR	7.6177	6.99917e−02
MEDIAN_MO-GENECI	8.2941	4.17035e−02
GENIE3_ET	8.5588	3.75826e−02
MRNETB	9.5000	1.01360e−02
TIGRESS	10.0588	4.40476e−03
MRNET	10.3235	3.08976e−03
GRNBOOST2	10.4412	2.80049e−03
LOCPCACMI	11.9118	1.28796e−04
RSNET	12.1176	8.74889e−05
PUC	12.5882	2.99879e−05
INFERELATOR	13.2647	5.53287e−06
PIDC	13.3529	4.71220e−06
CMI2NI	13.4412	3.98140e−06
BC3NET	14.8529	6.66146e−08
GRNVBEM	16.4118	3.91223e−10
ARACNE	17.0000	4.94054e−11
NARROMI	17.1471	3.02786e−11
PCACMI	17.3824	1.31674e−11
LEAP	18.0000	1.26040e−12
C3NET	18.1765	6.54808e−13
PLSNET	18.9706	2.61052e−14
KBOOST	20.0294	2.69102e−16
PCIT	20.2647	9.66022e−17
NONLINEARODES	24.2059	1.93437e−25
MEOMI	24.4412	5.32306e−26

Table 13
Friedman mean rank with Holm's adjusted p values (0.05) for the AUROC metric measured across all techniques in networks with 25 to 110 genes.

AUROC 25-110 genes		
Technique	$Friedman's Rank$	$Holm's Adj-p$
***BEST_MO-GENECI**	**3.73529**	–
GENIE3_RF	5.5000	4.09525e−01
GENIE3_ET	6.1765	4.09525e−01
MEDIAN_MO-GENECI	6.6471	3.91180e−01
GRNBOOST2	8.3529	6.58146e−02
TIGRESS	8.9118	3.58351e−02
CLR	9.1177	3.10503e−02
MRNETB	9.7941	1.15338e−02
CMI2NI	10.4118	4.19203e−03
MRNET	10.7353	2.48984e−03
LOCPCACMI	11.2353	9.78071e−04
PIDC	12.9118	2.05860e−05
PUC	13.1471	1.21583e−05
RSNET	13.5000	5.10683e−06
INFERELATOR	14.1176	9.68763e−07
LEAP	15.3529	2.38545e−08
PLSNET	16.2647	1.21400e−09
PCACMI	16.4412	6.97889e−10
BC3NET	18.4706	3.49502e−13
GRNVBEM	18.5000	3.27531e−13
ARACNE	19.2647	1.44156e−14
KBOOST	20.0882	4.16593e−16
NONLINEARODES	20.6471	3.43919e−17
NARROMI	21.0294	6.02465e−18
C3NET	21.4706	7.62403e−19
PCIT	21.5294	5.97097e−19
MEOMI	24.6471	4.50332e−26

Table 15 presents the AUROC results. In this case, it is observed that there are difficulties in confirming a statistically significant difference between the techniques, as nearly half of them are above the threshold. Nevertheless, it can be noted that BEST_MO-GENECI once again clearly leads both rankings. Furthermore, MEDIAN_MO-GENECI ranks second and third in the tables. It is surpassed by CLR in terms of AUPR, but this same technique ranks sixth for the AUROC metric.

Table 14
Friedman mean rank with Holm's adjusted p values (0.05) for the AUPR metric measured across all techniques in networks with 110 to 250 genes.

AUPR 110–250 genes		
Technique	$Friedman's Rank$	$Holm's Adj-p$
***BEST_MO-GENECI**	**2.83333**	–
CLR	5.0000	4.44534e−01
MEDIAN_MO-GENECI	6.5833	4.44534e−01
MRNET	6.6667	4.44534e−01
RSNET	7.4167	3.35301e−01
GENIE3_RF	7.9167	2.85960e−01
GRNBOOST2	8.0833	2.85960e−01
PCACMI	8.8333	1.65320e−01
MRNETB	9.1667	1.35140e−01
GENIE3_ET	10.1667	5.10336e−02
INFERELATOR	10.7500	2.82376e−02
ARACNE	11.5833	1.06108e−02
BC3NET	12.4583	3.39169e−03
PUC	13.1667	1.26136e−03
PLSNET	13.5000	8.02258e−04
PIDC	13.5833	7.51756e−04
C3NET	13.6667	7.00642e−04
LEAP	15.2500	4.78787e−05
NARROMI	17.9583	2.08922e−07
KBOOST	18.1667	1.38603e−07
MEOMI	18.9167	2.60800e−08
PCIT	21.3333	6.26486e−11

Table 15
Friedman mean rank with Holm's adjusted p values (0.05) for the AUROC metric measured across all techniques in networks with 110 to 250 genes.

AUROC 110–250 genes		
Technique	$Friedman's Rank$	$Holm's Adj-p$
***BEST_MO-GENECI**	**1.33333**	–
MEDIAN_MO-GENECI	4.1667	2.85168e−01
PCACMI	5.2500	2.79120e−01
GENIE3_ET	7.0000	1.20057e−01
GRNBOOST2	7.2500	1.20057e−01
CLR	7.3333	1.20057e−01
GENIE3_RF	7.5000	1.20057e−01
RSNET	7.8333	9.94704e−02
MRNETB	8.1667	7.95802e−02
MRNET	8.4167	6.78703e−02
PUC	10.0000	1.07847e−02
PIDC	10.0833	1.06108e−02
INFERELATOR	14.1667	1.55065e−05
LEAP	14.4167	1.04061e−05
PLSNET	14.9167	4.19078e−06
BC3NET	15.7917	7.39047e−07
ARACNE	16.3333	2.44677e−07
C3NET	17.2500	3.27207e−08
KBOOST	17.4167	2.3472e−08
NARROMI	18.7917	8.60938e−10
MEOMI	18.8333	8.15172e−10
PCIT	20.7500	5.04342e−12

It is also worth mentioning the emergence of techniques that had gone unnoticed until now and are beginning to gain momentum for the next size subgroup (of large-size nets). These are primarily MRNET and CLR, and their potential contribution to larger networks is becoming apparent.

The final subset for which this comparative study has been conducted consists of networks with a size between 250 and 2000 genes. The results for the AUPR metric are shown in Table 16, and for AUROC, they can be found in Table 17. As has been the case in the other scenarios, BEST_MO-GENECI emerges as the top performer for network inference in this case as well.

Moreover, as anticipated for the previous subset, CLR and MRNET seem to yield good results for larger networks. These techniques are the only ones that surpass the significance threshold for both tables. Furthermore, by using the small group of networks with more than

Table 16
Friedman mean rank with Holm's adjusted p values (0.05) for the AUPR metric measured across all techniques in networks with 250 to 2000 genes.

AUPR 250–2000 genes		
Technique	$Friedman's Rank$	$Holm's Adj - p$
***BEST_MO-GENECI**	**2.90323**	–
CLR	3.8065	4.44083e−01
ARACNE	4.2903	4.44083e−01
MRNET	4.8387	2.65207e−01
MEDIAN_MO-GENECI	6.1936	1.50890e−02
BC3NET	6.3548	1.42610e−02
C3NET	6.3548	1.42610e−02
RSNET	6.8065	4.13002e−03
MRNETB	7.9032	8.59231e−05
PIDC	9.9355	5.38791e−09
NARROMI	10.1290	2.00219e−09
PUC	10.5484	1.86184e−10
LEAP	11.8065	5.49889e−14
KBOOST	13.1613	2.21967e−18
PCIT	14.9677	3.33603e−25

Table 17
Friedman mean rank with Holm's adjusted p values (0.05) for the AUROC metric measured across all techniques in networks with 250 to 2000 genes.

AUROC 250–2000 genes		
Technique	$Friedman's Rank$	$Holm's Adj - p$
***BEST_MO-GENECI**	**2.51613**	–
MRNET	3.5161	4.44083e−01
MEDIAN_MO-GENECI	3.9032	4.44083e−01
CLR	4.3548	3.16541e−01
MRNETB	4.7097	2.18862e−01
RSNET	4.8065	2.18862e−01
PUC	7.2258	2.02893e−04
PIDC	7.4839	8.56653e−05
ARACNE	8.8387	2.08492e−07
BC3NET	10.4516	2.54677e−11
C3NET	10.5806	1.25188e−11
LEAP	11.5161	2.54894e−14
KBOOST	12.0968	4.00058e−16
NARROMI	13.2903	3.15211e−20
PCIT	14.7097	9.81590e−26

2000 genes as an extension of the current subset, it was observed in Table 9 how these techniques consistently achieved good results. In fact, they are highlighted in bold due to their high accuracy on a couple of occasions.

However, it should be remembered that CLR obtained a rather unfavorable position for the subset of small networks, and MRNET also did not excel until reaching larger network sizes (above 110 genes). Therefore, while they do not outperform BEST_MO-GENECI in this scenario, they also do not come close to covering the broad spectrum of domains that BEST_MO-GENECI excels in.

The fact that some techniques seem to exhibit a clear relationship between their performance and network size is a strong indication that size is a determining factor in specifying their domains of specialization. This does not mean that there are no other factors at play or that there are no techniques whose domains are independent of network size. It is important to emphasize that the way the networks have been grouped has allowed us to observe how techniques behave with respect to this specific factor. After careful analysis, it has been observed that some techniques clearly depend on network size, while others do not show a discernible pattern or trend related to this factor.

After statistically analyzing the results of this study, it can be rigorously confirmed that BEST_MO-GENECI provides the highest reliability when aiming for high-precision results, regardless of network size or specialization domain. Furthermore, it is worth mentioning that in cases where a poor choice is made on the front, MEDIAN_MO-GENECI has crossed the significance threshold in a total of 6 scenarios, falling just below it in the remaining two cases.

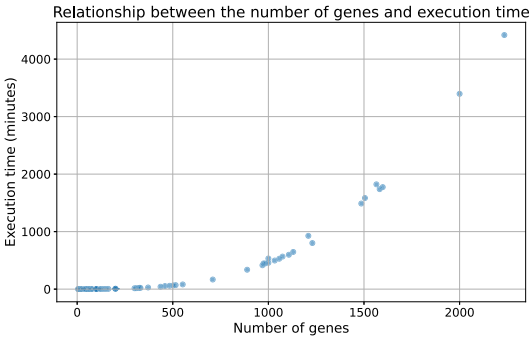


Fig. 13. Scatter plot illustrating the relationship between the number of genes and the execution time of the algorithm. Each point represents a specific execution instance. The plot provides insight into the performance characteristics of the algorithm across different gene counts.

This study has also made it possible to observe techniques such as PCIT, MEOMI, NARROMI, or KBOOST, which consistently yield low-precision results. Researchers may consider excluding these techniques when using the MO-GENECI approach, which could help eliminate noise and further enhance the evolutionary algorithm's results.

5.4. Computational complexity

It is evident that the complex and sophisticated strategy designed to optimally harmonize such a multitude of techniques incurs a computational cost equivalent to its precision. This computational cost depends on several factors. The first and most determining factor is the number of genes involved in the gene regulatory network that is intended to be inferred. Concerning this factor, Fig. 13 presents the execution times in minutes required by MO-GENECI to achieve 250,000 evaluations in each problem, ordered by their size. These times were recorded on a machine with 500 GB of RAM and 32 cores. It can be observed that the execution time is clearly exponential with respect to the size of the networks, starting to require significant time from 1000 nodes onwards. After analyzing certain executions with JProfiler [106], it has been noted that the most costly part and the one that most contributes to the slowdown of the algorithm in large networks is the objective function that counts different motifs. This is to be expected; however, the improvement in precision observed compared to the previous GENECI proposal makes this increase in the algorithm's computational cost worthwhile.

Additionally, there is the factor of the number of evaluations. In the experimentation of this work, we have set a total of 250,000 evaluations for all problems. However, this quantity was set on the high side, and it has been observed that in different problems, the algorithm achieves convergence much earlier. The point at which convergence occurs for a problem does not depend so much on the size of the network but rather on the coherence among the various techniques and their quantity. Therefore, these last two variables are also indirect factors that affect the execution time.

Finally, it is important to remember that MO-GENECI runs after collecting the results from all individual techniques. This means that comparing the computational cost of these individual techniques to this proposal is meaningless since they operate at different levels and have different purposes. The complexity of the individual techniques is relative to the task of inferring gene regulatory networks from expression data, while the complexity of MO-GENECI corresponds to the intelligent and optimized process of harmonizing a broad spectrum of inference techniques based on fitness functions specific to the biological domain to which this challenge belongs.

5.5. Real-world experimentation

After validating its efficacy, MO-GENECI has also been run on real-world gene expression data. Specifically, clinical data from melanoma patients were used. This data included gene expression levels collected from NanoString, specifically from the platform's immunological profiling panel. This panel was subjected to specific treatment and filtering techniques, which eliminated 35 genes showing less stable results.

This dataset was also employed in GENECI, and the approach taken then was to contrast the interactions with the highest confidence level of the solution with respect to the literature. However, in this case, since MO-GENECI is a multi-objective algorithm, we do not have a single solution, but a front of solutions. Therefore, the strategy used in this case has been to collect the most frequent and most trusted interactions taking into account their position after sorting each network by confidence level.

After carrying out this procedure, the following top 5 interactions were obtained. For each of them, both genes were entered into the STRING database [107] and the Co-Mentioned in Pubmed Abstracts section was accessed to obtain a list of publications. This list was subsequently filtered to find articles related to melanoma, cancer in general, or skin diseases.

1. **IL8 - CCL2**: In [108], the role of IL-8 as a potential predictive biomarker in head and neck cancer is studied. In their study, after analyzing patients before and after radiation therapy, it is shown that there is a strong relationship between IL-8 and CCL2 (MCP-1), which is statistically verified by Pearson's correlation coefficients. In addition, [109] establishes that CXCL8 (IL-8), CCL2, and CCL5 are three key chemokines in the invasion of tumor cells, providing evidence of their interactions through numerous experiments.
2. **OSM - IL8**: In [110], Oncostatin M (OSM) is shown to enhance skin squamous cell carcinoma (20% of skin cancer deaths) and that its expression in tumor lesions strongly correlates with that of the well-known neutrophil chemotactic factor IL-8. This relationship has been rigorously confirmed through experimentation and Spearman rank correlation calculation. Additionally, [111] studies the effect of ulipristal acetate (UPA) in the treatment of endometrial cancer. For this, the expression levels of the proinflammatory cytokines OSM, IL-6, and IL-8 were examined using quantitative real-time PCR, revealing a clear importance of the expression levels and interaction of these gene products in cancer.
3. **IL1R2 - ARG1**: This interaction was already validated in the GENECI publication, as it was the most reliable regulation reported by the preliminary algorithm of this project. Specifically, it was highlighted that several studies relate both genes within the context of cancer [112,113]. Specifically, in [114] it is concluded that the amebiasis pathway could be involved in melanoma metastasis through these genes.
4. **IL8 - OSM**: This concerns the inverse regulation of interaction 2, which points to a coregulation between both genes. The justification is the same as that for the inverse regulation, as validation through literature search does not provide a sufficient level of precision to determine the direction of the regulation.
5. **NFKBIA - TNFAIP3**: In [115], the role of KLF6 in gene regulation is analyzed, focusing on patients with glioblastoma. During their study, they demonstrate that NFKBIA and TNFAIP3 are the most potent negative regulators of NF-κB induced by KLF6. The findings of the previous article coincide with what was also demonstrated for colon cancer in [116]. It shows that once again NFKBIA and TNFAIP3 are two important feedback loops of NF-κB, and that samples from patients with colorectal cancer (n = 626) show much lower gene expression of NFKBIA (0.643 times) and TNFAIP3 (0.745 times) compared to healthy controls (n = 51) according to the TCGA database. Finally, in [117], this association is also confirmed for breast cancer. Therefore, despite the lack of specific studies for skin cancer, it suggests that this interaction is likely to play a relevant role in melanoma as well.

From these, it can be observed that only one appeared in the top 3 provided by GENECI in the previous publication. The other two are also in high positions but not notable, specifically IL18R1 - IL1RL1 is in seventh position and HLA-DQA1 - HLA-DQB1 in twelfth. This means that MO-GENECI continues to detect the most important interactions provided by GENECI but adds the inference of new interactions that have also been shown to be supported by the literature.

6. Conclusions

This work has developed an algorithmic proposal to address the initial uncertainty associated with choosing a truly suitable technique for inferring a specific gene regulation network. MO-GENECI has demonstrated its ability to provide reliable results for any specialization domain and network size, making it a secure and robust approach that guarantees high-quality results.

Its implementation has been carefully refined throughout this study. A series of new objective functions have been generated in this proposal and analyzed to select their most accurate versions, conducting rigorous statistical studies and considering a wide range of modifications.

Specific crossover and mutation operators have been also designed for this problem to ensure the feasibility of solutions after their execution, eliminating the need for repair mechanisms that could distort learning between generations. Furthermore, an extensive parameter setting procedure has been carried out to ensure the selection of the best possible combination of values.

MO-GENECI has been encapsulated into a Python software package available on PyPI,[5] complete with documentation and a list of commands. This enhances its accessibility and usability, making the experimentation undertaken fully reproducible.

It is important to emphasize that when presenting the results of MO-GENECI and its accuracy in analyzing a benchmark consisting of 106 gene regulation networks, its potential to uncover new interactions in previously unexplored real-world biological networks is highlighted. This advancement promises to significantly contribute to our understanding of genetic mechanisms, offering opportunities for future research in medicine and molecular biology.

Furthermore, MO-GENECI is presented as a proposal fully compatible with the emergence of new individual inference techniques. Far from becoming obsolete, the strategy presented in this work allows for the seamless integration of networks inferred by new methods, enhancing the quality of consensus as the precision of considered techniques increases.

In this sense, future work contemplates the investigation of new objective functions that address more complex characteristics of these types of networks. The continuous publication of methods for inferring gene regulation networks from expression data indicates that the scientific community continues to explore new properties of these networks and their corresponding reflection in input data.

CRediT authorship contribution statement

Adrián Segura-Ortiz: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **José García-Nieto:** Visualization, Supervision, Project administration, Methodology. **José F. Aldana-Montes:** Supervision, Funding acquisition. **Ismael Navas-Delgado:** Validation, Supervision, Resources, Project administration.

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

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.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors utilized the ChatGPT tool to enhance the text's clarity and coherence. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgments

This work has been partially funded by grant (funded by MCIN/AEI/10.13039/501100011033/) PID2020-112540RB-C41, AETHER-UMA (A smart data holistic approach for context-aware data analytics: semantics and context exploitation) and the Junta de Andalucia, Spain, under contract QUAL21 010UMA. Funding for open access charge: Universidad de Málaga/CBUA. Adrián Segura-Ortiz is supported by Grant FPU21/03837 (Spanish Ministry of Science, Innovation and Universities).

References

[1] E. Davidson, M. Levine, Gene regulatory networks, Proc. Natl. Acad. Sci. 102 (2005) 4935.
[2] V.A. Huynh-Thu, A. Irrthum, L. Wehenkel, P. Geurts, Inferring regulatory networks from expression data using tree-based methods, PLoS One 5 (9) (2010) e12776.
[3] V.A. Huynh-Thu, G. Sanguinetti, Combining tree-based and dynamical systems for the inference of gene regulatory networks, Bioinformatics 31 (10) (2015) 1614–1622.
[4] X. Zhang, X.-M. Zhao, K. He, L. Lu, Y. Cao, J. Liu, J.-K. Hao, Z.-P. Liu, L. Chen, Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information, Bioinformatics 28 (1) (2012) 98–104.
[5] 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, in: BMC Bioinformatics, vol. 7, Springer, 2006, pp. 1–15.
[6] P.E. Meyer, K. Kontos, F. Lafitte, G. Bontempi, Information-theoretic inference of large transcriptional regulatory networks, EURASIP J. Bioinform. Syst. Biol. 2007 (2007) 1–9.
[7] P. Meyer, D. Marbach, S. Roy, M. Kellis, Information-theoretic inference of gene networks using backward elimination, in: BioComp, 2010, pp. 700–705.
[8] A. Reverter, E.K. Chan, Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks, Bioinformatics 24 (21) (2008) 2491–2497.
[9] G. Altay, F. Emmert-Streib, Inferring the conservative causal core of gene regulatory networks, BMC Syst. Biol. 4 (1) (2010) 1–13.
[10] 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 (1) (2007) e8.
[11] 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, Nature Methods 9 (8) (2012) 796–804.
[12] M. Bansal, V. Belcastro, A. Ambesi-Impiombato, D. Di Bernardo, How to infer gene networks from expression profiles, Mol. Syst. Biol. 3 (2007).
[13] J.M. Raser, E.K. O'shea, Noise in gene expression: origins, consequences, and control, Science 309 (5743) (2005) 2010–2013.
[14] H. Khojasteh, A. Khanteymoori, M.H. Olyaee, EnGRNT: Inference of gene regulatory networks using ensemble methods and topological feature extraction, Inform. Med. Unlock. 27 (2021) 100773.
[15] 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.
[16] 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.
[17] 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.
[18] Y. Wang, T. Joshi, X.S. Zhang, D. Xu, L. Chen, Inferring gene regulatory networks from multiple microarray datasets, Bioinformatics (Oxford, England) 22 (2006) 2413–2420.
[19] A. Segura-Ortiz, J. García-Nieto, J.F. Aldana-Montes, I. Navas-Delgado, GENECI: A novel evolutionary machine learning consensus-based approach for the inference of gene regulatory networks, Comput. Biol. Med. 155 (2023) 106653.
[20] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Trans. Evol. Comput. 6 (2) (2002) 182–197.
[21] A. Pinna, N. Soranzo, I. Hoeschele, A. de la Fuente, Simulating systems genetics data with SysGenSIM, Bioinformatics 27 (17) (2011) 2459–2462.
[22] 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. 10 (1) (2018) 1–15.
[23] 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, Pharm. J. 14 (2) (2014) 93–106.
[24] 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.
[25] 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.
[26] Y. Watanabe, S. Seno, Y. Takenaka, H. Matsuda, An estimation method for inference of gene regulatory net-work using bayesian network with uniting of partial problems, BMC Genomics 13 (2012) S12.
[27] J. Wu, X. Zhao, Z. Lin, Z. Shao, Large scale gene regulatory network inference with a multi-level strategy, Mol. Biosyst. 12 (2016) 588–597.
[28] S. Hurtado, J. Garcia-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.
[29] 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.
[30] Y. Gan, X. Hu, G. Zou, C. Yan, G. Xu, Inferring gene regulatory networks from single-cell transcriptomic data using bidirectional rnn, Front. Oncol. 12 (2022).
[31] N. Kizaki, H. Yoshino, H. Kurokawa, The inference method of the gene regulatory network with a majority rule, Nonlinear Theory Appl., IEICE 6 (2015) 226–236.
[32] A. Ghazikhani, T. Akbarzadeh, R. Monsefi, Genetic regulatory network inference using recurrent neural networks trained by a multi agent system, in: 2011 1st International eConference on Computer and Knowledge Engineering, ICCKE, 2011.
[33] H. Yasuki, M. Kikuchi, H. Kurokawa, Inferring method of the gene regulatory networks using neural networks adopting a majority rule, in: The 2011 International Joint Conference on Neural Networks, 2011.
[34] N. Zarayeneh, E. Ko, J.H. Oh, S.C. Suh, C. Liu, J. Gao, D. Kim, M. Kang, Integration of multi-omics data for integrative gene regulatory network inference, Int. J. Data Min. Bioinform. 18 (2017) 223.
[35] J.D. Finkle, J. Wu, N. Bagheri, Windowed granger causal inference strategy improves discovery of gene regulatory networks, Proc. Natl. Acad. Sci. 115 (2018) 2252–2257.
[36] B. Yang, Y. Xu, Reconstructing gene regulation network based on conditional mutual information, in: Proceedings of the 2017 International Conference on Mechanical, Electronic, Control and Automation Engineering, MECAE 2017, 2017.
[37] R. de Matos Simoes, F. Emmert-Streib, Bagging statistical network inference from large-scale gene expression data, PLoS One 7 (3) (2012) e33624.
[38] T. Moerman, S. Aibar Santos, C. Bravo González-Blas, J. Simm, Y. Moreau, J. Aerts, S. Aerts, GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks, Bioinformatics 35 (12) (2018) 2159–2161.
[39] L.F. Iglesias-Martinez, B. De Kegel, W. Kolch, KBoost: a new method to infer gene regulatory networks from gene expression data, Sci. Rep. 11 (1) (2021) 15461.
[40] A.-C. Haury, F. Mordelet, P. Vera-Licona, J.-P. Vert, TIGRESS: trustful inference of gene regulation using stability selection, BMC Syst. Biol. 6 (1) (2012) 1–17.
[41] X. Zhang, J. Zhao, J.-K. Hao, X.-M. Zhao, L. Chen, Conditional mutual inclusive information enables accurate quantification of associations in gene regulatory networks, Nucleic Acids Res. 43 (5) (2014) e31.
[42] M. Sanchez-Castillo, D. Blanco, I.M. Tienda-Luna, M.C. Carrion, Y. Huang, A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data, Bioinformatics 34 (6) (2017) 964–970.
[43] R. Bonneau, D.J. Reiss, P. Shannon, M. Facciotti, L. Hood, N.S. Baliga, V. Thorsson, The inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo, Genome Biol. 7 (2006) 1–16.
[44] A.T. Specht, J. Li, LEAP: constructing gene co-expression networks for single-cell RNA-sequencing data using pseudotime ordering, Bioinformatics 33 (5) (2017) 764–766.

[45] X. Chen, M. Li, R. Zheng, S. Zhao, F.-X. Wu, Y. Li, J. Wang, A novel method of gene regulatory network structure inference from gene knock-out expression data, Tsinghua Sci. Technol. 24 (4) (2019) 446–455.
[46] J. Lei, Z. Cai, X. He, W. Zheng, J. Liu, An approach of gene regulatory network construction using mixed entropy optimizing context-related likelihood mutual information, Bioinformatics 39 (1) (2023) btac717.
[47] X. Zhang, K. Liu, Z.-P. Liu, B. Duval, J.-M. Richer, X.-M. Zhao, J.-K. Hao, L. Chen, NARROMI: a noise and redundancy reduction technique improves accuracy of gene regulatory network inference, Bioinformatics 29 (1) (2013) 106–113.
[48] B. Ma, M. Fang, X. Jiao, Inference of gene regulatory networks based on nonlinear ordinary differential equations, Bioinformatics 36 (19) (2020) 4885–4893.
[49] T.E. Chan, M.P. Stumpf, A.C. Babtie, Gene regulatory network inference from single-cell data using multivariate information measures, Cell Syst. 5 (3) (2017) 251–267.e3.
[50] S. Guo, Q. Jiang, L. Chen, D. Guo, Gene regulatory network inference using PLS-based methods, BMC Bioinformatics 17 (1) (2016) 1–10.
[51] X. Jiang, X. Zhang, RSNET: inferring gene regulatory networks by a redundancy silencing and network enhancement technique, BMC Bioinformatics 23 (1) (2022) 1–18.
[52] M. Zhao, W. He, J. Tang, Q. Zou, F. Guo, A hybrid deep learning framework for gene regulatory network inference from single-cell transcriptomic data, Brief. Bioinform. 23 (2) (2022) bbab568.
[53] T. Hillerton, D. Seçilmiş, S. Nelander, E.L. Sonnhammer, Fast and accurate gene regulatory network inference by normalized least squares regression, Bioinformatics 38 (8) (2022) 2263–2268.
[54] W. Liu, Y. Jiang, L. Peng, X. Sun, W. Gan, Q. Zhao, H. Tang, Inferring gene regulatory networks using the improved Markov blanket discovery algorithm, Interdiscip. Sci. Comput. Life Sci. (2022) 1–14.
[55] C. Skok Gibbs, C.A. Jackson, G.-A. Saldi, A. Tjärnberg, A. Shah, A. Watters, N. De Veaux, K. Tchourine, R. Yi, T. Hamamsy, et al., High-performance single-cell gene regulatory network inference at scale: the Inferelator 3.0, Bioinformatics 38 (9) (2022) 2519–2528.
[56] N. Namura, Surrogate-assisted reference vector adaptation to various Pareto front shapes for many-objective Bayesian optimization, in: 2021 IEEE Congress on Evolutionary Computation, CEC, 2021, pp. 901–908.
[57] P. Meyer, J. Saez-Rodriguez, Advances in systems biology modeling: 10 years of crowdsourcing dream challenges, Cell Syst. 12 (6) (2021) 636–653.
[58] P. Schmitt, B. Sorin, T. Frouté, N. Parisot, F. Calevro, S. Peignier, GReNaDIne: A data-driven Python library to infer gene regulatory networks from gene expression data, Genes 14 (2) (2023).
[59] M. Aluru, H. Shrivastava, S.P. Chockalingam, S. Shivakumar, S. Aluru, En-GRaiN: a supervised ensemble learning method for recovery of large-scale gene regulatory networks, Bioinformatics 38 (5) (2022) 1312–1319.
[60] D.M. Alawad, A. Katebi, M.W.U. Kabir, M.T. Hoque, AGRN: accurate gene regulatory network inference using ensemble machine learning methods, Bioinform. Adv. 3 (1) (2023) vbad032.
[61] Q. Zhang, H. Li, MOEA/D: a multiobjective evolutionary algorithm based on decomposition, IEEE Trans. Evol. Comput. 11 (6) (2007) 712–731.
[62] C.A.C. Coello, G.T. Pulido, M.S. Lechuga, Handling multiple objectives with particle swarm optimization, IEEE Trans. Evol. Comput. 8 (3) (2004) 256–279.
[63] S. Kukkonen, J. Lampinen, GDE3: The third evolution step of generalized differential evolution, in: 2005 IEEE Congress on Evolutionary Computation, Vol. 1, IEEE, 2005, pp. 443–450.
[64] A.J. Nebro, J.J. Durillo, J. Garcia-Nieto, C.C. Coello, F. Luna, E. Alba, SMPSO: A new PSO-based metaheuristic for multi-objective optimization, in: 2009 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making, MCDM, IEEE, 2009, pp. 66–73.
[65] S. Tsutsui, M. Yamamura, T. Higuchi, Multi-parent recombination with simplex crossover in real coded genetic algorithms, in: Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation-Volume 1, 1999, pp. 657–664.
[66] D. Hadka, MOEA framework-a free and open source java framework for multiobjective optimization, 2012.
[67] R. Eisinga, T. Heskes, B. Pelzer, M. Te Grotenhuis, Exact p-values for pairwise comparison of friedman rank sums, with application to comparing classifiers, BMC Bioinformatics 18 (1) (2017) 1–18.
[68] R. Albert, Scale-free networks in cell biology, J. Cell Sci. 118 (21) (2005) 4947–4957.
[69] S. Wuchty, Scale-free behavior in protein domain networks, Mol. Biol. Evol. 18 (9) (2001) 1694–1702.
[70] A.-L. Barabási, Linked: The New Science of Networks, American Association of Physics Teachers, 2003.
[71] M. Nazarieh, A. Wiese, T. Will, M. Hamed, V. Helms, Identification of key player genes in gene regulatory networks, BMC Syst. Biol. 10 (2016) 1–12.
[72] J. Åkesson, Z. Lubovac-Pilav, R. Magnusson, M. Gustafsson, ComHub: Community predictions of hubs in gene regulatory networks, BMC Bioinformatics 22 (2021) 1–12.
[73] S. Gulati, S. Shapiro, Goodness-of-fit tests for Pareto distribution, Stat. Models Methods Biomed. Tech. Syst. (2008) 259–274.
[74] U. Alon, Network motifs: theory and experimental approaches, Nature Rev. Genet. 8 (6) (2007) 450–461.
[75] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, Science 298 (5594) (2002) 824–827.
[76] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, U. Alon, Superfamilies of evolved and designed networks, Science 303 (5663) (2004) 1538–1542.
[77] R.J. Prill, P.A. Iglesias, A. Levchenko, Dynamic properties of network motifs contribute to biological network organization, PLoS Biol. 3 (11) (2005) e343.
[78] D. Michail, J. Kinable, B. Naveh, J.V. Sichi, JGraphT—A java library for graph data structures and algorithms, ACM Trans. Math. Softw. 46 (2) (2020) 1–29.
[79] R. Pinho, V. Garcia, M. Irimia, M.W. Feldman, Stability depends on positive autoregulation in Boolean gene regulatory networks, PLoS Comput. Biol. 10 (11) (2014) e1003916.
[80] 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.
[81] O. Liska, B. Bohár, A. Hidas, T. Korcsmáros, B. Papp, D. Fazekas, E. Ari, TFLink: an integrated gateway to access transcription factor–target gene interactions for multiple species, Database 2022 (2022) baac083.
[82] V.H. Tierrafría, C. Rioualen, H. Salgado, P. Lara, S. Gama-Castro, P. Lally, L. Gómez-Romero, P. Peña-Loredo, A.G. López-Almazo, G. Alarcón-Carranza, et al., RegulonDB 11.0: Comprehensive high-throughput datasets on transcriptional regulation in escherichia coli K-12, Microb. Genom. 8 (5) (2022) 000833.
[83] Z.-P. Liu, C. Wu, H. Miao, H. Wu, RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse, Database 2015 (2015).
[84] R. Oughtred, J. Rust, C. Chang, B.-J. Breitkreutz, C. Stark, A. Willems, L. Boucher, G. Leung, N. Kolas, F. Zhang, et al., The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions, Prot. Sci. 30 (1) (2021) 187–200.
[85] L. Fang, Y. Li, L. Ma, Q. Xu, F. Tan, G. Chen, GRNdb: decoding the gene regulatory networks in diverse human and mouse conditions, Nucleic Acids Res. 49 (D1) (2021) D97–D103.
[86] T. Van den Bulcke, K. Van Leemput, B. Naudts, P. van Remortel, H. Ma, A. Verschoren, B. De Moor, K. Marchal, SynTReN: a generator of synthetic gene expression data for design and analysis of structure learning algorithms, BMC Bioinformatics 7 (2006) 1–12.
[87] S. Rogers, M. Girolami, A Bayesian regression approach to the inference of regulatory networks from gene expression data, Bioinformatics 21 (14) (2005) 3131–3137.
[88] T. Schaffter, D. Marbach, D. Floreano, GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods, Bioinformatics 27 (16) (2011) 2263–2270.
[89] 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.
[90] P. Bellot, C. Olsen, P. Meyer, Grndata: synthetic expression data for gene regulatory network inference, 2018, R Package Version 1.
[91] L. Garcia-Alonso, C.H. Holland, M.M. Ibrahim, D. Turei, J. Saez-Rodriguez, Benchmark and integration of resources for the estimation of human transcription factor activities, Genome Res. (2019).
[92] S. Kolmykov, I. Yevshin, M. Kulyashov, R. Sharipov, Y. Kondrakhin, V.J. Makeev, I.V. Kulakovskiy, A. Kel, F. Kolpakov, GTRD: an integrated view of transcription regulation, Nucleic Acids Res. 49 (D1) (2020) D104–D111.
[93] L. Bovolenta, M. Acencio, N. Lemke, Htridb: an open-access database for experimentally verified human transcriptional regulation interactions, Nat. Proc. (2012) 1.
[94] J.A. Castro-Mondragon, R. Riudavets-Puig, I. Rauluseviciute, R. Berhanu Lemma, L. Turchi, R. Blanc-Mathieu, J. Lucas, P. Boddie, A. Khan, N. Manosalva Pérez, O. Fornes, T.Y. Leung, A. Aguirre, F. Hammal, D. Schmelter, D. Baranasic, B. Ballester, A. Sandelin, B. Lenhard, K. Vandepoele, W.W. Wasserman, F. Parcy, A. Mathelier, JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles, Nucleic Acids Res. 50 (D1) (2021) D165–D173.
[95] R. Lesurf, K.C. Cotto, G. Wang, M. Griffith, K. Kasaian, S.J.M. Jones, S.B. Montgomery, O.L. Griffith, T.O.R.A. Consortium, ORegAnno 3.0: a community-driven resource for curated regulatory annotation, Nucleic Acids Res. 44 (D1) (2015) D126–D132.
[96] S.V. Keränen, A. Villahoz-Baleta, A.E. Bruno, M.S. Halfon, REDfly: An integrated knowledgebase for insect regulatory genomics, Insects 13 (7) (2022) 618.
[97] F. Hammal, P. de Langen, A. Bergon, F. Lopez, B. Ballester, ReMap 2022: a database of human, mouse, drosophila and arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments, Nucleic Acids Res. 50 (D1) (2022) D316–D325.
[98] C. Jiang, Z. Xuan, F. Zhao, M.Q. Zhang, TRED: a transcriptional regulatory element database, new entries and other development, Nucleic Acids Res. 35 (suppl_1) (2007) D137–D140.

[99] H. Han, J.-W. Cho, S. Lee, A. Yun, H. Kim, D. Bae, S. Yang, C.Y. Kim, M. Lee, E. Kim, et al., TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions, Nucleic Acids Res. 46 (D1) (2018) D380–D386.

[100] M.C. Teixeira, R. Viana, M. Palma, J. Oliveira, M. Galocha, M.N. Mota, D. Couceiro, M.G. Pereira, M. Antunes, I.V. Costa, P. Pais, C. Parada, C. Chaouiya, I. Sá-Correia, P.T. Monteiro, YEASTRACT+: a portal for the exploitation of global transcription regulation and metabolic model data in yeast biotechnology and pathogenesis, Nucleic Acids Res. 51 (D1) (2022) D785–D791.

[101] N. Guelzim, S. Bottani, P. Bourgine, F. Képès, Topological and causal structure of the yeast transcriptional regulatory network, Nature Genet. 31 (1) (2002) 60–63.

[102] E. Zitzler, L. Thiele, M. Laumanns, C.M. Fonseca, V.G. Da Fonseca, Performance assessment of multiobjective optimizers: An analysis and review, IEEE Trans. Evol. Comput. 7 (2) (2003) 117–132.

[103] D.A. Van Veldhuizen, G.B. Lamont, Multiobjective Evolutionary Algorithm Research: A History and Analysis, Technical Report, Citeseer, 1998.

[104] E. Zitzler, L. Thiele, Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach, IEEE Trans. Evol. Comput. 3 (4) (1999) 257–271.

[105] H. Ishibuchi, H. Masuda, Y. Nojima, A study on performance evaluation ability of a modified inverted generational distance indicator, in: Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation, 2015, pp. 695–702.

[106] E.-y. Cho, Jprofiler: Code coverage analysis tool for omp project, 2006, Technical Report: CMU 17–654 & 17, Tech. Rep.

[107] D. Szklarczyk, R. Kirsch, M. Koutrouli, K. Nastou, F. Mehryary, R. Hachilif, A.L. Gable, T. Fang, N.T. Doncheva, S. Pyysalo, et al., The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest, Nucleic Acids Res. 51 (D1) (2023) D638–D646.

[108] S. Principe, E. Zapater-Latorre, L. Arribas, E. Garcia-Miragall, J. Bagan, Salivary IL-8 as a putative predictive biomarker of radiotherapy response in head and neck cancer patients, Clin. Oral Investig. (2022) 1–12.

[109] N. Erlichman, T. Baram, T. Meshel, D. Morein, B. Da'adoosh, A. Ben-Baruch, Tumor cell-autonomous pro-metastatic activities of PD-L1 in human breast cancer are mediated by PD-L1-S283 and chemokine axes, Cancers 14 (4) (2022) 1042.

[110] M. Simonneau, E. Frouin, V. Huguier, C. Jermidi, J.F. Jégou, J. Godet, A. Barra, I. Paris, P. Levillain, S. Cordier-Dirikoc, et al., Oncostatin m is overexpressed in skin squamous-cell carcinoma and promotes tumor progression, Oncotarget 9 (92) (2018) 36457.

[111] R. Kanda, Y. Miyagawa, O. Wada-Hiraike, H. Hiraike, K. Nagasaka, E. Ryo, T. Fujii, Y. Osuga, T. Ayabe, Ulipristal acetate simultaneously provokes antiproliferative and proinflammatory responses in endometrial cancer cells, Heliyon 8 (1) (2022).

[112] 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).

[113] 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).

[114] 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.

[115] A. Masilamani, R. Ferrarese, E. Kling, N. Thudi, H. Kim, D. Scholtens, F. Dai, M. Hadler, T. Unterkircher, L. Platania, et al., KLF6 depletion promotes NF-κb signaling in glioblastoma, Oncogene 36 (25) (2017) 3562–3575.

[116] H.-C. Huang, B.-H. Cai, C.-S. Suen, H.-Y. Lee, M.-J. Hwang, F.-T. Liu, R. Kannagi, BGN/TLR4/NF-κB mediates epigenetic silencing of immunosuppressive siglec ligands in colon cancer cells, Cells 9 (2) (2020) 397.

[117] S.-L. Chau, J.H.-M. Tong, C. Chow, J.S.-H. Kwan, R.W.-M. Lung, L.-Y. Chung, E.K.-Y. Tin, S.S.-Y. Wong, A.H.-K. Cheung, R.W.-H. Lau, et al., Distinct molecular landscape of Epstein–Barr virus associated pulmonary lymphoepithelioma-like carcinoma revealed by genomic sequencing, Cancers 12 (8) (2020) 2065.