Artificial Intelligence In Medicine xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

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

Computational methods for Gene Regulatory Networks reconstruction and analysis: A review

Fernando M. Delgado, Francisco Gómez-Vela*

Division of Computer Science, Pablo de Olavide University, ES-41013 Seville, Spain

ARTICLE INFO

Keywords:
Gene Network
Systems biology
Networks validation
Gene Regulatory Network
Gene Network inference

ABSTRACT

In the recent years, the vast amount of genetic information generated by new-generation approaches, have led to the need of new data handling methods. The integrative analysis of diverse-nature gene information could provide a much-sought overview to study complex biological systems and processes. In this sense, Gene Regulatory Networks (GRN) arise as an increasingly-promising tool for the modelling and analysis of biological processes. This review is an attempt to summarize the state of the art in the field of GRNs. Essential points in the field are addressed, thereof: (a) the type of data used for network generation, (b) machine learning methods and tools used for network generation, (c) model optimization and (d) computational approaches used for network validation. This survey is intended to provide an overview of the subject for readers to improve their knowledge in the field of GRN for future research.

1. Background

The great amount and variety of gene expression information generated in the last few years, have led to the need for processing and interpreting such information. In this sense, Gene Networks (GNs) have become a key tool for the understanding and modelling of complex biological processes. The term Gene Network, also called, Gene Regulatory Network (GRN), is used to describe complicated functional pathways in a given cell or tissue, which represent living processes such as metabolism, gene regulation, transport mechanisms or signal transduction.

GRNs are models used to describe and predict dependencies between molecular entities [1]. These are composed of nodes, representing genes, proteins, metabolites or RNA; and edges, which represent molecular relations, e.g. protein–DNA, protein–protein interactions or other relationships of several kind [2] (see Fig. 1). In Fig. 2, a schematic representation of how the abstraction in the modelling process looks like, is shown.

GRNs are a ground-breaking tool for the discovery of new interactions between biological entities, helping scientists in research and making easier hypotheses formulation. They have been successfully used in diagnostics, as in the case of Liang et al. [3]. Many predicted interactions have been confirmed experimentally, which confirms GRN's reliability [4]. The inference of GRNs has also been proven to be relevant in the study of fundamental processes occurring in living organisms [5], ranging from development to nutrition and metabolic coordination. Multiple applications in fields such as human health or agronomy have been developed thanks to the implementation of GRN models. Moreover, GRNs make easier the manipulation, control and coordination of cell physiologic events related to GRN activity: diseases, biotechnological applications or crop production among others. For example in Yan et al. [6], advances on network reconstruction, analysis and interpretation of GRN reliably allowed the identification of molecular biomarkers for monitoring cancer progression and treatment. Also, GRNs have contributed to the representation of developmental processes, as they can generate developmental patterns [7].

Reverse Engineering deals with the process of network inference or GRN reconstruction out of experimental results. In particular, computational GRN inference process relies on the well-known Knowledge Database Discovery (KDD) workflow. KDD goes from input data preprocessing to the validation of generated models, often performed by data base search and comparison with prior experimental data (see Fig. 3).

The process starts with the input data. This usually consists of gene expression datasets, which can be obtained either experimentally or from databases such as NCBI GEO [8] (Step 1 of Fig. 3). After the input dataset is selected, it may be preprocessed by any computational method in order to improve the quality of the study, either because it does not affect the intended network, or because it shows bad quality (Step 2 of Fig. 3). Then, the preprocessed data is used as input for a

* Corresponding author.
E-mail addresses: fmdelcha@alu.upo.es (F.M. Delgado), fgomez@upo.es (F. Gómez-Vela).

https://doi.org/10.1016/j.artmed.2018.10.006
Received 21 June 2018; Received in revised form 23 October 2018; Accepted 23 October 2018

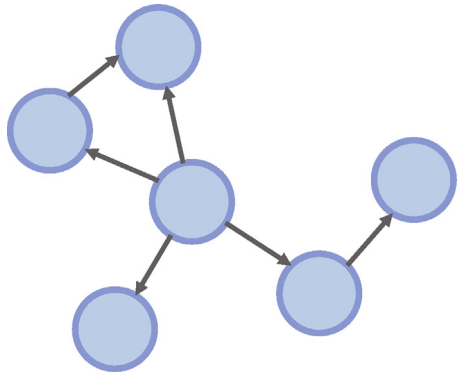

Fig. 1. An example of GRN topology, where nodes represent biological compounds and edges the relations between them.

computational inference algorithm, which provides the resulting network (Step 3 of Fig. 3). Used algorithms may be based on different machine learning approaches. Finally, the obtained model (network) is optimized and validated so true biological insights can be obtained from it, by a comparison with real biological knowledge (Steps 4 and 5 in Fig. 3).

In this work, we present a review of the whole process of GRN reconstruction, based on the KDD workflow (steps in Fig. 3), in an attempt to address the latest and most relevant advances in the field of computational-based GRNs reconstruction. Aiming to offer the most reliable and relevant review, works cited in this paper were classified according to the main taxonomies presented in the field, e.g. the works by Hecker et al. [9] or Dougherty [10] and selected from the main repositories such as Scopus, PubMed or Google Scholar. Besides, most relevant papers, based on citations, presented results and impact of the publication, were given additional priority.

The rest of the paper follows the description of each step of the GRN reconstruction workflow (KDD process). In Section 2, the kind of data and computational methods for data preprocessing used for GRN inference are described (Steps 1 and 2, Fig. 3). Section 3 introduces the main computational approaches for network inference (Step 3). Next, in Section 4, optimisation proposals for the inferred networks are described (Step 4). In Section 5, the different GRNs validation approaches are described (Step 5). Last, final considerations are shown in the Conclusions section.

2. Biological data: basic input for GRN inference

The rapid development of GRNs is linked to the increasing amount of high-throughput technologies, which provide with huge data sets to be managed. These constantly-updated technologies, like Next Generation Sequencing (NGS) [11], which has acquired significant quality, robustness and low noise during the last decade, allow a leader view into RNA and DNA samples. Then, sequencing has become a standard approach, since the Genome is often considered the cornerstone in the study of organisms [12].

Although GRNs have traditionally based on microarray technology [13], depending on the case, NGS may be more efficient than these as a primary source of expression data [14]. Novel techniques like RNA-Seq (RNA sequencing) [15], use NGS to reveal the presence and quantity of RNA in a biological sample at a given moment. RNA-Seq is then used to analyse the continuously-changing cell transcriptome and facilitates the quest for alternatively-spliced transcripts, post-transcriptional modifications, gene fusions, mutations/SNPs and a sort of changes in gene expression. Additionally, RNA-Seq is able to look at different RNA populations including small RNAs (like miRNAs), tRNAs, and ribosomal profiling, thus providing a quite complete overview of the cell state.

Another powerful application of NGS is the study of protein–DNA interactions by looking at protein binding sites in the Genome. This is especially useful in the case of transcription factors (TFs), for which Chromatin Immunoprecipitation (ChIP) techniques are used.

The integration of heterogeneous biological information, e.g. multiple *omics*, may enhance the capabilities of GRN inference. Prior to this points, one needs to know the basics on main *omics*, which are outlined in the following subsection. A general scheme of the process of GRN inference is shown on Fig. 4.

2.1. Omics and related technology

Although genes can be regulated at several levels of integration (transcription factors, co-factors, post-translational modifications of proteins, proteins and transcripts degradation or epigenetics among others), a key step is gene transcription [16]. Then, the choice of expression data for GRN reconstruction is often considered of preference. This is why many GRN approaches, termed *influential* GRNs, only consider transcript levels and try to establish direct or indirect relations between them. Ideally, a robust model, closest to the actual biological system, could be created by the integration of *omics* data sets (whole-genome data sets, transcriptomics, proteomics, interactomics, metabolomics, epigenomics or exomics among others) together with other previous biological knowledge.

In this subsection, the two main data sources for GRN reconstruction, Genome and Transcriptome, are addressed. Nevertheless, GRN inference is shifting towards the integration of heterogeneous data, and models become more complex and closer to reality.

2.1.1. The Genome

The term Genome refers to the collection of genes comprised in a biological system. In the past, these collections were limited to protein-coding genes, but the field has been extended to many other elements such as TF-binding regions, microRNAs, or evolutionarily-conserved regions [17]. In the case of TF, the goal is to detect potential links between these and differential gene expression in the cell/tissue.

At the primary archives level, the most important nucleotide sequence databases are: GenBank (USA) [18], EMBL (Europe) and Data Bank of Japan Center, DDBJ [19]. Also, ENSEMBL database reunites information related to mammals' genomes [20] and seeks for a centralized resource for geneticists, molecular biologists and other genomes researchers studying of our own species and other vertebrates and model organisms.

Furthermore, fields like Epigenetics (i.e. the study of influential factors out of classic genetics) enlarge our understanding of the Genome [21]. As an example, in Ramsey et al. [22], the available public data stored at The Cancer Genome Atlas, an important database for human cancer research, is used to construct the underlying GRN and to identify the key role of the RUNX-1 transcription factor in adenocarcinoma.

2.1.2. The Transcriptome

Functional Genomics or Transcriptomics refers to the analysis of gene expression patterns and tries to find relationships between them and their biological background [23]. Transcription is considered the main control mechanism in gene expression, so GRN reconstruction using expression levels is generally of preference [24].

Interestingly, a lot of RNA transcripts do not code for proteins (non-coding RNA, ncRNA), e.g. tRNA. On the other hand, ncRNA play a key role in multiple processes, including gene regulation. These transcripts are single-stranded RNA folded into structured molecule. Prediction of ncRNA (secondary) structure is possible as well as predicting where such ncRNA genes are located in the genome. Microarray-obtained Gene Expression data can be found at databases like Gene Expression Omnibus (GEO) [25] or ArrayExpress [26].

As an example of this, Kang et al. [27] used genome-wide expression data to generate a model able to predict acute rejection responses in kidney transplantation, using clinical trial data as an input.

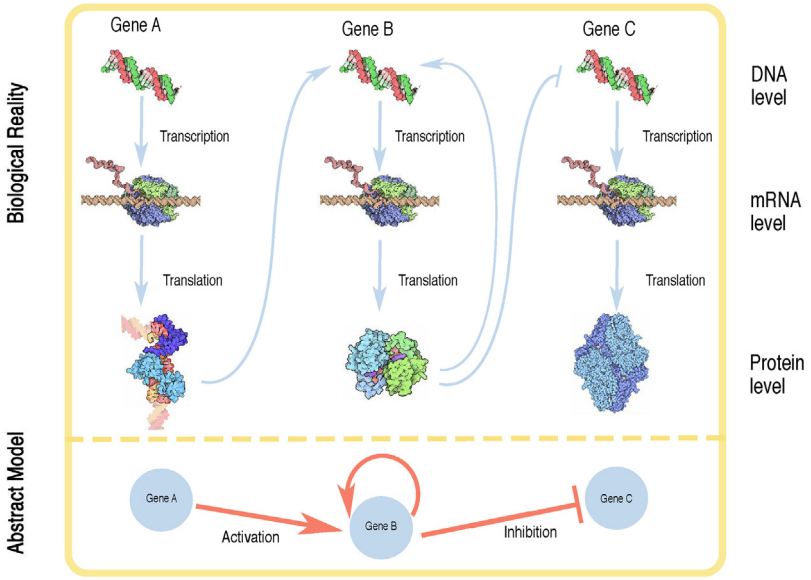

Fig. 2. The main goal in GRNs inference is to generate abstract models for actual biological processes. These models tries to represent complex interactions between molecular entities such as gene activation, inhibition or feedback loops.

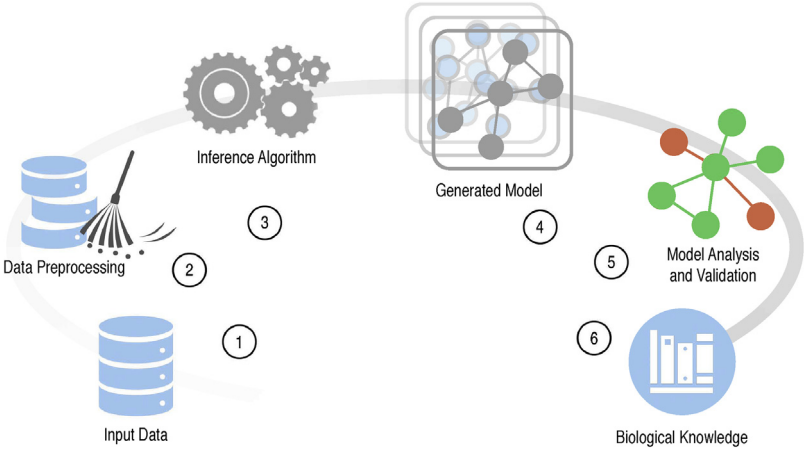

Fig. 3. GRN reconstruction steps based on the KDD workflow.

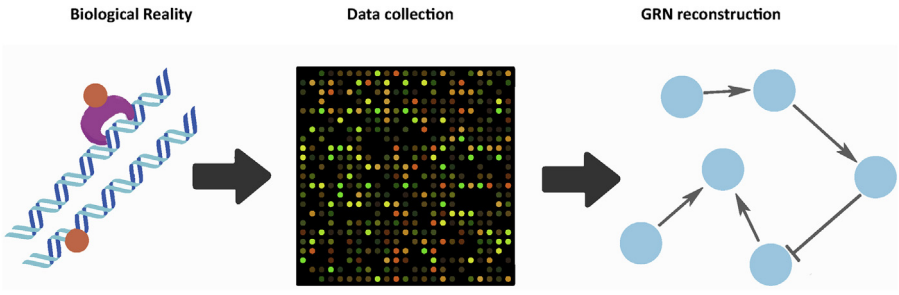

Fig. 4. A global representation of how the GRN are inferred from the biological data to abstract models.

2.1.3. Reading software for the collection of prior available knowledge

Aside from the huge amount of data stored in databases, there is knowledge generated experimentally which can also be integrated into the process of GRN reconstruction. Latest findings in biological data are directly incorporated to databases as mentioned above, some others are described in the literature. For this reason, software has been developed to automatically and reliably extract pieces of information about relationships between molecular elements from the literature [28].

This section does not intend to provide a complete and detailed description of the diversity of available biological data. Additionally, it illustrates the potential benefit as well as the challenge of integrating such diverse and complementary types of biological information to reliably infer GRNs. In the following subsections, some indications on experimental set-up for the obtaining of experimental data are addressed.

2.2. Experimental design to obtain biological data for GRN inference

On a frequent basis, gene expression experiments are performed for GRN reconstruction. Depending on the used approach, quality and quantity of the generated data may vary. Upon the performance of these experiments, two aspects must be taken into account: perturbation and observation of the biological system.

Experimental design includes usually systematic perturbations e.g. shift between different environmental conditions, interventions at the genetic, transcriptomic, proteomic or metabolomic level. As a result, differential expression patterns can be found under imposed conditions. Resultant gene expression patterns can be compared to a non-perturbed profile provided by a presumed model of the network. Thus, estimated *goodness* of the model can be then assessed [29].

Transcriptome level perturbations can also be used for GRN inference, for this, molecular techniques are often applied, e.g. RNA interference (iRNA) can knock-down other RNAs to their degradation in order to study the effect of their loss. Measuring altered gene expression levels provides some insight on the influence of the molecular elements involved, and it is key to model construction. For example, REDuction of UnCertain Edges (REDUCE) algorithm finds optimal gene knock-out experiment for inferring directed graphs of GRNs [30].

Measurements on perturbation experiments can be performed in a static (steady state) or time-course situation, the latter involves the use of dynamic programming. Depending on the knowledge to be achieved, the experimental set-up will vary and so will the choice between a static or a dynamic GRN architecture:

- Generation of static data comes with the assumption of an equilibrium or steady-state situation of the biological system. Depending on the case, the steady-state choice may miss critical dynamic events for reliable GRN construction i.e. dynamic changes occurring with time, as in Kim et al. [31].
- On the other hand, time-series experiments, where samples are taken in a series of time-points after perturbation, constitute the dynamic approach [32]. The experimental set-up determines the number of time-point measurements, thus, the data amount.

2.3. Data requirements

Reliable GRN reconstruction requires a considerable minimum quantity of accurate data. On the other hand, experimental costs and efforts also need to be minimized. Generally, the more nodes involved in the network, the more data will be needed.

In order to provide real biological insights, networks are conventionally based on experimental data. However, experimental data may not be useful for this aim for two main reasons: (i) The data collected from the experiment shows bad quality and (ii) data is unavailable.

Regarding the first case, model quality is proportional to biological data set quality. Alterations on the biological input (measurement noise or inappropriate experimental design) may lead to unreliable GRNs. Models could either estimate high-confidence gene regulatory interactions or just speculative dependencies, but depending on the complexity of the model, they may consist of many parameters and be more data-demanding. However, depending on the information one aims to obtain from the model, the precision required for the data may vary [9]. To cope with this, the inference strategy may use external prior knowledge from databases and literature, so experimental data required will depend on this knowledge and the ability of the used algorithm to integrate this information in the modelling process.

If data is not available, e.g. it is not possible or difficult to obtain, some approaches make use of fuzzy logic to infer missing data. In Bordon et al. [33], fuzzy logic is applied to assess quantitative values to an incomplete kinetic dataset for gene repression. Besides, GRN can still be reconstructed out of in silico generated data. For example, SiGNet (Signal Generator for Networks) is a Cytoscape app that simulates experimental data for a signalling network of known structure [34]. Despite not providing biologically-meaningful networks, this approach is often used for algorithm testing and training, so it can be put to work on an actual database afterwards.

Furthermore, model quality does not only depend on data quality, but also on the inference algorithm, which could vary the efficiency of GRN inference out of the same data set. It is necessary to find precise model parameters using heuristic approaches, which may perform sub-optimally.

As explained before, there is a direct relation between the complexity of the model and the required amount of data, but this certainly has a limit. Dimensionality problems come with difficulty of finding an accurate model (dynamic, large-scale, complex), when the size of the search space increases dramatically [35].

2.4. Data pre-processing

Data pre-processing, prior to GRN inference, is a key step for GRN reconstruction and quality of outcome. Methods for this aim will depend on the type of data and the experimental design.

As mentioned by Hecker et al. [9], there are two main sources of variability in GRN reconstruction: systematic errors (bias) and stochastic effects (noise). To ensure quality of the data and GRN outcome, fundamental analysis is applied, including noise filtering, system affect detection, etc. [36].

Systematic effects can be nearly removed through data normalization, since some genes expression can be very variable in one cell/tissue type. For this, housekeeping genes (relativity constant expression patterns) are often taken into account [37]. On the other hand, replicates performance provides with repeated measurements to reduce stochastic effects.

Finally, further data manipulation may be necessary, whose specificity will depend on the network inference methods e.g. dynamic programming requires the estimation of time derivatives for each measurement point of the time-series. Another example is the case of Boolean networks (see Section 3), which needs the conversion of measured expression levels into binary data prior to network inference. In the following section, the main algorithms for GRN reconstruction are explained.

3. Computational approaches for GRN construction

The term model architecture refers to the logical thinking underneath the GRN reconstruction. In the form of an algorithm, model architecture describes the behaviour of the regulatory dependencies between the biological components involved, basing on multiple other parameters.

The choice of a certain model architecture will shape the resulting GRN. Biological data has to be analysed so network structure (interactions between components) and model parameters (intensity and type of interaction) can be learned from it. Together with the increasing amount of biological data, the needs of analysis have led to the proposition of novel algorithms for GRN reconstruction.

In the following subsections, the main GRN inference methods are exposed, in an attempt to explain the basic thinking underneath each approach. Main GRN inference methods can be summarized in: (a) Information theory, (b) Boolean networks, (c) Differential equations models, (d) Bayesian and (e) Neural models. Also, a comparison between these methods is shown.

3.1. Information theory models

Information theory-based networks are the most common type of networks due to their computational simplicity. They are also called co-expression networks since they establish gene–gene relationships if the

dependence level between both gene is above a threshold set in advance. The higher the threshold, the sparser inferred GRN [38]. The main measures to determine the dependencies between genes are the correlation coefficients like Pearson, Spearman or Kendall coefficients. However, different measures like Euclidean distances or mutual information, were also applied for the inference of GRN. These models are suitable to cover different aspects of cells, which are here understood as time-varying living systems which perform complicated processes inside and between them.

An advantage of information theory models is the discovery of large GRN from low expression data due to their low computational cost. Among the main proposals based on this model for GRN reconstruction we can highlight: REVEAL (The REVerse Engineering ALgorithm) [39]; RELEVANCE [40]; ARACNE (Algorithm for the Reverse engineering of Accurate Cellular Network) [41] or ARACNE-based algorithms [42]; CLR (Context Likelihood of Relatedness) [43]; MRNET (based on the maximum relevance/minimum redundancy) [44] or FyNE [45]. These approaches apply correlation coefficients, Euclidean distances or information theory scores such as mutual information and conditional mutual information for the identification of gene interactions. Also, supervised-DTI and non-supervised-DTI approaches, based on the directed information (DTI) metric, are used [46]. Other algorithms use the minimum description length (best explanation of the data gathered with limited number of observations) to determine a threshold value [47].

Algorithms often undergo an improvement process for a more precise data analysis. This is how novel algorithms work, as: MI3 [48], based on three-way mutual information; SRI-CLR [49] adding the synergy index to CLR algorithms; MRNETB [50] combination of backward selection and sequential search; C3NET [51] which selects the highest mutual information value among the neighbors of each gene; or CMIP [52], which applies conditional mutual information from the lower order to the higher order.

Analogously, graphical Gaussian models (GGMs) are used to represent conditional dependencies between variables and allow distinguishing direct form indirect associations [53]. Also, mutual information models infer regulatory gene interactions basing on pairwise mutual information [54]. Two algorithms commonly used for network inference based on pairwise mutual information are ARACNe [55] and CLR [49].

Overall, main advantages of these models are the simplicity, low computational complexity and low number of required samples [56]. Therefore, this kind of network fits perfectly with the construction of large gene networks. On the other hand, these models are static and do not take into account multiple genes participating in a regulation. As an example, in Wang et al. [57], a total of 14 genes were identified as hubs/nodes in the regulation of postmenopausal osteoporosis disease, by means of an information theory-based approach.

3.2. Boolean networks

Boolean networks are easy to implement and allow capturing the actual dynamical behaviour of GRN. They are able to describe biological phenomena such as oscillations, multi-stationary events, long-range correlations, switch-like behaviour stability and hysteresis [58].

Boolean networks represent genes by variables and their expression level is discretized into Boolean binary values (by clustering and thresholding [59]): '0' for low values (silenced or nearly-silenced genes) or '1' for high values (activated genes) [29]. Thus, a gene i is represented by *on/off* expression values. Operators of logic (and, $\wedge$; or, $\vee$; and not, $\neg$), link Boolean variables in a Boolean function.

Boolean functions reconstructing the network compose directed graphs $G(X, F)$, where X represents a variable, which is associated to other variables by means of the function F [58]. The so-called state of the network (S) for a time t is represented by the values of all different nodes:

$$S(t) = \mathrm{xi}_1(t), \mathrm{xi}_2(t), \ldots \mathrm{xi}_n(t)$$

Although straightforward and simple, Boolean networks find their main limitation in the discretization step. Gene expression is rarely a matter of fully-activation or fully-silencing, since there are often uncountable different gene states in between. Thus, important details of system behaviour might be lost. Boolean networks also need to cope with noisy data problems [60], as the accuracy of thresholding will certainly determine network topology. Nevertheless, Boolean networks are easy to interpret and they offer a simple dynamic approach for GRN.

Considered as the simplest GRN inference approach, Boolean networks have been proven useful in many cases. In Simak et al. [61], regulatory relations and potential candidate biological functions are explored from *Saccharomyces cerevisiae* transcriptomic data. Boolean network function performance was validated using *Saccharomyces cerevisiae* time course data, and its outcome was consistent with all different cell cycle stages. Also, in Claussen et al. [62] a dynamic Boolean network model is used to infer interactions between low-abundance species in human gut microbiome, which are often overlooked. The model uncovered synergistic and competitive interactions between these species. In Polak et al. [63], regulation of immune responses in primary Langerhans cells were analysed using a Boolean GRN model, whose predictions were experimentally validated afterwards. In Moignard et al. [64], a GRN model for blood development was reconstructed and used to predict the role of some transcription factors, later validated experimentally. Finally in Orlando et al. [65], a Boolean network is used to model yeast cell cycle. This model works as an oscillator reflecting a sequenced transcriptional program.

3.3. Differential equations models

Ordinary differential equation (ODE) approaches use continuous instead of discrete variables. This leads to a more accurate model, and enables the dynamic modelling of gene regulation. Differential equations will then represent changes in gene expression as a function of other genes expression and taking into account environmental factors, allowing a quantitative modelling (closer to actual behaviour of the biological system) [58].

Modelling of gene expression dynamics through ODE often can be represented by the equation:

$$\frac{\mathrm{dx}_i}{\mathrm{dt}} = f_i(x_1, x_2, \ldots, x_N, p, u)$$

where x represents the expression level of a certain gene i at a given time t. N represents the number of genes involved, u refers to external perturbations of the system and p to the parameter set of that system. Then, f is the function, which describes the change rate of the state variables x_{1-n} depending on p. In ODE models, continuous-time variables with constraints are used, and negative values are not allowed, i.e. the assumption of protein and mRNA molecules degradation being unregulated [58].

There are multiple solutions to ODE systems if no constraints are assessed. This is why, specifications of the f function and constraints representing prior knowledge (simplification, approximations or educated guesses among others) are required for the identification of model structure and parameters. A disadvantage of many ODE models is that these consider only linear models or just specific types of non-linear functions [66,9], while regulatory processes are often characterized by complex non-linear dynamics. More complex variants of ODE-based models are stochastic differential equations models which take into account the stochastic nature of GRNs [67]. Moreover, ODE models cannot cope with large GRN modelling and value estimation for model parameters results hard in some cases due to their computational complexity.

ODE-based models have provided outstanding results in GRN inference. As an example, in Matsumoto et al. [68], SCODE algorithm

(Single Cell Ordinary Differential Equation) was applied to provide some insight in GRNs related to differentiation processes. Single cell RNA-Seq is performed on individual cells so differences on their expression patterns can be evaluated. Deng et al. [69], proved the efficiency of ODE models in dynamic GRNs reconstruction. Improvement on calculation scheme of the derivatives and data pre-filtration lead to improved scalability to large GRNs. However, combinations between different methods have been proven to be useful depending on the case.

3.4. Bayesian networks

Bayesian networks are one of the most used GRN inference architectures. They make use of the Bayes theorem of probability, then combining probability and graph theory to qualitatively model the properties of GRNs [58].

Bayesian networks are generally directed and acyclic graphs (DAGs) $G = (X, A)$ [70], which characterize the joint distribution of nodes as a series of local probability distributions (P). X represents genes/nodes ($x_1, x_2, \ldots, x_n$), or gene variables and A refers to the directed rod corresponding to probabilistic dependency interactions between these genes. The joint distribution of the variables in a Bayesian network is described in Chai et al. [58] by:

$$P(x_1, x_2, x_n) = \prod_{i=1}^{n} P(x_i|\text{parents}(x_i))t$$

where x_i is a node, n refers to the total number of involved genes, Parents (x_i) are all *parent* genes regulating the *child node* (gene x_i). Parameter P describes the Conditional Probability Distribution (CPD) or local distribution for the node i. Bayesian networks make use of the Markov assumption [71], which refers to the memoryless property of a stochastic process: given its parents, each node is independent of its non-descendants.

Bayesian networks imply a set of conditional dependencies so the algorithm is able to infer a DAG from them. Since many DAGs can be inferred from the data set (D), the algorithm has to find the best DAG (G) describing the data set and each graph (G) is evaluated through a Bayesian score.

$$S(G: D) = \log P(G|D) + \log P(G) + \text{constant}$$

In Larjo et al. [72], methods for learning Bayesian networks are detailed, having this learning three essential parts: (i) *Model Selection*, DAGs evaluation as candidate graphs of relationships; (ii) *Parameter Fitting*: given graphs and experimental data sets, find the best conditional probabilities for each node; (iii) *Fitness Rating*: scoring of each candidate model so the higher the score, the better the model fits to the data. The model with the highest score represents the GRN resulting from the inference. Learning of Bayesian networks can be based either on discrete (often Boolean) or continuous expression levels. Thus, the probabilistic model underneath could be a multinomial or a Gaussian distribution [73]. Multinomial variables take a finite number of possible values but in the case of continuous variables, in a classical approach, data would be discretized. Bayesian networks are hard to infer from continuous data since it requires large computational power. On the other hand, data does not require discretization.

Flexibility is the main advantage for Bayesian networks, since they combine different types of data as well as prior knowledge for reliable GRN inference [74]. Bayesian approaches are often considered of preference while inferring dynamic GRNs and have shed some light in many fields ranging from evolutionary-development to medicine. In Acerbi et al. [75], continuous time Bayesian networks were used for the successful revelation of well-known regulatory mechanisms in Th17 cells differentiation, also providing some new biological insights. Chekouo et al. [76], proposed a Bayesian model to identify micro RNAs and their target genes obtaining a valid algorithm for kidney cancer biomarkers identification. Also in Chudasama et al. [77], novel cancer biomarkers were identified thanks to Bayesian networks, which may support clinical practice and improve long term outcomes.

3.5. Neural networks

Inspired by animal central nervous systems, these models comprise two main approaches: Artificial Neural Network (ANN) and Recurrent Neural Networks (RNN). The first is purely neural, whereas the second also involves fuzzy logic [78]. RNN is a successful method for GRN inference, since it enables modelling of non-linear and dynamic interaction among genes [79]. Neural models allow continuous variables and their outcome looks similar to the neural connections observed in natural processes.

As described in Chai et al. [58], neural network models can be represented by:

$$\frac{\mathrm{d}e_i}{\mathrm{d}t} = \frac{1}{\tau_i}\left(g\left(\sum_{j=1}^{N} w_{ij}e_j + \beta_i\right) - \lambda_i e_i\right)$$

where w_{ij} refers to the type and concentration of the relation between genes at the ith and jth position. The reaction decay rate parameter is represented by λ_i, the basal expression level is indicated by β_i and e_i represents the gene expression level for the ith gene. Function g indicates the regulatory effect on each involved gene, defined by a set of weights such as w_{ij}. The computed weighted sum of all potential regulating genes is considered as the regulatory effect on a particular gene. A scoring function is also applied for evaluation of the outcomes, network performance optimization and error minimization.

In Ling et al. [80], a RNN was used to study p53/Mdm2-mediated response to DNA damage. Tong et al. [81], used ANNs to study gene–gene interactions for biomarkers in childhood sarcomas. Siddens et al. [82], used fuzzy neural network models to predict polycyclic aromatic hydrocarbon-mediated perturbations of the Cyp1b1 transcriptional regulatory network in mouse skin. Finally, in Rubiolo et al. [83], a supervised neural model called Extreme Learning Machine (ELM) was successfully used to reconstruct GRNs, even surpassing many commonly-used approaches.

3.6. Additional network architectures

Although GRN reconstruction is usually tackled by means of one of the models above (or a combination of them), there are some other GRN inference approaches that clearly differ from these methods. As an example of this, in Liu et al. [85], fuzzy cognitive maps and a dynamical multi-agent genetic algorithm with the decomposition-based model were used to deal with large-scale GRN inference. Also in Thiagarajan et al. [90], the power of Graphics Processing Units and parallel reverse engineering algorithm was used for the identification and simulation of genome-scale GRN, which pose computer intensive problems. Finally in Ud-Dean et al. [86], an algorithm called TRaCE+ is used for GRN reconstruction, being able to indicate positive or negative regulations, and reduce uncertain edges.

To sum up for this section, a comparison between all 5 exposed inference methods is shown on Table 1.

In Section 4, some important issues on the application of the inference algorithm are explained such as: feature selection, parameter estimation, structure optimization and integration of prior knowledge.

4. Network optimization based on machine learning algorithms

A naïve approach upon the GRN inference process would be the one of enumerating all possible DAGs for a given number of nodes, which is deemed brute-force search. However, the amount of possible DAGs for a given number of nodes grows exponentially, making this search problematic. Therefore, heuristics or/and constraints need to be applied to make the process more efficient.

Table 1
Comparison between different model architectures with some examples.

Computational approach	Strengths	Weaknesses
Information-theory models	• Large GRNs, even out of low expression genes • Mutual and conditional mutual information approach • Not computationally-demanding • Low number of samples	• Regulation by multiple genes is not considered • Static, only suitable for steady-state data
Examples:	REVEAL [36], RELEVANCE [40,84], ARACNE [42], CLR [43], MRNET [44]	
Boolean models	• Capable of inferring large networks • Generally easy to interpret • Simplify underlying complex biological phenomena • Allow supervised learning methods	• Deterministic nature • Discretization bottleneck (only on/off states) • Problems in handling incomplete or inconsistent expression data • High computing time • Most of them use small number of genes
Examples:	RCGA [85], TRaCE+ [86], CABeRNET [87]	
ODE models	• Directed signed graphs • Realistic dynamics • Suitable for both steady-state and time series expression data • Simplification of the system by means of linear functions • Allow prediction of the behaviour of the network under different conditions once parameters are known	• Not suitable for large networks • Linear functions also constrain the dynamic behaviour of cell regulatory functions (e.g. oscillations, multistationarity) • Hard to find appropriate values for model parameters • Noisy data leads to qualitative instead of quantitative GRN inference
Examples:	SCODE [68], HiDi [69]	
Bayesian models	• Noise and uncertainty handling • Do not require a large number of involved variables • Integration of prior knowledge and allowance of enrichment analyses • Statistical inference of gene network	• Feedback loops are not allowed • Fail in the inference using time series expression data • Cannot cope with large GRNs • Inherent combinatorial learning
Examples:	F-MAP [88], MDP [89], POMDP [71], QMR-DT [73]	
Neural models	• Recognize an input pattern • Model any functional relationship inferable from the data • Suitable for both steady-state and time series expression profiles • Noise handling and biologically plausible • Manage non-linear and dynamic behaviour	• Machine training experiments are hard to perform since every situation requires a different learning rate definition • Computational complexity makes them more suitable for very small systems
Examples:	ANN [27], RNN [78], ELM [83]	

Leaving the model algorithm aside, the inference approach needs further adjustment for reliable network reconstruction. First, a dimensionality reduction has to be performed on the large gene expression data in order to reduce computational cost [91,92]. The term *curse of dimensionality* was introduced by Bellman [93] to refer to the exponential growth of required amount of data to provide a reliable analysis. Some strategies to face this issue are:

- Reducing the number of nodes focusing only on features like genes or proteins of interest employing feature selection or feature mapping methods.
- Restricting the number of model parameters: using simple models and network connectivity constraints such as sparseness. This minimizes the number of edges in the network thus reducing the number of model parameters.
- Integrating specific prior knowledge about network structure which increases the model and provides higher quality results.

4.1. Feature selection and feature mapping

Modelling biological system requires assumption-making to focus only on those specific aspects which are important for aim of the study. Feature selection and feature mapping help reducing model complexity by excluding non-relevant features in GRN inference [58].

Upon feature selection, non-responsive or not well measured genes are removed from the data. Machine learning algorithms are often affected by data noise, which should also be reduced to avoid unnecessary model complexity. On the other hand, in feature mapping redundant information is removed. For this, different molecular entities in a functional element are combined when they represent the common behaviour that reflects a particular biological function, thus clustering nodes and reducing complexity [94].

Dimensionality has to be decreased using an appropriate approach so the network is still large enough to provide a significant result on the biological phenomena under study. Some of the approaches consist in the filtering of differentially expressed genes and the clustering those genes that are co-expressed. Differentially expressed genes, are often the only ones considered in some experiments, and are identified thanks to methods like DESeq2 (Love et al. [95]) or Limma (Ritchie et al. [96]). Moreover, these techniques try to reduce the number of model variables.

Feature selection seeks for the minimization of the number of estimated parameters in order to improve performance and generalizability of GRNs, solely using the data to deduce dependencies [94]. A general approach is to consider only genes showing significant expression changes under the studied experimental conditions.

According to Hira and Gillies [91], there are three main categories of feature selection strategies to be considered: (i) *filters*, which extract features from the data (without any involved learning); (ii) *wrappers*, which select useful features through learning techniques; and (iii) *embedded techniques*, combining feature selection and classifier construction. Clustering of co-expressed and/or co-regulated genes or proteins has been conventionally used for the reduction of network components. These are ultimately genes or proteins showing similar expression patterns or genes belonging to the same pathway or biological function.

As stated before, experimental gene expression data is often complemented with prior knowledge obtained from biological databases. In this enrichment analysis, genes are selected and compared with previously-determined functional gene groups. Enrichment analyses normally make use of GO terms, Gene Ontology Annotations (GOA) and databases like KEGG Pathways, OMIM and Gene Prospector [97].

Alternatively, Knowledge Driven Variable Selection (KDVS), performs discovery methods and enrichment analyses at the same time, thus enhancing results interpretability. An example of knowledge-driven enrichment analysis can be found in Sun et al. [98], where an enrichment analysis of a gene/protein network for primary

myelofibrosis is performed, basing on gene expression using KEGG pathways and OMIM database.

4.2. Biological network parameters

GRN structure has been proven to be sparse, and following a scale-free topology [99]. The latter means that nodal degree distribution of the GRN is a power law distribution.

Model structure and model parameters estimation are two main tasks of network inference. The idea behind these tasks is to apply a learning algorithm to fit the output of the mathematical model to the provided experimental data [29]. The choice of a network topology is another important step in GRN reconstruction. Herein, Radcliffe [100] studied the application of form analyses to GRN topology related problems.

Structure optimization is thus the process which finds the network topology (connectivity) that best explains the experimental data, taking into account constraints imposed by the available knowledge. A key point here is network sparseness. On the other hand, parameter optimization seeks for the identification of the best model parameters for a given model structure. As detailed in Valverde et al. [99], several GRN properties have to be taken into account for network reconstruction:

1 *Sparseness*: a gene is usually regulated by a small and limited number of genes. Then, regulatory inputs per node are limited. Exceptionally, the so-called master genes control large parts of the network (high out-degree) [16,101]. Generally, as network connectivity increases, the data will be better fitted by the model. Nevertheless, higher-connected networks bring several difficulties. Thus, there is a compromise between model quality and model complexity. Sparseness helps finding the most likely combination of regulators [102].
2 *Scale-freeness*: the distribution of node degrees in a GRN tends to have the form of a power law. In those scale-free networks most genes are sparsely connected whereas a few are highly connected.
3 *Highly-structured*: networks can be decomposed in modular components that consist of only a few genes, which follow regular hierarchies.
4 *Modularity*: this means, when an ensemble of genes cooperates within a same specific function [103]. This can be observed by clustering co-expressed genes.

4.3. Scoring functions for structure optimization

Explicit structure optimization methods compare different topologies of GRN models by means of a scoring function, which helps achieving network sparseness. Several scoring criteria have been developed for the different inference methods [73]. Gene interactions are added or removed in order to obtain a better-scored topology.

Brute-force search, where all possible combinations of interactions are tested, is only possible for small networks, on which strong restrictions are also applied. The total number of combinations for each node can be estimated as $2^n - 1$, where n is the number of involved nodes. A different optimization strategy has to be applied if this is not the case.

Heuristic methods apply educated guesses to lead the search to the most likely solution [78]. Some examples of search techniques are: breadth first search [104], beam search [105] and hill-climbing [106]. Search techniques add or remove connections in the network. The three main search techniques are: *Forward Selection* (Growing), which starts from a simple model and most important interactions are added first up to a certain limit; *Backward Elimination* (pruning), which starts from a highly connected model and less significant interactions are removed; and *Stepwise Selection*, which combines both previous approaches. As an example, in Gómez-Vela et al. [107], a greedy approach is used for structure optimization of densely-connected networks.

Finally, for a given modelling architecture and a network structure optimization strategy, one can infer the model from the data. Besides, the optimization strategy may be supported by prior knowledge (see Section 4.4).

4.4. Integration of diverse biological information

Modelling of GRNs has to be comprehensive and integrative, thus, as repeated throughout this review, the use of prior available knowledge (other experiments, databases, scientific literature, etc.) is a key step to reduce combinatorial complexity [14].

Knowledge-based modelling approaches are the most robust for realistic GRN inference [14]. Prior biological knowledge hampers biologically plausible assumptions, supporting the reverse engineering process. For example, many inference algorithms make use of the publicly-available data stored at GO and/or KEGG databases, as in the case of Zhu et al. [108].

Biological knowledge can be integrated into mathematical modelling and taken into account for network reconstruction, thus providing accuracy to the process. Methods integrating information from ChIP, microarrays or RNA-Seq experiments are considered of preference. There are ChIP-Seq databases from specific cell types providing with information about transcription factors or epigenetic and transcriptional landscapes. DNA data sets providing with DNA structures, sequence conservation and patterns can also be integrated in GRN inference. Most biological processes involve combinatorial contributions of transcriptional regulation, alternative splicing, post-translational modifications or protein-protein interactions. The Biological General Repository for Interaction Data sets (BioGRID) is a database of reference aiming to annotate protein, genetic and chemical interactions for all model organism species and humans [109].

However, this is a challenging process and the comprehensive integration of the available heterogeneous data may turn into a complex horizon. One of the attempts for this aim is The Cancer Genome Atlas (TCGA), which catalogues and discover major cancer-causing genomic alterations [110]. This platform is based on microarrays and NGS methods such as: RNAseq, MicroRNAseq, DNAseq, SNP-based platforms, array-based DNA methylation sequencing or reverse-phase protein array. As expected, for the integrated-multi-dimensional data analysis, comprehensive exploration is required, for which promising tools are arising. In Pineda et al. [111], several omics data are merged in an integrative analysis for a more comprehensive study of bladder cancer, by using a Global-LASSO. Also in the work by Salehzadeh-Yazdi et al. [112], metabolomic and epigenetic data are integrated in GRN modelling. Finally, in the work by Sinha [113] an improvement on inference accuracy is shown when integrating epigenetic information on a Bayesian model.

Integrative learning approaches start from a template network built out of databases and literature (real network topology) and then, an inference strategy is applied fitting the model to the data and taking into account the template [29]. The template information also known as training data can also be incorporated into the GRN inference process, and it can also be used to constrain explicit search methods [37].

Learning models use training data sets in order to build generalizable models. However, an important parameter to take into account is the data set shift, since most algorithms assume training and test data to be drawn from the same distribution. Data set shift may result in suboptimal fitting of the model. Data normalization and batch correction techniques help to cope with data set shifts. To make sure that two data sets are drawn from the same distribution, quantile normalization is applied, which normalizes target distributions to a reference distribution [37].

5. GRN validation and appraisal of inference methods

Once the final network is obtained, its biological significance has to

be tested. Not all GRN-predicted interactions are biologically meaningful. In general terms, an interaction between a gene and its presumed regulator is considered biologically meaningful when the disruption of one or both of these elements triggers a change on gene expression. The identification of meaningful interactions requires network validation, so further support can be obtained for them [114]. Also the other way round, models should be predictive, meaning this able to generate plausible biological interactions which may be afterwards proved right [10]. However, lack of validation does not necessarily mean 'not-biologically-meaningful' interactions, since the validation methodology plays a crucial role and/or many interactions may have not been described yet.

According to Dougherty and Qian [115], there are two main issues regarding network validation: (i) whether the inferred network provides good predictions on the experimental data (scientific validation) and (ii) whether the applied inference algorithm within a certain network model framework yields networks that are accurate relative to some criterion of goodness (inference validation). The boundaries between both approaches actually blur in practice, since validation of an inference model requires then scientific validation of the inference, and results of the later may be used to improve the inference method [10].

Network validation assesses the quality of the inferred GRN supporting on available knowledge. For this aim, scoring methodologies are often applied to obtain a quantitative evaluation of the model with respect to the information used for its generation (internal validation) and other information (external validation). Also, web-based frameworks have been developed to support GRN validation, as in the case of Genotet [116].

5.1. Quantitative evaluation of inference performance

GRNs can be evaluated using scoring methodologies which allow the comparison between different networks. The process is described in Fig. 5, where the inferred network is compared with a reference network (Gold-Standard, closest to reality to the general knowledge) obtaining a quality measure [117,118]. A Gold Standard enables the estimation of several metrics which would jointly provide an evaluation of model's goodness. This is certainly a key point in GRN inference, since data may provide a massive amount of possible interaction and only a few of them are deemed true [119].

GRN inference relies on the data set handling method, which often cannot depict all interactions. For example, ChIP techniques may cause only the strongest (or most represented) interactions to be considered in the model, leaving many (less represented) others behind. Aside from the experimental techniques limitation, relationships may be missed due to high thresholding, but also due to low expression levels or non-variant behaviour with respect to target genes. Thus, many equally-biologically-meaningful interactions are missed, false negatives (FN) [114].

Conversely, false positives (FP) are also incorporated into GRNs. FPs are deemed technical when the inferred interaction is just sporadic in nature and it is not retrieved even with the same assay and conditions. Technical FPs, depend on assay robustness so the used approach has to be optimized. FPs are deemed biological in case of not biologically meaningful interactions, which are robustly detected. This can be the case of both a TF and its target being regulated by another TF which does not change its expression. Sub-optimal specificity of antibodies in ChIP experiments, which results in binding to lower affinity or non-specific sites, may also result in a FPs gain.

Finally, true positives (TP) and true negatives (TN) are described according to previous knowledge, since these are only considered when a particular interaction has been proved experimentally.

In order to evaluate model quality, it is necessary to analyse if the model correctly predicts the GRN behaviour or if the model represents the true structure of the system. Statistical measurements are used to compare inferred models with the actual behaviour of the network [120]. This is the case of supervised network inference, where part of the actual network is used for model training and optimization.

According to Schrynemackers et al. [121], when the true Gold Standard is known, the inferred networks structure is compared to the first one using several metrics:

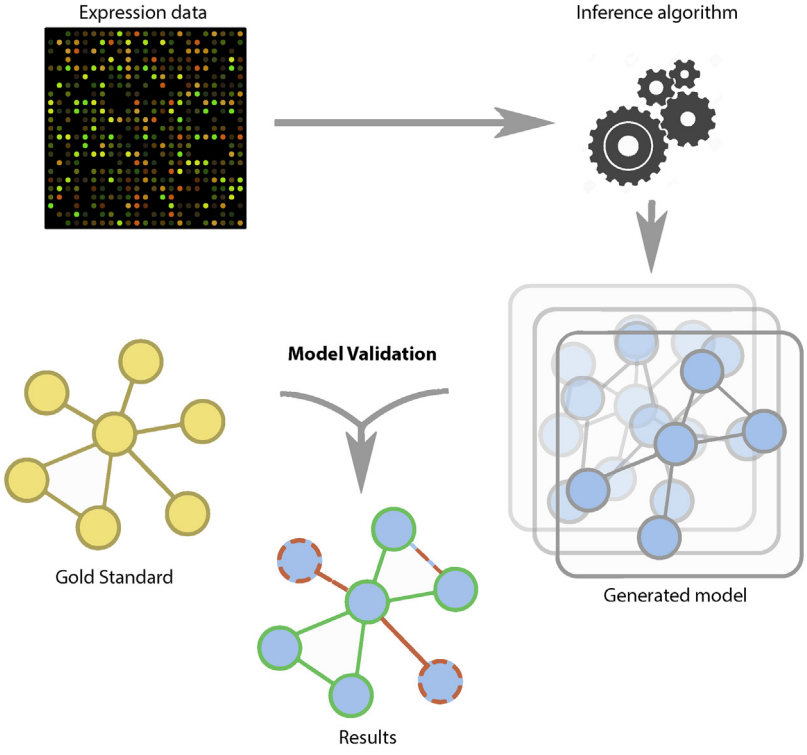

Fig. 5. Validation process scheme. This process usually offer a quality measure of the validated network.

- *True positive rate (TPR)*, *sensitivity* or *recall* is the number of TP divided by the actual number of positives: TP/(TP + FN).
- *True negative rate (TNR)* or *specificity* is the number of true negatives (TNs) divided by the actual number of negatives: TN/(TN + FP).
- *False positive rate (FPR)* also deemed as 1 − *specificity* is the number of FP divided by the number of actual negatives: FP/(FP + TN).
- *False negative rate (FNR)* or *miss* is FN divided by the number of actual negatives: FN/(FN + TP).
- *Precision* is the number of TP divided by the number of predicted positives: TP/(TP + FP).
- *Rate of positive predictions (RPP)* is the number of predicted positives divided by the total number of negatives and positives: (FP + TP)/(N + P).
- *F-score* is a measure of model's accuracy, and it is calculated as the harmonic mean of precision and recall:

$$F = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

These metrics are combined in the analysis of the inference performance. For this aim, several curves are displayed:

- *Receiver operating characteristic (ROC) curves:* plot TPR by means of the FPR, when the confidence threshold is varied from maximum to minimum confidence score [122]. For network comparison, ROC curves are integrated single real number by measuring the area under the ROC curve (AUROC). AUROC directly compares the inference quality against a random prediction. An AUROC of 1 corresponds to a perfect classifier and an AUROC of 0.5 corresponds to a random classifier. The higher the AUROC, the better the predictions. It is important to achieve low FPR for better precision [120,123].
- *Precision-Recall (PR) curves:* represent precision by means of the recall, when the confidence threshold is varied. An ideal classifier would provide a curve which passing trough (1,1), whereas a random classifier whereas a random classifier would rather provide a plateau [124].
  Analogously to ROC-curves, PR-curves are summarized in a single real number estimated by integrating the area under the curve (AUPR), which represents the average precision of the inference algorithm. The higher the AUPR the better the classifier is [125].

An advantage of ROC curves is that they do not depend on the ration between positives and negatives, whereas PR-curves do. On the other hand, PR curves are better to assess the method performance whilst the P/N ratio is close to the expected ratio when practically applying the model [126].

ROC and PR curves are both used to compare the performance of different inference algorithms, as in the case of De Smet and Marchal [119], Prill et al. [127] or Hase et al. [128].

It is worth mentioning that the presented scoring methodology is used mainly in Synthetic/Biological data-based network validation, which is described in the following subsections.

5.2. GRN inference algorithms performance evaluation

When a new GRN inference method is released, it is normally compared with the main former ones. Thus, GRN inference methods comparison and evaluation are required to find outperforming methods.

Former methods may include well-known, which have been proven useful at many applications. E.g., although outdated, ARACNE is still today often considered a method of reference.

Nevertheless, comparative studies are delicate and their results may be misleading. As an example, input data may vary between them (time series, large microarray or subset of gene expression levels datasets among others), thus impeding a standardized comparison methodology.

However, important contributions have been made in the literature in order to provide a reliable comparison between different inference methods. For example, in Bellot et al. [129], a good part of the above GRN inference methods are compared. In this work the authors present a software (a Bioconductor package called netbenchmark) that is able to carry out comparative between different inference methods. Despite this, they found that no single method is the best across different sources of data as mentioned above.

Also in Chen and Mar [130], a comparison between several GRN inference methods is performed demonstrating these methods, which were originally designed for bulk samples, do not suit biological relating when dealing with single cell expression data. Finally in Marbach et al. [118], the performance of different inference approaches was also analysed, revealing differences between model architectures depending on the case.

There are also worth to mention the DREAM (Dialogue for Reverse Engineering Assessments and Methods) challenges 3–5. In Marbach et al. [131], the authors present a method for the realistic performance assessment of GRN inference methods. Also in Prill et al. [127], an evaluation of GRN inference methods performance is carried out highlighting best strategies. Another interesting result is presented in [132] where the authors present a method and tool for generating in silico benchmark and performance profiling of GRN inference methods called GeneNetWeaver. Finally in Marbach et al. [118], it is shown that single GRN inference methods perform sub-optimally across multiple data sets, when compared to the integration multiple GRN inference methods.

5.3. Synthetic data validation

Some techniques consider the generation of synthetic data, also used to analyse GRN inference algorithm performance like the afore mentioned GeneNetWeaver. Gene expression values are simulated and then used as input for the GRN inference algorithm. The performance of the algorithm is tested by comparing the output network with the actual network inferred from the literature. However, this approach has some drawbacks since it cannot be used to assess biological significance, which requires making use of actual prior knowledge. Novel multimodel approaches integrate experiments based on both real-life biological and synthetic data sets which provide a higher precision for the inference [71]. Some examples of these tools are RegNet [133] or SynTReN [134] are used for this aim.

5.4. Biological data validation

Model predictions are subjected to external information that was not used in the modelling process, which can be found in the literature and databases. External validation often employs text-mining approaches and it is used to compare different network inference methods. Some tools like GeneNetVal [135], GNC [136], GFD-Net [137] or RefNet Builder [138] use the biological information stored in databases like KEGG Biogrid or GO to assess the model goodness through a direct comparison between the computed network with a gold-standard.

However, GRN predictive capability, can be influenced by experimentally-found interactions that are retrieved [114]. Frequently, several interactions would not have been described yet. Thus, experimental approaches for verification should be taken into account, since these have some limitations and scientists need to design a method for a particular interaction to be verified [10].

6. Conclusions

As discussed throughout this manuscript, GRN inference basing on

large-scale data is a major challenge for systems biology which has gained relevance over the recent past years. More complete biological insight will be gained in the upcoming years, when new techniques are developed for the integration of complex omics such as the epigenome or microbiome, which are trends of current research.

There is a huge variety of approaches, inference methods and evaluation metrics for reliable GRN reconstruction. Even if modelling architectures rely on different mathematical formalisms, they all provide similar networks which require some simplifications. Notably, usefulness of GRN reconstruction depends on both its application and the available data. GRN inference methods have their own advantages and disadvantages depending on the available data and the purpose of the inference. Many efforts have been made for their comparison by using these algorithms to infer a GRN from a single data set and then assessing their validity. These comparisons require appropriate evaluation methods to satisfactorily determine algorithm performance.

The curse of dimensionality still makes suffer inference methods dealing with large data sets, thus novel and more efficient algorithms that are highly-scalable are still required. Dimensionality problems usually come with the integration of large prior biological knowledge, and model parameters such as sparseness do little to solve these problems. Together with sparseness, feature selection is important for the inference of GRN from large data sets. These parameters limit the number of regulators per gene and penalize model complexity.

Also, the integration of omics data from single cells is still challenging, so there is a need for standardized methods. Besides this, the integration of multiple-source biological knowledge makes easier data insufficiency problems, and it is a major focus in GRN research.

To sum up, GRN models are certainly a powerful tool for the understanding of biological systems, and their improve is ligated to the advances in biotechnology and bioinformatics, which will enable the characterization of complex relations between biological entities.

Conflict of interest

The authors have no affiliations with or involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

Acknowledgements

This work was supported in part by the collaboration scholarship of the Ministry of Education, Culture and Sports of the Government of Spain. This scholarship was granted for the collaboration research with DATAi group (Intelligent Data Analysis – TIC200) of the Division of Computer Science of Pablo de Olavide University.

References

[1] McCall MN. Estimation of gene regulatory networks. Postdoc J 2013;1:60.
[2] Emmert-Streib F, Dehmer M, Haibe-Kains B. Untangling statistical and biological models to understand network inference: the need for a genomics network ontology. Front Genet 2014;5:299.
[3] Liang L, Gao L, Zou X-P, Huang M-L, Chen G, Li J-J, et al. Diagnostic significance and potential function of miR-338-5p in hepatocellular carcinoma: a bioinformatics study with microarray and RNA sequencing data. Mol Med Rep 2018;17:2297–312.
[4] Huang R, He Y, Sun B, Liu B. Bioinformatic analysis identifies three potentially key differentially expressed genes in peripheral blood mononuclear cells of patients with Takayasu's arteritis. Cell J (Yakhteh) 2018;19:647.
[5] Ogundijo OE, Elmas A, Wang X. Reverse engineering gene regulatory networks from measurement with missing values. EURASIP J Bioinform Syst Biol 2016;2017:2.
[6] Yan W, Xue W, Chen J, Hu G. Biological networks for cancer candidate biomarkers discovery. Cancer Inform 2016;15:CIN–39458.
[7] Levine M, Davidson EH. Gene regulatory networks for development. Proc Natl Acad Sci U S A 2005;102:4936–42.
[8] Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res 2013;41:D991–5.
[9] Hecker M, Lambeck S, Toepfer S, Van Someren E, Guthke R. Gene regulatory network inference: data integration in dynamic models—a review. Biosystems 2009;96:86–103.
[10] Dougherty ER. Validation of gene regulatory networks: scientific and inferential. Brief Bioinform 2010;12:245–52.
[11] Buermans H, Den Dunnen J. Next generation sequencing technology: advances and applications. Biochim Biophys Acta (BBA)-Mol Basis Dis 2014;1842:1932–41.
[12] Cereb N, Kim HR, Ryu J, Yang SY. Advances in DNA sequencing technologies for high resolution HLA typing. Hum Immunol 2015;76:923–7.
[13] Li Y, Liu L, Bai X, Cai H, Ji W, Guo D, et al. Comparative study of discretization methods of microarray data for inferring transcriptional regulatory networks. BMC Bioinform 2010;11:520.
[14] Pataskar A, Tiwari VK. Computational challenges in modeling gene regulatory events. Transcription 2016;7:188–95.
[15] Monger C, Kelly PS, Gallagher C, Clynes M, Barron N, Clarke C. Towards next generation CHO cell biology: bioinformatics methods for RNA-Seq-based expression profiling. Biotechnol J 2015;10:950–66.
[16] Larvie JE, Sefidmazgi MG, Homaifar A, Harrison SH, Karimoddini A, Guiseppi-Elie A. Stable gene regulatory network modeling from steady-state data. Bioengineering 2016;3:12.
[17] Chaitankar V, Karakülah G, Ratnapriya R, Giuste FO, Brooks MJ, Swaroop A. Next generation sequencing technology and genomewide data analysis: perspectives for retinal research. Prog Retin Eye Res 2016;55:1–31.
[18] Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. Genbank. Nucleic Acids Res 2012;41:D36–42.
[19] Kodama Y, Mashima J, Kosuge T, Katayama T, Fujisawa T, Kaminuma E, et al. The DDBJ Japanese genotype-phenotype archive for genetic and phenotypic human data. Nucleic Acids Res 2014;43:D18–22.
[20] Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2015. Nucleic Acids Res 2014;43:D662–9.
[21] Tammen SA, Friso S, Choi S-W. Epigenetics: the link between nature and nurture. Mol Aspects Med 2013;34:753–64.
[22] Ramsey J, Butnor K, Peng Z, Leclair T, van der Velden J, Stein G, et al. Loss of RUNX1 is associated with aggressive lung adenocarcinomas. J Cell Physiol 2018;233:3487–97.
[23] Jiang Z, Zhou X, Li R, Michal JJ, Zhang S, Dodson MV, et al. Whole transcriptome analysis with sequencing: methods, challenges and potential solutions. Cell Mol Life Sci 2015;72:3425–39.
[24] Lappalainen T, Sammeth M, Friedländer MR, AC't Hoen P, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013;501:506.
[25] Clough E, Barrett T. The gene expression omnibus database. Stat Genom: Methods Protocols 2016:93–110.
[26] Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, et al. ArrayExpress—a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 2003;31:68–71.
[27] Kang T, Ding W, Zhang L, Ziemek D, Zarringhalam K. A biological network-based regularized artificial neural network model for robust phenotype prediction from gene expression data. BMC Bioinform 2017;18:565.
[28] Fluck J, Hofmann-Apitius M. Text mining for systems biology. Drug Discov Today 2014;19:140–4.
[29] Sverchkov Y, Craven M. A review of active learning approaches to experimental design for uncovering biological networks. PLoS Comput Biol 2017;13:e1005466.
[30] Ud-Dean SM, Gunawan R. Optimal design of gene knockout experiments for gene regulatory network inference. Bioinformatics 2015;32:875–83.
[31] Kim J-W, Yang H-J, Brooks MJ, Zelinger L, Karakülah G, Gotoh N, et al. NRL-regulated transcriptome dynamics of developing rod photoreceptors. Cell Rep 2016;17:2460–73.
[32] Li Y, Varala K, Coruzzi GM. From milliseconds to lifetimes: tracking the dynamic behavior of transcription factors in gene networks. Trends Genet 2015;31:509–15.
[33] Bordon J, Moškon M, Zimic N, Mraz M. Fuzzy logic as a computational tool for quantitative modelling of biological systems with uncertain kinetic data. IEEE/ACM Trans Comput Biol Bioinform 2015;12:1199–205.
[34] Coker EA, Mitsopoulos C, Workman P, Al-Lazikani B. Signet: a signaling network data simulator to enable signaling network inference. PLoS One 2017;12:e0177701.
[35] Wang M, Benedito VA, Zhao PX, Udvardi M. Inferring large-scale gene regulatory networks using a low-order constraint-based algorithm. Mol Biosyst 2010;6:988–98.
[36] Liang Y, Kelemen A. Computational dynamic approaches for temporal omics data with applications to systems medicine. BioData Min 2017;10:20.
[37] Thompson JA, Tan J, Greene CS. Cross-platform normalization of microarray and RNA-seq data for machine learning applications. PeerJ 2016;4:e1621.
[38] Kourilsky P. The natural defense system and the normative self model. F1000Research 2016;5.
[39] Liang S, Fuhrman S, Somogyi R. REVEAL, a general reverse engineering algorithm for inference of genetic network architectures. Pacific symposium on biocomputing, vol. 3 1998:18–29.
[40] Butte AJ, Kohane IS. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Biocomputing 2000. 1999. p. 418–29.
[41] Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinform 2006;7:S7.
[42] Montes RAC, Coello G, González-Aguilera KL, Marsch-Martínez N, de Folter S, Alvarez-Buylla ER. ARACNE-based inference, using curated microarray data, of *Arabidopsis thaliana* root transcriptional regulatory networks. BMC Plant Biol

2014;14:97.
[43] Madar A, Greenfield A, Vanden-Eijnden E, Bonneau R. DREAM3: network inference using dynamic context likelihood of relatedness and the inferelator. PLoS One 2010;5:e9803.
[44] Olsen C, Meyer PE, Bontempi G. On the impact of entropy estimation on transcriptional regulatory network inference based on mutual information. EURASIP J Bioinform Syst Biol 2008;2009:308959.
[45] Gómez-Vela F, Barranco CD, Díaz-Díaz N. Incorporating biological knowledge for construction of fuzzy networks of gene associations. Appl Soft Comput 2016;42:144–55.
[46] Rao A, Hero III AO, States DJ, Engel JD. Using directed information to build biologically relevant influence networks. Computational systems bioinformatics: vol. 6). 2007. p. 145–56.
[47] Meyer PE, Kontos K, Lafitte F, Bontempi G. Information-theoretic inference of large transcriptional regulatory networks. EURASIP J Bioinform Syst Biol 2007;2007.
[48] Luo W, Woolf PJ. Reconstructing transcriptional regulatory networks using three-way mutual information and Bayesian networks. Computational biology of transcription factor binding. Springer; 2010. p. 401–18.
[49] Watkinson J, Liang K-c, Wang X, Zheng T, Anastassiou D. Inference of regulatory gene interactions from expression data using three-way mutual information. Ann N Y Acad Sci 2009;1158:302–13.
[50] Guo W, Calixto CP, Tzioutziou N, Lin P, Waugh R, Brown JW, et al. Evaluation and improvement of the regulatory inference for large co-expression networks with limited sample size. BMC Syst Biol 2017;11:62.
[51] Altay G, Emmert-Streib F. Structural influence of gene networks on their inference: analysis of C3NET. Biol Direct 2011;6:31.
[52] Zheng G, Xu Y, Zhang X, Liu Z-P, Wang Z, Chen L, et al. CMIP: a software package capable of reconstructing genome-wide regulatory networks using gene expression data. BMC Bioinform 2016;17:535.
[53] Xie Y, Liu Y, Valdar W. Joint estimation of multiple dependent Gaussian graphical models with applications to mouse genomics. Biometrika 2016;103:493–511.
[54] Kiani NA, Zenil H, Olczak J, Tegnér J. Evaluating network inference methods in terms of their ability to preserve the topology and complexity of genetic networks. Semin Cell Dev Biol 2016;51:44–52.
[55] Trescher S, Münchmeyer J, Leser U. Estimating genome-wide regulatory activity from multi-omics data sets using mathematical optimization. BMC Syst Biol 2017;11:41.
[56] Jakub NAKHZ, Jesper O. Evaluating network inference methods in terms of. networks 2016;19:2271–82.
[57] Wang X-L, Liu Y-M, Zhang Z-D, Wang S-S, Du Y-B, Yin Z-S. Utilizing benchmarked dataset and gene regulatory network to investigate hub genes in postmenopausal osteoporosis. J Cancer Res Ther 2018.
[58] Chai LE, Loh SK, Low ST, Mohamad MS, Deris S, Zakaria Z. A review on the computational approaches for gene regulatory network construction. Comput Biol Med 2014;48:55–65.
[59] Melkman AA, Cheng X, Ching W-K, Akutsu T. Identifying a probabilistic Boolean threshold network from samples. IEEE Trans Neural Netw Learn Syst 2018;29:869–81.
[60] Maheshri N, O'Shea EK. Living with noisy genes: how cells function reliably with inherent variability in gene expression. Ann Rev Biophys Biomol Struct 2007;36.
[61] Simak M, Yeang C-H, Lu HH-S. Exploring candidate biological functions by Boolean function networks for *Saccharomyces cerevisiae*. PLoS One 2017;12:e0185475.
[62] Claussen JC, Skiecevičienė J, Wang J, Rausch P, Karlsen TH, Lieb W, et al. Boolean analysis reveals systematic interactions among low-abundance species in the human gut microbiome. PLoS Comput Biol 2017;13:e1005361.
[63] Polak ME, Ung CY, Masapust J, Freeman TC, Ardern-Jones MR. Petri Net computational modelling of Langerhans cell interferon regulatory factor network predicts their role in T cell activation. Sci Rep 2017;7:668.
[64] Moignard V, Woodhouse S, Haghverdi L, Lilly AJ, Tanaka Y, Wilkinson AC, et al. Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat Biotechnol 2015;33:269.
[65] Orlando DA, Lin CY, Bernard A, Wang JY, Socolar JE, Iversen ES, et al. Global control of cell-cycle transcription by coupled CDK and network oscillators. Nature 2008;453:944.
[66] Voit EO. Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists. Cambridge University Press; 2000.
[67] Rosenfeld S. Mathematical descriptions of biochemical networks: stability, stochasticity, evolution. Prog Biophys Mol Biol 2011;106:400–9.
[68] Matsumoto H, Kiryu H, Furusawa C, Ko MS, Ko SB, Gouda N, et al. SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. Bioinformatics 2017;33:2314–21.
[69] Deng Y, Zenil H, Tegnér J, Kiani NA. HiDi: an efficient reverse engineering schema for large-scale dynamic regulatory network reconstruction using adaptive differentiation. Bioinformatics 2017;33:3964–72.
[70] Kaderali L, Radde N. Inferring gene regulatory networks from expression data. Computational intelligence in bioinformatics. Springer; 2008. p. 33–74.
[71] Erdogdu U, Polat F, Alhajj R. Employing decomposable partially observable Markov decision processes to control gene regulatory networks. Artif Intell Med 2017;83:14–34.
[72] Larjo A, Shmulevich I, Lähdesmäki H. Structure learning for Bayesian networks as models of biological networks. Data mining for systems biology. Springer; 2013. p. 35–45.
[73] Pineda AL, Gopalakrishnan V. Novel application of junction trees to the interpretation of epigenetic differences among lung cancer subtypes. AMIA Jt Summits Transl Sci Proc 2015;2015:31.
[74] Deeter A, Dalman M, Haddad J, Duan Z-H. Inferring gene and protein interactions using PubMed citations and consensus Bayesian networks. PLoS One 2017;12:e0186004.
[75] Acerbi E, Zelante T, Narang V, Stella F. Gene network inference using continuous time Bayesian networks: a comparative study and application to Th17 cell differentiation. BMC Bioinform 2014;15:387.
[76] Chekouo T, Stingo FC, Doecke JD, Do K-A. miRNA-target gene regulatory networks: a Bayesian integrative approach to biomarker selection with application to kidney cancer. Biometrics 2015;71:428–38.
[77] Chudasama D, Bo V, Hall M, Anikin V, Jeyaneethi J, Gregory J, et al. Identification of novel cancer biomarkers of prognostic value using specific gene regulatory networks (GRN): a novel role of RAD51AP1 for ovarian and lung cancers. Carcinogenesis 2018;39:407–17.
[78] Biswas S, Acharyya S. Neural model of gene regulatory network: a survey on supportive meta-heuristics. Theory Biosci 2016;135:1–19.
[79] Kordmahalleh MM, Sefidmazgi MG, Harrison SH, Homaifar A. Identifying time-delayed gene regulatory networks via an evolvable hierarchical recurrent neural network. BioData Min 2017;10:29.
[80] Ling H, Samarasinghe S, Kulasiri D. Novel recurrent neural network for modelling biological networks: oscillatory p53 interaction dynamics. Biosystems 2013;114:191–205.
[81] Tong DL, Boocock DJ, Dhondalay GKR, Lemetre C, Ball GR. Artificial neural network inference (ANNI): a study on gene–gene interaction for biomarkers in childhood sarcomas. PLoS One 2014;9:e102483.
[82] Siddens LK, Tilton SC, Williams DE, Krueger SK, Larkin A, Waters KM, et al. Application of a fuzzy neural network model in predicting polycyclic aromatic hydrocarbon-mediated perturbations of the Cyp1b1 transcriptional regulatory network in mouse skin. Toxicol Appl Pharmacol 2012.
[83] Rubiolo M, Milone D, Stegmayer G. Extreme learning machines for reverse engineering of gene regulatory networks from expression time series. Bioinformatics 2017;1:8.
[84] Jiao Y, Lawler K, Patel GS, Purushotham A, Jones AF, Grigoriadis A, et al. DART: Denoising algorithm based on relevance network topology improves molecular pathway activity inference. BMC Bioinform 2011;12:403.
[85] Liu J, Chi Y, Zhu C, Jin Y. A time series driven decomposed evolutionary optimization approach for reconstructing large-scale gene regulatory networks based on fuzzy cognitive maps. BMC Bioinform 2017;18:241.
[86] Ud-Dean SM, Heise S, Klamt S, Gunawan R. Trace+: ensemble inference of gene regulatory networks from transcriptional expression profiles of gene knock-out experiments. BMC Bioinform 2016;17:252.
[87] Paroni A, Graudenzi A, Caravagna G, Damiani C, Mauri G, Antoniotti M. CABeRNET: a Cytoscape app for Augmented Boolean models of gene Regulatory NETworks. BMC Bioinform 2016;17:64.
[88] Shahdoust M, Pezeshk H, Mahjub H, Sadeghi M. F-map: a Bayesian approach to infer the gene regulatory network using external hints. PLoS One 2017;12:e0184795.
[89] Wang YR, Huang H. Review on statistical methods for gene network reconstruction using expression data. J Theor Biol 2014;362:53–61.
[90] Thiagarajan R, Alavi A, Podichetty JT, Bazil JN, Beard DA. The feasibility of genome-scale biological network inference using graphics processing units. Algorithms Mol Biol 2017;12:8.
[91] Hira ZM, Gillies DF. A review of feature selection and feature extraction methods applied on microarray data. Adv Bioinform 2015.
[92] Sanchez-Osorio I, Ramos F, Mayorga P, Dantan E. Foundations for modeling the dynamics of gene regulatory networks: a multilevel-perspective review. J Bioinform Comput Biol 2014;12:1330003.
[93] Bellman R. Dynamic programming. Courier Corporation; 2013.
[94] Kerr WT, Douglas PK, Anderson A, Cohen MS. The utility of data-driven feature selection: Re: Chu et al. 2012. NeuroImage 2014;84:1107–10.
[95] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
[96] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
[97] Squillario M, Barbieri M, Verri A, Barla A. Enhancing interpretability of gene signatures with prior biological knowledge. Microarrays 2016;5:15.
[98] Sun C, Cao X, Zhou C, Liu L, Feng F, Liu R, et al. Construction of gene/protein interaction networks for primary myelofibrosis and KEGG pathway-enrichment analysis of molecular compounds. Genet Mol Res 2015;14:16126–32.
[99] Valverde S, Ohse S, Turalska M, West BJ, Garcia-Ojalvo J. Structural determinants of criticality in biological networks. Front Physiol 2015;6:127.
[100] Radcliffe NJ. Genetic set recombination and its application to neural network topology optimisation. Neural Comput Appl 1993;1:67–90.
[101] Jia B, Wang X. Regularized EM algorithm for sparse parameter estimation in nonlinear dynamic systems with application to gene regulatory network inference. EURASIP J Bioinform Syst Biol 2014;2014:5.
[102] Gui S, Rice AP, Chen R, Wu L, Liu J, Miao H. A scalable algorithm for structure identification of complex gene regulatory network from temporal expression data. BMC Bioinform 2017;18:74.
[103] di Bernardo D, Belcastro V. Reverse engineering gene regulatory networks for elucidating transcriptome organisation, gene function and gene regulation in mammalian systems. 2007.
[104] Dechter R, Pearl J. Generalized best-first search strategies and the optimality of A^*. J ACM 1985;32:505–36.
[105] Someren Ev, Wessels L, Backer E, Reinders M. Genetic network modeling.

Pharmacogenomics 2002;3:507–25.
[106] Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. Mach Learn 2006;65:31–78.
[107] Gómez-Vela F, Rodriguez-Baena DS, Vázquez-Noguera JL. Structure optimization for large gene networks based on greedy strategy. Comput Math Methods Med 2018;2018.
[108] Zhu X, Chen S, Jiang Y, Xu Y, Zhao Y, Chen L, et al. Analysis of miRNA expression profiles in melatonin-exposed GC-1 spg cell line. Gene 2018;642:513–21.
[109] Chatr-Aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, Kolas NK, et al. The BioGRID interaction database: 2017 update. Nucleic Acids Res 2017;45:D369–79.
[110] Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol 2015;19:A68.
[111] Pineda S, Steen K, Malats N. Integrative eQTL analysis of tumor and host omics data in individuals with bladder cancer. Genet Epidemiol 2017;41:567–73.
[112] Salehzadeh-Yazdi A, Asgari Y, Saboury AA, Masoudi-Nejad A. Computational analysis of reciprocal association of metabolism and epigenetics in the budding yeast: a genome-scale metabolic model (GSMM) approach. PLoS One 2014;9:e111686.
[113] Sinha S. Integration of prior biological knowledge and epigenetic information enhances the prediction accuracy of the Bayesian Wnt pathway. Integr Biol 2014;6:1034–48.
[114] Walhout AJ. What does biologically meaningful mean? A perspective on gene regulatory network validation. Genome Biol 2011;12:109.
[115] Dougherty E, Qian X. Validation of gene regulatory network inference based on controllability. Front Genet 2013;4:272.
[116] Yu B, Doraiswamy H, Chen X, Miraldi E, Arrieta-Ortiz ML, Hafemeister C, et al. Genotet: an interactive web-based visual exploration framework to support validation of gene regulatory networks. IEEE Trans Vis Comput Graphics 2014;20:1903–12.
[117] Jansen R, Gerstein M. Analyzing protein function on a genomic scale: the importance of gold-standard positives and negatives for network prediction. Curr Opin Microbiol 2004;7:535–45.
[118] Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, et al. Wisdom of crowds for robust gene network inference. Nat Methods 2012;9:796.
[119] De Smet R, Marchal K. Advantages and limitations of current network inference methods. Nat Rev Microbiol 2010;8:717.
[120] Haibe-Kains B, Emmert-Streib F. Quantitative assessment and validation of network inference methods in bioinformatics. Front Genet 2014;5:221.
[121] Schrynemackers M, Küffner R, Geurts P. On protocols and measures for the validation of supervised methods for the inference of biological networks. Front Genet 2013;4:262.
[122] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29–36.
[123] Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn 1997;30:1145–59.
[124] Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. European conference on information retrieval 2005. 2005. p. 45–359.
[125] Boyd K, Eng KH, Page CD. Area under the precision-recall curve: point estimates and confidence intervals. Joint European conference on machine learning and knowledge discovery in databases. 2013. p. 451–66.
[126] Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd international conference on Machine learning. 2006. p. 233–40.
[127] Prill RJ, Marbach D, Saez-Rodriguez J, Sorger PK, Alexopoulos LG, Xue X, et al. Towards a rigorous assessment of systems biology models: the DREAM3 challenges. PLoS One 2010;5:e9202.
[128] Hase T, Ghosh S, Yamanaka R, Kitano H. Harnessing diversity towards the reconstructing of large scale gene regulatory networks. PLoS Comput Biol 2013;9:e1003361.
[129] Bellot P, Olsen C, Salembier P, Oliveras-Vergés A, Meyer PE. NetBenchmark: a bioconductor package for reproducible benchmarks of gene regulatory network inference. BMC Bioinform 2015;16:312.
[130] Chen S, Mar JC. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. BMC Bioinform 2018;19:232.
[131] Marbach D, Schaffter T, Mattiussi C, Floreano D. Generating realistic in silico gene networks for performance assessment of reverse engineering methods. J Comput Biol 2009;16:229–39.
[132] Schaffter T, Marbach D, Floreano D. GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods. Bioinformatics 2011;27:2263–70.
[133] Seifert M, Beyer A. regNet: an R package for network-based propagation of gene expression alterations. Bioinformatics 2017;34:308–11.
[134] Van den Bulcke T, Van Leemput K, Naudts B, van Remortel P, Ma H, Verschoren A, et al. SynTReN: a generator of synthetic gene expression data for design and analysis of structure learning algorithms. BMC Bioinform 2006;7:43.
[135] Gómez-Vela F, Díaz-Díaz N. Gene network biological validity based on gene–gene interaction relevance. Sci World J 2014;2018.
[136] Gómez-Vela F, Lagares JA, Díaz-Díaz N. Gene network coherence based on prior knowledge using direct and indirect relationships. Comput Biol Chem 2015;56:142–51.
[137] Diaz-Montana JJ, Diaz-Diaz N. Development and use of the Cytoscape app GFD-Net for measuring semantic dissimilarity of gene networks. F1000Research 2014;3.
[138] Li Y, Gong P, Perkins EJ, Zhang C, Wang N. RefNetBuilder: a platform for construction of integrated reference gene regulatory networks from expressed sequence tags. BMC Bioinform 2011;12:S20.