Overview of Computational Approaches for Inference of MicroRNA-Mediated and Gene Regulatory Networks

Blagoj Ristevski[1]
Faculty of Information and Communication Technologies—Bitola, St. Kliment Ohridski University, Bitola, Macedonia
[1]Corresponding author: e-mail address: blagoj.ristevski@uklo.edu.mk

Contents

Abstract

This chapter describes biological backgrounds of regulatory relationships in living cells, high-throughput experimental technologies, and application of computational approaches in reverse engineering of microRNA (miRNA)-mediated and gene

Advances in Computers
ISSN 0065-2458
http://dx.doi.org/10.1016/bs.adcom.2014.12.001

regulatory networks (GRNs). The most commonly used models for GRNs inference based on Boolean networks, Bayesian networks, dynamic Bayesian networks, association networks, novel two-stage model using integration of *a priori* biological knowledge, differential and difference equations models are detailed and their inference capabilities are compared. The regulatory role of miRNAs and transcription factors (TFs) in miRNAs-mediated regulatory networks is described as well. Additionally, commonly used methods for target prediction of miRNAs and TFs are described as well as most commonly used biological regulatory relationships databases and tools are listed. The mainly validation criteria used for assessment of inferred regulatory networks are explained. Finally, concluding remarks and further directions for miRNA-mediated and GRNs inference are given.

1. INTRODUCTION

Many biological, physiological, and biochemical molecular processes occur simultaneously in living cells. Regulation of these processes is conducting by inherited information contained in the organisms' genome. Inference of the mutual interactions between numerous components of biological systems based on available experimental data for interactions between DNAs, RNAs, proteins, and metabolites is needed to clarify and represent existing regulatory mechanisms. These components and their mutual interactions compose complex networks named as gene regulatory networks (GRNs) [1]. Generally speaking, there are two approaches for inferring of GRNs [2]:

- Mechanistic (or physical) networks that employ protein–DNA and protein–protein interactions (PPIs) data are usually named as transcription or protein networks. The aim of this static networks modeling is to reveal regulatory interactions on physical level, and
- Influence networksthat refer to the reverse engineering of GRNs based on gene expression data and the inferred networks regard to gene–gene interactions.

The GRNs structure is depicted by a graph consisted of *nodes* representing the genes, proteins, metabolites, their complexes or even modules, and *edges* that represent direct or indirect interactions between nodes. In the influence GRNs, proteins and metabolites appear as hidden variables in GRNs, while the only observable variables are gene expression data. These hidden variables might cover unobserved results that are not measured. Figure 1 illustrates the projection of interactions from the space of metabolites and proteins into the space of genes. Dashed lines represent gene regulatory

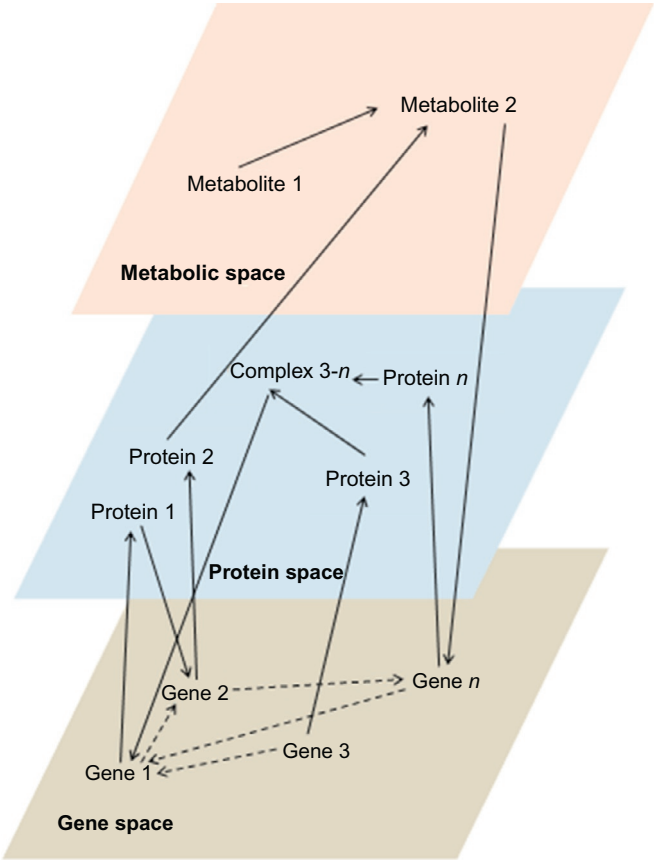


Figure 1 Projection of GRNs in different spaces: gene, protein, and metabolic space.

relationships, whereas the full lines depict the interactions among genes, proteins, metabolites, and their complexes [3].

This chapter systematizes the different models for GRNs inference such as Boolean, Bayesian, dynamic Bayesian, association networks, and other models comparing their advantages and shortcomings. In addition, the role of microRNAs (miRNAs) in posttranscriptional regulation and computational approaches for reconstruction of miRNA-mediated regulatory networks are explained. In addition, network topology and network validation of the inferred regulatory networks are depicted.

This chapter is organized as follows. In Section 2, the biological and experimental backgrounds of the cells' regulatory mechanisms are presented. Section 3 describes the computational backgrounds of the inference of

miRNA-mediated and GRNs. The following section provides more detail description of most commonly used models for GRNs inference such as Boolean networks, Bayesian networks (BNs) and dynamic Bayesian networks (DBNs), differential and difference equations system model, association networks, and several other models. Computational approaches for inference of miRNA-mediated regulatory networks and miRNA and transcription factors (TFs) target prediction algorithms are depicted in Section 4.5. The subsequent section describes the commonly used validation criteria for assessment of applied models for inference of regulatory relationships. Finally, the last section gives the concluding remarks and further directions toward inference of miRNA-mediated and GRNs.

2. BIOLOGICAL BACKGROUNDS OF CELL REGULATORY MECHANISMS AND EXPERIMENTAL TECHNOLOGIES

Genes are fundamental physical and functional inheritance units of all living organisms. The coding genes are templates for protein synthesis. Other genes might specify RNA templates as machines for production of different types of RNAs.

The process in which DNA is transcribed into messenger RNA (mRNA) and proteins are produced by translation represents the well-known central dogma in molecular biology. The first stage of gene expression is DNA transcription into RNA. Resulted RNA can be mRNA, if the expressed gene is protein coding, otherwise it is noncoding RNA. Then the second stage follows, in which mRNA translates into a sequence of amino acids that composes a protein. When a protein is synthesized, the matching protein-coding gene is expressed.

Only a small part of RNAs is coding RNAs, whereas the bigger part from the genome of eukaryotes transcribes into noncoding RNAs. In the last decade, several small noncoding RNAs such as miRNAs and small inferring RNAs (siRNAs) are revealed [4]. The length of miRNAs is about 18–25 nucleotides [5]. To date, there are more than 1800 human miRNAs listed in the miRBase database [6].

A regulatory relationship between a miRNA and an mRNA denotes that a change in the miRNA expression level will effect a change in the expression level of the target mRNA. Each of these miRNAs might regulate expression of hundreds or even thousands of target mRNAs. MiRNAs regulate expression by more than 30% of coding genes [7,8]. MiRNAs cause transcription cleavage or translation repression by connecting to their target

mRNA [9]. MiRNAs regulate gene expression by inhibiting mRNA translation and/or lightening mRNA degradation. Recent *in vitro* and *in vivo* studies have shown that miRNAs can inhibit translation initiation and support decay of target mRNAs. There are three different potentials not mutually exclusive manners of miRNA-mediated repression by destabilization of target mRNAs, inhibition of translation initiation, or blocking of translation after initiation.

The gene expression regulation on the posttranscriptional level taking place by mRNA cleavage or translation repression with binding of miRNAs to the 3′-untranslated regions (3′-UTRs) of target mRNA [10]. To identify mRNAs regulated by silencing or overexpression of a specific miRNA, quantitative real-time polymerase chain reaction is used [11].

Recent studies have shown that miRNAs are one of the key participants of regulation in many biological processes in metabolism, proliferation, differentiation, development, apoptosis, cellular signaling, cancer development, and metastasis. MiRNAs are involved in cancer, rheumatic, infectious cardiovascular and neuronal diseases, metabolic disorders (glucose and lipid metabolisms), epigenetics (mitotically and meiotically heritable gene expression changes not involving a change in the DNA sequence) [7,8,12,13].

One of the most important regulatory functions of proteins is transcription regulation. Proteins, which bind to DNA sequences and regulate the transcription of DNAs and hence gene expression, are called TFs. TFs can inhibit or activate gene expression of the target genes [14].

The gene expression levels indicate the approximate number of produced RNA copies from the corresponding gene, which means that gene expression level corresponds to the amount of synthesized proteins. To obtain gene expression data experimentally of many genes in a sample, high-throughput technologies are used, such as DNA microarray, serial analysis of gene expression (SAGE), quantitative polymerase chain reaction (qPCR), as well as next-generation sequencing technology like RNA-Sequencing (RNA-Seq).

The essential basis of DNA microarrays is hybridization between two strands of DNA. This technology is used to measure the expression levels of large numbers of genes simultaneously or to determine genotype of multiple regions of a genome. It is known as DNA chip or biochip used for DNA detection, or to detect RNA that may or may not be translated into proteins. SAGE technology produces a snapshot of the mRNA population in a sample of interest. SAGE sampling is based on sequencing mRNA

output, not on hybridization of mRNA output to probes, so transcription levels are measured more quantitatively than using DNA microarray. qPCR technology amplifies and quantifies a targeted DNA simultaneously, whereas RNA-Seq is able to identify and quantify transcripts, perform robust whole-transcriptome analysis on a wide range of samples at a given time moment.

Besides gene expression data, other data types such as protein–DNA, PPIs data, and miRNAs targets are needed to reveal regulatory relationships.

Experimentally, identifying of TF binding sites (TFBSs) on the genome for particular proteins and to reveal protein–DNA interactions, chromatin immunoprecipitation (ChIP)-based methods are used [15]. ChIP-chip technology uses ChIP with hybridization microarrays (chips) to identify the protein binding sites and their locations throughout the genome. ChIP-chip technology uses short DNA sequences as probes. A population of immunoprecipitation-enriched DNA fragments is produced and enrichment of each probe from produced population is measured [16]. Differently, ChIP-Sequencing (ChIP-Seq) technology uses secondary sequencing of DNA instead of microarray [15]. The sequencing technology such as ChIP-Seq, RNA-Seq, and miRNA-Seq are very well-established technologies.

TFs and miRNAs are in mutual interaction with more *cis*-regulatory elements. Similarly to TFs, genes also contain binding sides for other TFs that may be targeted by miRNAs. Thus, the mutual interactions between miRNAs and TFs make miRNAs very important factors in the gene regulation.

High-throughput techniques and data such as proteomics, transcriptomics, and miRNAomics lighten inference of large-scale miRNA-mediated and GRNs. Integration of these different types of biological data can significantly improve the accuracy and the reliability of the inferred miRNA-mediated and GRNs [17].

3. COMPUTATIONAL BACKGROUNDS OF THE INFERENCE OF MiRNA-MEDIATED AND GRNs

Theoretical studies of GRNs have started in the 1960s. The emergence of experimental high technologies for discovering regulatory mechanisms such as DNA microarrays, ChIP-chip, ChIP-Seq, RNA-Seq, and qPCR has provided huge amounts of gene expression, protein–protein, and protein–DNA interactions data. Because the experimental wet lab

technologies cannot measure mutual influences among all genes from one organism's genome simultaneously, computational methods are needed to infer and reveal regulatory interactions between genes, miRNAs, TFs, and other constituents that compose complex regulatory networks.

GRNs reconstruction is useful to elucidate disease-causing perturbations in two different manners: changes in the interactions of the component genes and changes in the cell type in which a gene is expressed, the magnitude of gene expression, the beginning time and time span of the transcriptional activity [18].

Several models for GRNs inference have been developed, which are based on the basic reverse engineering methods. However, these models handle with only certain data types and inferred networks do not largely match the real regulatory mechanisms. The reverse engineered networks might contain many erroneous regulatory relationships. This shortcoming is a motivation for developing of new models that can include *a priori* knowledge and could be able to integrate heterogeneous data. Such inferred GRNs should elucidate gene regulatory mechanisms more correctly and more reliably.

Finding more accurate and reliable structures of GRNs from gene expression data is a problem of machine learning known as structure learning, while the parameter learning aims to find parameters of inferred networks that match best to the true regulatory relationships. Both structure and parameter learning of the reconstructed networks are challenging topics that bring together learning techniques from artificial intelligence with bioinformatics, functional genomics, and biostatistics [1].

In miRNA-mediated regulatory networks, TFs and miRNAs have very important role. Determination of the TFs of the given genes is by using TF binding matrix (motif)-based methods. TFBSs are usually short (around 5–15 basepairs) and regularly degenerate sequences. The sequence pattern is presented by a matrix (motif), whose entries give the probability distribution of DNA nucleobases adenine (A), cytosine (C), thymine (T), and guanine (G) at each site. The motifs of TFs are concluded from known TFBSs determined experimentally [19]. TRANSFAC [20] and JASPAR [21] provide major collections of currently annotated TF binding motifs. Predicted TF targets can be determined by scanning promoter regions of given genes with motifs.

In order to understand the multiple functions of miRNAs in biological processes, it is crucial to determine their targets. Prediction of miRNA targets by using computational methods is often imprecise because the

miRNA–mRNA interactions are relied on a limited sequence length of the miRNAs seed region. Additionally, mRNA recognition is affected by the sequence context around the target as well by factors that might halt miRNA binding.

By inference of miRNAs-mediated and GRNs, several networks' properties should be taken in consideration such as sparseness, scale-free topology, modularity, and structure of the inferred networks [2].

The inferred regulatory networks should be sparse. It means that a limited number of nodes regulate the other nodes. Some nodes in the network called hubs can have regulatory relationships to many targets, i.e., the outdegree of the nodes is not limited.

Another important feature of the inferred regulatory networks is their topology, which should be scale free. Scale-free networks have the power distribution function of the nodes connectivity degrees [22]. This property provides the robustness of the inferred networks considering the random topology perturbations.

Structures with small connectivity follow the regulatory hierarchy. The networks structure allows decomposition of a network into basic modular units composed of several nodes, called network motifs [2]. Modularity of the networks regards to the presence of clusters of highly coexpressed genes/miRNAs and/or genes/miRNAs with similar function.

4. MODELS FOR GRNs INFERENCE

A plethora of models such as Boolean networks, BNs, DBNs, Petri nets, graphical Gaussian models (GGMs), linear and nonlinear differential and difference equations, information theory approach, state space models, fuzzy logic models, two-stage model that integrates biological *a priori* knowledge, and many other models are utilized to reverse engineer GRNs. GRN models extend from maps of genetics interactions, physical interaction graphs to models that cover the gene expression kinetics, and hence network dynamics.

4.1 Boolean Networks

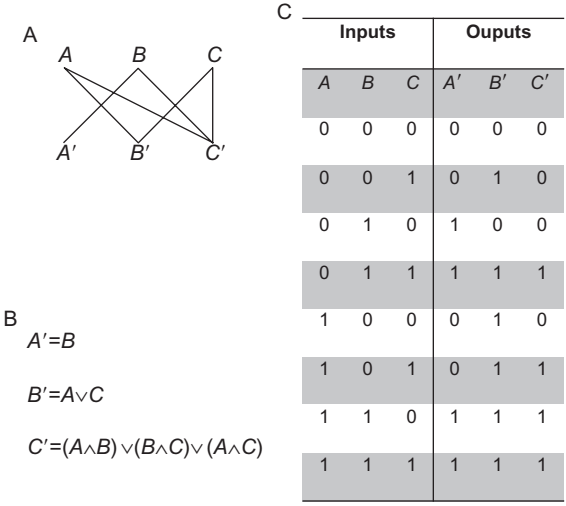
The model based on Boolean networks is one of the simplest models for GRNs inference. A Boolean network is presented by graph whose nodes present the genes and the edges between nodes represent the regulatory interactions between genes. In this model, gene expression data are

discretized and presented by two values: 1 or 0. If the gene expression is above a set threshold, the corresponding state is 1, otherwise 0.

The network diagram shown in Fig. 2A is not sufficient to understand logical dependencies between genes. The aim of the reverse engineering in Boolean networks is to find Boolean functions of every gene in the network, so discretized values of gene expression can be explained by the model as shown in Fig. 2B. Alternatively, Boolean networks can be represented by state transitions table presented in Fig. 2C.

The small changes in gene expression time series data cannot be encompassed by two-level discretization, because it leads to information loss. Thus, inferred regulatory networks can be unrealistic and with erroneous interactions between nodes. Another weakness of Boolean networks is the super-exponential number of all possible networks. If n is the number of genes, then the number of all Boolean functions depends on n super-exponentially and it is equal to 2^{2^n}.

Several extended models based on Boolean networks have been proposed. A REVerse Engineering ALgorithm (REVEAL) constructs a Boolean network of given gene expression data by setting the genes in-degree values to k[23]. This algorithm derives minimal network structures from the state transition tables of the Boolean network by using the mutual information approach. If n is the number of nodes, the number of all possible networks is given by:



Inputs			Ouputs		
A	B	C	A′	B′	C′
0	0	0	0	0	0
0	0	1	0	1	0
0	1	0	1	0	0
0	1	1	1	1	1
1	0	0	0	1	0
1	0	1	0	1	1
1	1	0	1	1	1
1	1	1	1	1	1

Figure 2 A Boolean network presented by (A) a wiring diagram, (B) Boolean functions, and (C) a state transition table.

$$\left(2^{2^k}\frac{n!}{(n-k)!}\right)^n \tag{1}$$

REVEAL can be applied on gene expression data discretized on multiple-value levels. On the other hand, multiple discretization levels increase the number of potential state transitions. Thus, the number of all possible networks is significantly greater than the number of networks derived from two-level Boolean networks. REVEAL has better inference capabilities for smaller in-degree value k. For greater in-degree k, parallel processing and efficiency increasing of the search space of all possible networks are needed [23].

The models based on Boolean network simplify the structure and dynamics of gene regulatory relationships. They are deterministic, i.e., the state space is restricted and these networks reach the steady state or fall into dynamic attractor [24]. The inferred networks provide only a quantitative measure of gene regulatory relationships.

Probabilistic Boolean networks model is another model composed of several Boolean networks that work simultaneously [25]. All networks share information about the states of whole system. When a network transits in next state, the remaining networks are synchronized.

4.2 Bayesian Networks

BNs are among the most effective models for GRNs inference. A BN is special type of graph defined as a triple (G, F, q), where G denotes the graph structure, F is the set of probability distributions, and q is the set of parameters for F[26]. The graph structure G includes a set of n nodes $X_1, X_2, \ldots, X_n$ and a set of directed edges between nodes. The nodes correspond to the random variables while directed edges show the conditional dependences between the random variables.

A directed edge from the node X to node Y is denoted as $X \rightarrow Y$, which means that X is a parent node of Y, denoted as pa(Y), and Y is a child node of X. If the node Z can be attained by following a directed path starting from node X, then the node Z is a descendant of X, and X is ancestor of Z. Nodes and edges together have to make a directed acyclic graph (DAG). One directed graph is acyclic if there is no directed path $X_1 \rightarrow X_2 \rightarrow \ldots \rightarrow X_n$ such as $X_1 = X_n$, i.e., there is no pathway starting and ending at the same node.

The joint probability distribution of all network nodes is calculated by the following equation:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\mathrm{pa}(X_i)) \quad (2)$$

where multipliers are local probability distributions. This factorization of the joint probability distribution on multipliers facilitates its computing as a product of simpler conditional probability distributions.

GRNs inference using BNs is accompanied by structure and parameter learning. Structure learning aims to find network structure that fits best the real regulatory relationships. Similar to Boolean networks, the number of possible DAGs also super-exponentially depends on the number of nodes of the BNs. For a given network structure, the parameter learning includes estimation of the unknown model parameters for each gene. Parameter learning determines conditional dependencies between network nodes. Because the network inference using BNs is an NP-hard problem, BNs are the most suitable when they are applied to small-scale networks composed of tens to hundred genes [27].

It is possible to infer GRNs by BNs based on static, dynamic, discrete, or continuous gene expression data. If the node variables are continuous, then network inference is more complicated to perform because of the additional complex computations concerning learning of BNs.

BNs are able to deal with stochastic nature of gene expression profiles as well as with their incompleteness and noise. The main difficulty in BNs learning is the higher number of genes compared to the number of conditions and incapability to handle feedback loops that exist in the real GRNs.

Friedman *et al.* [28] have introduced a framework for discovering interactions between genes based on microarray data using BNs by modeling of each variable with conditional probability distribution function related to other variables. In the proposed approach, two comparative experiments are conducted for different probability distributions: multinomial distribution and linear Gaussian distribution. The main shortcomings of this model are nonconstraints search heuristics on the search space and nonusing *a priori* biological knowledge.

4.3 Dynamic Bayesian Networks

BNs can represent probabilistic relations between variables without considering time lags and they cannot deal with time series data [29]. However, regulatory interactions in the real GRNs do not occur simultaneously, so there is a particular time lagging. Another disadvantage of BNs is that they

cannot represent real biological systems in which exist mutual interactions between entities of biological systems such as feedback loops [30].

These shortcomings make BNs inappropriate for GRNs inference from time series gene expression data, because it is necessary to include dynamic (temporal) features of gene regulatory relationships. Thus, BNs are extended to cover temporal features by introduction of DBNs. Supposing that the changes in time series gene expression data occur in a limited number of discrete time intervals T. Let $X = \{X_1, X_2, \ldots, X_n\}$ is a set of time-dependent variables, where $X_i[t]$ is a random variable representing the value of X_i at the time point t and $0 \leq t \leq T$. A DBN is a BN that contains the T random vectors $X_i[t]$ [31], an initial BN, a transitional BN consisted of transition DAG $G_{\rightarrow}$, and transitional probability distribution $P_{\rightarrow}$:

$$P_{\rightarrow}(X[t+1] = x[t+1] | X[t] = x[t]) \tag{3}$$

Then, joint probability distribution of the DBN is computed by:

$$P(x[0], \ldots, x[T]) = P_0(x[0]) \prod_{t=0}^{T-1} P_{\rightarrow}(x[t+1] | x[t]) \tag{4}$$

From Eq. (4), for each x at each time point t, the following is obtained:

$$P(x[t+1] | x[0], \ldots, x[t]) = P(x[t+1] | x[t]) \tag{5}$$

Equation (5) denotes that the variables values at time point t depend on the values of variables at the previous time point $t-1$ and no other information is required, i.e., DBNs have Markov property [32].

For probabilistic inference of DBNs, the standard algorithms used in BNs inference can be used, too. However, in the case of large-scale networks, learning of DBNs becomes too complex.

DBNs are effective for GRNs inference when they combine other types of biological data. An example for that is the proposed method that integrates gene expression data with *a priori* biologic knowledge about TFBSs using DBNs and structural expectation–maximization algorithm [33].

Daly *et al.* had used high-order DBNs to model time lag gene regulatory interactions based on time series gene expression data [34].

Figure 3 illustrates a DBN that describes cyclic regulation between gene 1 and gene 2 in different time points (red arrow lines; dark gray in the print version), although the graph does not contain obviously cyclic pathway.

A manner how DBNs can be employed for network inference and how they can be learned, their relationship with the hidden markov model

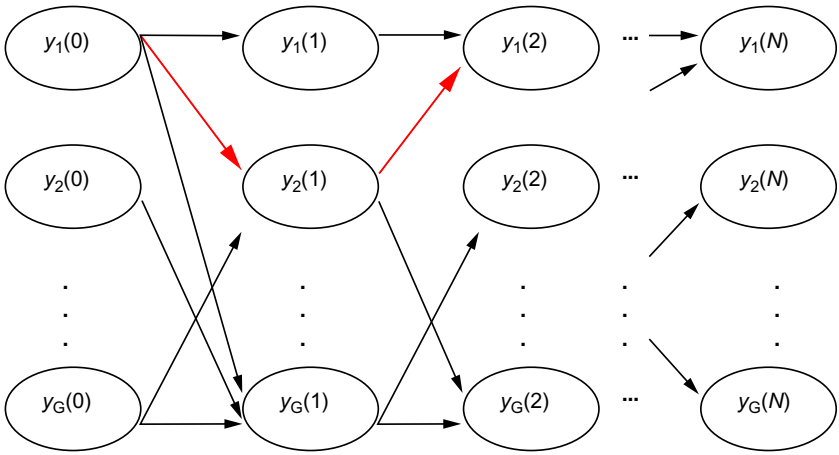


Figure 3 A GRN inferred using DBNs applied on an input time series gene expression data for *G* genes whose expression is measured in *N* time points.

(HMM), Boolean and stochastic Boolean networks, and DBNs with continuous variables is presented in Ref. [35]. Moreover, Boolean networks, linear and nonlinear equations models can be considered as a special case of DBNs.

To overcome the high complexity needed for GRNs inference by DBNs, a model with constraint of the potential regulators has been proposed [36]. This constraint considers those genes that have changes in gene expression level at the previous or at the same time points regarding their target genes. The proposed model uses the time lag of changes in expression levels of regulator and target genes, which increases the accuracy of the inferred networks. The time points of initial over- or under-regulation of the genes are determined. The genes with changes in previous and current time points are denoted as potential regulators to those genes with expression changes in the following time points. In such a way, a subset of potential regulators for every target gene is chosen.

Another approach named as Bayesian Orthogonal Least Squares (BOLS) for inference of GRNs is proposed in Ref. [37]. This approach combines the orthogonal least squares, second-order derivatives for network pruning, and BN. In the inferred networks, that are sparse, only limited number of genes regulates every gene and the number of false inferred regulatory relationships is small.

An effective algorithm for structure learning as an extension of K2 algorithm is proposed in Ref. [38]. This algorithm is utilized for learning of large-scale BNs by introducing sorting of the genes to improve the efficiency

in the large-scale network inference. The assessment of the efficiency of the proposed algorithm is performed by performing Monte Carlo simulation and then comparing to the greedy hill-climbing algorithm.

4.4 Association Networks

Association networks are static networks that can describe the possible structure of the GRNs by an undirected graph. They can be applied on time series and steady-state gene expression data. If two nodes/genes are connected by an edge, then it is not possible to determine which gene is regulated and which one is regulator. To identify which genes are coexpressed and where should be an edge, similarity metrics such as Pearson coefficient or mutual information is applied and additionally a threshold is set. For higher set threshold values, the inferred GRNs are sparser.

Although the association networks do not determine the edges directions in the networks, they are suitable to be employed in inference of large-scale networks because of their low computational costs [2]. The directions of the regulatory interactions can be determined by computing the similarity between genes, their possible regulators, and with additional knowledge about interaction with TFs.

A proposed algorithm ARACNE is based on mutual information between gene expression data [39]. It defines the network edges as statistical dependences, whereby the directed regulatory relationships using data about TFs and their TFBSs can be determined. Using ARACNE, the number of falsely predicted gene interactions in the networks can be reduced significantly. The computational complexity of this algorithm is $O(N^3 + N^2M^2)$, where N is the number of genes and M is the number of samples. The lower complexity makes this algorithm to be suitable for inference of large-scale GRNs [39].

Very popular models are GGMs that use partial correlation coefficients to determine the conditional dependencies between genes [40]. GGMs can distinguish directed or undirected interactions between genes, unlike the correlation networks where the edges correspond to correlation between genes.

Let X is a gene expression data matrix with n rows and p columns, where n is the number of experimental conditions and p is the number of genes. Supposing that data entries from matrix X follow the normal distribution $N_P(\mu, \Sigma)$, where $\mu = \left(\mu_1, \ldots, \mu_p\right)^{\mathrm{T}}$ is the means vector and $\Sigma = (\sigma ij)_{1 \le i,j \le p}$ is a positive definite covariance matrix. The matrix Σ is

decomposed of two parts: variance components σ_i^2 and Brevis–Pearson correlation matrix $P = \left(\rho_{ij}\right)$. A partial correlation matrix $Z = \left(\varsigma_{ij}\right)$ is composed of the correlation coefficients between any two genes i and j with regard to all other genes. The matrix Z is associated with the inverse matrix of the standard correlation coefficients matrix P. Their relationship is given by the following equations [41]:

$$\Omega = \left(\omega_{ij}\right) = P^{-1} \tag{6}$$

and

$$\varsigma_{ij} = -\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}} \tag{7}$$

In Eq. (6), instead of correlation matrix P, the covariance matrix Σ can be used. Partial correlation coefficients ς_{ij} are correlation coefficients of conditional bivariant normal distributions of the genes i and j. Two variables that follow normal distribution are conditionally independent if and only if their partial coefficients are equal to zero. The conditional independence of the random variables is determined by the zero entries in the inverse correlation matrix Ω.

To infer a GRN by using the GGMs from a data set, the correlation matrix P is estimated by unbiased sampling of the covariance matrix, given by:

$$\hat{\Sigma} = \left(\hat{\sigma}_{ij}\right) = \frac{1}{n-1}(X - \overline{X})^T(X - \overline{X}) \tag{8}$$

The estimation of partial correlation coefficients is calculated by Eqs. (6) and (7) from the sample correlation matrix. The entries from estimated correlation matrix $\hat{Z}$, which differentiate from zero, are determined by statistical tests. The network inference finishes with a visualization of correlation structure by a graph, whose edges correspond to nonzero partial correlation coefficients.

The main shortcoming of the described classical GGMs is that they can be applied when the number of experimental conditions n is greater than the number of genes p, because they cannot calculate the partial correlations. The commonly used statistical tests for GGMs selection are valid only for data with large number of samples [40]. If $p > n$ then covariance matrix is not positive definite, so its inverse matrix cannot be calculated. Also, the

existence of an additional linear dependence between variables leads to multicollinearity.

Therefore, to obtain positive definite covariance matrix, covariance matrix estimation is performed by shrinkage estimators and thus, its inverse matrix could be found [42,43]. Then, the edges in the graph are determined by model selection of the network graphs.

4.5 Differential and Difference Equations Models

Concentration of RNAs, proteins, and other metabolites changes through time. Therefore, differential equations systems can be an appropriate model to describe gene regulatory relationships [44]. Ordinary differential equations (ODEs) systems use continuous gene expression data directly and can easily cover positive and negative feedback loops.

The main constraint of this model based on ODEs is the assumed constant or linear changes of the concentration of regulators, although their changes through time are actually nonlinear.

The dynamics of gene expression data changes is described by the following differential equation:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x, p, u, t) \tag{9}$$

where $x(t) = (x_1(t), x_2(t, \ldots, x_n(t)))$ is a vector of gene expression data for n genes at time t, f is the function describing the changes of variables x_i depending on the model parameters p and external perturbations u. GRNs inference aims to determine the function f and parameters p for given measured signals x and u at the time t[2].

Equation (9) can have more solutions, so structure and parameters identification of the model requires identification of the function f based on *a priori* knowledge or some approximations. The function f can be linear or nonlinear and when this function is nonlinear, to simplify, its linearity is supposed and Eq. (9) transforms into the following equation:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \sum_{j=1}^{N} w_{ij}x_j + b_i u, \quad i = 1, \ldots, N \tag{10}$$

where w_{ij} are entries of weight matrix W and parameters b_i determine the external disturbance u to gene expression. This model is also called a model of regulatory matrices composed of weight coefficients w_{ij} that present the regulatory dependences. If a weight coefficient is positive, then

corresponding gene has activating role, while a weight coefficient is negative, then that gene has inhibitory role. If weight coefficients values are zeros, then genes do not interact mutually.

In linear models, network inference from small number of samples is easier to carry out. The identification of function f and the parameters in the nonlinear models is more difficult because the number of samples in gene expression data is smaller than the number of genes and finding numerical solutions is more complex.

Another way to describe the changes of gene expression is by using *S*-systems with activating and inhibitory components, given by following equation:

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{N} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{N} X_j^{h_{ij}} \tag{11}$$

where α_i and β_i are positive constants and h_{ij} and g_{ij} are kinetic exponents [2]. In these models, there are many parameters whose identification requires carrying out numerous experiments, and therefore to find solutions, these differential equations are approximated.

An optimized model for GRNs inference that uses known *a priori* biological knowledge from available databases for genome, proteome, transcriptome, and scientific literature has been proposed in Ref. [45]. This model is based on differential equations whose particular solutions are obtained by singular values decomposition. The obtained solutions are optimized by using mathematical programming.

A special case of differential equations system is the model of pairwise linear differential equations, proposed in Ref. [46]. In this model, it is supposed that gene regulation can be represented by pairwise linear equations. This model uses only gene expression data and neglects regulation on posttranscriptional level.

Beside ODEs, difference equations can describe the dynamics of GRNs. Unlike the differential equations models that deal with continuous variables, the variables in the difference equations model are discrete. Discretization of the gene expression data leads to information loss [44]. However, difference equations are more suitable when time series gene expression data are available. The change of gene expression data is described by the following equation:

$$\frac{x_i(t+\Delta t) - x_i(t)}{\Delta t} = \sum_{j=1}^{N} w_{ij} x_j(t) + b_i u, \quad i = 1, \ldots, N \tag{12}$$

Difference equations model can be transformed into a system of linear algebraic equations that can be solved easily by linear algebra methods [2].

4.6 Other Models for Inference of GRNs

Besides the above-mentioned models, numerous models for GRNs reconstruction are proposed.

Collateral-Fuzzy Gene Regulatory Network Reconstruction (CF-GeNe), proposed by Sehgal *et al.*, applies collateral assessment of the missing values [47]. This model utilizes the fuzzy nature of gene coregulation determined by fuzzy *c*-means clustering algorithm. This clustering algorithm allows one gene to belong to several clusters, i.e., biological processes. CF-GeNe can handle noisy data and missing values and it determines the optimal number of clusters.

Fujita *et al.* had proposed a model of GRNs using sparse autoregressive vector in Ref. [48]. This model can infer gene regulatory relationships when the number of samples is lesser than the number of genes without using *a priori* knowledge and it can deal with the feedback loops.

The linear model in the finite state space proposed in Ref. [49] infers gene regulatory relationships including discrete and continuous aspects of the gene regulation. The model assumptions are that gene activity is determined by the state of the TFBSs, each binding sites can be located in one of the final number of states, genes may be repressed or they can have some activity and the state of the binding sides depends on the concentration of TFs.

Li *et al.* had proposed another model that uses the state space with hidden variables for the GRNs reconstruction [50]. This model is dynamic and it is consisted of observations and states. The observations $(O_1, O_2, \ldots, O_T)$ are generated from the vector of states $(S_1, S_2, \ldots, S_T)$ according to the formal model given by:

$$S_t = AS_{t-1} + W_t, \quad O_t = BS_t + V_t \tag{13}$$

where A denotes the transition state probability $P(S_t|S_{t-1})$ from the state at time $t-1$ to t, B denotes the probability $P(O_t|S_t)$ of observation to be determined by the state in the same time point, while W_t and V_t represent the perturbations of the states and the observations, respectively. This model can be considered as a subtype of DBNs. The hidden variables include the regulatory motifs such as feedback loops and autoregulation.

A qualitative model for GRNs reconstruction employing Petri nets is proposed in Ref. [51]. This model, which is based on Boolean networks, uses minimization logic to transform Boolean rules into Petri nets. It overcomes the super-exponential number of states in the Boolean networks that depends on the number of nodes.

For hierarchical reconstruction of GRNs, Lee and Yang have proposed a model, which uses the clustering of gene expression data [52]. This model can infer regulatory relationships in large-scale networks. This method uses the recurrent neural networks to infer GRNs and applies the learning algorithm to update the main network parameters in discrete time steps.

Another method called FBN, applies the clustering of gene expression data to obtain modules (clusters) and then, it infers the gene regulatory relationships between clusters [53]. This method uses fuzzy clustering to reduce the search space for BNs learning.

4.7 Recent Models for Inference of GRNs by Integration of *A Priori* Knowledge

The GRNs inference based on gene expression data is very complex and difficult task, particularly because the present technical biological noise in microarray data should not be ignored. Furthermore, the number of experiments or conditions is lesser than the number of genes whose expression profiles are measured. Such shortcomings of the microarray data lead to unsatisfactory precision and accuracy of inferred networks, i.e., erroneous edges in inferred networks. To increase the accuracy and precision, employing other types of biological data and *a priori* knowledge such as knowledge obtained from scientific literature, protein–DNA interactions data, and other available databases is needed [54,55]. The capabilities of these models to reveal complex systems come from the model extensions by including *a priori* knowledge and using complementary and diverse data types [56].

One such method is suggested by Li, which combines qualitative and quantitative biological data for prediction of GRNs [57]. This method uses parallel processing and multiprocessor system to speed up the structural learning of BNs.

Based on comparison of the inference capabilities in Refs. [58,59], Ristevski and Loskovska [60] have suggested a novel model for GRNs inference, which performs in two stages. The first stage of the proposed model uses GGMs, because they are a good starting point to reveal the

"hub" genes. The GRNs structure G is represented by an adjacency matrix, whose entries G_{ij} can be either 1 or 0, which means presence or absence of a directed edge between ith and jth node of the network G, respectively. As a result of the first phase of the proposed model, a matrix of *a priori* knowledge *Gprior* is obtained, whose elements are computed by:

$$Gprior_{ij} = \begin{cases} \dfrac{1}{2}\dfrac{|pcor_{ij}| - pcor_{\min}}{pcor_{\max} - pcor_{\min}} + \dfrac{1}{2} \\ 0, \text{ if } |pcor_{ij}| < pcor_{\min} \text{ or edge direction is from } j \text{ to } i \end{cases} \tag{14}$$

where $pcor_{\max}$ and $pcor_{\min}$ are the maximum and minimum (set threshold) partial correlation coefficient, respectively [60]. This matrix of *a priori* knowledge *Gprior*, whose entries $Gprior_{ij} \in [0, 1]$, presents a basis for the second phase of the proposed model.

To integrate the *a priori* knowledge obtained in first phase, the second phase uses a function *Gprior′* as a measure of matching between the given network G and the obtained *a priori* knowledge *Gprior*[55]. The integration of *a priori* knowledge *Gprior* is according to prior distribution of the network structure G, which follows Gibbs distribution, given by the following equation [54,55]:

$$P(G|\beta) = \frac{e^{-\beta Gprior'(G)}}{Z(\beta)} \tag{15}$$

where the denominator is normalization constant calculated from all possible network structures Γ by the formula $Z(\beta) = \sum_{G \in \Gamma} e^{-\beta Gprior'(G)}$. In the second stage of the proposed model structure Bayesian learning using Markov chain Monte Carlo simulations is performed [60]. The flow chart of this model is illustrated in Fig. 4. This model has shown even better inference capabilities of networks inference, compared to Boolean networks, GGMs, and DBNs in the case when it was applied on experimental data sets as well as simulated datasets [59].

Beside gene expression data, the network inference using available heterogeneous -omics data, like transcriptomics, proteomics, interactomics, and metabolomics data, becoming more flexible. Integration of these data and using *a priori* knowledge can contribute to achieve more reliable comprehension of the regulatory relationships.

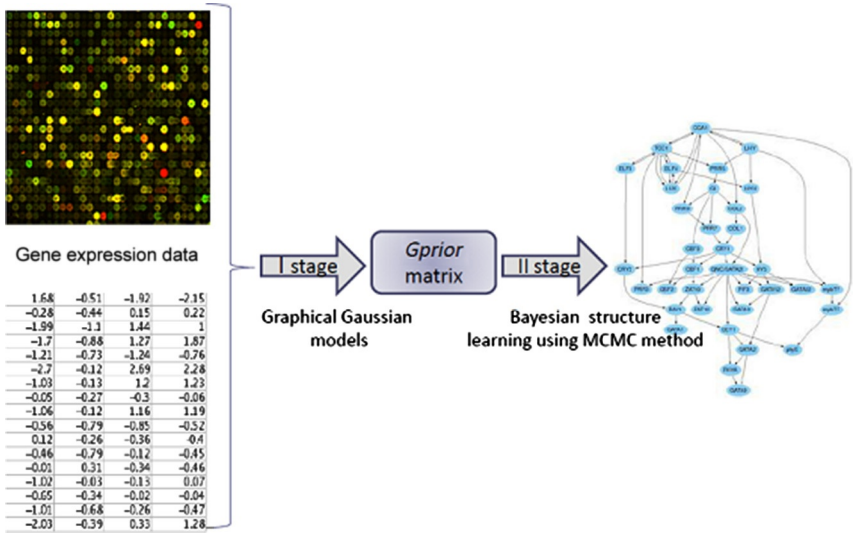


Figure 4 The flowchart of the two-stage inference model that integrates *a priori* knowledge [61].

5. COMPUTATIONAL APPROACHES FOR INFERENCE OF MicroRNA-MEDIATED REGULATORY NETWORKS

5.1 MicroRNA-Mediated Regulatory Networks

miRNAs repress translation of thousands of genes including TFs and hence significantly affect many types of cellular processes. Mostly, miRNAs and TFs regulate their mutual targets synchronously, composing complex networks named as *miRNA-mediated regulatory networks* [19]. The network is structured by the nodes and edges as connections between nodes. The nodes can represent TFs, miRNAs, target genes of miRNAs and TFs, and regulators of miRNAs, while the edges representing the regulatory relationships between the nodes. Figure 5 shows an miRNA-mediated regulatory networks, where full lines present transcriptional level regulation, whereas dashed edges present regulation on posttranscriptional level.

5.2 Types of Regulatory Relationships

There are several types of regulatory mechanisms [62]:

- TF regulates gene expression, denoted as *TF–gene* interaction. TFs can regulate gene expression by repressing or activating their target genes forming transcriptional regulatory networks.

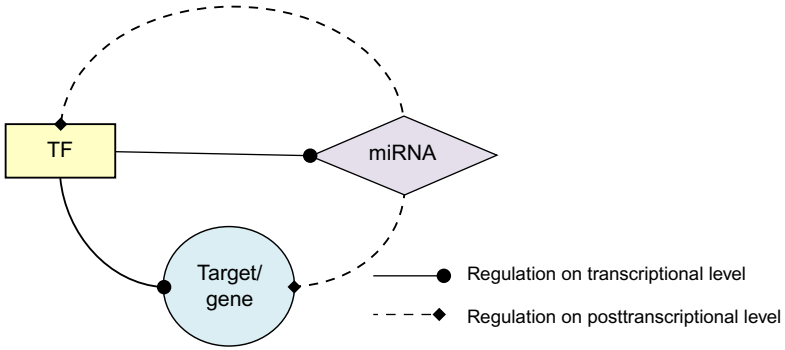


Figure 5 An miRNA-mediated regulatory network.

- TF regulates miRNA, denoted as *TF–miRNA* relationship. TFs are playing the key role in regulation of expression of miRNAs. Identifying the regulators of miRNAs is crucial to fully understand the miRNA-mediated regulatory networks.
- *MiRNA–TF* denotes that miRNA regulates TF. MiRNA–TF regulatory networks play an important role as functional entity in the cell regulatory processes in different cell types including cancer.
- MiRNA regulation on gene expression is denoted as *miRNA–gene*. Single miRNA targets hundreds of genes and even more, miRNAs often acting together by overlapping their targets. This makes miRNAs key elements in complex regulatory networks.
- *Gene–gene* interactions, which are covered in above-described GRNs.
- *TF–TF* or PPIs, forming PPI networks.

The most commonly used repositories, computational models, and tools for inferring/revealing of mutual regulatory interactions between TFs, miRNAs, and genes are listed in Table 1.

The most commonly used algorithms for the miRNAs target predictions are TargetScan 6.2 [63], Diana-microT [64], miRanda—microRNA Target Detection [65], miRTarbase [66], etc. These algorithms use different combination of features to identify whether particular sequence is a target of miRNAs or not. These algorithms exhaustively search for structural indication suggesting the presence of an interaction and they are often uncorrelated, which can lead to the nonoverlapping prediction results. Impossibility to integrate in a single model all possible interplaying options, which can influence the results of the miRNAs targeting and different type of sequence resources used as reference data set, is a shortcoming of the above-mentioned methods.

Table 1 Computational Models/Tools Used to Reveal Regulatory Relationships Between miRNAs, TFs, and Genes

Type of Interaction	Repositories/Databases or Computational Models/Tools
TF–gene	Match [88] CircuitsDB [89] Transcriptional Regulatory Element Database (TRED) [71] Human Transcriptional Regulation Interactions database (HTRIdb) [72] TRANScription FACtor database (TRANSFAC®) [20] mirGen 2.0 [90] ChIPBase [91]
TF–miRNA	Match [88] CircuitsDB [89] TransmiR [92] ChIPBase [91]
Gene–gene	Models listed in Section 4 of this chapter
miRNA–TF	TargetScan 6.2 [63] CircuitsDB [89]
miRNA–gene	Diana-microT [64] Tarbase 6.0 [73] miRecords [74] miRTarbase [66] TargetScan 6.2 [63] miRanda—microRNA Target Detection [65] CircuitsDB [89] doRiNa [93] miRNA—Target Gene Prediction at EMBL [94]
TF–TF/PPI	Biological General Repository for Interaction Datasets (BioGRID) [76] Human Protein Reference Database (HPRD) [77] IntAct [78] MIPS-Mammalian Protein–Protein Interaction Database (MIPS-MPPI) [79] KEGG [75]

Besides simple training-prediction model utilized in miRNAs targets prediction, diverse machine learning algorithms had also been employed such as support vector machines (SVMs) and random walk [19].

There are several computational methods to identify miRNA targets by using sequence analysis, which do not deal with the temporal dynamics of miRNA-regulated networks.

Schulz *et al.* had developed a tool named as MIRna Dynamic Regulatory Events Miner (mirDREM) [67]. Reconstructing of dynamic regulatory networks by mirDREM covers effects of TFs and miRNAs on their targets over time by integrating time series gene expression data with protein–DNA interaction data and the miRNAs expression level and their binding activities to predict their targets.

In Ref. [68], a feature dependency analysis across samples is performed in order to determine regulators (miRNAs and TFs) that significantly describe common and subtype-specific gene expression changes. To rank subtype-specific features, a score based on increasing in squared loss on samples, which belong to a subtype excluding the regulator from the learned model, is used.

Liao *et al.* had developed a data decomposition method—network component analysis (NCA) for reconstruction of regulatory signals and control strengths using partial and qualitative network connectivity information [69]. This method is applied on transcription regulatory network.

Ripoli *et al.* had proposed fuzzy logic approach to reveal miRNAs role to gene expression regulation [70].

Sun *et al.* developed an integrative framework for miRNA–TF regulatory networks construction for glioblastoma multiforme (GBM) [62]. The aim of this model is to identify GBM-specific miRNA–TF regulatory network and critical miRNA elements in a given pathway. Although GRNs modeling including TFs and miRNAs is very complex task, it will clarify both directly and indirectly regulatory mechanisms and mutual interactions between regulators and their targets. Computational models are very useful to uncover these complex regulatory relationships, particularly because the biological experiments' processes are very expensive and time demanding.

Lai *et al.* presented a systems' biology approach that combine data-driven modeling and model-driven experiments to examine the role of miRNA-mediated repression in GRNs [11]. Experimental approaches have constraints when handling complex biological systems regulations on transcriptional and posttranscriptional level by TFs and miRNAs. For mathematical modeling of GRNs, the authors integrated data from literature, biological databases, and experiments from different resources. Extracting experimental verified data for transcriptional level regulators is from literature or databases such as TRED [71], HTRIdb [72], TRANSFAC [20] or miRGen 2.0. Tarbase 6.0 [73], miRecords [74], and miRTarBase [66] provide information about miRNA–gene interactions, while KEGG [75], BioGRID [76], HPRD [77], IntAct [78], and MIPS-MPPI [79] are repositories for PPIs [11].

Similar to protein-coding genes, the miRNA transcription is closely controlled by upstream TFs. To discover the biological roles of the miRNAs, their targets' functions should be clarified. The miRNA expression and microarray data can be employed to examine specific TF–miRNA regulations. Self-regulation of miRNA genes is a mechanism that works as a buffering system for their expression modulation [80]. Feedback loops between specific miRNAs and the upstream TFs are another type of regulatory relationships in the miRNA-mediated regulatory networks [80]. These feedback circular connections are very important units in balancing TF and miRNA expression in the living organisms. Integration of more feedback circuits that exist between miRNAs and TFs increases the biological meaning, reliability, and complexity of the miRNA-mediated regulatory networks.

Le *et al.* suggested framework for complex regulatory network construction with three components: genes, miRNAs, and TFs [10]. Their model performed in three steps: data preparation, BN learning and integration, and network inference. The aim of the network inference is to find the global network for the subgraphs that show the interactions among miRNAs and TFs and network motifs composed of at least two regulators using network motifs finding algorithms.

Most of the existing computational models and tools find out statistically correlation and association between miRNAs and mRNAs. Correlation and association are not necessarily measures that provide an insight into causal gene regulation. Le *et al.* [81] refer the causal effects discovery as miRNA causal regulatory relationships. Their method uses miRNAs and mRNAs expression data and validates the causal effects of miRNAs on mRNAs assuming that miRNAs and mRNAs mutually interact. This method can be used to identify which gene set is casually regulated by particular miRNAs. Moreover, this method can theoretically infer the causal associations between every pair variables in the data set, for instance TFs, miRNAs, and mRNAs.

Pio *et al.* suggested semi-supervised learning approach for miRNA target prediction [82]. Their approach, beside the positive examples, utilizes unlabeled examples. The nontraditional classifier is learned using SVMs algorithm to learn to combine the results of several prediction algorithms. For validation of suggested model, a set of experimentally verified miRNA–miRNA interaction from miRTarBase and a set of miRNA target prediction are used.

Examination of protein changes following miRNA knockout (technique in which one of an miRNA is inoperative/"knocked out" of the

organism) or knockdown (technique by which the miRNAs expression of an organism's genome is reduced) is a helpful basis to assign the specific interaction function [83]. The possibility of indirect effects can be difficult to eliminate, because each miRNA can have many targets.

Laboratory experiments for characterization of individual regulatory interactions can reveal much more information about these interactions. By using expression data, target predictions and biological knowledge, candidates for regulatory interactions can be obtained by pointing on the messages predicted to be most responsive to the miRNA, coexpressed with the miRNA in the related cells, in those regulatory networks' nodes that are subject of interest.

5.3 Robustness of miRNA-Mediated Regulatory Networks

A good procedure is to disturb only particular interaction and monitor the phenotypic effects. A beneficial technique is disrupting a single miRNA–target interaction by using antisense reagents that hybridize to the target site, thus to disallow miRNA pairing. The phenotypic effects of these preserved interactions are very challenging task, especially their detection in the wet lab, although the simultaneously perturbation of all miRNA interactions by their knocking out usually does not have considerable phenotypic effects [83]. One of the more reasons of toleration of such disturbances for miRNA targets, which are gene regulatory proteins, is the regulatory network buffering. Many regulatory interactions, including many miRNA–target interactions, belong to complex regulatory networks with bifurcating pathways and feedback control enabling accurate reaction regardless of an inoperative node in the network. With this ability to buffer the effects of missing a node, such networks must be disturbed somewhere else before the missing miRNA interaction has evident phenotypic effects [83]. Perturbation of the miRNA node is expected to make the network susceptive to discover the importance of the rest of regulatory nodes.

Recent studies have uncovered that target hub genes, which carry vast number of TFBSs, are possible subject to massive regulation by many miRNAs. It means that nodes with more connections will more probably obtain new connections during time. The top genes with big number of both miRNA binding sites and TFBS are boosted in the functions related to development and differentiation of cells. Many of these target hub genes are transcription regulators, proposing a crucial pathway for miRNAs to indirectly regulate genes by repressing TFs [19].

miRNAs could be also target of hub genes. There is a class of miRNAs regulated by a large number of TFs, while the others are regulated by only a few TFs. miRNA expression profiles in embryonic developmental stages and adult tissues or cancer samples had disclosed that the miRNAs from the first class have higher expression levels in embryonic developmental stages, while the second class miRNAs are more expressed in adult tissues or cancer samples.

Regulator hub genes are more likely to have interactions with miRNAs, because they regulate large number of targets. miRNAs together with master TFs prefer to coregulate their targets. Regulator hub genes are very important constituents in the GRNs, since perturbations on them can disturb functions of numerous target genes. As miRNAs buffer stochastic perturbations, their preference to regulator hub genes could provide robustness of the regulatory network [19].

6. MODEL VALIDATION

Validation of inferred miRNA-mediated and GRNs represents an assessment of the accuracy of the inferred networks, compared to the available knowledge in so-called "gold standard" networks. To validate regulatory relationships inferred by using computational models, reliable biological experimental data are needed. In order to ensure more accurate and reliable model prediction, validation of the model with the data that are not included in the parameter estimation is required.

As commonly used validation criteria, receiver operating characteristic (ROC) curve and area under the ROC curve (AUC) are used.

In a graph between two nodes, it might be or not an edge, or using the formalism of machine learning, each edge (instance) of the network belongs to either positive (p) or negative (n) class, and classifier results belong to either class p or class n [84].

For a given two-class classifier and test samples, four cases can occur:

- True positive (TP), if the instance is positive and it is classified as positive;
- False negative (FN), if the instance is positive and it is classified as negative;
- True negative (TN), if the instance is negative and it is classified as negative;
- False positive (FP), if the instance is negative and it is classified as positive.

Based on the defined TP, FN, TN, and FP, following variables are defined [85,86]:

tpr (true-positive rate) (recall):

$$\text{tpr} = \frac{\text{TP}}{P} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{16}$$

fpr (false-positive rate):

$$\text{fpr} = \frac{\text{FP}}{N} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{17}$$

precision:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{18}$$

accuracy:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{P + N} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{19}$$

An ROC curve is a plot of a function where on the x-axis, the fpr—and on the y-axis—the tpr are applied, as shown in Fig. 6. In other words, an ROC curve represents the ratio between sensitivity and (1 - specificity) [87]. If the ROC curve is more above the line $y = x$, then classification is better.

To facilitate comparison of inference capabilities, the area under ROC curve (AUC) can also be used. The AUC is the area covered by the ROC curve with the x-axis, as shown in Fig. 6. The statistical meaning of the

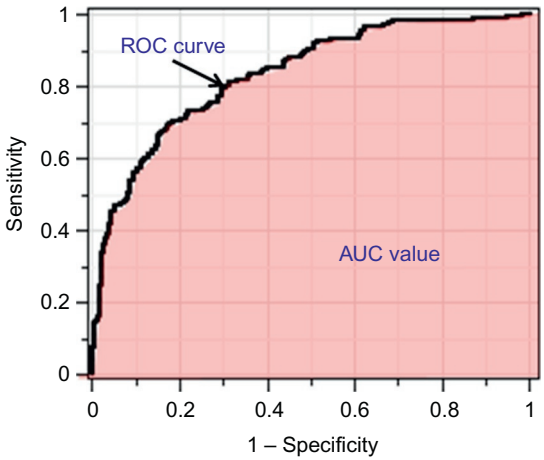


Figure 6 The ROC curve and AUC value.

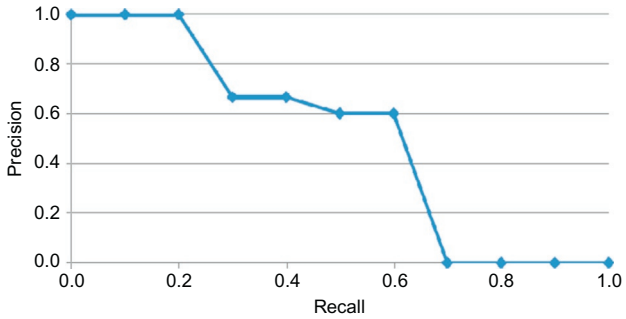


Figure 7 Precision–recall operating characteristics (P-ROC) curve.

AUC corresponds to the probability that the classifier will rank a randomly selected positive instance higher than a selected negative instance [84].

Alternatively, precision–recall operating characteristic (P-ROC, PR) curve can be used as validation criterion, where on the *x*-axis recall and on the *y*-axis precision are applied, as shown in Fig. 7.

7. CONCLUSION AND FURTHER WORKS

This overview of the most commonly used models for inference of miRNA-mediated and GRNs has shown that there is still a need for development of models that can integrate *a priori* knowledge in order to increase models inference capabilities. Such *a priori* knowledge significantly can improve the accuracy and biological reliability of the inferred regulatory networks. Inferred network edges, which are not present in the biological regulatory repositories and database, are indications for further experimental analysis to confirm or reject their presence as true regulatory relationships. Nevertheless, despite the advantages of described models, it can be concluded that there is no "silver bullet" inference model, which will have highly effective and accurate inference capabilities.

To assign the specific interaction function, a helpful base is protein changes examination of following miRNA knockout or knockdown. Laboratory experiments for characterization of individual regulatory interactions can reveal much more information about regulatory interactions.

By validation of the inferred networks, the main problem is the lack of reference "gold standard" networks. In addition, greater efforts should be made toward upgrading of existing regulatory databases with confirmed regulatory relationships between genes, miRNAs, TFs, and the other components involved in the cell regulatory processes.

Various -omics data uncover diverse perspectives of regulatory networks. Hence, integration of heterogeneous data and using biological *a priori* knowledge remains still challenging and partially unsolved issue in the inference of regulatory networks.

REFERENCES

[1] B. Ristevski, A survey of models for inference of gene regulatory networks, Nonlinear Anal. Modell. Control 18 (4) (2013) 444–465.
[2] M. Hecker, S. Lambeck, S. Toepfer, E. van Someren, R. Guthke, Gene regulatory network inference: data integration in dynamic models—a review, Biosystems 96 (2009) 86–103.
[3] P. Brazhnik, A. de la Fuente, P. Mendes, Gene networks: how to put the function in genomics, Trends Biotechnol. 20 (11) (2002) 467–472.
[4] Z. Shuang, L. Mo-Fang, Mechanisms of microRNA-mediated gene regulation, Sci. China C Life Sci. 52 (12) (2009) 1111–1116.
[5] Y. Nikolsky, J. Bryant (Eds.), Protein Networks and Pathway Analysis, Humana Press, Clifton, NJ, 2009, pp. 303–352.
[6] http://www.mirbase.org.
[7] Z. Wang, MicroRNA Interference Technologies, Springer-Verlag, New York, 2009.
[8] C. Li, Y. Feng, G. Coukos, L. Zhang, Therapeutic microRNA strategies in human cancer, AAPS J. 11 (4) (2009) 747–757.
[9] J.G. Joung, Z. Fei, Identification of microRNA regulatory modules in Arabidopsis via a probabilistic graphical model, Bioinformatics 25 (3) (2009) 383–393.
[10] T.D. Le, L. Liu, B. Liu, A. Tsykin, G.J. Goodall, K. Satou, J. Li, Inferring microRNA and transcription factor regulatory networks in heterogeneous data, BMC Bioinformatics 14 (2013) 92.
[11] X. Lai, A. Bhattacharya, U. Schmitz, M. Kunz, J. Vera, O. Wolkenhauer, A systems' biology approach to study microRNA-mediated gene regulatory networks, Biomed. Res. Int. 2013 (2013) Article ID 703849.
[12] A. Drakaki, D. Iliopoulos, MicroRNA gene networks in oncogenesis, Curr. Genomics 10 (2009) 35–41.
[13] S.K. Shenouda, S.K. Alahari, MicroRNA function in cancer: oncogene or a tumor suppressor? Cancer Metastasis Rev. 28 (3–4) (2009) 369–378.
[14] N.A. Kolchanov, T.I. Merkulova, E.V. Ignatieva, et al., Combined experimental and computational approaches to study the regulatory elements in eukaryotic genes, Brief. Bioinform. 8 (4) (2007) 266–274.
[15] P.J. Park, ChIP-seq: advantages and challenges of a maturing technology, Nat. Rev. Genet. 10 (10) (2009) 669–680.
[16] P. Collas (Ed.), Chromatin Immunoprecipitation Assays, Humana Press, Totowa, NJ, 2009, pp. 133–143.
[17] W. Zhao, E. Serpedin, E.R. Douherty, Recovering genetic regulatory networks from chromatin immunoprecipitation and steady-state microarray data, EURASIP J. Bioinform. Syst. Biol. 2008 (2008) 248747.
[18] H. Bolouri, Modeling genomic regulatory networks with big data, Trends Genet. 30 (5) (2014) 182–191.
[19] Y. Zhou, Interconnection of microRNA and transcription factor-mediated regulatory networks, in: Y. Gusev (Ed.), MicroRNA Profiling in Cancer—A Bioinformatics Perspective, Pan Stanford Publishing Pte. Ltd., Singapore, 2010
[20] E. Wingender, et al., The TRANSFAC system on gene expression regulation, Nucleic Acids Res. 29 (1) (2001) 281–283.

[21] A. Sandelin, W. Alkema, P. Engström, W.W. Wasserman, B. Lenhard, JASPAR: an open-access database for eukaryotic transcription factor binding profiles, Nucleic Acids Res. 32 (2004) D91–D94.
[22] M. Nicolau, M. Schoenauer, On the evolution of scale-free topologies with gene regulatory network model, Biosystems 98 (3) (2009) 137–148.
[23] S. Liang, S. Fuhrman, R. Somogyi, REVEAL, a general reverse engineering algorithm for inference of genetic network architectures, Pac. Symp. Biocomput. 3 (1998) 18–19.
[24] Y.K. Kwon, K.H. Cho, Analysis of feedback loops and robustness in network evolution based on Boolean models, BMC Bioinformatics 8 (2007) 430.
[25] I. Shmulevich, E.R. Dougherty, W. Zhang, From Boolean to probabilistic Boolean networks as models of genetic regulatory networks, Proc. IEEE 90 (11) (2002) 1778–1792.
[26] N. Friedman, M. Goldszmidt, Learning Bayesian networks with local structure, in: Proceedings of the 12th International Conference on Uncertainty in Artificial Intelligence, 1996, pp. 252–262.
[27] M. Grzegorczyk, D. Husmeier, Improving the structure MCMC sampler for Bayesian networks by introducing a new edge reversal move, Mach. Learn. 71 (2008) 265–305.
[28] N. Friedman, M. Linian, I. Nachman, D. Pe'er, Using Bayesian networks to analyze expression data, J. Comput. Biol. 7 (2000) 601–620.
[29] D. Husmeier, R. Dybowskiand, S. Roberts (Eds.), Probabilistic Modeling in Bioinformatics and Medical Informatics, Springer-Verlag, London, 2005.
[30] L. Kaderali, N. Radde, Inferring gene regulatory networks from expression data, Stud. Comput. Intell. 94 (2008) 33–74.
[31] R.E. Neapolitan, Learning Bayesian Networks, Prentice Hall, New York, 2003.
[32] A. Shermin, M.A. Orgun, Using dynamic Bayesian networks to infer gene regulatory networks from expression profiles, in: SAC '09, Proceedings of the 2009 ACM Symposium on Applied Computing, 2009, pp. 799–803.
[33] Y. Zhang, Z. Deng, H. Jiang, P. Jia, Inferring Gene Regulatory Networks from Multiple Data Sources via a Dynamic Bayesian Networks with Structural EM, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 204–214.
[34] R. Daly, K.D. Edwards, J.S. O'Neill, S. Aitken, A.J. Millar, M. Girolami, Using higher-order dynamic Bayesian networks to model periodic data from the circadian clock of *Arabidopsis thaliana*, in: PRIB 2009, 2009, pp. 67–78.
[35] K. Murphy, S. Mian, Modelling gene expression data using dynamic Bayesian networks, in: Technical Report, Computer Science Division, University of California, Berkeley, CA, 1999.
[36] M. Zou, S.D. Conzen, A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data, Bioinformatics 21 (1) (2005) 71–79.
[37] C.S. Kim, Bayesian Orthogonal Least Squares (BOLS) algorithm for reverse engineering of gene regulatory networks, BMC Bioinformatics 8 (2007) 251.
[38] K. Numata, S. Imoto, S. Miyano, A structure learning algorithm for inference of gene networks from microarray gene expression data using Bayesian networks, in: 7th IEEE International Conference on Bioinformatics and Bioengineering, BIBE 2007, 2007, pp. 1280–1284.
[39] A.A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R.D. Favera, A. Califano, ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context, BMC Bioinformatics 7 (Suppl. 1) (2006) S7.
[40] J. Schäfer, K. Strimmer, Learning large-scale graphical Gaussian models from genomic data, CPP776, in: Science of Complex Networks: From Biology to the Internet and WWW; CNET 2004, 2004, pp. 263–276.

[41] J. Schäfer, K. Strimmer, A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics, Stat. Appl. Genet. Mol. Biol. 4 (1) (2005) Article 32.
[42] F. Jaffrezic, G.T. Klopp, Gene network reconstruction from microarray data, BMC Proc. 3 (Suppl. 4) (2009) S12.
[43] J. Schäfer, R. Opgen-Rhein, K. Strimmer, Reverse engineering genetic networks using the GeneNet package, R. News 6/5 (2006) 50–53.
[44] L.F.A. Wessels, E.P. van Someren, M.J.T. Reinders, A comparison of genetic network models, Pac. Symp. Biocomput. 6 (2001) 508–519.
[45] J. Li, X.S. Zhang, An optimization model for gene regulatory networks reconstruction with known biological information, in: 1st International Symposium on Optimization and Systems Biology (OSB'07), Beijing, China, 2007, pp. 35–44.
[46] J. Gebert, N. Radde, G.-W. Weber, Modeling gene regulatory networks with piecewise linear differential equations, Eur. J. Oper. Res. 181 (3) (2007) 1148–1165.
[47] M.S.B. Sehgal, I. Gondal, L.S. Dooley, CF-GeNe: fuzzy framework for robust gene regulatory network inference, J. Comput. 1 (7) (2006) 1–8.
[48] A. Fujita, J.R. Sato, H.M. Garay-Malpartida, R. Yamaguchi, S. Miyano, M.C. Sogayar, C.E. Ferreira, Modeling gene expression regulatory networks with the sparse vector autoregressive model, BMC Syst. Biol. 1 (2007) 39.
[49] A. Brazma, T. Schlitt, Reverse engineering of gene regulatory networks: a finite state linear model, Genome Biol. 4:P5 (2003) 1–31.
[50] Z. Li, S.M. Shaw, M.J. Yedwabnick, C. Chan, Using a state-space model with hidden variables to infer transcription factor activities, Bioinformatics 22 (6) (2006) 747–754.
[51] L.J. Steggles, R. Banks, A. Wipat, Modelling and analysing genetic networks: from Boolean networks to Petri nets: Technical Report, University of Newcastle upon Tyne, Computing Science, 2006, pp. 127–141.
[52] W.-P. Lee, K.-C. Yang, A clustering-based approach for inferring recurrent neural networks as gene regulatory networks, Neurocomputing 71 (2008) 600–610.
[53] F. Wang, D. Pan, J. Ding, A new approach combined fuzzy clustering and Bayesian networks for modeling gene regulatory networks, in: Proceedings of the 2008 International Conference on BioMedical Engineering and Informatics, 1, 2008, pp. 29–33.
[54] A.V. Werhli, D. Husmeier, Reconstructing gene regulatory networks with Bayesian networks by combining expression data with multiple sources of priori knowledge, Stat. Appl. Genet. Mol. Biol. 6 (2007) Article 15.
[55] S. Imoto, T. Higuchi, T. Goto, K. Tashiro, S. Kuhara, S. Miyano, Combining microarrays and biological knowledge for estimating gene networks via Bayesian networks, in: Proceedings of the Computational Systems Bioinformatics (CSB '03), 2003, pp. 104–113.
[56] T. Ideker, J. Dutkowski, L. Hood, Boosting signal-to-noise in complex biology: prior knowledge is power, Cell 144 (6) (2011) 860–863.
[57] S. Li, Integrate qualitative biological knowledge to build gene networks by parallel dynamic Bayesian network structure learning, in: Bioinformatics and Bioengineering, BIBE 2007, 2007, pp. 87–92.
[58] B. Ristevski, S. Loskovska, ROC curves comparison of inferred gene regulatory networks, in: 13th International Multiconference Information Society 2010, Intelligent Systems, Ljubljana, Slovenia, 2010, pp. 39–42.
[59] B. Ristevski, S. Loshkovska, A comparison of models for gene regulatory networks inference, in: 2nd International Conference, ICT Innovations 2010, Ohrid, Republic of Macedonia, 2010, pp. 59–68.
[60] B. Ristevski, S. Loskovska, A novel model for inference of gene regulatory networks, HealthMed J. 5 (6) (2011) 2024–2033.

[61] B. Ristevski, Inference of gene regulatory networks by integrating prior knowledge, in: R. Hofestädt, N. Kolchanov (Eds.), German/Russian Network of Computational Systems Biology: Report 2012, Berichte aus der Medizinischen Informatik und Bioinformatik, Shaker-Verlag, Aachen, 2012.
[62] J. Sun, X. Gong, B. Purow, Z. Zhao, Uncovering microRNA and transcription factor mediated regulatory networks in glioblastoma, PLoS Comput. Biol. 8 (7) (2012) e1002488.
[63] B.P. Lewis, C.B. Burge, D.P. Bartel, Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets, Cell 120 (1) (2005) 15–20.
[64] M. Maragkakis, T. Vergoulis, et al., DIANA-microT web server upgrade supports fly and worm miRNA target prediction and bibliographic miRNA to disease association, Nucleic Acids Res. 39 (2011) W145–W148.
[65] D. Betel, M. Wilson, A. Gabow, D.S. Marks, C. Sander, The microRNA.org resource: targets and expression, Nucleic Acids Res. 36 (Database Issue) (2008) D149–D153.
[66] S.D. Hsu, Y.T. Tseng, S. Shrestha, et al., miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions, Nucleic Acids Res. 42 (Database Issue) (2014) D78–D85.
[67] M.H. Schulz, K.V. Pandit, C.L.L. Cardenas, N. Ambalavanan, N. Kaminskiand, Z.B. Joseph, Reconstructing dynamic microRNA-regulated interaction networks, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 15686–15691.
[68] M. Setty, et al., Inferring transcriptional and microRNA-mediated regulatory programs in glioblastoma, Mol. Syst. Biol. 8 (2012) Article 605.
[69] J.C. Liao, et al., Network component analysis: reconstruction of regulatory signals in biological systems, Proc. Natl. Acad. Sci. U.S.A. 100 (26) (2003) 15522–15527.
[70] A. Ripoli, et al., The fuzzy logic of microRNA regulation: a key to control cell complexity, Curr. Genomics 11 (5) (2010) 350–353.
[71] F. Zhao, Z. Xuan, L. Liu, M.Q. Zhang, TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies, Nucleic Acids Res. 33 (Database Issue) (2005) D103–D107.
[72] L.A. Bovolenta, M.L. Acencio, N. Lemke, HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions, BMC Genomics 13 (Database) (2012) 405.
[73] T.I. Vergoulis, P. Vlachos, et al., Tarbase 6.0: capturing the exponential growth of miRNA targets with experimental support, Nucleic Acids Res. 40 (D1) (2012) D222–D229.
[74] F. Xiao, Z. Zuo, G. Cai, S. Kang, X. Gao, T. Li, miRecords: an integrated resource for microRNA-target interactions, Nucleic Acids Res. 37 (Database Issue) (2009) D105–D110.
[75] M. Tanabe, M. Kanehisa, Using the KEGG database resource, Curr. Protoc. Bioinformatics 38(1) (2012) 12:1.12.1–1.12.43.
[76] C. Stark, B.J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, M. Tyers, BioGRID: a general repository for interaction datasets, Nucleic Acids Res. 1 (34) (2006) 535–539.
[77] T.S.K. Prasad, et al., Human protein reference database—2009 update, Nucleic Acids Res. 37 (2009) D767–D772.
[78] S. Orchard, et al., The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases, Nucleic Acids Res. 42 (2014) D358–D363.
[79] P. Pagel, S. Kovac, M. Oesterheld, et al., The MIPS mammalian protein-protein interaction database, Bioinformatics 21 (6) (2005) 832–834.
[80] Y. Meng, C. Shao, M. Chen, Toward microRNA-mediated gene regulatory networks in plants, Brief. Bioinform. 12 (6) (2011) 645–659.

[81] T.D. Le, L. Liu, A. Tsykin, G.J. Goodall, B. Liu, B.-Y. Sun, J. Li, Inferring microRNA–mRNA causal regulatory relationships from expression data, Bioinformatics 29 (6) (2013) 765–771.
[82] G. Pio, D. Malerba, D. D'Elia, M. Ceci, Integrating microRNA target predictions for the discovery of gene regulatory networks: a semi-supervised ensemble learning approach, BMC Bioinformatics 15 (Suppl. 1) (2014) S4.
[83] D.P. Bartel, MicroRNA target recognition and regulatory functions, Cell 136 (2) (2009) 215–233.
[84] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (2006) 861–874.
[85] P. Baldi, S. Brunak, Bioinformatics: The Machine Learning Approach, The MIT Press, London, 2001.
[86] J. Huang, C.X. Ling, Using AUC and accuracy in evaluating learning algorithms, IEEE Trans. Knowl. Data Eng. 17 (3) (2005) 299–310.
[87] L.M. de Campos, A scoring function for learning Bayesian networks based on mutual information and conditional independence tests, J. Mach. Learn. Res. 7 (2006) 2149–2187.
[88] A.E. Kel, E. Gossling, I. Reuter, E. Cheremushkin, O.V. Kel-Margoulis, et al., MATCH: a tool for searching transcription factor binding sites in DNA sequences, Nucleic Acids Res. 31 (2003) 3576–3579.
[89] O. Friard, A. Re, D. Taverna, M. De Bortoli, D. Corá, CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse, BMC Bioinformatics 11 (2010) 435.
[90] P. Alexiou, T. Vergoulis, et al., miRGen 2.0: a database of microRNA genomic information and regulation, Nucleic Acids Res. 38 (Database Issue) (2010) D137–D141.
[91] J.H. Yang, J.H. Li, S. Jiang, H. Zhou, L.H. Qu, ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data, Nucleic Acids Res. 41 (2013) D177–D187.
[92] J. Wang, M. Lu, C. Qiu, Q. Cui, TransmiR: a transcription factor-microRNA regulation database, Nucleic Acids Res. 38 (Suppl. 1) (2010) D119–D122.
[93] G. Anders, S.D. Mackowiak, M. Jens, et al., doRiNA: a database of RNA interactions in post-transcriptional regulation, Nucleic Acids Res. 40 (Database Issue) (2012) D180–D186.
[94] J. Brennecke, A. Stark, R.B. Russell, S.M. Cohen, Principles of microRNA-target recognition, PLoS Biol. 3 (3) (2005) e85.

ABOUT THE AUTHORS

Blagoj Ristevski is currently an assistant professor at the St. Kliment Ohridski University in Bitola, Republic of Macedonia. He is a head of the Department of Information Systems at Faculty of Information and Communication Technologies—Bitola. From 2002 to 2007, he worked as a teaching and research assistant at Faculty of Technical Sciences in Bitola, afterward until 2014 he worked as an assistant professor at Faculty of Administration and Information Systems Management—Bitola, Republic of Macedonia.

He received his Ph.D. degree in Computer Science and Engineering from the University Ss. Cyril and Methodius in Skopje, Faculty of Electrical

Engineering and Information Technology, Macedonia with the doctoral thesis "Modelling, Analysis and Validation of Gene Regulatory Networks" in 2011. He received Magister degree in Computer Science and Engineering and Graduate Engineer degree in Electrical Engineering from the same Faculty in 2007 and 2001, respectively.

He is an author and a coauthor of many research papers published in international journals and conference proceedings. Since 2009, he has done several research stays at research centers at University of Göttingen and University of Bielefeld, Germany; University of Groningen, The Netherlands; and University of Sofia, Bulgaria. His current research interests are Bioinformatics, Gene Regulatory Networks, Reverse Engineering, Database Management Systems, Algorithms, and Data Mining and Cryptography.