

## REVIEW

## Open Access



# Analysis of miRNA, mRNA, and TF interactions through network-based methods

Pietro H Guzzi<sup>1\*†</sup>, Maria Teresa Di Martino<sup>2†</sup>, Piersandro Tagliaferri<sup>2</sup>, Pierfrancesco Tassone<sup>2,3</sup> and Mario Cannataro<sup>1</sup>

## Abstract

Recent findings have elucidated that the regulation of messenger RNA (mRNA) levels is due to the synergistic and antagonist actions of transcription factors (TFs) and microRNAs (miRNAs). Mutual interactions among these molecules are easily modeled and analyzed using graphs whose nodes are molecules, and directed edges represent the associations among them. In particular, small subgraphs having three nodes also referred to as feed-forward loops (FFLs) or regulatory loops play a crucial role in many different diseases, such as cancer. Available technological platforms enable the investigation of only a single aspect of these mechanisms, e.g., the quantification of levels of mRNA or miRNA. Consequently, there exist different data sources for investigating some aspects of this problem, e.g., miRNA-mRNA or TF-mRNA associations. The comprehensive analysis is made possible only by the integration and the analysis of these data sources. Currently, the interest of researchers in this area is growing, the number of projects is increasing, and the number of challenges and issues for computer scientists is considerable. The need for an introductory survey from a computer science point of view consequently arises. This survey starts by discussing general concepts related to production of data. Then, main existing approaches of analysis are presented and discussed. Future improvements and challenges are also discussed.

**Keywords:** mRNA; miRNA; Transcription factor; Network analysis; Data integration

## Review

### Introduction

The development of novel technological platforms in molecular biology has produced a large amount of data about different aspects of the *omic* world [1]. Consequently, the need for the development of novel approaches and methods to manage, store, and analyze this data arose [2–4]. In particular, this has caused the rise of a novel discipline, often referred to as *computational systems biology* or *network systems biology*, in which computer science, bioinformatics, and mathematical modeling play a synergistic role in the interpretation of large datasets belonging to different data sources [5, 6]. Network systems biology aims to discover basic principles of mutual interactions (or interplay) among different biological molecules (such as proteins, genes, or small fragments of non-coding nucleic

acids) under the assumption that the information gathered from integrated analysis is higher than in the separate study of any data source [7, 8].

The flow of information in this field starts from technological platforms that produce different data about molecular biology as depicted in Fig. 1. Examples of such platforms are microarray for studying the expression of messenger RNA (mRNA) [9, 10] and microRNA (miRNA) [11], genomic microarrays for studying copy number variations (CNV) or single nucleotide polymorphisms (SNP), novel microarrays for studying non-coding RNAs (e.g., miRNA), genomic arrays for pharmacogenomics studies [12, 13], and novel next-generation sequencing (NGS) techniques. Classical approaches of analysis have produced a lot of information about the role of single class of molecules, but there is a lack of introduction of novel techniques aiming to analyze the interplay of molecules by integrating these data sources into a single comprehensive one [14, 15].

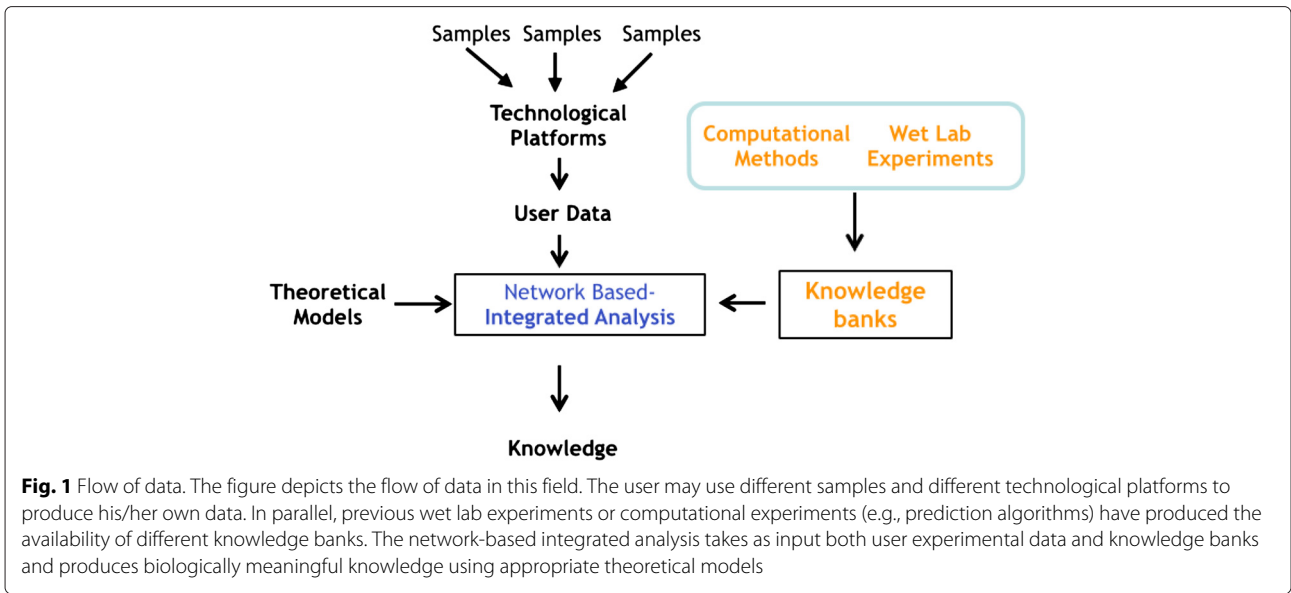
Here we focus on the study of complex mechanisms of the regulation of gene expression. Recent results have

\*Correspondence: [hguzzi@unicz.it](mailto:hguzzi@unicz.it)

†Equal contributors

<sup>1</sup>Department of Medical and Surgical Sciences, Magna Graecia University, Catanzaro, Italy

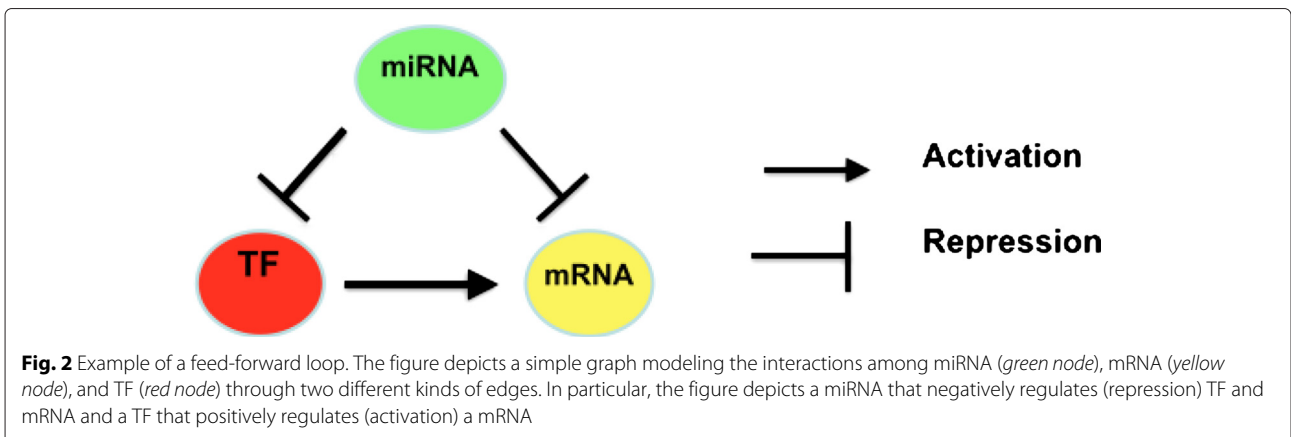
Full list of author information is available at the end of the article



confirmed that the transcription of mRNA into proteins is a multi-step process in which different molecules play a synergistic role [16]. In particular, miRNAs and transcription factors (TFs) play a direct role in the regulation of gene expression that results in variable levels of gene transcripts and proteins. Since there is not a direct technological platform to investigate these complex interactions, the integration of different datasets will become increasingly important as elucidated in the work by Muniategui et al. [17]. The integration of these datasets may be easily made by using models from graph theory. Consequently, it is possible to build comprehensive graphs in which nodes are miRNAs, mRNAs, and TFs, and directed edges connecting them represent the action of the molecules as depicted in Fig. 2. Edges are subdivided into (i) activation edges which represent a molecule whose action results in an increasing of the level of another one, and (ii) inhibition edges which connect a molecule whose action

results in a decreasing of the levels of another one. Usually, edges connect a miRNA to a mRNA or a TF, or a TF to a gene [18]. Starting from this formalism, it is possible to extract small connected subgraphs with three different classes of nodes, representing feedback loops and feed-forward loops (FFLs) in which miRNAs participate together with transcription factors as depicted in Fig. 2.

The efforts of the scientific community have produced a set of projects regarding integrated data analysis based on graph theory. Because in recent years much work has been made in the study of TF and miRNA co-regulation, we think that there is a need to present in a systematic catalogue all the available methods. In this review, we summarize the types of regulatory networks. Future challenges and perspectives on TF-miRNA co-regulation are also discussed. Moreover, as a specific contribution of the presented work, we extended the work of



[19] by discussing some recent approaches and by using a computer science perspective.

## Background

### *mRNA, miRNA, and transcription factor interactions*

As stated in the central dogma of molecular biology, genes guide protein synthesis through mRNA molecules. Since the information contained in genes cannot be directly translated into proteins, information is at first transcribed into mRNA molecules. Each molecule of mRNA encodes the information for one protein. The mRNA molecules migrate through the nuclear envelope to the cytoplasm, where they are translated by the rRNA of ribosomes. Finally, each mRNA is translated into a polymer of amino acids: a protein. In an ideal case, the quantity of mRNA molecules should be directly related to the quantity of the related protein. In such a way, the investigation of the quantity of mRNA through microarray technology should enable the investigation of the quantity of produced proteins. Unfortunately, as suggested by experimental evidences, this process is made complex by the presence of regulatory mechanisms that directly influence the production of proteins. In particular, recent findings have elucidated the role of two main classes of molecules that influence positively and negatively the protein synthesis: miRNA and TF [20].

miRNA refers to a set of small RNA molecules composed of 21–23 nucleotides that do not encode any protein but participate as regulators in protein formation [21]. Recent studies demonstrated that miRNAs play an essential role in carcinogenesis because the disgregation of their activity may cause the development of tumor invasion and migration [22]. miRNAs also act as a possible new target for molecular target therapy of various cancers [23, 24]. Thus, there is an increasing interest for miRNA studies in clinical applications such as in serological diagnosis and molecular-targeted therapeutics [25]. TFs are modular proteins that regulate gene transcription through binding to the promoter region of target genes by their DNA-binding domains. In such a way, TFs may increase the gene expression levels and the consequent level of produced proteins.

### *Interaction databases*

The interaction databases used by the works here surveyed fall into three main classes:

- Databases storing associations among miRNA and genes, i.e., storing which genes are targeted by miRNAs
- Databases storing the associations among TF and genes, i.e., storing which genes are targeted by TFs
- Databases storing the associations among TF and miRNA, i.e., storing which TFs are targeted by miRNAs

All of these databases may store both confirmed associations, i.e., associations supported by experimental evidences, and predicted associations, i.e., associations that are predicted by computational methods. The current scenario presents some main characteristics: (i) the number of confirmed associations is in general less than that of predicted ones, (ii) the number of false positives (i.e., not real associations) is considerable, and (iii) the level of overlap among databases is low. Consequently, all the approaches consider different data sources and integrate them in order to enhance the quality of considered associations.

The association among miRNAs and their target genes, i.e., genes up- or downregulated, is currently an increasing research area. Currently, there exist different prediction softwares, i.e., softwares that can predict possible genes regulated by a miRNA through machine learning approaches, and different technological platforms that are able to confirm these results in wet lab experiment [26]. As a result of the joint effort (both in silico and wet lab experiments), several databases that store the association among miRNAs and mRNAs are now available. Examples of these databases are Microcosm [27], microRNA.org [28], DIANA-microT [29], miRDB [30], PicTar [31], PITA [32], RNA22 [33], and TargetScan [34].

Similar to miRNAs, the complete enumeration of all the interactions among TF and genes is far to be complete. Thus, information stored into databases is quite incomplete. Main experiments used for discovering TF-gene relations are immunoprecipitations (ChIP) followed by sequencing (ChIP-seq) or by microarray hybridization (ChIP-chip) [35]. Both techniques enable a high-throughput discovery of relations, but usually, they also generate a large number of false positives [36]. In parallel to these techniques, we should recall main computational approaches for predicting TF and for retrieving resulting information from databases.

For instance, the TRANSFAC database [37] is one of the main resources of experimentally verified TF targets from publications or databases. Similarly, CHEA [38] stores ChIP-seq and ChIP-chip data related to TF targets generated by different projects. The availability of different data sources with different reliabilities causes the need of integration of several of these methods and data to obtain comprehensive and accurate TF targets [18].

The third main knowledge source used by the works discussed in this survey is represented by databases storing information related to the regulation of miRNAs by TFs. The number of TF-miRNA regulation databases is lower than the number of the other two kinds of databases, because this approach is the youngest area of research. Examples of databases are TransmiR [39], TransFac [40], TargetScan [34], and PicTar [31].

### Network-based approaches for integrated analysis

#### A general model for integrating miRNA, mRNA, and TF data

All the approaches here discussed present some main characteristics. They have an internal knowledge base of associations extracted from literature and databases. The knowledge base is a comprehensive graph of associations. Nodes of these graphs fall into three classes representing respectively miRNAs, mRNAs, and TFs. Edges fall into two classes: activation and inhibition edges. A directed activation/inhibition edge connects a molecule that increases/decreases the level of another one. Main differences among the approaches are represented by the association databases that are used. This internal knowledge base is used for guiding the analysis of experimental data. Usually, experimental data are both miRNA and mRNA expression data taken from a pool of samples extracted from patients in case-control or time series experiments. For each patient, both mRNA and miRNA data are produced. Consequently, those experiments produced two expression vectors from each mRNA  $m_i$  and each miRNA  $m_{ij}$ . Then, the expression vectors are correlated using some relatedness measures, such as Pearson correlation  $\rho(m_i, m_{ij})$  for each mRNA-miRNA pair.

Then, data of knowledge bases are used to build the association graph from experimental data. This association graph is then mined to find small graphs representing FFL. The rest of the section presents some main approaches currently available for academic users. We should note that the literature also reports an approach of integration available for Ingenuity Pathway Analysis

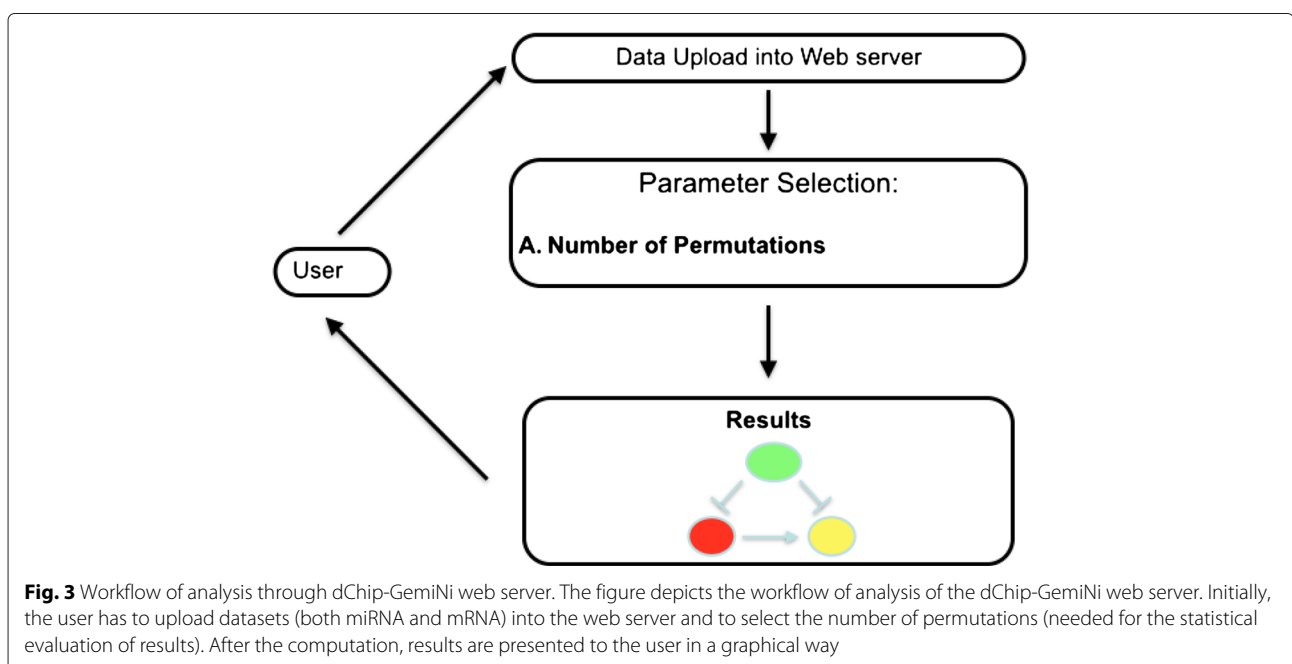
software that we do not report here since it is not freely available [41].

#### dChip-GemiNi (Gene and miRNA Network-based Integration)

dChip-GemiNi (Gene and miRNA Network-based Integration) [42] is a web server freely available for academic users which is able to integrate and analyze paired miRNA-mRNA expression data. The server side is written using the R programming language. Users may also download the source code for running it in a local environment. The ability of dChip-GemiNi has been tested by using some paired miRNA-mRNA datasets of solid cancers (liver, kidney, prostate, lung, and germ cell), and results are discussed in [42].

The workflow of analysis that has been used to build dChip-GemiNi contains four steps:

1. Initially, publicly available databases (e.g., TargetScan [34] for miRNA-mRNA association and data from TRANSFAC [40] for TF binding sites) have been used to construct TF-miRNA-gene networks, i.e., networks in which nodes are miRNA, genes, and TF, and edges represent the *regulates* relationship among them (e.g., a miRNA is connected to the target genes and a TF is connected to the target genes).
2. Then, experimental data (i.e., gene and miRNA expression profiles) are collected from publicly databases (e.g., GEO [43]).
3. Resulting networks (obtained in steps 1 and 2) are mined to extract significant motifs referred to as FFL motifs, i.e., small connected graphs in which there



**Fig. 3** Workflow of analysis through dChip-GemiNi web server. The figure depicts the workflow of analysis of the dChip-GemiNi web server. Initially, the user has to upload datasets (both miRNA and mRNA) into the web server and to select the number of permutations (needed for the statistical evaluation of results). After the computation, results are presented to the user in a graphical way

exist three different nodes (TF, miRNA, and mRNA) (see Fig. 2 for an example of FFL motifs).

- Data of step 1 are used to further validate the statistical relevance of results through an ad hoc defined network motif score (NMS). The NMS is a function of multiple scores, including TF and miRNA binding scores to their target sequences, differential expression  $P$  values of the FFL components between normal and cancer tissues, and TF and miRNA's target enrichment in differentially expressed genes and miRNAs.

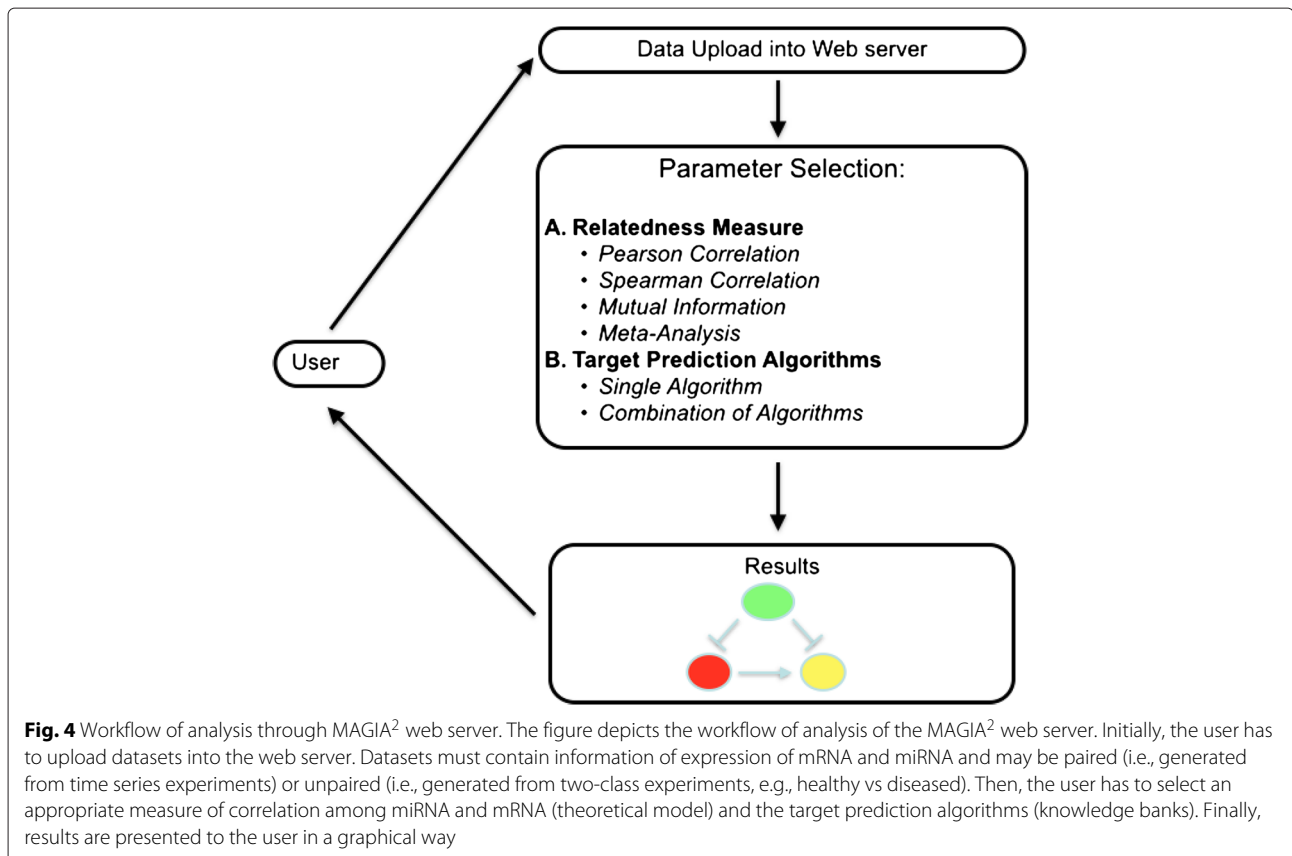
As depicted in Fig. 3, when the user has to analyze experimental data, he/she has to start from two vectors of expression levels (one for mRNA and one for miRNA) obtained from experiments analyzing two conditions, e.g., normal and cancer. Data may be paired (i.e., for each sample, there exist both mRNA and miRNA) or non-paired (i.e., data belong to the same class but not to the same samples). Then, the user has to upload them into the web server and he/she receives as output a list of significant FFLs that are altered with respect to those used as the null model. dChip-GemiNi is also able to individuate FFLs consisting of TFs (i.e., genes that are able to regulate the expression of other genes), miRNAs, and their common

target genes. In such a way, it can discover knowledge that cannot be discovered by the classical analysis. Experimental data are compared with respect to known associations among miRNAs, mRNAs, and TFs obtained from the literature and stored into the web server. TFs derived from literature are used as a null model to statistically rank predicted FFLs from the experimental data.

#### MAGIA<sup>2</sup> web server

MAGIA<sup>2</sup> [44] is the evolution of the MAGIA web tool for the integrated analysis of both genes and microRNA. MAGIA<sup>2</sup> is deployed as a freely available web server. To build association networks, MAGIA<sup>2</sup> uses eight different databases of miRNA/mRNA associations: Microcosm [27], mirorna.org [28], DIANA-microT [29], miRDB [30], PicTar [31], PITA [32], RNA22 [33], and TargetScan [34]. Such predictors are used to build the null models, i.e., associations that are known by literature. Regarding TFs, MAGIA<sup>2</sup> uses experimentally validated TF-miRNA interactions reported in mirGen2.0 [45] and TransmiR [39], whereas TF-gene interactions are obtained from ECRbase database [46].

The analysis through the MAGIA<sup>2</sup> web server starts by uploading data into the web server, usually a matrix for gene/transcripts and one for miRNA expression data.



**Fig. 4** Workflow of analysis through MAGIA<sup>2</sup> web server. The figure depicts the workflow of analysis of the MAGIA<sup>2</sup> web server. Initially, the user has to upload datasets into the web server. Datasets must contain information of expression of mRNA and miRNA and may be paired (i.e., generated from time series experiments) or unpaired (i.e., generated from two-class experiments, e.g., healthy vs diseased). Then, the user has to select an appropriate measure of correlation among miRNA and mRNA (theoretical model) and the target prediction algorithms (knowledge banks). Finally, results are presented to the user in a graphical way

Data may belong to time series experiments in which for each sample there exists a pair miRNA/mRNA experiment (referred to as matched data), or a two-class experiment (referred to as un-matched data). Then, users have to select an association measure among mRNA and miRNA, i.e., a measure of relatedness among expression values. For matched experiments, MAGIA<sup>2</sup> offers the following measures: Pearson linear correlation, Spearman rank-based correlation, and an association measure based on information theory for time series experiments (referred to as matched). Diversely, for un-matched design, only a meta-analysis is possible.

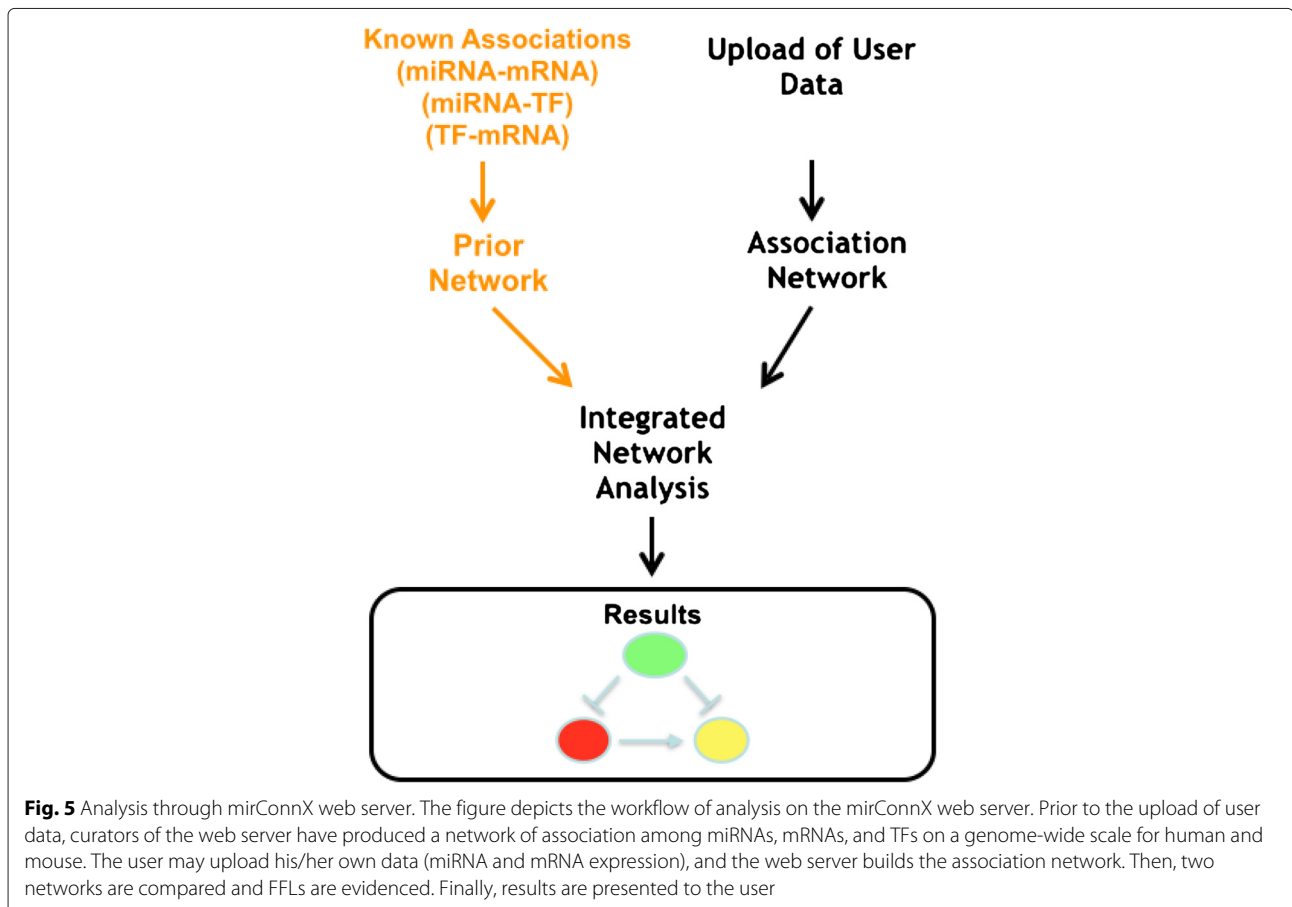
The choice among measures is strictly dependent on the characteristics of data: for non-normally distributed data and/or small sample size experiments (e.g., 3–5), it is suggested to use Spearman correlation, which is a non-parametric rank-based linear measure, whereas for normally distributed data and medium-large sample size (more than 5 samples), authors suggest the use of the Pearson linear correlation measure; finally, for large sample size (more than 20 samples), it is suggested to use mutual information that is an information measure quantifying the mutual dependence of variables.

Diversely, for un-matched experiments, i.e., experiments in which samples are subdivided into two classes, the web server offers the meta-analysis approach that is based on the combination of *P* values of differential expression, separately for genes and miRNAs across sample classes. The user may also choose which databases are used to extract associations from those explained so far. In case of choice of multiple databases, search results may contain their union or intersection. Finally, experimentally derived associations are compared to those contained in the databases, and two kinds of networks are derived as depicted in Fig. 4.

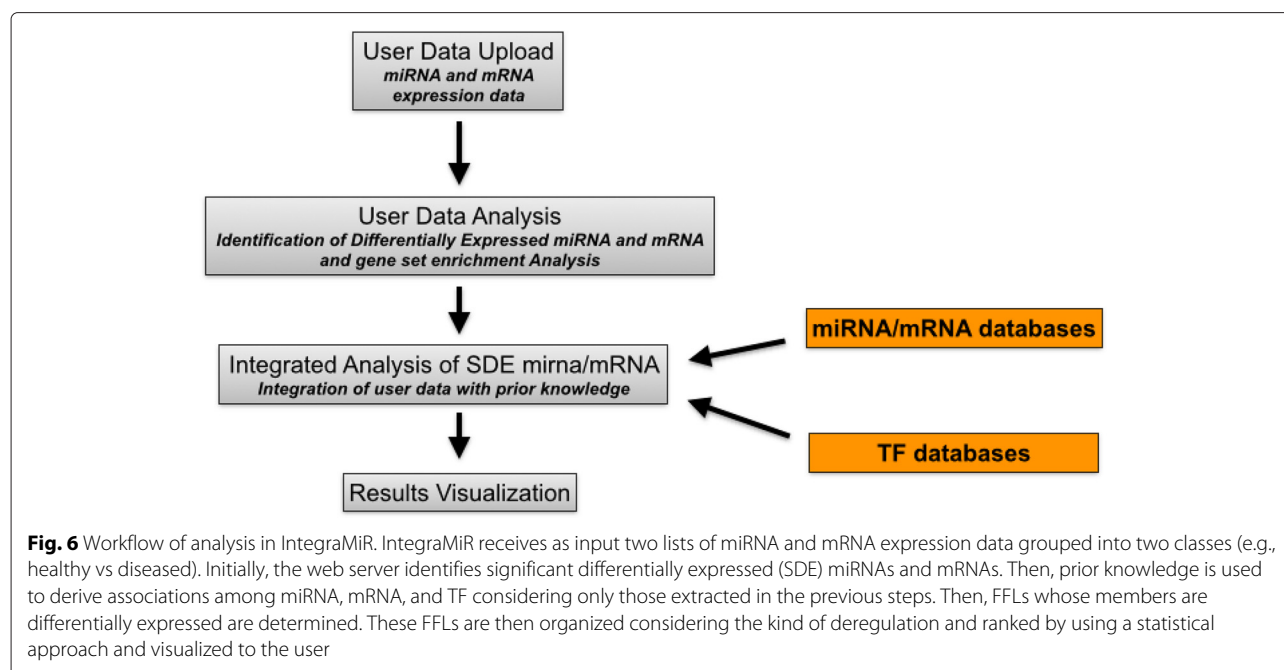
#### *mirConnX*

mirConnX [47] is based on a broader perspective with respect to the previous approaches since it uses a genome-wide approach. Unfortunately, it enables only the analysis of data of two organisms: human and mouse. The workflow of analysis is based on the comparison of two networks of associations among genes, TFs, and mRNAs, as depicted in Fig. 5.

The first network, used as a null model, is derived from the analysis of databases and literature. In this network, nodes are miRNAs, TFs, and genes, and an edge connects







two nodes when an association has been found. Examples of associations are (i) a miRNA that regulates a gene or a TF, or (ii) a TF that regulates a gene. Edges are weighted, and the weight reflects the strength of the association. miRNA targets are derived by integrating results stored in PITA [32], miRANDA [48], TargetScan 5.0 [34], RNAhybrid [49], Pictar [31], TarBase [50], and miRecords [51] databases. Similarly, associations among TF and genes are derived by integrating predictions stored in JASPAR [52] and TRANSFAC [37]. The integration step is based on a mathematical model which is able to derive a value of confidence for each prediction that is used as a weight for the resulting edge.

The network built from experimental data uploaded by the user is obtained by analyzing all the possible pairwise interactions between TFs, miRNAs, and genes across the samples/replicates. The user may choose different measures of associations, both parametric and non-parametric (e.g., Pearson, Spearman, and Kendall).

Finally, the software integrates the two networks via a simple weighted sum function ( $S$ ) producing a novel network in which edges, which are found in both networks, have a greater weight. Results are finally visualized by using a Cytoscape-based interface [53] and all feed-forward loops, and their neighbors are evidenced. In addition, other simple analyses can be executed (e.g., an ontology-based analysis).

#### IntegraMiR

IntegraMiR [54] is a novel approach of integration of data that is based on the workflow depicted in Fig. 6. It receives as input mRNA and miRNA expression data, obtained from samples that are subdivided into two classes (e.g., controls vs. cases). It starts by searching for differentially expressed genes and mRNAs between two conditions by using the Bioconductor package LIMMA [55]. This step produces two lists, one for differentially expressed genes and one for differentially expressed miRNAs. Moreover,

**Table 1** Comparison of network-based analysis approaches considering availability and input data

Tool	Implementation	Input data	Input data grouping	Format of input data
dChip-GemiNi [42]	Web server - R script	mRNA-miRNA expression data grouped into two classes	Two-class data. Paired and not paired	Textual matrices
MAGIA <sup>2</sup> [44]	Web server	mRNA-miRNA expression data. Time series and two-class data	Two-class and time series data	Textual matrices
mirConnX [47]	Web server	mRNA-miRNA expression data	Two-class data	Textual matrices. <i>Only human and mouse</i>
IntegraMiR [54]	R script	mRNA-miRNA expression data grouped into two classes	Two-class data	Textual matrices

**Table 2** Comparison of network-based analysis approaches considering algorithmic approach and parameters of analysis

Tool	Algorithmic approach	Parameters
dChip-GemiNi [42]	Identification of SDE genes. Building and analysis of experimental network	Permutations
MAGIA <sup>2</sup> [44]	Building and analysis of experimental network	Relatedness measure. Target databases
mirConnX [47]	Comparison of networks	Association measure for experimental data. Weight of networks
IntegraMiR [54]	Building and analysis of experimental network	

IntegraMiR uses LIMMA package to perform gene set enrichment analysis (GSEA), taking into account known biological knowledge about these transcripts to derive biological significance of both changed and unchanged transcripts. Then, associations among mRNA and miRNA are derived considering their individual expression levels (i.e., considering pairs of mRNA-miRNA whose regulation is inversely correlated) or through their target interactions—via functional analysis through literature and databases. Once this step is finished, IntegraMiR uses the TRANSFAC database [37] to derive associations among TFs and mRNAs and the TransmiR database [39] to derive associations among TFs and miRNAs. In particular, it focuses only on differentially expressed miRNA and mRNA. Thus, it can reconstruct FFLs whose members are differentially expressed. These FFLs are then organized considering the kind of deregulation and ranked by using a statistical approach and visualized to the user (see the original publication for a complete list of results). The software is available for download at (see the original publication for a complete list of results [54]).

#### Further analysis approaches

The current state of the art of research includes some other approaches of analysis that have been developed in different moments. Some of these approaches are not implemented in a single tool although they present a fully reproducible way to analyze miRNA-TF relationships [56].

For instance, Henriksen et al. [57] applied an integrated approach of analysis to identify miRNA-mRNA regulatory networks that are involved in glioma, a primary

brain tumor. They identified miRNA functional targets during glioma malignant progression by combining the paired expression profiles of miRNAs and mRNAs of patients.

Nazarov et al. [58] developed an integrated analysis approach based on the use of different tools, both academic and commercial. The workflow of analysis is structured into different steps. They start from paired miRNA and mRNA data obtained from microarray experiments. In the first step, they pre-process miRNA and mRNA data using the Partek GS<sup>®</sup> platform in order to filter out non-relevant or out-of-quality data. Then, they use the LIMMA package of Bioconductor [59] to identify significant differentially expressed miRNA and mRNA. Then, they use Ingenuity Pathway Analysis (IPA)<sup>®</sup> to build regulatory networks of miRNA, mRNA, and transcription factors. In particular, they identify upstream regulators by using IPA. The IPA platform enables the reconstruction of causal networks constructed from individual relationships by providing a set of tools for inferring and scoring upstream regulators of gene expression data [41].

#### Discussion

We here compare the so far discussed approaches by considering the following parameters:

- *Input and implementation:* We consider (i) the format of input (e.g., textual files or raw data), (ii) the experimental platforms (miRNA or mRNA), (iii) the design of the experiments (e.g., two-class experiments or time series), and (iv) the availability as a web server or as a stand-alone tool.

**Table 3** Comparison of network-based analysis approaches considering internal knowledge bases

Tool	miRNA-mRNA	TF-genes	miRNA-TF
dChip-GemiNi [42]	TargetScan and PicTar	TRANSFAC matrices v7.0	miRBase - TRANSFAC
MAGIA <sup>2</sup> [44]	Microcosm, microrna.org, DIANA-microT, miRDB, PicTar, PITA, RNA22, and TargetScan	ECRbase	mirGen2.0 and TransmiR
mirConnX [47]	PITA, miRANDA, TargetScan 5.0, RNAhybrid, and PicTar	JASPAR and TRANSFAC	CoreBoost_HM
IntegraMiR [54]	mSigDB, mirTarBase, miRecords	TRANSFAC	TransMiR

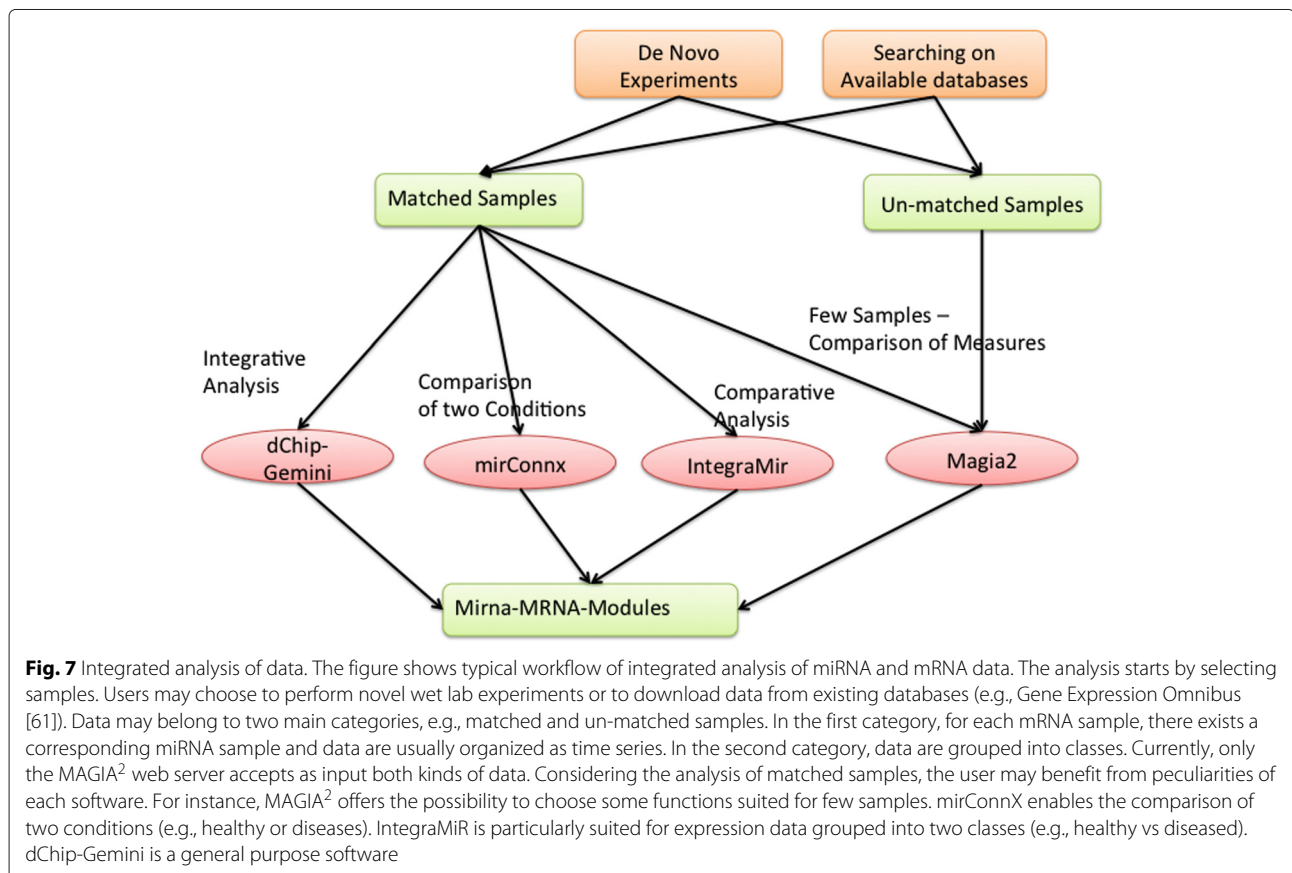


**Table 4** Comparison of network-based analysis approaches considering output information

Tool	Kind	Format	Link to external knowledge bases
dChip-GemiNi [42]	Statistics of association. Visualization of network	Static	Not available
MAGIA <sup>2</sup> [44]	Statistics of association. Visualization of network	Dynamic and exportable in Cytoscape	Available
mirConnX [47]	Statistics of association. Visualization of network	Dynamic	Available
IntegraMiR [54]	Lists of associated mRNA-miRNA and TF	Static	Not available

- **Analysis:** We consider the algorithmic approach (i.e., main characteristics of the analysis) and the main parameters customizable by the users.
- **Knowledge bases:** We consider which data sources have been used to derive associations among molecules, i.e., (i) miRNA-mRNA associations, (ii) TF-genes associations, and (iii) miRNA-TF associations.
- **Output:** We consider the characteristics of the output, its format (i.e., graphic or textual), as well as the possibility to link results to external knowledge bases (i.e., ontologies or semantic analysis [60]).

Considering Table 1, we should note at first that software available as web server (dChip-GemiNi, MAGIA<sup>2</sup>, and mirConnX) are more user-friendly from a biological corner since the installation and running of R scripts is not easy without a bioinformatics support. Moreover, the MAGIA<sup>2</sup> web server enables the use of both two class and time series data, enhancing the possibility of analysis. All the softwares enable the use of different identifiers for genes, and some of them (e.g., dChip-GemiNi) have the possibility to use ad hoc identifiers. mirConnX has a main limitation on the input species since it may analyze only human and mouse data.



Considering Table 2, we report that the MAGIA<sup>2</sup> web server is more flexible than the others since it gives to the user the possibility to choose different correlation measures and several target databases. Moreover, the user may intersect different databases. All the approaches compare experimental data with respect to knowledge bases, and in particular, mirConnX enables to weigh the influence of knowledge bases.

Considering Table 3, we report that the MAGIA<sup>2</sup> web server used the largest number of association databases. In particular, we note that the most popular databases are TargetScan and Pictar (used by dChip-GemiNi, MAGIA<sup>2</sup>, and mirConnX) for miRNA-mRNA associations and TRANSFAC for TF-gene association (used by dChip-GemiNi, mirConnX, and IntegraMiR).

Finally, considering the presentation of results, we note that the best performances are in generally achieved by using external visualizers (e.g., the Cytoscape web interface used by mirConnX or MAGIA<sup>2</sup>). Moreover, mirConnX provides the possibility to link results to external databases (e.g., for enrichment analysis or search) (Table 4).

Figure 7 reports some short examples of typical case studies by discussing main options and choices that are available to researchers.

## Conclusions

As evidenced before, the TF-miRNA-mRNA association represents undoubtedly a main resource for elucidating gene expression regulation at a systems level. The complete determination of miRNA and TF targets will enable a more powerful and reliable analysis. Consequently, from a technological point of view, the miRNA and TF target prediction and validation is still an urgent issue. In parallel, from a computational point of view, the integration of more data sources may improve the quality of analysis, since computational TF-miRNA regulatory networks are available for some genomes and diseases. Moreover, integrating TF-miRNA regulatory networks with other networks, such as functional networks (e.g., signaling pathways, metabolic pathways, protein-protein interaction networks) or semantic networks, will be an important improvement. This integration will aid in explaining how these networks regulate the biological processes and diseases at the systems level.

## Competing interests

Authors declare that they have no competing interests.

## Authors' contributions

PHG and MTD conceived the main ideas of this paper. MC led the bioinformatics aspect of this research. PST and PFT led the clinical and biological aspects. All authors read and approved the manuscript.

## Acknowledgements

This work has been supported by the Italian Association for Cancer Research (AIRC), PI: PT. "Special Program Molecular Clinical Oncology - 5 per mille" n.

9980, 201015 and the DICET-INMOTO-ORCHESTRA Project (PON04a2\_D) funded by the Italian Ministry of Education and Research (MIUR).

## Author details

<sup>1</sup>Department of Medical and Surgical Sciences, Magna Graecia University, Catanzaro, Italy. <sup>2</sup>Department of Experimental and Clinical Medicine, Magna Graecia University, Salvatore Venuta University Campus, Catanzaro, Italy. <sup>3</sup>Sbarro Institute for Cancer Research and Molecular Medicine, Center for Biotechnology, College of Science and Technology, Temple University, Philadelphia, PA, USA.

Received: 14 January 2015 Accepted: 18 May 2015

Published online: 04 June 2015

## References

1. M Wilm, Quantitative proteomics in biological research. *Proteomics*. **9**(20), 4590–4605 (2009). doi:10.1002/pmic.200900299
2. M Cannataro, PH Guzzi, A Sarica, Data mining and life sciences applications on the grid. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. **3**(3), 216–238 (2013)
3. M Cannataro, PH Guzzi, P Veltri, Protein-to-protein interactions: technologies, databases, and algorithms. *ACM Comput Surveys (CSUR)*. **43**(1), 1 (2010)
4. M Mina, PH Guzzi, Improving the robustness of local network alignment: design and extensive assessment of a Markov clustering-based approach. *Comput. Biol. Bioinformatics, IEEE/ACM Trans.* **11**(3), 561–572 (2014). doi: 10.1109/TCBB.2014.2318707
5. A Schratzenholz, K Groebe, V Soskic, in *Systems Biology in Drug Discovery and Development. Methods in Molecular Biology™*, ed. by JM Walker, Q Yan. Systems biology approaches and tools for analysis of interactomes and multi-target drugs, vol. 662 (Humana Press Totowa, NJ, 2010), pp. 29–58. Chap. 2. doi:10.1007/978-1-60761-800-3\_2. [http://dx.doi.org/10.1007/978-1-60761-800-3\\_2](http://dx.doi.org/10.1007/978-1-60761-800-3_2)
6. A-L Barabasi, ZN Oltvai, Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**(2), 101–113 (2004). doi: 10.1038/nrg1272
7. NJ Martinez, AJ Walhout, The interplay between transcription factors and microRNAs in genome-scale regulatory networks. *Bioessays*. **31**(4), 435–445 (2009)
8. A Pujol, R Mosca, J Farrés, P Aloy, Unveiling the role of network and systems biology in drug discovery. *Trends Pharmacol. Sci.* **31**(3), 115–123 (2010)
9. MT Di Martino, V Campani, G Misso, MEG Cantafio, A Gullà, U Foresta, PH Guzzi, M Castellano, A Grimaldi, V Gigantino, et al, In vivo activity of miR-34a mimics delivered by stable nucleic acid lipid particles (SNALPs) against multiple myeloma. *PLoS One*. **9**(2), 90005 (2014)
10. MT Di Martino, A Gullà, MEG Cantafio, M Lionetti, E Leone, N Amodio, PH Guzzi, U Foresta, F Conforti, M Cannataro, et al, In vitro and in vivo anti-tumor activity of miR-221/222 inhibitors in multiple myeloma. *Oncotarget*. **4**(2), 242 (2013)
11. M Lionetti, P Musto, MT Di Martino, S Fabris, L Agnelli, K Todoerti, G Tuana, L Mosca, MEG Cantafio, V Grieco, et al, Biological and clinical relevance of miRNA expression signatures in primary plasma cell leukemia. *Clin. Cancer Res.* **19**(12), 3130–3142 (2013)
12. MT Di Martino, M Arbitrio, PH Guzzi, E Leone, F Baudi, E Piro, T Pranterà, I Cucinotto, T Calimeri, M Rossi, et al, A peroxisome proliferator-activated receptor gamma (PPARG) polymorphism is associated with zoledronic acid-related osteonecrosis of the jaw in multiple myeloma patients: analysis by DMET microarray profiling. *Br. J. Haematol.* **154**(4), 529–533 (2011)
13. MT Di Martino, M Arbitrio, E Leone, PH Guzzi, M Saveria Rotundo, Single nucleotide polymorphisms of ABC5 and ABCG1 transporter genes correlate to irinotecan-associated gastrointestinal toxicity in colorectal cancer patients: a DMET microarray profiling study. *Cancer biology & therapy*. **12**(9), 780–787 (2011)
14. T Venkatesh, HB Harlow, Integromics: challenges in data integration. *Genome Biol.* **3**(8), 1–3 (2002)
15. DB Searls, Data integration: challenges for drug discovery. *Nat. Rev. Drug Discov.* **4**(1), 45–58 (2005)
16. MV Iorio, CM Croce, microRNA involvement in human cancer. *Carcinogenesis*. **33**(6), 1126–1133 (2012). doi:10.1093/carcin/bgs140. <http://carcin.oxfordjournals.org/content/33/6/1126.full.pdf+html>

17. A Muniategui, J Pey, FJ Planes, A Rubio, Joint analysis of miRNA and mRNA expression data. *Brief. Bioinform.* **14**(3), 263–278 (2013)
18. K Chen, N Rajewsky, The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.* **8**(2), 93–103 (2007)
19. H-M Zhang, S Kuang, X Xiong, T Gao, C Liu, A-Y Guo, Transcription factor and microRNA co-regulatory loops: important regulatory motifs in biological processes and diseases. *Briefings in Bioinformatics* (2013). doi: 10.1093/bib/bbt085. <http://bib.oxfordjournals.org/content/early/2013/12/04/bib.bbt085.full.pdf+html>
20. DJ Burgess, Molecular evolution: decoupled transcription factor output? *Nat. Rev. Genet.* **16**(1), 4–5 (2015)
21. M Garofalo, CM Croce, Role of microRNAs in maintaining cancer stem cells. *Adv. Drug. Deliv. Rev.* **81**(0), 53–61 (2015). doi:10.1016/j.addr.2014.11.014
22. GA Calin, CM Croce, MicroRNA signatures in human cancers. *Nat. Rev. Cancer.* **6**(11), 857–866 (2006)
23. M Rossi, N Amodio, MT Di Martino, D Caracciolo, P Tagliaferri, From target therapy to miRNA therapeutics of human multiple myeloma: theoretical and technological issues in the evolving scenario. *Current drug targets.* **14**(10), 1144–1149 (2013)
24. M Rossi, MT Di Martino, E Morelli, M Leotta, A Rizzo, A Grimaldi, Molecular targets for the treatment of multiple myeloma. *Current cancer drug targets.* **12**(7), 757–767 (2012)
25. N Amodio, MT Di Martino, A Neri, P Tagliaferri, P Tassone, Non-coding RNA: a novel opportunity for the personalized treatment of multiple myeloma. *Expert opinion on biological therapy.* **13**(S1), S125–S137 (2013)
26. N Rajewsky, microRNA target predictions in animals. *Nat. genet.* **38**, 8–13 (2006)
27. S Griffiths-Jones, HK Saini, S van Dongen, AJ Enright, miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36**(suppl 1), 154–158 (2008)
28. D Betel, M Wilson, A Gabow, DS Marks, C Sander, The microRNA.org resource: targets and expression. *Nucleic Acids Res.* **36**(suppl 1), 149–153 (2008)
29. M Maragkakis, M Reczko, VA Simossis, P Alexiou, GL Papadopoulos, T Dalamagas, G Giannopoulos, G Goumas, E Koukis, K Kourtis, et al., DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acids Res.* **29**(2) (2009)
30. X Wang, miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA.* **14**(6), 1012–1017 (2008)
31. A Krek, D Grün, MN Poy, R Wolf, L Rosenberg, EJ Epstein, P MacMenamin, I da Piedade, KC Gunsalus, M Stoffel, et al., Combinatorial microRNA target predictions. *Nat. Genet.* **37**(5), 495–500 (2005)
32. M Kertesz, N Iovino, U Unnerstall, U Gaul, E Segal, The role of site accessibility in microRNA target recognition. *Nat. Genet.* **39**(10), 1278–1284 (2007)
33. KC Miranda, T Huynh, Y Tay, Y-S Ang, W-L Tam, AM Thomson, B Lim, I Rigoutsos, A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell.* **126**(6), 1203–1217 (2006)
34. A Grimson, KK-H Farh, WK Johnston, P Garrett-Engele, LP Lim, DP Bartel, MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. cell.* **27**(1), 91–105 (2007)
35. MJ Buck, JD Lieb, ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics.* **83**(3), 349–360 (2004)
36. J Qin, MJ Li, P Wang, MQ Zhang, J Wang, ChIP-Array: combinatory analysis of ChIP-seq/chip and microarray gene expression data to discover direct/indirect targets of a transcription factor. *Nucleic Acids Res.* **39**(suppl 2), 430–436 (2011)
37. E Wingender, The transfac project as an example of framework technology that supports the analysis of genomic regulation. *Brief. Bioinform.* **9**(4), 326–332 (2008)
38. A Lachmann, H Xu, J Krishnan, SI Berger, AR Mazloom, A Ma'ayan, ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics.* **26**(19), 2438–2444 (2010)
39. J Wang, M Lu, C Qiu, Q Cui, TransmiR: a transcription factor–microRNA regulation database. *Nucleic Acids Res.* **38**(suppl 1), 119–122 (2010)
40. B Lenhard, WW Wasserman, TFBS: Computational framework for transcription factor binding site analysis. *Bioinformatics.* **18**(8), 1135–1136 (2002). doi:10.1093/bioinformatics/18.8.1135. <http://bioinformatics.oxfordjournals.org/content/18/8/1135.full.pdf+html>
41. A Kramer, J Green, J Pollard Jr, S Tugendreich, Causal analysis approaches in Ingenuity Pathway Analysis (IPA). *Bioinformatics.* **30**, 523–530 (2013)
42. Z Yan, PK Shah, SB Amin, MK Samur, N Huang, X Wang, V Misra, H Ji, D Gabuzda, C Li, Integrative analysis of gene and miRNA expression profiles with transcription factor–miRNA feed-forward loops identifies regulators in human cancers. *Nucleic Acids Res.* **39**(5) (2012)
43. T Barrett, SE Wilhite, P Ledoux, C Evangelista, IF Kim, M Tomashevsky, KA Marshall, KH Phillippy, PM Sherman, M Holko, et al., NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.* **41**(D1), 991–995 (2013)
44. A Bisognin, G Sales, A Coppe, S Bortoluzzi, C Romualdi, MAGIA2: from miRNA and genes expression data integrative analysis to microRNA–transcription factor mixed regulatory circuits (2012 update). *Nucleic Acids Res.* **46**(2012)
45. P Alexiou, T Vergoulis, M Gleditsch, G Prekas, T Dalamagas, M Megraw, I Grosse, T Sellis, AG Hatzigeorgiou, miRGen 2.0: a database of microRNA genomic information and regulation. *Nucleic Acids Res.* **888** (2009)
46. G Loots, I Ovcharenko, ECRbase: database of evolutionary conserved regions, promoters, and transcription factor binding sites in vertebrate genomes. *Bioinformatics.* **23**(1), 122–124 (2007)
47. GT Huang, C Athanassiou, PV Benos, mirConnX: condition-specific mRNA–microRNA network integrator. *Nucleic Acids Res.* (2011). doi: 10.1093/nar/gkr276. <http://nar.oxfordjournals.org/content/early/2011/05/10/nar.gkr276.full.pdf+html>
48. AJ Enright, B John, U Gaul, T Tuschl, C Sander, DS Marks, et al, MicroRNA targets in Drosophila. *Genome Biol.* **5**(1), 1–1 (2004)
49. J Krüger, M Rehmsmeier, RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* **34**(suppl 2), 451–454 (2006)
50. GL Papadopoulos, M Reczko, VA Simossis, P Sethupathy, AG Hatzigeorgiou, The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Res.* **37**(suppl 1), 155–158 (2009)
51. F Xiao, Z Zuo, G Cai, S Kang, X Gao, T Li, miRecords: an integrated resource for microRNA–target interactions. *Nucleic Acids Res.* **37**(suppl 1), 105–110 (2009)
52. X Xie, J Lu, E Kulbokas, TR Golub, V Mootha, K Lindblad-Toh, ES Lander, M Kellis, Systematic discovery of regulatory motifs in human promoters and 3 UTRs by comparison of several mammals. *Nature.* **434**(7031), 338–345 (2005)
53. ME Smoot, K Ono, J Ruscheinski, P-L Wang, T Ideker, Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics.* **27**(3), 431–432 (2011)
54. AS Afshar, J Xu, J Goutsias, Integrative identification of deregulated miRNA/TF-mediated gene regulatory loops and networks in prostate cancer. *PLoS ONE.* **9**(6), 100806 (2014). doi:10.1371/journal.pone.0100806
55. GK Smyth, in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Statistics for Biology and Health*, ed. by R Gentleman, V Carey, W Huber, R Irizarry, and S Dudoit. *limma: Linear models for microarray data* (Springer New York, 2005), pp. 397–420. Chap. 23. doi: 10.1007/0-387-29362-0\_23. [http://dx.doi.org/10.1007/0-387-29362-0\\_23](http://dx.doi.org/10.1007/0-387-29362-0_23)
56. H-M Zhang, S Kuang, X Xiong, T Gao, C Liu, A-Y Guo, Transcription factor and microRNA co-regulatory loops: important regulatory motifs in biological processes and diseases. *Briefings in Bioinformatics.* **16**(1), 45–58 (2015). doi:10.1093/bib/bbt085. <http://bib.oxfordjournals.org/content/16/1/45.full.pdf+html>
57. M Henriksen, KB Johnsen, HH Andersen, L Piilgaard, M Duroux, MicroRNA expression signatures determine prognosis and survival in glioblastoma multiforme—a systematic overview. *Mol. neurobiol.* **50**(3), 896–913 (2014)
58. PV Nazarov, SE Reinsbach, A Muller, N Nicot, D Philippidou, L Vallar, S Kreis, Interplay of microRNAs, transcription factors and target genes: linking dynamic expression changes to function. *Nucleic Acids Res.* **41**(5), 2817–2831 (2013)
59. JM Wettenhall, GK Smyth, limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics.* **20**(18), 3705–3706 (2004)
60. PH Guzzi, M Mina, C Guerra, M Cannataro, Semantic similarity analysis of protein data: assessment with biological features and issues. *Brief. Bioinform.* **13**(5), 569–585 (2012)
61. R Edgar, M Domrachev, AE Lash, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**(1), 207–210 (2002)