

# Non-coding RNA regulatory networks

Simona Panni<sup>(1)\*</sup>, Ruth C. Lovering<sup>(2)</sup>, Pablo Porras<sup>(3)</sup>, Sandra Orchard<sup>(3)</sup>

<sup>(1)</sup> Università della Calabria, Dipartimento di Biologia, Ecologia e Scienze della Terra, Via Pietro Bucci Cubo 6/C, Rende, CS, Italy      [simona.panni@unical.it](mailto:simona.panni@unical.it).

<sup>(2)</sup> Institute of Cardiovascular Science, University College London, Rayne Building, 5 University Street, London, WC1E 6JF, UK.

<sup>(3)</sup> European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Genome Campus, Hinxton, CB10 1SD, UK.

\*Corresponding author [simona.panni@unical.it](mailto:simona.panni@unical.it).

Authors email address:

Simona Panni: [simona.panni@unical.it](mailto:simona.panni@unical.it)

Ruth C. Lovering : [r.lovering@ucl.ac.uk](mailto:r.lovering@ucl.ac.uk)

Pablo Porras : [pporras@ebi.ac.uk](mailto:pporras@ebi.ac.uk)

Sandra Orchard : [orchard@ebi.ac.uk](mailto:orchard@ebi.ac.uk)

# Non-coding RNA regulatory networks

## Abstract

It is well established that the vast majority of human RNA transcripts do not encode for proteins and that non-coding RNAs regulate cell physiology and shape cellular functions. A subset of them are involved in gene regulation at different levels, from epigenetic gene silencing to post-transcriptional regulation of mRNA stability. Notably, the aberrant expression of many non-coding RNAs has been associated with aggressive pathologies. Rapid advances in network biology indicates that the robustness of cellular processes is the result of specific properties of biological networks such as scale-free degree distribution and hierarchical modularity, suggesting that regulatory network analyses could provide new insights on gene regulation and dysfunction mechanisms.

In this study we present an overview of public repositories where non-coding RNA-regulatory interactions are collected and annotated, we discuss unresolved questions for data integration and we recall existing resources to build and analyse networks.

## 1. Introduction: non-coding RNAs involved in transcriptional regulation

A regulatory network is composed of a complex web of molecular factors that interact with each other and with genes in order to control gene expression. In the simplest form of gene-regulatory network, transcription factors bind to regulatory sequences and their interaction results in the induction or inhibition of the gene [1].

Recent advances in molecular biology however, have shown that gene expression is largely regulated not only by proteins but also by non-coding RNAs (ncRNAs). In the last decade, an increasing number of studies have reported on the involvement of ncRNAs in various physiological processes, and next generation sequencing technologies together with bioinformatics analyses, have revealed the pervasive transcription of an unexpected variety of RNA molecules. Among them, microRNAs, long non-coding RNAs (lncRNAs) and circular RNAs (circRNAs) have been shown to be involved in transcriptional regulation at different levels (Figure 1).

## 1.1 MicroRNAs

MicroRNAs are small RNA molecules, that regulate the expression of the target messenger RNAs (mRNAs), usually by binding to a short complementary sequence often located in the 3' UTR region of the mRNA [2]. They are characteristically an average length of 22 nucleotides. In mammalian cells, microRNAs are incorporated in the effector complex RISC (Figure 2A-B), where they bind to a member of the Argonaute protein family and guide it to cleave the complementary mRNA or to inhibit translation and thus protein synthesis [3]. A complete collection of the so-far predicted and verified microRNAs is available at the miRBase database (<http://www.mirbase.org/>) and includes about 1915 human microRNA precursors (of which at least 725 are high-confidence identifications, as defined by the database curators), some of which produce active microRNAs with both strands, while with others only one of the two strands is active [4,5]. Functional characterization through large-scale experimental techniques to unravel their functions is expensive and time consuming. Thus, several bioinformatics approaches have been developed to identify their mRNA targets, based on the complementarity between the 3' UTR and the "seed" sequence i.e. 6-8 nucleotides at the 5' of the mature microRNA (Figure 2C). As an example, TargetScan predicts microRNA targets from the presence of 6-8 nucleotides in the 3' UTR of the mRNA, matching nucleotides 2-7 of the microRNA. It scores the targets according to the sequence conservation in 5 genomes [6] and other context features, such as the predicted structure accessibility of the site [7]. Pairing to nucleotides 13-16 of the microRNA (3' supplementary sites) and unconventional binding, such as 12 contiguous bases complementary to the centre of the microRNA (centred sites), are also considered [8,9]. Other algorithms based on similar considerations are PicTar [10] and miRanda [11]. Similarly, RNAhybrid [12] provides a variation of RNA secondary structure prediction methods, and calculates the most favourable hybridization site between the microRNA and the mRNA. RNA22 instead, is based on a procedure which first discovers statistically significant patterns in mRNAs, then search complementary microRNAs [13]. Machine-learning based predictors have been also developed with the help of validated interacting pairs [14]. A complete overview of

prediction algorithms and their specific characteristics is given in [15]. These algorithms are very often used as the preliminary screening to identify genes regulated by a specific microRNA. However, networks have also been built and analysed directly from the prediction, without experimental validation of the binding [16,17].

## **1.2 lncRNAs, circRNAs and enhancer RNAs**

MicroRNAs also interact with the second family of regulative RNAs: the long-non-coding RNAs (lncRNAs). lncRNAs are a highly heterogeneous class of RNAs: when transcribed from intergenic regions they are described as long intergenic non-coding RNAs (lincRNAs), but they can be transcribed from any part of protein coding genes, both in the sense or antisense direction. The key defining features of lncRNAs is that they have to be longer than 200 nucleotides and not encode a protein. However, as extensively discussed in [18], the lack of specific features on lncRNAs to help with the classification may lead to misannotations, even though several strategies can be employed to discriminate between protein-coding and non-coding transcripts. Moreover, some well-characterized lncRNAs, such as Xist or H19, contain putative ORFs, and bifunctional transcripts might be more common than previously supposed [18].

This very general definition means that the lncRNA class probably includes molecules with very different functions. Consequently, unlike for microRNAs, it has not been possible to identify a common mechanism of action. Several studies have shown that lncRNAs, such as Hotair, exert their function by binding to transcription regulation proteins, including histone modification enzymes and chromatin remodelling factors [19], but they can also regulate mRNAs by complementary binding (antisense lncRNAs) and binding to microRNAs (Figure 1). This latter function is often referred to as the “sponge effect” meaning that lncRNAs may sequester microRNAs and thus regulate their ability to inhibit gene expression. The interaction between microRNAs and lncRNA can be predicted by the same algorithms and platforms that predict microRNA-mRNA interactions (Table 1) or by lncRNA/mRNA correlation analyses [20]. A subset of lncRNAs, termed immune-gene-priming lncRNAs (IPLs), activate immune genes in endothelial cells and other immune cells during immune-responses, recruiting the H3K4 histone-methyltransferase WDR5-MLL1 to the target promoters. Interestingly, the IPLs act *in cis*, and are correctly positioned in

proximity of the promoters as a consequence of the three-dimensional chromatin looping topology [21].

Recently, a new class of ncRNAs has been identified which have a role in gene regulation. Circular RNAs are single-stranded RNA molecules, in which the 3' and 5' ends are covalently linked giving rise to circular molecules. They originate in a process called back-splicing and usually contain exons, introns, or even non-coding intergenic regions. High-throughput experiments have identified thousands of circRNAs [22–24], and these are annotated in the circBase and circRNADB databases [25,26]. In these studies, the identification of circRNAs is obtained with computational tools from RNA-Seq data, while only a small percentage of the circRNAs have been experimentally verified. To rule out the potential artefacts, candidates can be validated testing their sensitivity to RNase R: the exonuclease is supposed to digest linear RNAs, leaving circular RNA molecules unaffected [22–24]. An accurate analysis of the digestion profiles has provided evidences that some circBase entries may be false positives [27]. Ambiguous alignment of transcripts originating from repetitive sequences also may result in false positives, therefore the annotation of reliable circRNAs should be carefully evaluated [27]. The functions of circRNAs are still being characterized, however it has been shown that some of them can bind to specific proteins [28,29] or can act as a microRNA sponge [23,30]. As we will describe below, only predicted networks involving circRNAs are computationally available at the moment, while the few direct interactions experimentally demonstrated are still scattered in the literature. The functions of enhancer RNAs (eRNAs) are even less well characterized: these transcripts originate from the enhancer regions and their transcription positively correlates with activated histone marks and other parameters of active enhancer elements [31]. This suggests that they may have a role in target-gene activation, however, although several studies confirm this hypothesis, the debate concerning their functional relevance is still an open issue [32]. More than 40.000 bidirectional capped eRNA candidates have been identified in human cells [31], and different potential mechanisms of action have been proposed, including the interesting observation that eRNAs are able to bind to the transcriptional co-activator CREB binding protein (CBP) and to activate its core enzyme histone-acetyltransferase [33]. This study hypothesizes that eRNAs exert their regulative role stimulating chromatin modifying enzymes such as histone-acetyltransferases, thereby promoting gene expression.

### 1.3 Examples of regulatory networks

Countless examples of regulatory networks involving ncRNAs are annotated in the literature. One of the first global transcriptional regulatory maps was modelled for the eukaryotic organism *Saccharomyces cerevisiae* in 2002 with a genome-wide location analysis obtained by tagging each transcription factor gene with an epitope and identifying target promoters using chromatin immunoprecipitation [1]. More recently, a first draft of human transcriptional regulatory network was deciphered from the genome-wide binding analysis of 119 regulatory factors generated by chromatin immunoprecipitation followed by sequencing analysis, and integrated with PPI data and a microRNA-ncRNA network [34]. The study provided an initial interesting analysis of human gene regulation, including the observation that transcription factors may prefer one allele over another. The network has since improved with more data and compared to gene-regulatory networks from *C. elegans* and *D. melanogaster* [35].

Notably, regulatory networks have been inferred from RNA expression profiles, when the abundance of one RNA affects the abundance of a second RNA. The first RNA may code for a transcription factor, but the regulative effects may also occur through signalling cascades [36]. These methods are often referred to as reverse-engineering in the literature. One of the most used network-inference algorithms is ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks), which has been widely used to reconstruct tissue and cell-type specific networks [37–40].

In general, many transcription factors have been reported to regulate ncRNAs and to exert their function through the activity of microRNAs or lncRNAs. One of the most-studied transcription factor is the tumour suppressor p53, which regulates key cellular processes and exerts its activity primarily at the transcriptional level [41]. In basal conditions the protein is maintained at low concentration by proteasomal degradation, but upon cellular stress (such as DNA damage), p53 is activated and acts on a large set of coding genes. It usually activates them, but repression has also been reported [42]. Several microRNAs and lncRNAs are direct or indirect targets of p53 activation: the expression levels of miR34a, miR15, miR200, miR145 and miR107 are negatively affected by p53 deletion [43]. Interestingly it has been shown that p53 activates miR145, which suppresses MYC, which is also directly suppressed by p53 (Figure 3) [43,44]. Two lncRNAs transcriptionally activated by p53 were suggested to act as mediators of the p53 repression: the lncRNA lincRNA-p21, also known as Trp53cor1 (tumour protein p53 pathway corepressor 1),

interacts with hnRNP K and the DNA methyltransferase DNMT1 to promote the silencing of the target genes, while lincMkln1 forms a complex with the Polycomb Repressive Complex 2 to exert the repression [42]. Several other lncRNAs have been reported as p53 targets, and in a genome-wide analysis of the DNA-damage, the authors identified 18 lncRNAs as high-confidence p53 transcriptional targets [45]. The p53 protein in turn is regulated by non-coding RNAs: miR125b and miR504, among others, bind to the 3' UTR of the mRNA to downregulate p53 level and function [46]. Other microRNAs downregulate p53 regulators, including miR145, which regulates MDM2 (the ubiquitin ligase that maintains the basal level of p53), but is regulated by p53 [46]. The example clearly shows the complexity of regulatory networks and the high interconnectivity of proteins, genes and non-coding RNAs.

## **2. How to retrieve data to build ncRNA regulatory networks**

To represent and analyse the unprecedented amount of information about transcriptional regulation, regulatory networks can be modelled as graphs, where different nodes (microRNA, lncRNAs, circRNAs, mRNAs, proteins and genes) are linked by edges that depict the interaction (undirected graph) or the causal relation (directed graph) between one node and the other.

RNA interactions can be retrieved from freely accessible databases that collect RNA-RNA and RNA-protein interactions, as described below (Table 2). However, no standardized common criteria have been developed to compare these data, so it is very important to consider the source of the data. One of the main differences among RNA-RNA interaction datasets is whether the interactions are predicted, i.e. based on the algorithms listed in the previous section (Table 1), have been identified by text mining, or are experimentally verified complexes, manually curated.

### **2.1 Predicted RNA-RNA Networks**

Unlike protein-protein interactions, the prediction of RNA-RNA interactions takes advantage of the foreseen sequence complementarity between the interacting pairs. Predicted networks of microRNA-mRNA and microRNA-lncRNA interactions can be downloaded from several platforms that compare the prediction of different algorithms (Table 1). Data on gene

expression downregulation after microRNA transfection can be applied to rank hypothetical target sites in these predicted networks [47,48].

miRecords [49] starBase [50] and miRWalk [51], are resources which integrate target interactions predicted by up to 11 programs. In fact, different algorithms often result in different predictions and it is a common practice to focus on the intersection of at least 3-4 algorithms, although this can lead to some true targets being missed [49]. An example is illustrated in Figure 4: the prediction for miR-17 obtained from starBase selecting “TargetScan” and “very high stringency” results in more than 700 interactions with 646 mRNAs. If the intersection of 4 algorithms is selected the number of positives mRNAs falls to 349. Comparing these two lists of potential miR-17 targets with 39 true-positive interactors, 13 are present in the first list and 12 in the second. Notably, if one of the 4 prediction algorithms is RNA22 [13], the number of the predicted miR-17 target mRNAs decreases to 40. This would tempt the user to believe that RNA22 should be included in the prediction, however, in the case of miR-17, only 3 of the true positives are present among the 40 predicted targets. This example points to the importance of setting up the predictive method with well-known examples before performing the analyses. It is also important to note that different algorithms may refer to different sets of sequences, for example a prediction on the same mRNA may map to different sequence positions depending on the Ensembl or GENCODE version used [15,52,53]. More generally, the lack of an unambiguous identifier (ID) for each RNA sequence (especially for less characterized organisms) seriously complicates data merging and analysis. For non-coding RNAs a platform named RNACentral has been recently developed to ensure unequivocal identification of ncRNAs [54].

An increasing number of circRNA interactors have been recently experimentally identified in the literature, unfortunately very little of this data has been collected by any interaction database to date. However, a few resources have been developed to predict potential targets of circRNAs enabling preliminary analysis on their functions. CircInteractome utilizes data from CLIP experiments collected in starBase to predict potential circRNA-protein interactions and TargetScan to predict microRNA binding [55]. CircNet offer a predicted regulatory network of circRNA-microRNA based on the occurrence of the target seed sequence on the circRNA [56].

## **2.2 Experimentally verified Networks**

Many techniques are available for the detection of RNA containing complexes or microRNA targets, and they can be divided into low throughput assays, returning few interactors with relatively high confidence, and high-throughput approaches. The first type of experiments includes luciferase reporter assays, gel Electrophoresis Mobility Shift Assay (EMSA), and RNA immunoprecipitation. In brief, the luciferase assay is widely used for microRNA interactions and allows the detection of RNA-RNA pairs, when the binding of one RNA influences the stability or the function of the other. This interaction can be measured by fusing the binding region of the regulated RNA to a reporter gene. If the binding occurs, the expression of the reporter gene will be decreased when the microRNA is co-transfected together with the wild-type form of the fusion, but not when the binding region is mutated in the putative binding sequence. EMSA is an in-vitro method to detect nucleic acid-protein complexes, based on the observation that the nucleic acid migrates faster in a non-denaturing electrophoresis assay if compared to the protein-bound counterpart. With RNA immunoprecipitation (RNA IP), a protein of interest is captured with a specific antibody and the interacting RNA is identified by Northern blot. A number of methods based on the co-immunoprecipitation of RNA-protein complexes have been recently developed to detect RNA interactions at the transcriptomic level. In these methods, RNA molecules are first linked to the interacting proteins through ultraviolet crosslinking to form covalent bonds, then the protein is immunoprecipitated and RNA is analysed through RT-PCR (Cross-Linking ImmunoPrecipitation, CLIP) or High-Throughput Sequencing (HITS-CLIP or CLIP-Seq). If photoreactive ribonucleoside analogs are incorporated into nascent RNA molecules, binding sites can be mapped (PhotoActivatable Ribonucleoside-enhanced Cross-Linking and ImmunoPrecipitation, PAR-CLIP) [57–59]. Similarly, RNA-RNA interactions can be retrieved by a method called CLASH (Crosslinking, Ligation And Sequencing of Hybrids): after the protein immunoprecipitation, the interacting RNA molecules are ligated and the hybrids are sequenced: chimeric reads represent RNA-RNA interactions [60]. An overview of these methods is given in [61]. The Chromatin Interaction Analysis by Paired-End Tag sequencing (ChIA-PET) was developed to generate chromatin interaction data, but it was recently applied to detect lncRNAs-genes interactions [21]. It combines Chromatin ImmunoPrecipitation (ChIP) with Chromosome Conformation Capture technology (3C) to detect chromatin interactions (or lncRNA-chromatin interactions) associated with a specific protein of interest. Proximal DNA fragments that are part of the same protein-DNA complex are ligated and sequenced [62].

Several databases collect these high-throughput interactions and present lists of binding RNAs in a user-friendly interface. Among them starBase (<http://starbase.sysu.edu.cn/>) annotates data from more than 700 CLIP experiments, including both coding and non-coding RNA-protein interactions [50]. Another useful database from which RNA interactions can be downloaded is RAID (RNA Association Interaction Database [www.rna-society.org/raid](http://www.rna-society.org/raid)). In RAID [63], experimental evidence extracted from the literature and computational predictions are integrated with data from 18 other resources. More than 3 million RNA-RNA interactions are predictions (the majority of them), others are derived from weak experimental proofs (i.e. high-throughput CLASH, CLIP-Seq, ChIP-Seq etc) and only a minor percentage are supported by strong experimental evidence (for example RNA IP or luciferase reporter assay). However, for the convenience of the user each interaction is clearly annotated with the source of the data, and a score is calculated from the number of experiments and the reliability of the evidence, which allow the user to set a case-specific threshold. Other databases add a score of reliability, however scores from different dataset cannot be used comparatively. The RNA-protein Association and Interaction Networks database (RAIN, <https://rth.dk/resources/rain/>) is a resource designed to integrate non-coding RNA interactions into the protein database STRING (<https://string-db.org/>, [64]). Similar to RAID, it collects both experimental data, predictions and text-mining associations and adds probabilistic scores to rank the interacting pairs [65]. In addition, there are specialized repositories that focus on just one class of RNAs, as LnChrom and LIVE that contain only lncRNA interaction data [66,67].

Causal pathways databases, which represent data as directed causal relationships, also increasingly captures RNA interactions [68–70]. Examples of such resources include Signalink2 (<http://signalink.org/>) which combines major pathways, scaffold proteins, transcription factor-gene and microRNA-mRNA interactions from different sources within a multi-layered network to facilitate the modelling of signalling systems [68]. SIGNOR (<https://signor.uniroma2.it>) offers a collection of experimentally-validated causal relationships between two (or more) entities involved in signalling, where one entity is annotated as the “regulator” and the other as “regulated” [69]. In the open source pathways resource Reactome (<https://reactome.org>), molecular entities are organized in manually curated pathways peer-reviewed by external experts [70]. All of these manually curated resources are essential sources of information, whereas both text-mining and predictions may retrieve a huge number of false positives that can be filtered only if gold standards of true interactors, strongly supported from experiments, exist. Recently, members of the Gene

Ontology (GO) Consortium have been capturing experimentally validated microRNA interactions [71]. Specific guidelines have been developed to ensure consistent annotation of the experimental data [72] and the majority of these annotations include the GO term 'mRNA binding' (or child terms) with information about the target RNA (as an Ensembl gene ID) included in the 'annotation extension' field [73]. These annotations are available from GO browsers (AmiGO and QuickGO [74], RNAcentral [54], the PSICQUIC web service [75] [www.ebi.ac.uk/Tools/webservices/psicquic/view/home.xhtml](http://www.ebi.ac.uk/Tools/webservices/psicquic/view/home.xhtml), file name EBI-GOA-miRNA) and by ftp from EMBL-EBI ([www.ebi.ac.uk/GOA/downloads](http://www.ebi.ac.uk/GOA/downloads)).

Inconsistency of the choice of gene and gene product IDs and metadata descriptions are serious obstacles to data merging and interpretation. The Molecular Interaction (MI) workgroup of the Human Proteome Organization-Proteomics Standards Initiative (HUPO-PSI) has been working for more than 15 years to create common rules and standard formats to combine data from different resources and optimize manual curation of molecular interactions [76,77]. It is worth mentioning that recently a pilot project has been started to extend this effort to ncRNA interactions, in order to assure the quality of the annotations according to the PSI-MI standards and to allow the integration of consistently curated data [78]. This project resulted in the representation of around 900 yeast RNA-RNA interactions in the IntAct database (<https://www.ebi.ac.uk/intact/>) [79], a member of the HUPO PSI-MI. IntAct has in fact expanded its remit to molecular entities other than proteins, including both genes and non-coding RNAs, and now features over 42,000 interactions involving some form of nucleic acid. The well-established PSI-MI-XML 2.5 data format enables the capture of molecular features relevant to the binding, such as post-transcriptional and translational modifications, mutations, regions involved in the interaction and so on, that are not captured in other ncRNA databases [80–82]. Although only experimental evidences of physical associations are annotated in IntAct and indirect causal relations are not included in this repository, annotated regulatory networks are automatically integrated with the huge amount of protein interaction data already collected. However, although both the GO consortium and IntAct are capturing only the highest quality experimentally verified microRNA:mRNA interaction data, they are using different gene product identifiers in these annotations so mapping step is required to enable data integration [72].

### **3. Network analysis**

Different types of analysis using interaction or even pathway data can be applied to analyse regulatory networks [83,84]. In many of these approaches, building and visualizing an interaction network is a required step, and there are several tools currently available for this task, using both command line or Graphical User Interface (GUI). Most of these tools can also address the need of integrating external information into them [85]. The igraph software package [86] and Cytoscape [87] have been recently used to explore the structural properties of the yeast RNA interactome, finding that RNA networks seem to be scale-free networks, showing properties similar to PPI networks [78].

Several network representations can be used for regulatory interactions, and we will briefly introduce activity flow diagrams, process descriptions and entity relationship diagrams. Activity flow diagrams represent the flow from one activity to another, where the activities may not be characterized and may represent, for instance, the phosphorylation ability of a kinase which is stimulated by another activity, or the activation of a gene by the activity of a transcription factor. Activities are linked by stimulating or inhibitory directed edges. Process descriptions are bipartite graphs with two sets of nodes: molecular entities and reactions with associated attributes, linked by directed edges. Entity relationship diagrams show the relations between molecular entities (nodes) that influence properties of the entities. The relation is asymmetric and one entity is the regulator, while the other is the regulation-target. Examples and critical discussion of network representations can be found in [88]. Regulatory networks can also be described as undirected interaction networks, where molecular entities are the nodes and the effect of the interaction is not described.

Networks from different species or cell types can be compared through appropriate computational tools: network alignment compares node characteristics (such as sequences) and neighbourhood topologies among cross-species networks to reveal conserved subgraphs or missing cellular processes. The first alignment was proposed by Kelley et al. [89] and the procedure, named PathBLAST, searches for conserved pairs of pathways between yeast and bacteria, allowing “gaps” and “mismatches” similar to those available in sequences alignments. More recently, several strategies have been developed to perform global alignment, local alignment or network querying [90–93].

Global alignment methods search for the best full-scale correspondence between two or more networks, while a local alignment returns small subgraphs representing complexes or pathways retained among the input networks. Several of these methods can map each node of one graph to a single node on the other, or with many nodes, and vice-versa, according

to their sequence similarity, so that a dictionary of similarities is needed for the algorithm to proceed. Most of the alignment methods use dictionaries focused on protein sequences and these dictionaries will need to be implemented for RNA and genes to be applied to ncRNA regulatory networks. Similar approaches can be used for querying a target network with a specific module of interest, to see if similar patterns are conserved in the target. This approach is referred to as “Network querying” and again it relies on node similarity and on the topology of the module [94,95]. For a summary of alignment see [96,97]. Computational strategies have also been developed to identify recurring, significant patterns of directed interactions, known as “motifs”, that may correspond to a specific function [98]. The simplest motif is the so-called feed-forward loop (FFL): a three-node motif where each node is linked to the others by directed edges. Each of the three regulatory interactions can be either activation or repression, leading to different types of FFL [99]. An FFL may represent the directed interaction between a transcription factor, a gene and a ncRNA that, in turn, regulates the transcription factor or the target gene (see Figure 3B). In the human transcription network 2377 microRNA-mediated FFL have been identified, where a transcription factor regulates a target gene and a microRNA, which regulates the target gene [34]. A four-node motif termed bi-fan also frequently appears in networks. The bi-fan motif is composed by two source nodes (transcription factors or ncRNAs for instance) that cross-regulate two target nodes [98]. NetMatch and NetMatchStar are Cytoscape plugins which search for a query motif, drawn by the user using the NetMatch editor or chosen from a list of available motifs, and return subgraphs where nodes have the same connection geometry as the query. Nodes can be labelled with the associated GO terms to query for specific molecular function of the motifs [100,101].

In general, ontology terms have been widely used to assist in the description and analysis of data for proteins and genes [102,103]. The standardized controlled vocabulary of GO is hierarchical and specifically designed for supporting computational analyses: the biological process, molecular function and cellular localization of each entity are represented with three classes of ontologies in a structured syntax of identifiers that can be used to filter networks or to label the nodes [104]. Recently, terms representing transcription factors have been significantly revisited in collaboration with the Gene Regulation Consortium (<http://thegreco.org> [103]) and the annotation effort has expanded to include ncRNA entities [71]. Several resources, such as GOrize, BiNGO and ClueGO have been developed to visualize enriched GO terms in a list of gene products [105–109]. GO annotations are now

available also for microRNAs and a few lncRNAs in the GO Consortium annotation files [71]. The GO terms associated with the microRNAs describe more than just the binding activity of the RNA. Many of the associated GO terms describe the downstream effect of the microRNA, such as the regulation of angiogenesis or the regulation of leukocyte adhesion. Consequently, it is possible to conduct a functional enrichment on a microRNA regulatory network and highlight the enriched GO terms associated with both the microRNA and the target genes [71]. For more information about how to browse, download and use bio-ontologies please see [110].

## 4. Conclusion

In the last few years, hundreds of ncRNAs have been described, involved in several processes such as RNA maturation, regulation of transcription, chromatin remodelling and post-transcriptional RNA modifications, therefore it is unquestionable that ncRNAs have a crucial role in gene regulatory networks. The well-known regulative function of microRNAs however is juxtaposed with the plethora of transcripts without any documented function, many of which are classified as lncRNAs. Even the abundance of lncRNAs in human genome is still an open issue and estimations are merely based on transcriptional evidence combined with the lack of coding potential, leading to different estimates depending, among other factors, on the method used to identify putative ORFs [18]. Nevertheless, there is considerable evidence to suggest lncRNAs as optimal candidates to exert epigenetic control and markedly different mechanisms have been documented, ranging from the association with chromatin-remodelling complexes to the recruitment of transcription factors to enhancers or the muster of activators away from the promoter [111,112]. Despite the number of proposed models, a minimal percentage of the lncRNA transcripts have been confidently assigned to specific targets and few have a convenient common name. For the majority of lncRNAs, information derives only from high-throughput experiments. As described above, advances in technology have made it possible to obtain results for thousands of transcripts in a single experiment, and indeed the focus of the biological research has moved towards a holistic approach with the “system biology” that gives an integrative interpretation of the data, by means of computational modelling [113]. Results of bioinformatics predictions and models are, however, heavily dependent on the quality of the

data used in the training. In this prospect, regulative ncRNA networks have three points of weakness:

*Identifiers:* One of the major obstacles in merging RNA data is the choice of identifiers made by the different resources. RNACentral [54] was launched in 2014 to solve this issue for ncRNAs, however, the use of miRBase entries [4] for microRNAs and of genomic location for lncRNA is still largely preferred. Although mRNA IDs are available from both Ensembl and NCBI, these entities are often identified by the NCBI, HGNC or Ensembl gene IDs [52]. This is most likely to be due to the fact that there is no ‘parent’ or ‘canonical’ mRNA ID to group the large number of transcripts associated with a single gene, as is the case for protein isoforms within UniProt records. In addition, very few experiments provide sufficient evidence for a specific transcript ID to be assigned. As a consequence, information about the effect of a specific RNA mutation on a disease is very difficult to map from one resource to another. In this context, it is worth mentioning the project named “Functional Annotation of Mammalian Genomes 5” (FANTOM 5) (<http://fantom.gsc.riken.jp/5/>), that aims at mapping the human and mouse transcriptome on the respective genomes with a method based on the cap analysis of gene expression (CAGE) that correctly identify the 5’ ends of capped RNAs. This project has provided atlases of mRNAs, lncRNAs and microRNAs [114]. As we have described above, regulative networks can consider physical binding between the ncRNA and the regulated gene/messenger, or the causal interaction that produce the downregulation/upregulation of the expression. The use of the gene ID is consistent with the causal information, but can be misleading in the representation of a physical binding.

*GO terms:* Thanks to the work of the GO consortium (<http://geneontology.org>) proteins and genes entries are associated with a detailed set of computable ontology terms supported by the literature, that can be used to classify, filter and analyse the networks. The development of terms, to describe the activities of ncRNAs, is still in progress and the majority of ncRNAs are not yet annotated.

*Standards:* according to the FAIR principles (<https://fairsharing.org>), to enhance the value of the digital resources, integration of data and metadata will be facilitated by the adoption of common protocols and quality standards [115]. Detailed information of ncRNA physical or functional interactions is still largely scattered in the literature or annotated in resources with very different quality standards. Indeed, most of the collected interactions derive from

large-scale experiments or text mining import, result in an urgent need for sets of gold standard interactions, manually curated from the literature. Appropriate standards and data formats to enable this are supplied by the HUPO PSI-MI data exchange formats and associated controlled vocabulary [81].

Experimental characterization of ncRNAs remains a major challenge but the amount of available data is slowly increasing and the implementation of standards by databases and repositories should improve the reproducibility of analyses. Despite this, studies on regulatory networks of proteins, genes and ncRNAs have already provided new insights into the mechanism of both physiological and dis-regulated systems [116,117] and efforts from more integrated studies will contribute in the near future to the better identification of molecular driver events and help to in the search for potential therapeutics or predictors of disease progression and drug efficacy [118,119].

## Acknowledgements

This article is based upon work from COST Action CA15205 (Gene Regulation Ensemble Effort for the Knowledge Commons) supported by COST (European Cooperation in Science and Technology, [www.cost.eu](http://www.cost.eu)). R.LC was supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre and Alzheimer's Research UK grant ARUK-NAS2017A-1.

## References

- [1] T.I. Lee, N.J. Rinaldi, F. Robert, D.T. Odom, Z. Bar-Joseph, G.K. Gerber, N.M. Hannett, C.T. Harbison, C.M. Thompson, I. Simon, J. Zeitlinger, E.G. Jennings, H.L. Murray, D.B. Gordon, B. Ren, J.J. Wyrick, J.-B. Tagne, T.L. Volkert, E. Fraenkel, D.K. Gifford, R.A. Young, Transcriptional regulatory networks in *Saccharomyces cerevisiae*, *Science*. 298 (2002) 799–804. doi:10.1126/science.1075090.
- [2] V. Ambros, The functions of animal microRNAs, *Nature*. 431 (2004) 350–355. doi:10.1038/nature02871.
- [3] G. Meister, T. Tuschl, Mechanisms of gene silencing by double-stranded RNA, *Nature*. 431 (2004) 343–349. doi:10.1038/nature02873.
- [4] A. Kozomara, S. Griffiths-Jones, miRBase: annotating high confidence microRNAs using deep sequencing data, *Nucleic Acids Res.* 42 (2014) D68–73. doi:10.1093/nar/gkt1181.
- [5] A. Kozomara, M. Birgaoanu, S. Griffiths-Jones, miRBase: from microRNA sequences to function, *Nucleic Acids Res.* 47 (2019) D155–D162. doi:10.1093/nar/gky1141.

- [6] 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 (2005) 15–20. doi:10.1016/j.cell.2004.12.035.
- [7] V. Agarwal, G.W. Bell, J.-W. Nam, D.P. Bartel, Predicting effective microRNA target sites in mammalian mRNAs, *Elife*. 4 (2015). doi:10.7554/eLife.05005.
- [8] R.C. Friedman, K.K.-H. Farh, C.B. Burge, D.P. Bartel, Most mammalian mRNAs are conserved targets of microRNAs, *Genome Res.* 19 (2009) 92–105. doi:10.1101/gr.082701.108.
- [9] C. Shin, J.-W. Nam, K.K.-H. Farh, H.R. Chiang, A. Shkumatava, D.P. Bartel, Expanding the microRNA targeting code: functional sites with centered pairing, *Mol. Cell*. 38 (2010) 789–802. doi:10.1016/j.molcel.2010.06.005.
- [10] A. Krek, D. Grün, M.N. Poy, R. Wolf, L. Rosenberg, E.J. Epstein, P. MacMenamin, I. da Piedade, K.C. Gunsalus, M. Stoffel, N. Rajewsky, Combinatorial microRNA target predictions, *Nat. Genet.* 37 (2005) 495–500. doi:10.1038/ng1536.
- [11] B. John, A.J. Enright, A. Aravin, T. Tuschl, C. Sander, D.S. Marks, Human MicroRNA targets, *PLoS Biol.* 2 (2004) e363. doi:10.1371/journal.pbio.0020363.
- [12] J. Krüger, M. Rehmsmeier, RNAhybrid: microRNA target prediction easy, fast and flexible, *Nucleic Acids Res.* 34 (2006) W451–454. doi:10.1093/nar/gkl243.
- [13] K.C. Miranda, T. Huynh, Y. Tay, Y.-S. Ang, W.-L. Tam, A.M. Thomson, B. Lim, I. Rigoutsos, A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes, *Cell*. 126 (2006) 1203–1217. doi:10.1016/j.cell.2006.07.031.
- [14] N. Wong, X. Wang, miRDB: an online resource for microRNA target prediction and functional annotations, *Nucleic Acids Res.* 43 (2015) D146–152. doi:10.1093/nar/gku1104.
- [15] Á.L. Riffo-Campos, I. Riquelme, P. Brebi-Mieville, Tools for Sequence-Based miRNA Target Prediction: What to Choose?, *Int J Mol Sci.* 17 (2016). doi:10.3390/ijms17121987.
- [16] Y. Zhang, Y. Xu, L. Feng, F. Li, Z. Sun, T. Wu, X. Shi, J. Li, X. Li, Comprehensive characterization of lncRNA-mRNA related ceRNA network across 12 major cancers, *Oncotarget*. 7 (2016) 64148–64167. doi:10.18632/oncotarget.11637.
- [17] R. Shalgi, D. Lieber, M. Oren, Y. Pilpel, Global and local architecture of the mammalian microRNA-transcription factor regulatory network, *PLoS Comput. Biol.* 3 (2007) e131. doi:10.1371/journal.pcbi.0030131.
- [18] M.E. Dinger, K.C. Pang, T.R. Mercer, J.S. Mattick, Differentiating protein-coding and noncoding RNA: challenges and ambiguities, *PLoS Comput. Biol.* 4 (2008) e1000176. doi:10.1371/journal.pcbi.1000176.
- [19] M.-C. Tsai, O. Manor, Y. Wan, N. Mosammaparast, J.K. Wang, F. Lan, Y. Shi, E. Segal, H.Y. Chang, Long noncoding RNA as modular scaffold of histone modification complexes, *Science*. 329 (2010) 689–693. doi:10.1126/science.1192002.
- [20] P. Paci, T. Colombo, L. Farina, Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer, *BMC Syst Biol.* 8 (2014) 83. doi:10.1186/1752-0509-8-83.
- [21] S. Fanucchi, E.T. Fok, E. Dalla, Y. Shibayama, K. Börner, E.Y. Chang, S. Stoychev, M. Imakaev, D. Grimm, K.C. Wang, G. Li, W.-K. Sung, M.M. Mhlanga, Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments, *Nat. Genet.* 51 (2019) 138–150. doi:10.1038/s41588-018-0298-2.
- [22] J. Salzman, C. Gawad, P.L. Wang, N. Lacayo, P.O. Brown, Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types, *PLoS ONE*. 7 (2012) e30733. doi:10.1371/journal.pone.0030733.
- [23] S. Memczak, M. Jens, A. Elefsinioti, F. Torti, J. Krueger, A. Rybak, L. Maier, S.D. Mackowiak, L.H. Gregersen, M. Munschauer, A. Loewer, U. Ziebold, M. Landthaler, C. Kocks, F. le Noble, N. Rajewsky, Circular RNAs are a large class of animal RNAs with regulatory potency, *Nature*. 495 (2013) 333–338. doi:10.1038/nature11928.
- [24] W.R. Jeck, J.A. Sorrentino, K. Wang, M.K. Slevin, C.E. Burd, J. Liu, W.F. Marzluff, N.E. Sharpless, Circular RNAs are abundant, conserved, and associated with ALU repeats, *RNA*. 19 (2013) 141–157. doi:10.1261/rna.035667.112.
- [25] P. Glažar, P. Papavasileiou, N. Rajewsky, circBase: a database for circular RNAs, *RNA*. 20 (2014) 1666–1670. doi:10.1261/rna.043687.113.
- [26] X. Chen, P. Han, T. Zhou, X. Guo, X. Song, Y. Li, circRNADb: A comprehensive database

- for human circular RNAs with protein-coding annotations, *Sci Rep.* 6 (2016) 34985. doi:10.1038/srep34985.
- [27] C.-Y. Chen, T.-J. Chuang, Comment on “A comprehensive overview and evaluation of circular RNA detection tools,” *PLoS Comput. Biol.* 15 (2019) e1006158. doi:10.1371/journal.pcbi.1006158.
- [28] T. Schneider, L.-H. Hung, S. Schreiner, S. Starke, H. Eckhof, O. Rossbach, S. Reich, J. Medenbach, A. Bindereif, CircRNA-protein complexes: IMP3 protein component defines subfamily of circRNPs, *Sci Rep.* 6 (2016) 31313. doi:10.1038/srep31313.
- [29] W.W. Du, W. Yang, Y. Chen, Z.-K. Wu, F.S. Foster, Z. Yang, X. Li, B.B. Yang, Foxo3 circular RNA promotes cardiac senescence by modulating multiple factors associated with stress and senescence responses, *Eur. Heart J.* 38 (2017) 1402–1412. doi:10.1093/eurheartj/ehw001.
- [30] T.B. Hansen, T.I. Jensen, B.H. Clausen, J.B. Bramsen, B. Finsen, C.K. Damgaard, J. Kjems, Natural RNA circles function as efficient microRNA sponges, *Nature.* 495 (2013) 384–388. doi:10.1038/nature11993.
- [31] M.T.Y. Lam, W. Li, M.G. Rosenfeld, C.K. Glass, Enhancer RNAs and regulated transcriptional programs, *Trends Biochem. Sci.* 39 (2014) 170–182. doi:10.1016/j.tibs.2014.02.007.
- [32] Y.J. Kim, P. Xie, L. Cao, M.Q. Zhang, T.H. Kim, Global transcriptional activity dynamics reveal functional enhancer RNAs, *Genome Res.* 28 (2018) 1799–1811. doi:10.1101/gr.233486.117.
- [33] D.A. Bose, S.L. Berger, eRNA binding produces tailored CBP activity profiles to regulate gene expression, *RNA Biol.* 14 (2017) 1655–1659. doi:10.1080/15476286.2017.1353862.
- [34] M.B. Gerstein, A. Kundaje, M. Hariharan, S.G. Landt, K.-K. Yan, C. Cheng, X.J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A.P. Boyle, P. Cayting, A. Charos, D.Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Fietze, Y. Fu, J. Gertz, F. Grubert, A. Harman, P. Jain, M. Kasowski, P. Lacroute, J.J. Leng, J. Lian, H. Monahan, H. O’Geen, Z. Ouyang, E.C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T.E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K.Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P.J. Farnham, R.M. Myers, S.M. Weissman, M. Snyder, Architecture of the human regulatory network derived from ENCODE data, *Nature.* 489 (2012) 91–100. doi:10.1038/nature11245.
- [35] A.P. Boyle, C.L. Araya, C. Brdlik, P. Cayting, C. Cheng, Y. Cheng, K. Gardner, L.W. Hillier, J. Janette, L. Jiang, D. Kasper, T. Kawli, P. Kheradpour, A. Kundaje, J.J. Li, L. Ma, W. Niu, E.J. Rehm, J. Rozowsky, M. Slatery, R. Spokony, R. Terrell, D. Vafeados, D. Wang, P. Weisdepp, Y.-C. Wu, D. Xie, K.-K. Yan, E.A. Feingold, P.J. Good, M.J. Pazin, H. Huang, P.J. Bickel, S.E. Brenner, V. Reinke, R.H. Waterston, M. Gerstein, K.P. White, M. Kellis, M. Snyder, Comparative analysis of regulatory information and circuits across distant species, *Nature.* 512 (2014) 453–456. doi:10.1038/nature13668.
- [36] D. Hurley, H. Araki, Y. Tamada, B. Dunmore, D. Sanders, S. Humphreys, M. Affara, S. Imoto, K. Yasuda, Y. Tomiyasu, K. Tashiro, C. Savoie, V. Cho, S. Smith, S. Kuhara, S. Miyano, D.S. Charnock-Jones, E.J. Crampin, C.G. Print, Gene network inference and visualization tools for biologists: application to new human transcriptome datasets, *Nucleic Acids Res.* 40 (2012) 2377–2398. doi:10.1093/nar/gkr902.
- [37] A.A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. Dalla 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. doi:10.1186/1471-2105-7-S1-S7.
- [38] A. Lachmann, F.M. Giorgi, G. Lopez, A. Califano, ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information, *Bioinformatics.* 32 (2016) 2233–2235. doi:10.1093/bioinformatics/btw216.
- [39] M.J. Alvarez, Y. Shen, F.M. Giorgi, A. Lachmann, B.B. Ding, B.H. Ye, A. Califano, Functional characterization of somatic mutations in cancer using network-based inference of protein activity, *Nat. Genet.* 48 (2016) 838–847. doi:10.1038/ng.3593.
- [40] P. Rajbhandari, G. Lopez, C. Capdevila, B. Salvatori, J. Yu, R. Rodriguez-Barrueco, D. Martinez, M. Yarmarkovich, N. Weichert-Leahey, B.J. Abraham, M.J. Alvarez, A. Iyer, J.L. Harenza, D. Oldridge, K. De Preter, J. Koster, S. Asgharzadeh, R.C. Seeger, J.S. Wei, J. Khan, J. Vandesompele, P. Mestdag, R. Versteeg, A.T. Look, R.A. Young, A. Iavarone, A. Lasorella, J.M.

- Silva, J.M. Maris, A. Califano, Cross-Cohort Analysis Identifies a TEAD4-MYC Positive Feedback Loop as the Core Regulatory Element of High-Risk Neuroblastoma, *Cancer Discov.* 8 (2018) 582–599. doi:10.1158/2159-8290.CD-16-0861.
- [41] A.J. Levine, p53, the cellular gatekeeper for growth and division, *Cell.* 88 (1997) 323–331. doi:10.1016/s0092-8674(00)81871-1.
- [42] E. Grossi, Y. Sánchez, M. Huarte, Expanding the p53 regulatory network: lncRNAs take up the challenge, *Biochim. Biophys. Acta.* 1859 (2016) 200–208. doi:10.1016/j.bbarm.2015.07.011.
- [43] H. Hermeking, MicroRNAs in the p53 network: micromanagement of tumour suppression, *Nat. Rev. Cancer.* 12 (2012) 613–626. doi:10.1038/nrc3318.
- [44] M. Sachdeva, S. Zhu, F. Wu, H. Wu, V. Walia, S. Kumar, R. Elble, K. Watabe, Y.-Y. Mo, p53 represses c-Myc through induction of the tumor suppressor miR-145, *Proc. Natl. Acad. Sci. U.S.A.* 106 (2009) 3207–3212. doi:10.1073/pnas.0808042106.
- [45] Y. Sánchez, V. Segura, O. Marín-Béjar, A. Athie, F.P. Marchese, J. González, L. Bujanda, S. Guo, A. Matheu, M. Huarte, Genome-wide analysis of the human p53 transcriptional network unveils a lncRNA tumour suppressor signature, *Nat Commun.* 5 (2014) 5812. doi:10.1038/ncomms6812.
- [46] J. Liu, C. Zhang, Y. Zhao, Z. Feng, MicroRNA Control of p53, *J. Cell. Biochem.* 118 (2017) 7–14. doi:10.1002/jcb.25609.
- [47] P. Sood, A. Krek, M. Zavolan, G. Macino, N. Rajewsky, Cell-type-specific signatures of microRNAs on target mRNA expression, *Proc. Natl. Acad. Sci. U.S.A.* 103 (2006) 2746–2751. doi:10.1073/pnas.0511045103.
- [48] D. Betel, A. Koppal, P. Agius, C. Sander, C. Leslie, Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites, *Genome Biol.* 11 (2010) R90. doi:10.1186/gb-2010-11-8-r90.
- [49] F. Xiao, Z. Zuo, G. Cai, S. Kang, X. Gao, T. Li, miRecords: an integrated resource for microRNA-target interactions, *Nucleic Acids Res.* 37 (2009) D105–110. doi:10.1093/nar/gkn851.
- [50] J.-H. Li, S. Liu, H. Zhou, L.-H. Qu, J.-H. Yang, starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data, *Nucleic Acids Res.* 42 (2014) D92–97. doi:10.1093/nar/gkt1248.
- [51] C. Sticht, C. De La Torre, A. Parveen, N. Gretz, miRWalk: An online resource for prediction of microRNA binding sites, *PLoS ONE.* 13 (2018) e0206239. doi:10.1371/journal.pone.0206239.
- [52] D.R. Zerbino, P. Achuthan, W. Akanni, M.R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C.G. Girón, L. Gil, L. Gordon, L. Haggerty, E. Haskell, T. Hourlier, O.G. Izuogu, S.H. Janacek, T. Juettemann, J.K. To, M.R. Laird, I. Lavidas, Z. Liu, J.E. Loveland, T. Maurel, W. McLaren, B. Moore, J. Mudge, D.N. Murphy, V. Newman, M. Nuhn, D. Ogeh, C.K. Ong, A. Parker, M. Patricio, H.S. Riat, H. Schuilenburg, D. Sheppard, H. Sparrow, K. Taylor, A. Thormann, A. Vullo, B. Walts, A. Zadissa, A. Frankish, S.E. Hunt, M. Kostadima, N. Langridge, F.J. Martin, M. Muffato, E. Perry, M. Ruffier, D.M. Staines, S.J. Trevanion, B.L. Aken, F. Cunningham, A. Yates, P. Flicek, Ensembl 2018, *Nucleic Acids Res.* 46 (2018) D754–D761. doi:10.1093/nar/gkx1098.
- [53] A. Frankish, M. Diekhans, A.-M. Ferreira, R. Johnson, I. Jungreis, J. Loveland, J.M. Mudge, C. Sisu, J. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, S. Carbonell Sala, J. Chrast, F. Cunningham, T. Di Domenico, S. Donaldson, I.T. Fiddes, C. García Girón, J.M. Gonzalez, T. Grego, M. Hardy, T. Hourlier, T. Hunt, O.G. Izuogu, J. Lagarde, F.J. Martin, L. Martínez, S. Mohanan, P. Muir, F.C.P. Navarro, A. Parker, B. Pei, F. Pozo, M. Ruffier, B.M. Schmitt, E. Stapleton, M.-M. Suner, I. Sycheva, B. Uszczyńska-Ratajczak, J. Xu, A. Yates, D. Zerbino, Y. Zhang, B. Aken, J.S. Choudhary, M. Gerstein, R. Guigó, T.J.P. Hubbard, M. Kellis, B. Paten, A. Reymond, M.L. Tress, P. Flicek, GENCODE reference annotation for the human and mouse genomes, *Nucleic Acids Res.* 47 (2019) D766–D773. doi:10.1093/nar/gky955.
- [54] The RNAcentral Consortium, RNAcentral: a hub of information for non-coding RNA sequences, *Nucleic Acids Res.* 47 (2019) D1250–D1251. doi:10.1093/nar/gky1206.
- [55] D.B. Dudekula, A.C. Panda, I. Grammatikakis, S. De, K. Abdelmohsen, M. Gorospe, CircInteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs, *RNA Biol.* 13 (2016) 34–42. doi:10.1080/15476286.2015.1128065.
- [56] Y.-C. Liu, J.-R. Li, C.-H. Sun, E. Andrews, R.-F. Chao, F.-M. Lin, S.-L. Weng, S.-D. Hsu, C.-C. Huang, C. Cheng, C.-C. Liu, H.-D. Huang, CircNet: a database of circular RNAs derived from transcriptome sequencing data, *Nucleic Acids Res.* 44 (2016) D209–215. doi:10.1093/nar/gkv940.

- [57] J. Ule, K.B. Jensen, M. Ruggiu, A. Mele, A. Ule, R.B. Darnell, CLIP identifies Nova-regulated RNA networks in the brain, *Science*. 302 (2003) 1212–1215. doi:10.1126/science.1090095.
- [58] D.D. Licatalosi, A. Mele, J.J. Fak, J. Ule, M. Kayikci, S.W. Chi, T.A. Clark, A.C. Schweitzer, J.E. Blume, X. Wang, J.C. Darnell, R.B. Darnell, HITS-CLIP yields genome-wide insights into brain alternative RNA processing, *Nature*. 456 (2008) 464–469. doi:10.1038/nature07488.
- [59] M. Hafner, M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano, A.-C. Jungkamp, M. Munschauer, A. Ulrich, G.S. Wardle, S. Dewell, M. Zavolan, T. Tuschl, Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP, *Cell*. 141 (2010) 129–141. doi:10.1016/j.cell.2010.03.009.
- [60] G. Kudla, S. Granneman, D. Hahn, J.D. Beggs, D. Tollervey, Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast, *Proc. Natl. Acad. Sci. U.S.A.* 108 (2011) 10010–10015. doi:10.1073/pnas.1017386108.
- [61] J. König, K. Zarnack, N.M. Luscombe, J. Ule, Protein-RNA interactions: new genomic technologies and perspectives, *Nat. Rev. Genet.* 13 (2012) 77–83. doi:10.1038/nrg3141.
- [62] D. Thiel, N.D. Conrad, E. Ntini, R.X. Peschutter, H. Siebert, A. Marsico, Identifying lncRNA-mediated regulatory modules via ChIA-PET network analysis, *BMC Bioinformatics*. 20 (2019) 292. doi:10.1186/s12859-019-2900-8.
- [63] Y. Yi, Y. Zhao, C. Li, L. Zhang, H. Huang, Y. Li, L. Liu, P. Hou, T. Cui, P. Tan, Y. Hu, T. Zhang, Y. Huang, X. Li, J. Yu, D. Wang, RAID v2.0: an updated resource of RNA-associated interactions across organisms, *Nucleic Acids Res.* 45 (2017) D115–D118. doi:10.1093/nar/gkw1052.
- [64] D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N.T. Doncheva, J.H. Morris, P. Bork, L.J. Jensen, C. von Mering, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, *Nucleic Acids Res.* 47 (2019) D607–D613. doi:10.1093/nar/gky1131.
- [65] A. Junge, J.C. Refsgaard, C. Garde, X. Pan, A. Santos, F. Alkan, C. Anthon, C. von Mering, C.T. Workman, L.J. Jensen, J. Gorodkin, RAIN: RNA-protein Association and Interaction Networks, Database (Oxford). 2017 (2017). doi:10.1093/database/baw167.
- [66] F. Yu, G. Zhang, A. Shi, J. Hu, F. Li, X. Zhang, Y. Zhang, J. Huang, Y. Xiao, X. Li, S. Cheng, LnChrom: a resource of experimentally validated lncRNA-chromatin interactions in human and mouse, Database (Oxford). 2018 (2018). doi:10.1093/database/bay039.
- [67] G. An, J. Sun, C. Ren, Z. Ouyang, L. Zhu, X. Bo, S. Peng, W. Shu, LIVE: a manually curated encyclopedia of experimentally validated interactions of lncRNAs, Database (Oxford). 2019 (2019). doi:10.1093/database/baz011.
- [68] D. Fazekas, M. Koltai, D. Türei, D. Módos, M. Pálffy, Z. Dúl, L. Zsákai, M. Szalay-Bekő, K. Lenti, I.J. Farkas, T. Vellai, P. Csermely, T. Korcsmáros, Signalink 2 - a signaling pathway resource with multi-layered regulatory networks, *BMC Syst Biol.* 7 (2013) 7. doi:10.1186/1752-0509-7-7.
- [69] L. Perfetto, L. Briganti, A. Calderone, A. Cerquone Perpetuini, M. Iannuccelli, F. Langone, L. Licata, M. Marinkovic, A. Mattioni, T. Pavlidou, D. Peluso, L.L. Petrilli, S. Pirrò, D. Posca, E. Santonico, A. Silvestri, F. Spada, L. Castagnoli, G. Cesareni, SIGNOR: a database of causal relationships between biological entities, *Nucleic Acids Res.* 44 (2016) D548–554. doi:10.1093/nar/gkv1048.
- [70] A. Fabregat, F. Korninger, G. Viteri, K. Sidiropoulos, P. Marin-Garcia, P. Ping, G. Wu, L. Stein, P. D'Eustachio, H. Hermjakob, Reactome graph database: Efficient access to complex pathway data, *PLoS Comput. Biol.* 14 (2018) e1005968. doi:10.1371/journal.pcbi.1005968.
- [71] R.P. Huntley, B. Kramarz, T. Sawford, Z. Umrao, A. Kalea, V. Acquaah, M.J. Martin, M. Mayr, R.C. Lovering, Expanding the horizons of microRNA bioinformatics, *RNA*. 24 (2018) 1005–1017. doi:10.1261/rna.065565.118.
- [72] R.P. Huntley, D. Sitnikov, M. Orlic-Milacic, R. Balakrishnan, P. D'Eustachio, M.E. Gillespie, D. Howe, A.Z. Kalea, L. Maegdefessel, D. Osumi-Sutherland, V. Petri, J.R. Smith, K. Van Auken, V. Wood, A. Zampetaki, M. Mayr, R.C. Lovering, Guidelines for the functional annotation of microRNAs using the Gene Ontology, *RNA*. 22 (2016) 667–676. doi:10.1261/rna.055301.115.
- [73] R.P. Huntley, M.A. Harris, Y. Alam-Faruque, J.A. Blake, S. Carbon, H. Dietze, E.C. Dimmer, R.E. Foulger, D.P. Hill, V.K. Khodiyar, A. Lock, J. Lomax, R.C. Lovering, P. Mutowo-

- Meulenet, T. Sawford, K. Van Auken, V. Wood, C.J. Mungall, A method for increasing expressivity of Gene Ontology annotations using a compositional approach, *BMC Bioinformatics*. 15 (2014) 155. doi:10.1186/1471-2105-15-155.
- [74] M. Munoz-Torres, S. Carbon, Get GO! Retrieving GO Data Using AmiGO, QuickGO, API, Files, and Tools, *Methods Mol. Biol.* 1446 (2017) 149–160. doi:10.1007/978-1-4939-3743-1\_11.
- [75] N. del-Toro, M. Dumousseau, S. Orchard, R.C. Jimenez, E. Galeota, G. Launay, J. Goll, K. Breuer, K. Ono, L. Salwinski, H. Hermjakob, A new reference implementation of the PSICQUIC web service, *Nucleic Acids Res.* 41 (2013) W601-606. doi:10.1093/nar/gkt392.
- [76] H. Hermjakob, L. Montecchi-Palazzi, G. Bader, J. Wojcik, L. Salwinski, A. Ceol, S. Moore, S. Orchard, U. Sarkans, C. von Mering, B. Roechert, S. Poux, E. Jung, H. Mersch, P. Kersey, M. Lappe, Y. Li, R. Zeng, D. Rana, M. Nikolski, H. Husi, C. Brun, K. Shanker, S.G.N. Grant, C. Sander, P. Bork, W. Zhu, A. Pandey, A. Brazma, B. Jacq, M. Vidal, D. Sherman, P. Legrain, G. Cesareni, I. Xenarios, D. Eisenberg, B. Steipe, C. Hogue, R. Apweiler, The HUPO PSI's molecular interaction format--a community standard for the representation of protein interaction data, *Nat. Biotechnol.* 22 (2004) 177–183. doi:10.1038/nbt926.
- [77] E.W. Deutsch, S. Orchard, P.-A. Binz, W. Bittremieux, M. Eisenacher, H. Hermjakob, S. Kawano, H. Lam, G. Mayer, G. Menschaert, Y. Perez-Riverol, R.M. Salek, D.L. Tabb, S. Tenzer, J.A. Vizcaino, M. Walzer, A.R. Jones, Proteomics Standards Initiative: Fifteen Years of Progress and Future Work, *J. Proteome Res.* 16 (2017) 4288–4298. doi:10.1021/acs.jproteome.7b00370.
- [78] S. Panni, A. Prakash, A. Bateman, S. Orchard, The yeast noncoding RNA interaction network, *RNA*. 23 (2017) 1479–1492. doi:10.1261/rna.060996.117.
- [79] S. Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N.H. Campbell, G. Chavali, C. Chen, N. del-Toro, M. Duesbury, M. Dumousseau, E. Galeota, U. Hinz, M. Iannuccelli, S. Jagannathan, R. Jimenez, J. Khadake, A. Lagreid, L. Licata, R.C. Lovering, B. Meldal, A.N. Melidoni, M. Milagros, D. Peluso, L. Perfetto, P. Porras, A. Raghunath, S. Ricard-Blum, B. Roechert, A. Stutz, M. Tognolli, K. van Roey, G. Cesareni, H. Hermjakob, The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases, *Nucleic Acids Res.* 42 (2014) D358-363. doi:10.1093/nar/gkt1115.
- [80] S. Kerrien, S. Orchard, L. Montecchi-Palazzi, B. Aranda, A.F. Quinn, N. Vinod, G.D. Bader, I. Xenarios, J. Wojcik, D. Sherman, M. Tyers, J.J. Salama, S. Moore, A. Ceol, A. Chatr-Aryamontri, M. Oesterheld, V. Stümpflen, L. Salwinski, J. Nerothin, E. Cerami, M.E. Cusick, M. Vidal, M. Gilson, J. Armstrong, P. Woollard, C. Hogue, D. Eisenberg, G. Cesareni, R. Apweiler, H. Hermjakob, Broadening the horizon--level 2.5 of the HUPO-PSI format for molecular interactions, *BMC Biol.* 5 (2007) 44. doi:10.1186/1741-7007-5-44.
- [81] M. Sivade Dumousseau, D. Alonso-López, M. Ammari, G. Bradley, N.H. Campbell, A. Ceol, G. Cesareni, C. Combe, J. De Las Rivas, N. Del-Toro, J. Heimbach, H. Hermjakob, I. Jurisica, M. Koch, L. Licata, R.C. Lovering, D.J. Lynn, B.H.M. Meldal, G. Micklem, S. Panni, P. Porras, S. Ricard-Blum, B. Roechert, L. Salwinski, A. Shrivastava, J. Sullivan, N. Thierry-Mieg, Y. Yehudi, K. Van Roey, S. Orchard, Encompassing new use cases - level 3.0 of the HUPO-PSI format for molecular interactions, *BMC Bioinformatics*. 19 (2018) 134. doi:10.1186/s12859-018-2118-1.
- [82] IMEx Consortium Curators, N. Del-Toro, M. Duesbury, M. Koch, L. Perfetto, A. Shrivastava, D. Ochoa, O. Wagih, J. Piñero, M. Kotlyar, C. Pastrello, P. Beltrao, L.I. Furlong, I. Jurisica, H. Hermjakob, S. Orchard, P. Porras, Capturing variation impact on molecular interactions in the IMEx Consortium mutations data set, *Nat Commun.* 10 (2019) 10. doi:10.1038/s41467-018-07709-6.
- [83] S.-J. Schramm, V. Jayaswal, A. Goel, S.S. Li, Y.H. Yang, G.J. Mann, M.R. Wilkins, Molecular interaction networks for the analysis of human disease: utility, limitations, and considerations, *Proteomics*. 13 (2013) 3393–3405. doi:10.1002/pmic.201200570.
- [84] P. Creixell, J. Reimand, S. Haider, G. Wu, T. Shibata, M. Vazquez, V. Mustonen, A. Gonzalez-Perez, J. Pearson, C. Sander, B.J. Raphael, D.S. Marks, B.F.F. Ouellette, A. Valencia, G.D. Bader, P.C. Boutros, J.M. Stuart, R. Linding, N. Lopez-Bigas, L.D. Stein, Mutation Consequences and Pathway Analysis Working Group of the International Cancer Genome Consortium, Pathway and network analysis of cancer genomes, *Nat. Methods*. 12 (2015) 615–621. doi:10.1038/nmeth.3440.
- [85] S.K. Miryala, A. Anbarasu, S. Ramaiah, Discerning molecular interactions: A comprehensive review on biomolecular interaction databases and network analysis tools, *Gene*.

642 (2018) 84–94. doi:10.1016/j.gene.2017.11.028.

[86] N.T. Csardi G, The igraph software package for complex network research, InterJournal, Complex Systems 1695. (n.d.). <http://igraph.org>.

[87] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.* 13 (2003) 2498–2504. doi:10.1101/gr.1239303.

[88] N. Le Novère, Quantitative and logic modelling of molecular and gene networks, *Nat. Rev. Genet.* 16 (2015) 146–158. doi:10.1038/nrg3885.

[89] B.P. Kelley, R. Sharan, R.M. Karp, T. Sittler, D.E. Root, B.R. Stockwell, T. Ideker, Conserved pathways within bacteria and yeast as revealed by global protein network alignment, *Proc. Natl. Acad. Sci. U.S.A.* 100 (2003) 11394–11399. doi:10.1073/pnas.1534710100.

[90] R. Singh, J. Xu, B. Berger, Global alignment of multiple protein interaction networks with application to functional orthology detection, *Proc. Natl. Acad. Sci. U.S.A.* 105 (2008) 12763–12768. doi:10.1073/pnas.0806627105.

[91] N. Ferraro, L. Palopoli, S. Panni, S.E. Rombo, Asymmetric comparison and querying of biological networks, *IEEE/ACM Trans Comput Biol Bioinform.* 8 (2011) 876–889. doi:10.1109/TCBB.2011.29.

[92] R.A. Pache, A. Céol, P. Aloy, NetAligner--a network alignment server to compare complexes, pathways and whole interactomes, *Nucleic Acids Res.* 40 (2012) W157–161. doi:10.1093/nar/gks446.

[93] M. Mina, P.H. Guzzi, Improving the Robustness of Local Network Alignment: Design and Extensive Assessment of a Markov Clustering-Based Approach, *IEEE/ACM Trans Comput Biol Bioinform.* 11 (2014) 561–572. doi:10.1109/TCBB.2014.2318707.

[94] Q. Huang, L.-Y. Wu, X.-S. Zhang, An efficient network querying method based on conditional random fields, *Bioinformatics.* 27 (2011) 3173–3178. doi:10.1093/bioinformatics/btr524.

[95] H. Jeong, B.-J. Yoon, SEQUOIA: significance enhanced network querying through context-sensitive random walk and minimization of network conductance, *BMC Syst Biol.* 11 (2017) 20. doi:10.1186/s12918-017-0404-6.

[96] F.E. Faisal, L. Meng, J. Crawford, T. Milenković, The post-genomic era of biological network alignment, *EURASIP J Bioinform Syst Biol.* 2015 (2015) 3. doi:10.1186/s13637-015-0022-9.

[97] S. Panni, S.E. Rombo, Searching for repetitions in biological networks: methods, resources and tools, *Brief. Bioinformatics.* 16 (2015) 118–136. doi:10.1093/bib/bbt084.

[98] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, *Science.* 298 (2002) 824–827. doi:10.1126/science.298.5594.824.

[99] S. Mangan, U. Alon, Structure and function of the feed-forward loop network motif, *Proc. Natl. Acad. Sci. U.S.A.* 100 (2003) 11980–11985. doi:10.1073/pnas.2133841100.

[100] A. Ferro, R. Giugno, G. Pigola, A. Pulvirenti, D. Skripin, G.D. Bader, D. Shasha, NetMatch: a Cytoscape plugin for searching biological networks, *Bioinformatics.* 23 (2007) 910–912. doi:10.1093/bioinformatics/btm032.

[101] F. Rinnone, G. Micale, V. Bonnici, G.D. Bader, D. Shasha, A. Ferro, A. Pulvirenti, R. Giugno, NetMatchStar: an enhanced Cytoscape network querying app, *F1000Res.* 4 (2015) 479. doi:10.12688/f1000research.6656.2.

[102] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat. Genet.* 25 (2000) 25–29. doi:10.1038/75556.

[103] The Gene Ontology Consortium, The Gene Ontology Resource: 20 years and still GOing strong, *Nucleic Acids Res.* 47 (2019) D330–D338. doi:10.1093/nar/gky1055.

[104] S. Jain, G.D. Bader, An improved method for scoring protein-protein interactions using semantic similarity within the gene ontology, *BMC Bioinformatics.* 11 (2010) 562. doi:10.1186/1471-2105-11-562.

[105] S. Maere, K. Heymans, M. Kuiper, BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks, *Bioinformatics.* 21 (2005)

3448–3449. doi:10.1093/bioinformatics/bti551.

[106] O. Garcia, C. Saveanu, M. Cline, M. Fromont-Racine, A. Jacquier, B. Schwikowski, T. Aittokallio, Golorize: a Cytoscape plug-in for network visualization with Gene Ontology-based layout and coloring, *Bioinformatics*. 23 (2007) 394–396. doi:10.1093/bioinformatics/btl605.

[107] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pagès, Z. Trajanoski, J. Galon, ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, *Bioinformatics*. 25 (2009) 1091–1093. doi:10.1093/bioinformatics/btp101.

[108] D. Merico, R. Isserlin, G.D. Bader, Visualizing gene-set enrichment results using the Cytoscape plug-in enrichment map, *Methods Mol. Biol.* 781 (2011) 257–277. doi:10.1007/978-1-61779-276-2\_12.

[109] H. Mi, A. Muruganujan, X. Huang, D. Ebert, C. Mills, X. Guo, P.D. Thomas, Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0), *Nat Protoc.* 14 (2019) 703–721. doi:10.1038/s41596-019-0128-8.

[110] C. Dessimoz, N. Škunca, eds., *The Gene Ontology Handbook*, Springer New York, New York, NY, 2017. doi:10.1007/978-1-4939-3743-1.

[111] M. Ballarino, M. Morlando, A. Fatica, I. Bozzoni, Non-coding RNAs in muscle differentiation and musculoskeletal disease, *J. Clin. Invest.* 126 (2016) 2021–2030. doi:10.1172/JCI84419.

[112] J.R. Alvarez-Dominguez, H.F. Lodish, Emerging mechanisms of long noncoding RNA function during normal and malignant hematopoiesis, *Blood*. 130 (2017) 1965–1975. doi:10.1182/blood-2017-06-788695.

[113] A.-L. Barabási, Z.N. Oltvai, Network biology: understanding the cell's functional organization, *Nat. Rev. Genet.* 5 (2004) 101–113. doi:10.1038/nrg1272.

[114] M. Lizio, I. Abugessaisa, S. Noguchi, A. Kondo, A. Hasegawa, C.C. Hon, M. de Hoon, J. Severin, S. Oki, Y. Hayashizaki, P. Carninci, T. Kasukawa, H. Kawaji, Update of the FANTOM web resource: expansion to provide additional transcriptome atlases, *Nucleic Acids Research*. 47 (2019) D752–D758. doi:10.1093/nar/gky1099.

[115] S.-A. Sansone, P. Rocca-Serra, Review%3A Interoperability standards, Figshare. (2016). doi:10.6084/m9.figshare.4055496.

[116] N. Liu, E.N. Olson, MicroRNA regulatory networks in cardiovascular development, *Dev. Cell*. 18 (2010) 510–525. doi:10.1016/j.devcel.2010.03.010.

[117] D. Yang, Y. Sun, L. Hu, H. Zheng, P. Ji, C.V. Pecot, Y. Zhao, S. Reynolds, H. Cheng, R. Rupaimoole, D. Cogdell, M. Nykter, R. Broaddus, C. Rodriguez-Aguayo, G. Lopez-Berestein, J. Liu, I. Shmulevich, A.K. Sood, K. Chen, W. Zhang, Integrated analyses identify a master microRNA regulatory network for the mesenchymal subtype in serous ovarian cancer, *Cancer Cell*. 23 (2013) 186–199. doi:10.1016/j.ccr.2012.12.020.

[118] Y. Shen, M.J. Alvarez, B. Bisikirska, A. Lachmann, R. Realubit, S. Pampou, J. Coku, C. Karan, A. Califano, Systematic, network-based characterization of therapeutic target inhibitors, *PLoS Comput. Biol.* 13 (2017) e1005599. doi:10.1371/journal.pcbi.1005599.

[119] F. Cheng, I.A. Kovács, A.-L. Barabási, Network-based prediction of drug combinations, *Nat Commun.* 10 (2019) 1197. doi:10.1038/s41467-019-09186-x.

**Table 1.** Summary of miRNA-target prediction tools listed in this review

Name	Main Parameters	URL	Reference
TargetScan	Sequence complementarity Sequence conservation Context features	<a href="http://www.targetscan.org/vert_72/">http://www.targetscan.org/vert_72/</a>	[7]
PicTar	Sequence complementarity Sequence conservation	<a href="https://pictar.mdc-berlin.de/">https://pictar.mdc-berlin.de/</a>	[10]
MiRanda-mirSVR	Sequence complementarity Sequence conservation	<a href="http://www.microrna.org/microrna/home.do">http://www.microrna.org/microrna/home.do</a>	[48]
RNAhybrid	Minimum free energy for hybridization	<a href="https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid">https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid</a>	[12]
RNA22	Pattern Discovery	<a href="https://cm.jefferson.edu/rna22/">https://cm.jefferson.edu/rna22/</a>	[13]
MirTarget2	Machine-learning based	<a href="http://mirdb.org/">http://mirdb.org/</a>	[14]

**Table 2** Summary of the principal databases and resources described in the present study

Name	Featured Data	URL	Reference
<b>IntAct</b>	Interaction Database	<a href="https://www.ebi.ac.uk/intact/">https://www.ebi.ac.uk/intact/</a>	[79]
<b>RAID</b>	Interaction Database	<a href="http://www.rna-society.org/raid2/index.html">http://www.rna-society.org/raid2/index.html</a>	[63]
<b>RAIN</b>	Interaction Database	<a href="https://rth.dk/resources/rain/">https://rth.dk/resources/rain/</a>	[65]
<b>starBase</b>	Interaction Database	<a href="http://starbase.sysu.edu.cn/starbase2/index.php">http://starbase.sysu.edu.cn/starbase2/index.php</a>	[50]
<b>STRING</b>	Interaction Database	<a href="https://string-db.org/">https://string-db.org/</a>	[64]
<b>LnChrom</b>	lncRNA interaction	<a href="http://biocc.hrbmu.edu.cn/LnChrom/">http://biocc.hrbmu.edu.cn/LnChrom/</a>	[66]
<b>LIVE</b>	lncRNA interaction	<a href="https://live.bioinfotech.org/">https://live.bioinfotech.org/</a>	[67]
<b>SignalLink</b>	Causal interaction	<a href="http://signalink.org/">http://signalink.org/</a>	[68]
<b>SIGNOR</b>	Causal interaction	<a href="https://signor.uniroma2.it/">https://signor.uniroma2.it/</a>	[69]
<b>Reactome</b>	Causal interaction	<a href="https://reactome.org/">https://reactome.org/</a>	[70]
<b>CircInteractome</b>	circRNA interactions	<a href="https://circinteractome.nia.nih.gov/">https://circinteractome.nia.nih.gov/</a>	[55]
<b>CircNet</b>	circRNA interactions	<a href="http://syslab5.nchu.edu.tw/CircNet/">http://syslab5.nchu.edu.tw/CircNet/</a>	[56]
<b>miRecords</b>	microRNA interactions	<a href="http://c1.accurascience.com/miRecords/">http://c1.accurascience.com/miRecords/</a>	[49]
<b>miRWalk</b>	microRNA interactions	<a href="http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/">http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/</a>	[51]
<b>miRBase</b>	microRNA database	<a href="http://www.mirbase.org/">http://www.mirbase.org/</a>	[4]
<b>RNAcentral</b>	ncRNA database	<a href="https://rnacentral.org/">https://rnacentral.org/</a>	[54]

## Figure Legends

**Figure 1** Non-coding RNAs affect gene regulation through various mechanisms: microRNAs (green) and lncRNAs (red) can directly regulate gene expression by binding to the mRNA (microRNAs), to the gene/nascent transcript (lncRNAs) or to histone modifiers bringing them to a gene (lncRNAs). But lncRNAs and microRNAs can also regulate each other through the “sponge” effect. CircRNAs (orange circle) regulate microRNAs through the sponge effect and gene expression through mRNA splicing.

**Figure 2** A) MicroRNA biogenesis. MicroRNAs are transcribed in the nucleus as precursor molecules named pri-microRNAs which are folded into a hairpin structure. Pri-microRNAs are then processed by the RNase III Drosha and its cofactor DGCR8 into pre-microRNAs which are exported to the cytoplasm. The pre-microRNA is processed by the RNase III domain of Dicer into a double stranded microRNA of about 22 nucleotides with overhanging ends. The two strands of the mature microRNA are indicated as 3p and 5p. Occasionally both strands give rise to functional microRNAs but more often one of the two is degraded. The mature microRNA is incorporated into the RNA-induced-silencing-complex RISC and directed to the target mRNA to induce post-transcriptional gene silencing.

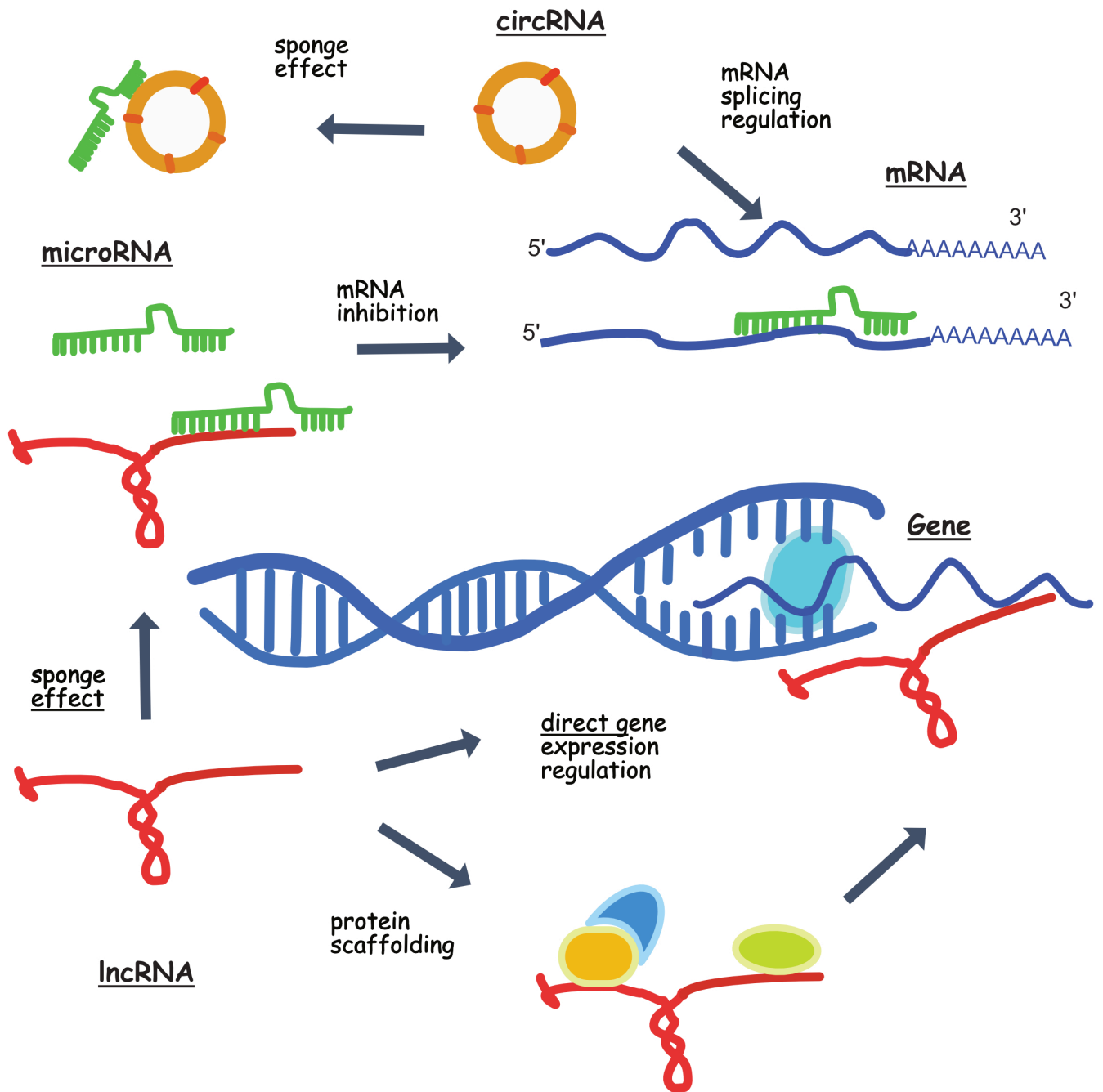
B) Pri-microRNA hairpin structure. Pri-microRNAs often have 5' cap and 3' poli-A as a result of Pol II transcription. The stem is 33 base-pairs long, with imperfect base pairing, and terminates with a loop of 10 nucleotides. The stem is composed of a basal section of 11 bp and a longer section of 22 bp which are separated by Drosha (black arrows indicate the positions where the pri-microRNA is cut).

C) The microRNA nucleotides that bind to the target mRNA (in green) typically map to the 5' region of the microRNA, position 2-7 (seed sequence). Occasionally an additional base-pairing site mapping to nucleotides 13-16 is present. Alternatively, an unconventional site mapping to nucleotides 4-15 can mediate the binding (centred site).

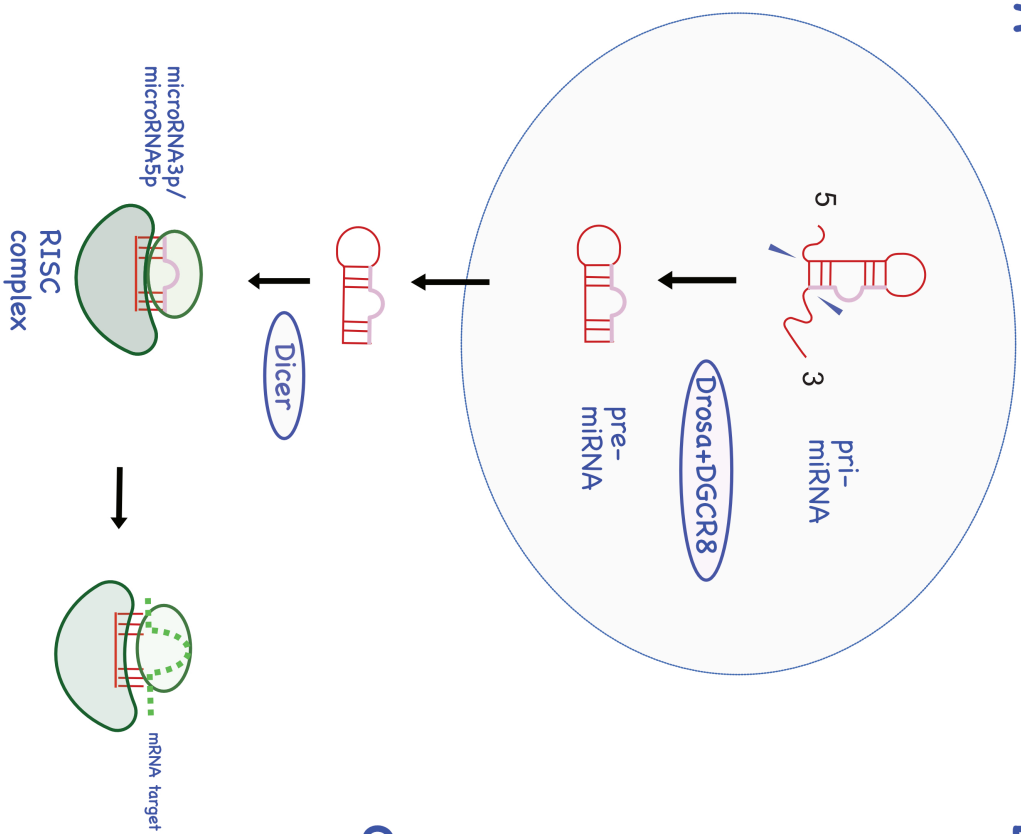
**Figure 3: Schematic representation of the p53 network.** A) The p53 transcription factor activates several microRNAs (listed on the right side), and lncRNAs, but it is also regulated by microRNAs (two examples listed on the left side of the picture). lincRNAp21 and lincMkn1 are activated by p53 and contribute to the repression of target genes. B) Examples of feed-forward-loops involving the p53 protein and miR145.

**Figure 4:** A set of 39 true-positives taken from the EBI-GOA-miRNA and IntAct files and from the literature was compared with target prediction obtained from starBase platform (<http://starbase.sysu.edu.cn/starbase2/mirMrna.php>).

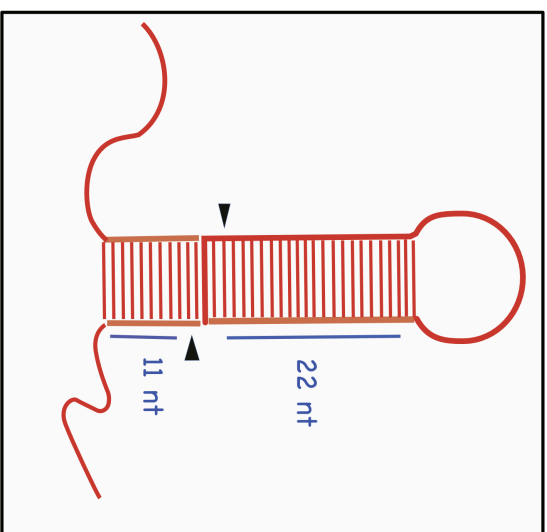
TargetScan predicts 646 potential interactors of which 13 are true-positives (bottom part of the figure, yellow ring). The intersection of TargetScan, picTar, PITA and miRanda retrieves 349 potential targets of which 12 are true-positives (upper part of the figure, blue ring). The intersection of TargetScan, picTar, miRanda and RNA22 retrieves 40 targets, of which 3 are true positives (red square).



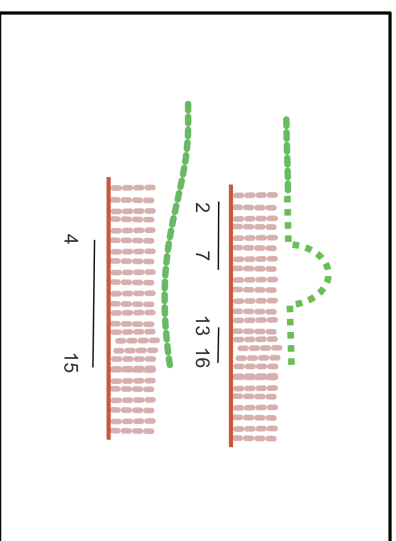
A



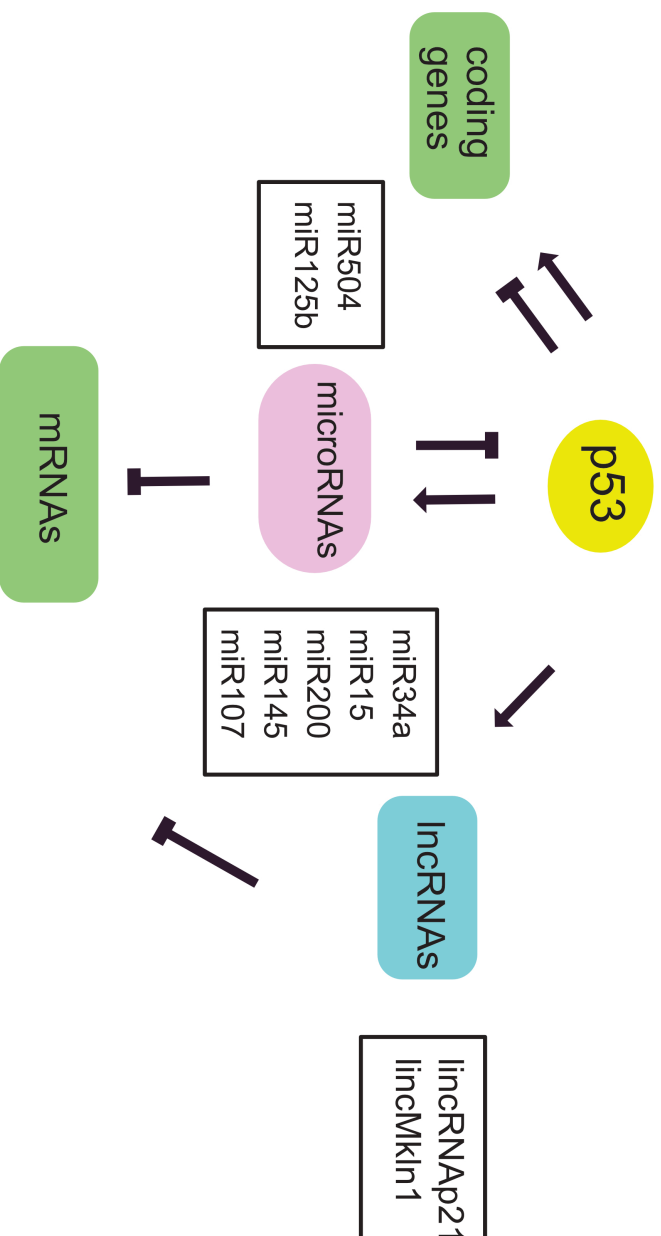
B



C

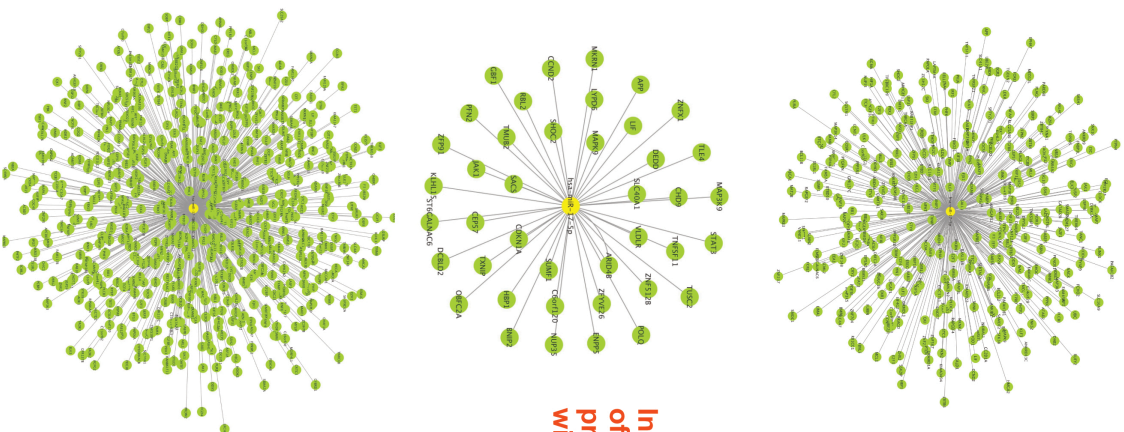


# A



# B





Intersection  
of four  
predictors

Intersection  
of four  
predictors  
with RNA22

Targets can  
prediction

