miRFFLDB - An integrated resource for human miRNA-Transcription Factor-Target Gene coregulatory networks and associated Feed Forward Loops.

Apoorv Gupta#, R Sugadev, YK Sharma, Bhuvnesh Kumar, Pankaj Khurana*#

Defence Institute of Physiology and Allied Scienes, Lucknow Road, Timapur, New Delhi

#Shared First authors
***Corresponding Author: pkhurana08@gmail.com**

Highlights

- microRNA (miRNA) and Transcriptional Factors (TFs) are two important regulatory biomolecules and a complex interplay between the two are known to control a Target Gene (TG).
- The tripartite motif of TF, miRNA regulating a common TG is known as Feed Forward Loop (FFL) or regulons.
- Identification of FFL motifs help to decode the interplay between miRNAs, TFs and TGs and get new mechanistic insights into complex molecular and cellular regulation processes in various multifactorial disorders.
- miRFFLDB is a comprehensive repository of miRNA:TF:TG co-regulatory network and associated FFLs of 2,596 human miRNAs.
- The server characterizes the miRNA–Target Interactions (MTIs) as miRNA:TG and miRNA:TF interactions. In addition, TF:TG and TF:miRNA interactions are added, which helps in exemplifying the network as miRNA:TF:TG co-regulatory directed network.
- The server further identifies miRNA-FFLs and TF-FFLs tripartite motifs from the co-regulatory networks using in-house scripts based on graph theory principles. The server also provides an interactive visualization of coregulatory networks and their associated FFLs.
- FFL motif identification is a powerful metanalysis tool for predicting novel miRNA, TF, genes candidates for prognostic, therapeutic, diagnostic purposes.

Abstract:

miRNA and Transcrition Factors (TFs) are fundamental transcriptional and post-transcriptional regulators of the gene expression respectively, controlling important biological processes. Gene regulation is viewed as a complex process where regulatory elements and their targets form highly complex network interactions. In recent years, many high throughput studies revealed that these complex regulatory interactions are mediated by the complex interplay between miRNA and TF regulating a common Target Gene (TG) in conjunction. Moreover, miRNAs

and TFs also known to regulate each other. This complex coregulatory mechanism may be represented in the form of miRNA:TF:TG coregulatory network, which can be used to identify several small recurring subgraphs known as motifs. One of these motifs, called the Feed-Forward Loops (FFLs) is a three-node pattern which is composed of a miRNA and a TF, one of which regulates the other and both jointly regulating a TG. FFLs have been proven useful in elucidating the complex interplay of gene regulation during many physiological and pathological conditions. Hence we have developed miRFFLDB, a comprehensive integrated resource for identification and visualization of human miRNA:TF:TG coregulatory directed networks and their associated FFLs. In house scripts based on the graph theory principles have been used to identify FFL motifs i.e. miRNA-FFL and TF-FFL. miRFFLDB also provides interactive visualization of these coregulatory networks and their associated FFLs. miRFFLDB can be used as a comprehensive ready-reference for human miRNA:TF:TG coregulatory networks and associated FFLs for decrypting complex cellular interactions of these regulatory biomolecules at the molecular level. miRFFLDB is available online at http://www.mirffldb.in.

Keywords: Coregulatory networks, Feed Forward Loops, miRNA-target interactions, Regulatory Networks

1. Introduction

Gene expression regulation is a complex process involving various regulatory biomolecules across numerous levels[1]. Transcription factors (TFs) and microRNAs (miRNAs) are the most common regulatory biomolecules that fine-tune gene expression by regulating at transcriptional level and post-transcriptional level respectively [2]. They are known to regulate gene expression independently; but now increasing evidence show that miRNAs and TFs also work synergistically in the form of complex networks to regulate the gene expression, which further modulates cellular and molecular processes e.g. cell signaling, RNA splicing, siRNA control mechanisms, and chromatin modifications etc[3]. Their combinatorial role in disease initiation, processes, and recurrence is also been studied [4-6]. These complex regulatory interactions can be best viewed using miRNA:TF:TG co-regulatory networks. These co-regulatory networks are responsible for the impressive degree of complexity in gene-regulation in higher eukaryots [7]. During the past ten years, studies on recurrent circuits, also known as network motifs, in gene co-regulatory network have significantly contributed to uncover this complexity, and in better understanding the regulatory roles of miRNAs and TFs[7]. Specifically, miRNAs and TFs have been shown to regulate shared TGs in Feed Forward Loops (FFLs) [8]. FFLs are tripartite motifs in which miRNA regulates TF and/or regulated by it, and both together regulate the TG in the network [9]. Based on type of interaction between miRNA and TF, the FFLs can be broadly classified into 2 types: TF-FFL and miRNA-FFL[10]. In TF-FFL, the TF regulates the miRNA and the TG while the miRNA represses the TG. In miRNA-FFL, the miRNA represses both the TF and TG expression while the TF also regulates TG expression (Supplementary Figure S1). At the network level, both miRNA-FFLs and TF-FFLs are important genetic overrepresented motif patterns that occur more often than by chance in

biological networks. Hence they serve as basic building blocks of a complex regulatory system [11, 12]. These FFLs are extensively studied to discover underlying genotypic and phenotypic relationship in complex disease conditions. e.g. TF: c-Myc; miR: miR-17-5p; target: E2F1 and c-Myc; miR: miR-20a; target: E2F1 were identified as a module tightly regulating proliferative signal of proto-oncogene c- Myc and promoting cell cycle progression[13, 14]. C-Myc was termed as mater regulator of cellular proliferation in most of the human cancer types [15]. Hence c-Myc regulated FFL motifs in human cancer is extensively studied and a separate catalogue of C-myc regulating FFL motifs is consolidated [16]. Apart from the c-Myc other important FFLs regulating major molecular functions during cancer include TF: P53; miR: miR-34; target: MET, TF: E2F; miR: miR-106b/93/25; target: CDK, TF: AP-1; miR: miR-101; target: MMP9. These are known to regulate the extent of cancer invasion, anti-proliferative activity and cell migration respectively in different cancer studies [7]. A database of Cancer-Specific MicroRNA And Transcription Factor Co-Regulatory Networks (CMTCN) was recently developed to identify miRNA:TF:TG coregulatory network and associated FFLs in various human cancer types[17]. miRNA:TF:TG coregulatory networks and FFLs have also been found to be very useful in identifying mechanisms involved in other multifactorial disease for disease-initiation, -progression and -recurrence e.g. TF: MAX; miR: miR-320; target: BMP6 and TF: IRF1; miR: miR-103a-3p; target: ACVR2B was overrepresented in myocardial infarction recurrence condition and both targets BMP6, ACVR2B have been identified as biomarker for myocardial infarction recurrence[18]. Similarly TF: TP53; miR: mir-34b; target: CAV1, was demonstrated to be only dysregulated module in cardiac hypertrophy patients [19]. 9 FFLs, four miRNA-FFLs and five co-regulatory FFLs were found to be regulating cell-cycle regulation during a hypoxia stress[20]. There are many resources that catalogue the miRNA:TF:TG coregulatory network and associated FFLs e.g. RegNetwork is a regulatory network repository for gene regulatory networks (GRNs) provides different types of regulatory interactions i.e. ‘TF:TG’,‘TF:miRNA’ , ‘miRNA:TG’ ‘miRNA:TF’[21]. CircuitsDB is a database that provides mixed miRNA/ TF-FFL circuits of 193 mature miRNAs and 130 pre-miRNAs of human and mouse[22]. TFmiR and MAGIA2 webservers use miRNAs/mRNAs expression profile to identify significantly enriched TF and miRNAs interactions for construction of miRNA:TF:TG coregulatory network[23, 24]. Similar to TFmiR and MAGIA2, dChip- Gene and miRNA Network-based Integration (GemiNI) is a webserver that uses predicted targets of miRNA and TF and miRNA/mRNA expression profiles for the identification of FFLs[25]. Though there are several databases, repositories, webservers and tools available for FFL identification but these resources have certain limitations. Some database are not comprehensive, whereas others require an expression profile of multiple miRNAs or TFs for identification of FFL. In such cases, some FFLs which are biologically relevant may be lost, during enrichment processes. Hence there is a need for a comprehensive global resource of human miRNA:TF:TG coregulatory networks and associated FFLs.

Here, we present miRFFLDB, a comprehensive user-friendly database of human miRNA: TF: TG coregulatory directed networks and their associated FFLs. The current version of miRFFLDB contains miRNA:TF:TG coregulatory networks of 2,596 human miRNAs. To construct the miRNA:TF:TG coregulatory networks, miRFFLDB has an exhaustive list of

5,02,652 experimentally-validated miRNA-Target Interactions (MTIs). The database stores experimentally validated MTIs with details like target name, target Entrez ID, experimental validation type, and support experimental validation (Functional/Functionally Weak) and hyperlink to corresponding publication. miRFFLDB characterizes these MTIs as miRNA:TG or miRNA:TF interactions. Further, TF:TG and TF:miRNA interactions are also added to construct miRNA:TF:TG directed coregulatory network. miRFFLDB offers interactive visualization of this network in which the nodes are color-coded for their easy characterization (TG as yellow, TF as green and miRNA as red). miRFFLDB uses in-house scripts based on graph theory principle to identify tripartite miRNA-FFL and TF-FFL motifs from the network. It also provides an interactive visualization of each tripartite graph. The MTIs, miRNA-FFLs and TF-FFLs, can be downloaded in excel and PDF format. The network file can also be downloaded in .SIF format, which can be used with popular visualization software like Cytoscape, Gephi, BINA, etc. Hence miRFFLDB can be considered as a comprehensive non-redundant catalogue of human miRNA: TF: TG coregulatory networks and its associated FFLs. We hope that miRFFLDB will catalyze research in understanding the complex crosstalk between miRNA, TF and TG. These motifs may offer mechanistic insights in the complex regulatory mechanisms. Combining it with experimental validation, these FFLs can identify novel players that can be used as diagnostic or prognostic markers and therapeutic targets for multifactorial complex disorders and pathophysiological conditions.

2. MATERIAL AND METHODS

2.1 Data Collection and Processing

To construct miRFFLDB database, a non-redundant list of human miRNA was retrieved from miRBase (version 22). To ensure uniformity in the nomenclature, the precursor miRNAs were mapped to their mature form using miRBase[26] and miRDB[27] repository. Also, the dead/obsolete entries were removed. Experimentally-validated targets of each human miRNA were fetched from public repositories i.e. mirTarBase[28] and miRecords[29] using multiMiR[30] R package. For each interaction, additional information like target name, target Entrez ID, support experiment details, support experimental validation (functional strong/functionally Weak) and the reference PMID were also recorded. Supplementary Figure S2 schematically illustrates the general workflow for the collection of resource data for miRFFLDB.

2.2 Data Enrichment

The data was further enriched by labelling MTIs as miRNA:TG and miRNA:TF interactions. For this, a comprehensive list of 3282 human TFs and 958 transcription co-factors (TcoFs) was collated. A list of 3462, 3240, 1769, 1758, 3360, 1538 TFs were fetched from TFcheckpoint[31], DBD[32], ORFeome[33], TcoF-DB V2[34], TFCat[35], TFClass[36],

TRANSFAC[37] databases respectively. A non-redundant list of 3282 unique TFs were obtained after removing duplicate entries. Further 958 TcoFs were also identified from TcoF-DB V2 database [34]. Thus a total 3738 unique TFs and TcoFs were obtained. The TF:TG interactions were fetched from TRANSFAC, OregAnno 3.0 [38] and TRRUST V2[39] databases. TF:miRNA interactions were fetched from TransmiR[40] and PuTmiR[41] databases. Finally, all the dataset files were loaded as JavaScript Object Notation (JSON) files format. These interactions were compiled to construct directed tripartite miRNA:TF:TG coregulatory networks for each miRNA. Vis.js was used to display these files as miRNA: TF:TG coregulatory networks[42].

2.3 FFL Motif Identification.

Each vertex in the network is labelled as V_m, V_{TF} or V_{TG}. Where, V_m refers to a miRNA node, V_{TF} refers to a TF node and V_{TG} refers to a TG node. The edges are annotated as E_{mt}, E_{mg}, E_{tg}, E_{tm} where E_{mt} refers to edge from V_m to V_{TF} , E_{mg} refers to edge from V_m to V_{TG} , E_{tg} refers to edge from V_{TF} to V_{TG} and E_{tm} refers to edge from V_{TF} to V_m.

For each V_m, its edges E_{mg} and E_{mt} were identified from the miRNA:TG interactions and miRNA:TF interactions datasets respectively and the corresponding V_{TF} and V_{TG} were listed. Thereafter, a corresponding edge E_{tg} that links V_{TF} to V_{TG} is searched in the TF:TG interaction dataset. If found, a tripartite graph containing three vertices (V_m , V_{TF}, V_{TG}) and edges(E_{mt}, E_{mg}, E_{tg}) was labeled as miRNA-FFL motif graph.

Similarly in order to identify TF-FFL motif, for each V_m, its edges E_{tm} and E_{mg} were identified from TF:miRNA interaction and miRNA:TG interaction datasets respectively and the corresponding V_{TF} and V_{TG} were listed. The program then scans the TF:TG interaction dataset to look for a Etg that connects the V_{TF} and V_{TG}. If found, a tripartite graph of three vertices (V_m , V_{TF}, V_{TG}) and edges(E_{tm}, E_{mg}, E_{tg}) was labeled as TF-FFL motif graph. This methodology was used to identify all miRNA-FFLs and TF-FFLs for each human miRNA.

2.4 Database development

All JSON dataset files were finally stored in MongoDB database[43]. MongoDB is an open-source document-oriented (NoSQL) database [43, 44]. The variables and query functions of the Application Program Interface (API) in the miRFFLDB web application are defined in the python. The database uses Asynchronous JavaScript and XML (AJAX) technique for API calls[45]. AJAX is a web development technique that is used for creating interactive web applications. It utilizes XHTML for content along with the document object model and JavaScript for dynamic content display.

3. RESULTS

miRFFLDB is a comprehensive resource for human miRNA:TF:TG coregulatory directed networks and their associated FFLs. It currently contains a non-redundant list of 2,596 human

miRNAs, 11,864 experimentally validated TGs and 3,738 TFs and TcoFs (Figure 1a). The database contains ~3,80,000 experimentally validated MTIs mined from mirTarBase[28], miRecords[29] and ~7200 corresponding reference PMIDs. These ~3,80,000 miRNA-target interactions are categorized as ~1,14,000 miRNA:TF interactions and ~2,66,000 miRNA:TG interactions. Additionally, ~49,000 TF:TG interactions were fetched from OregAnno[38] and TRRUST V2[39] databases and 1500 TF:miRNA interactions were fetched from TransmiR[40] and PuTmiR[41] (Figure 1b). These interactions were used to construct miRNA:TF:TG coregulatory directed network for each human miRNA.

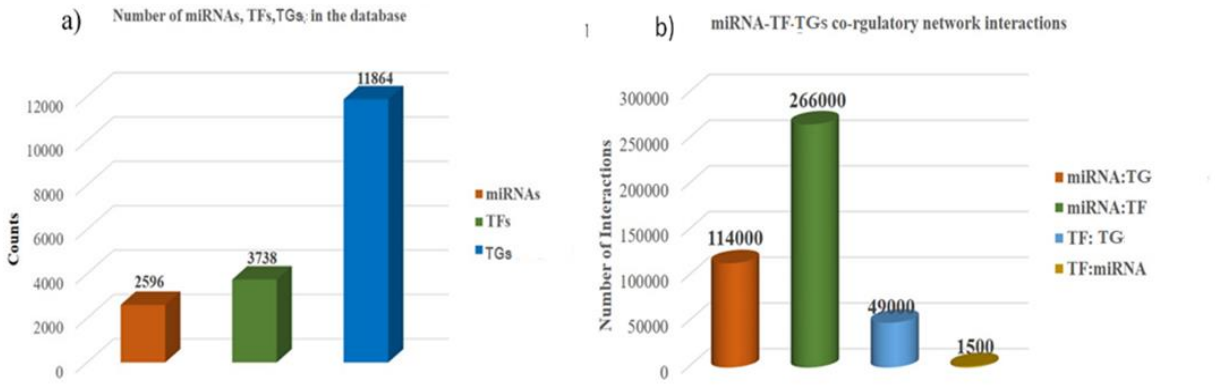


Figure 1 a) Graphical representation of the number of miRNAs, TFs and TGs in the miRFFLDB. b) Number of miRNA:TG, miRNA:TF, TF:TG, TF:miRNA interactions in the database.

3.1 Browse Interface

miRFFLDB browse option facilitates users to identify miRNA:TF: TG coregulatory networks and FFLs by three routes i.e. 'search by miRNA', 'search by TF(s)' or 'search by gene(s)' (Figure 2a). The first option of 'search by miRNA' allows the user to select miRNA from a pull down menu. On clicking the search button, the page connects to the detailed information page of the respective miRNA, which is discussed in detail in next section 3.2. The database also facilitates the user to search the database by selecting TF(s) and TG(s) from a pull down menu. On selecting a TF/TG, the program returns a list of FFLs containing the queried TF(s) or gene(s) in a tabular format (Figure 2b, 2c). The table can be downloaded in Excel /PDF format. The miRNAs in the table are further hyperlinked to the corresponding miRNA information page.

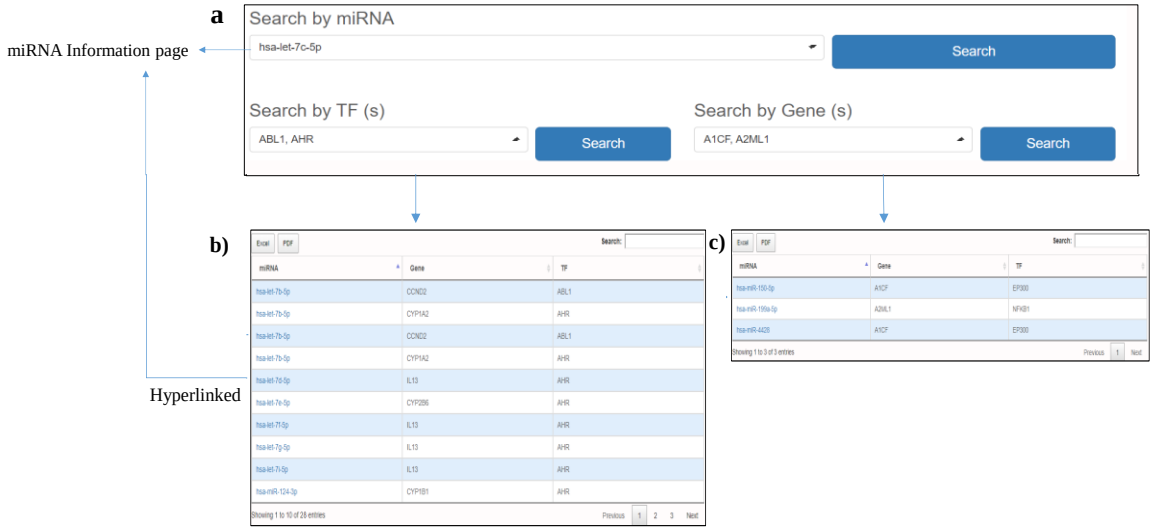


Figure 2: The figure shows the screenshot of the 'Browse' option. This option allow the users to browse the database using three different options: 'Search by miRNA', 'Search by TF(s)' and 'Search by Gene(s)'.a) On searching with miRNA, the page connects to the detailed information page of the respective miRNA. b) On searching by TF, a table representing the FFLs containing the queried TF(s) is displayed c) On searching by gene, a table representing the FFLs containing the queried gene(s) is generated.

3.2 miRNA Information Page

The miRNA information page for each human miRNA can be divided into two sections.
(i) The first section displays interactive visualization of miRNA:TF:TG coregulatory network and a tabular representation of miRNA-FFLs and TF-FFLs that are identified from the network.
(ii) The second section enlists the experimentally-validated MTIs

(i) miRNA:TF:TG co-regulatory network and FFLs

The top of the miRNA information page highlights the name of the mature miRNA which is hyperlinked to miRBase database[26]. miRBase database is a comprehensive resource of miRNA sequence data, annotation and predicted gene targets. Hence hyperlinking each human miRNA in the database serve as a ready-reference to get additional details about these miRNAs that includes mature and precursor miRNA sequence, miRNA stem loop structure, chromosome location, source organism etc.
The miRNA:TF:TG network is displayed as interactive network in which the nodes are color coded for easy characterization (miRNA in red and its interacting TFs and TGs in green and yellow respectively) (Figure 3a). The network is also equipped with zoom-in, zoom-out, and node-translation options. The network file can be downloaded in (.SIF) format and can be readily used by many open-source software's such as Cytoscape[46], Gephi[47], BINA[48], etc for further analysis.

The miRNA-FFL and TF-FFL are identified from the coregulatory network using in-house scripts and presented as separate tables. Each row of the table defines a FFL with a miRNA, TF and TG. The table can be downloaded as Excel/PDF. User can also visualize tripartite graph by clicking on the miRNA, which will appear in a pop-up window (Figure 3a). Both the miRNA-FFL and TF- FFL tables are provided with an integrated search option. This allows the user to search the lengthy tables by TF/gene symbol. Each TF and TG symbol in the FFL table are further hyperlinked to the GeneCards[49] database. This would serve as a ready-reference for the user to get additional details about the gene that includes aliases of genes, promoter and enhancer location of gene, protein coded by gene, functional characterization, cellular localization, pathway enrichment, gene-gene interaction network, drug-gene relationship, tissue specific gene expression profile, orthologs, paralogs, transcript variant etc.

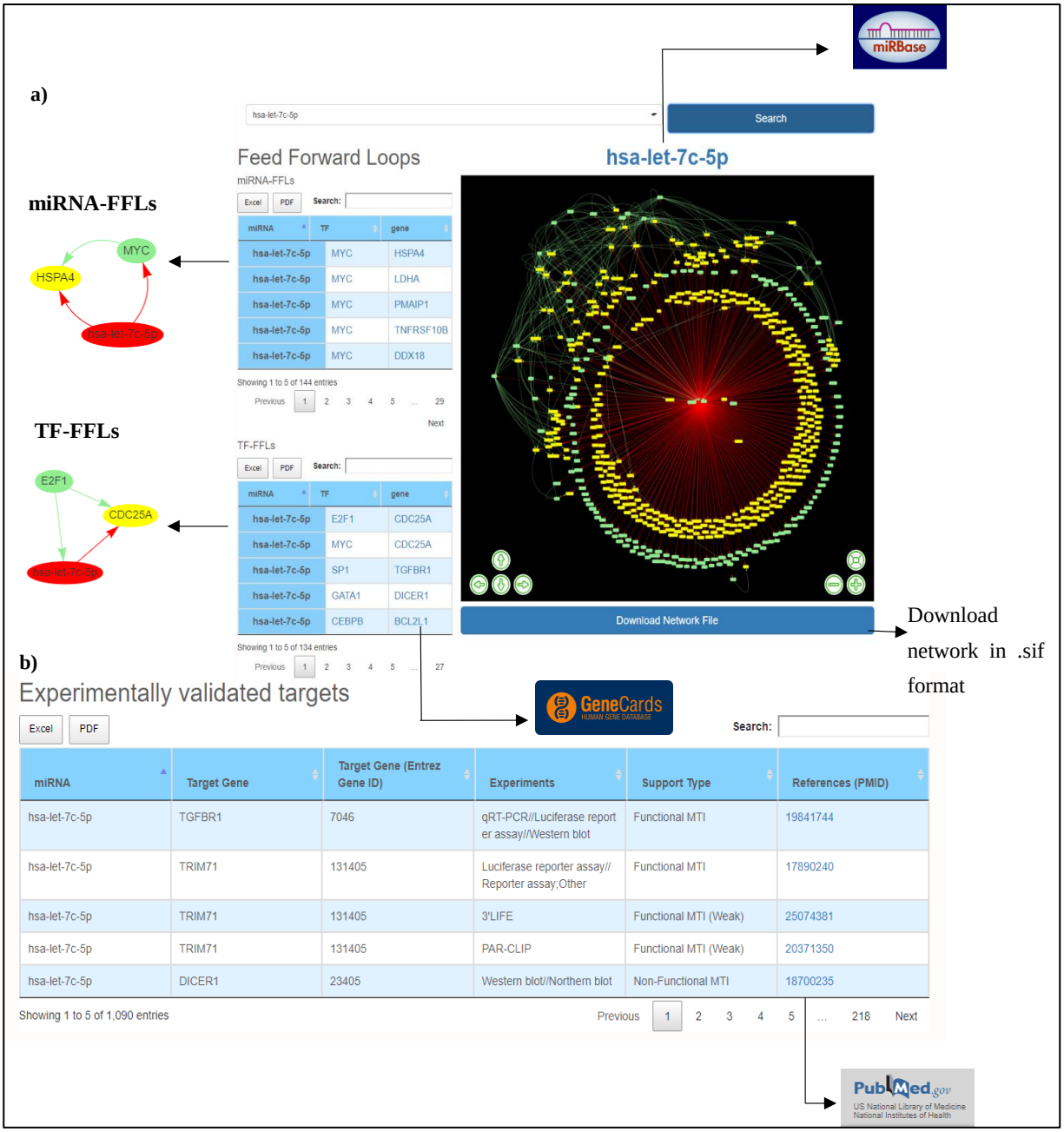


Figure 3: The web image of miRNA information page of hsa-let-7c-5p. a) The first section

shows the coregulatory network and the associated miRNA-FFLs and TF-FFLs. Clicking on the miRNA, a pop-up window displays the corresponding FFL. Searched miRNA name is hyperlinked to the miRBase database, which provides additional details of miRNA. Also, gene names are hyperlinked to GeneCards database, which provides additional details about the gene. b) The experimentally validated targets of the miRNA are listed along with the details like Target gene, its entrez ID, experiment, support type and the respective reference paper. The PMID is hyperlinked to the PubMed database for ready-access to the publication.

(ii) Experimentally validated MTIs

miRFFLDB also provide experimentally validated MTIs along with its associated information that includes target symbol, target Entrez ID, experimental validation type, and support experimental validation (Non-Functional/Functional/Functionally Weak) and the reference paper. The information is provided in a tabular format that can be downloaded in the form of excel/PDF file. The table is also provided with an integrated search option, which allows to search the lengthy tables. The user may search with target symbol, entrez ID, experiment type, support type or PMID (Figure 3b). The PMID is hyperlinked to the PubMed database for ready access to the publication.

3.3 Case Study

As a proof of concept, miRFFLDB was used to analyze the coregulatory network and FFLs of MYC. MYC is a proto-oncogene that has been implicated as the hallmark initiator of many human carcinomas [50]. When activated pathologically, it governs hallmark cancer responses such as uncontrolled cell proliferation, tumor synthesis, metabolic pathways integral to cell growth etc [50, 51]. Hence, studying the FFLs regulated by MYC can provide insights about the complex mechanism that control downstream signaling of MYC during cancer initiation. This would also help to identify other potential candidates for diagnostics, prognostics and therapeutics for human carcinomas. To study MYC-FFLs, MYC containing FFLs were retrieved from miRFFLDB. The miRNAs from these FFLs were retrieved and frequency distribution showed four outlier i.e. hsa-miR-16-5p, hsa-miR-335-5p, hsa-miR-92a-3p, hsa-let-7b-5p that have most number of occurrences in these FFLs (Supplementary Figure S3).
To validate these 4 miRNAs, textmining from biomedical literature was also performed to identify miRNAs that relate to MYC expression in cancer cells. Textmining was performed via keywords ‘MYC’ and ‘cancer’ using miRTex textmining tool [52]. Textming indicated that miR-92a as the most extensively studied miRNA that was related to MYC expression in cancer cells. miR-16 also appeared in top miRNAs related to MYC expression in cancer cells. However, miR-335 and let-7b have not been reported extensively in textmining analysis (Supplementary Figure S4). miRNAs hsa-miR-92a-3p and hsa-miR-16-5p interaction with MYC has been studied extensively but their coregulatory effect in cancer cells is still not reported. To infer the coregulatory effect of MYC,miR-92a-3p and hsa-miR-16-5p, pathway enrichment of their target genes in earlier identified MYC-FFLs was performed. The results show that ribosomal proteins are the most enriched set in both target sets (Supplementary Table

S1). The ribosomal proteins i.e. MRPL20, RPS6, RPL10, RPL12, RPL9, MRPL3, RPL5, RPS2, RPLP0, RPS12, RPL13, RPS25, RPL21, RPSA, RPS24, RPS5 are obtained in the most enriched set ((Supplementary Table S2). Thus analysis from mirFFLDB proposed that MYC, miR-92a-3p and hsa-miR-16-5p may interact and coregulate the genes associated with ribosomal pathway and it could be an interesting biological process to characterize and identify certain diagnostic/prognostic/therapeutic markers for cancers. Literature survey indeed identified that ribosomal proteins have been reported for diagnostics and prognostics implications in human carcinomas [53-55].

Hence as proven successfully for cancer, mirFFLDB could be used to identify specific lead molecular pathways that could be targeted for in-depth study and for identifying biomolecular markers in different pathological/pathophysiological conditions. A thorough analysis of the FFLs from miRFFLDB can also help in identification and understanding of complex molecular-interactions and also shed light on mechanistic details during complex disorders. Detailed examination of these coregulated molecular interactions would aid in the search for novel biomarkers for effective therapeutic, diagnostic and prognostic purposes.

4. DISCUSSION AND CONCLUSION

A decade earlier, Yitzhak Pilpel et al studied the global miRNA:TF:TG mammalian co-regulatory network and uncovered two network architecture of regulatory network [56]. The first network architecture contains interactions between pairs of miRNAs with shared TGs called MTIs. This architecture follows scale-free topology with some target genes knows as 'hubs' that are massively regulated by more number of miRNAs than most others. The second network architecture consists of recurring circuits of miR:TF pairs which combinatorially regulate large sets of common TGs. The network analysis of this global miRNA:TF mammalian co-regulatory network revealed several recurring motifs of miRNA and TF regulating a hub gene. These motifs help in fine-tuning of many complex molecular and cellular processes. Since then, studies on recurrent circuits in regulatory networks, also known as network motifs, have substantially contributed to addressing this complexity and in particular to better understand miRNAs and TFs exert their regulatory roles[5, 7, 9]. The most studied recurring motifs in miRNA:TF:TG coregulatory networks are FFLs. FFLs have been proven to be the best network analysis tool to study the combinatorial TG regulation by miRNA and TF in many complex pathologies. Hence FFLs in coregulatory networks are crucial in providing new insights into the logic and evolution of a new regulatory layer of the complex eukaryotic genome.

miRFFLDB offers a comprehensive and interactive collection of human miRNA:TF:TG coregulatory networks and associated FFLs. miRFFLDB pools miRNA:TG, miRNA:TF, TF:TG and TF:miRNA interactions for constructing complex miRNA:TF:TG coregulatory networks of 2,596 human miRNAs. These networks were used to identify the miRNA-FFL and TF-FFL motifs. Additionally, miRFFLDB also provides experimentally validated miRNA-Target Interactions (MTIs) resources for these human miRNAs.

A thorough literature survey suggested that though there are several server platforms, databases, and tools available for identification of gene regulatory networks (GRNs) and associated FFLs.

CircuitsDB is a database for mixed miRNA/ TF FFL circuits in human and mouse contains information of only 193 mature miRNAs and 130 pre-miRNAs FFLs. Whereas, currently more than ~2600 human mature miRNAs (miRBase release 22) and ~3,80,000 experimentally validated MTIs have already been identified[22, 28, 57]. RegNetwork is an integrated resource of transcriptional and post-transcriptional regulatory networks and provides a collection of regulatory interactions among TFs, miRNAs and target genes. [21]. The database, however, does not provide FFLs. Also during the thorough examination of the database, many interactions were found missing. A random example of hsa-miR-141-5p has been used to compare the regulatory interactions present in our database and RegNetwork (Supplementary Table S3). Server tools such as TFmiR, MAGIA2 require miRNAs/mRNAs expression profile as input to identify regulatory interactions based on the input deregulated genes and deregulated miRNAs [23, 24]. For identification of regulatory interactions, the program first identifies miRNAs, whose target genes and TFs, are significantly enriched within the input deregulated genes by using the hypergeometric distribution function. Hence only interactions among these miRNAs and significantly enriched TF, TGs are used to construct miRNA:TF:TG coregulatory network [23, 24]. This methodology reduces a significant number of interactions from the network. Subsequently, it would also lead to the loss of important FFLs identified from the network. Another similar webserver dChip- GemiNI identifies FFLs by using combine score of miRNA and TF target prediction and miRNA/mRNA expression level called Network Motif Score (NMS) [25]. This webserver only consider predictive score for identifying an edge between TF:TG, miRNA:TG. This further leads to the false FFLs discoveries as their edges are not experimentally-validated. Databases like 'CMTCN' [17] and 'curated database of miRNA mediated feed-forward loops including MYC as master regulator'[16] provide only the cancer-specific miRNA:TF:TG coregulatory networks and associated FFLs. Though these databases are recently published but their global coverage is limited to only cancer. However, FFLs have also been identified in many other multifactorial diseases and pathophysiological conditions like myocardial infarct, schizophrenia, and hypoxia respectively. So these databases, webservers, and tools cannot be considered universal repository for miRNA:TF:TG coregulatory networks and FFLs.

Herein, we developed miRFFLDB which is a comprehensive coregulatory network resource that integrates miRNA:TG interactions, miRNA:TF interactions, TF:TG interactions and TF:miRNA interactions to build co-regulatory networks. These complex networks provide a global representation of complex regulatory and target interactions of human miRNAs and TFs. It can advance our understanding of complex molecular mechanisms that are controlled by these regulatory molecules. miRFFLDB identifies FFLs that will help to decode the interplay between miRNAs, TFs and TGs to get new mechanistic-insights into specific molecular and cellular processes. Further analysis of these FFL motifs using network biology techniques could also help us in identifying potential disease-markers and therapeutic targets.

The resource can also be used for other network-based computational and integrated analysis. This would allow researchers to be more focused on analysis rather than going for manual curation of the interactions from different databases to construct these co-regulatory networks.

Also it is ready-reference of enriched FFLs for human miRNA pool. We believe that miRFFLDB will help researchers to design and also analyze miRNA/TF- based experiments.

Supplementary materials:

Figure S1: Schematic representation of a TF-FFL and miRNA-FFL.

Figure S2: Overview of data collection and annotation in miRFFLDB database

Figure S3: Shows quartile distribution of the occurrences of miRNAs in MYC -FFLs that are identified from miRFFLDB.

Figure S4: Frequency chart of textmining results of 'miRNAs interacting with MYC during cancer'.

Table S1: Shows result of KEGG pathway enrichment of MYC-miR-92a-3p and MYC-hsa-miR-16-5p TF-FFLs regulated gene targets.

Table S2: The table shows 16 MYC-hsa-miR-16-5p and 14 MYC/MYCN- hsa-miR-92a-3p TF-FFLs co-regulating ribosomal proteins targets obtained from miRFFLDB database.

Table S3: Tabular representation of comparison in regulatory interactions of the hsa-miR-141-5p present in RegNetwork database with miRFFLDB.

Author Contributions:

Pankaj Khurana: Conceptualization, Methodology, Software, validation, Reviewing and editing Apoorv Gupta.: Data curation, data integration, testing, Writing- Original draft preparation. R Sugadev, YK Sharma, B Kumar: Supervision.

Abbreviations

TF	Transcription factor
TG	Target Gene
FFL	Feed Forward Loop
MTI	miRNA-Target Interaction
TcoF	Transcription Cofactor

References

[1] Macneil LT, Walhout AJ. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome research. 2011;21:645-57.

[2] Martinez NJ, Walhout AJ. The interplay between transcription factors and microRNAs in genome-scale regulatory networks. BioEssays : news and reviews in molecular, cellular and developmental biology. 2009;31:435-45.

[3] Feng Z, Zhang W, Xu J, Gauron C, Ducos B, Vriz S, et al. Optical control and study of biological processes at the single-cell level in a live organism. Reports on progress in physics Physical Society. 2013;76:072601.

[4] Jiang W, Mitra R, Lin CC, Wang Q, Cheng F, Zhao Z. Systematic dissection of dysregulated transcription factor-miRNA feed-forward loops across tumor types. Briefings in bioinformatics. 2016;17:996-1008.

[5] Xu Y, Yue W, Yao Shugart Y, Li S, Cai L, Li Q, et al. Exploring Transcription Factors-microRNAs Co-regulation Networks in Schizophrenia. Schizophrenia bulletin. 2016;42:1037-45.

[6] Lin Y, Sibanda VL, Zhang HM, Hu H, Liu H, Guo AY. MiRNA and TF co-regulatory network analysis for the pathology and recurrence of myocardial infarction. Scientific reports. 2015;5:9653.

[7] Cora D, Re A, Caselle M, Bussolino F. MicroRNA-mediated regulatory circuits: outlook and perspectives. Physical biology. 2017;14:045001.

[8] Yan Y, Shi Y, Wang C, Guo P, Wang J, Zhang CY, et al. Influence of a high-altitude hypoxic environment on human plasma microRNA profiles. Scientific reports. 2015;5:15156.

[9] Zhao Q, Liu H, Yao C, Shuai J, Sun X. Effect of Dynamic Interaction between microRNA and Transcription Factor on Gene Expression. BioMed research international. 2016;2016:2676282.

[10] Li R, Jiang S, Li W, Hong H, Zhao C, Huang X, et al. Exploration of prognosis-related microRNA and transcription factor co-regulatory networks across cancer types. RNA biology. 2019;16:1010-21.

[11] Anastasiadou E, Jacob LS, Slack FJ. Non-coding RNA networks in cancer. Nature reviews Cancer. 2018;18:5-18.

[12] Guo Y, Alexander K, Clark AG, Grimson A, Yu H. Integrated network analysis reveals distinct regulatory roles of transcription factors and microRNAs. Rna. 2016;22:1663-72.

[13] Pickering MT, Stadler BM, Kowalik TF. miR-17 and miR-20a temper an E2F1-induced G1 checkpoint to regulate cell cycle progression. Oncogene. 2009;28:140-5.

[14] Wu Q, Qin H, Zhao Q, He XX. Emerging role of transcription factor-microRNA-target gene feed-forward loops in cancer. Biomedical reports. 2015;3:611-6.

[15] Miller DM, Thomas SD, Islam A, Muench D, Sedoris K. c-Myc and cancer metabolism. Clinical cancer research : an official journal of the American Association for Cancer Research. 2012;18:5546-53.

[16] El Baroudi M, Cora D, Bosia C, Osella M, Caselle M. A curated database of miRNA mediated feed-forward loops involving MYC as master regulator. PloS one. 2011;6:e14742.

[17] Li R, Chen H, Jiang S, Li W, Li H, Zhang Z, et al. CMTCN: a web tool for investigating cancer-specific microRNA and transcription factor co-regulatory networks. PeerJ. 2018;6:e5951.

[18] Zhang G, Shi H, Wang L, Zhou M, Wang Z, Liu X, et al. MicroRNA and transcription factor mediated regulatory network analysis reveals critical regulators and regulatory modules in myocardial infarction. PloS one. 2015;10:e0135339.

[19] Qu W, Shi S, Sun L, Zhang F, Zhang S, Mu S, et al. Construction of a microRNAassociated feedforward loop network that identifies regulators of cardiac hypertrophy and acute myocardial infarction. International journal of molecular medicine. 2018;42:2062-70.

[20] Gupta A, Ragumani S, Sharma YK, Ahmad Y, Khurana P. Analysis of Hypoxiamir-Gene Regulatory Network Identifies Critical MiRNAs Influencing Cell-Cycle Regulation Under Hypoxic Conditions. MicroRNA. 2019;8:223-36.

[21] Liu ZP, Wu C, Miao H, Wu H. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. Database : the journal of biological databases and curation. 2015;2015.

[22] Friard O, Re A, Taverna D, De Bortoli M, Cora D. CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. BMC bioinformatics. 2010;11:435.

[23] Hamed M, Spaniol C, Nazarieh M, Helms V. TFmiR: a web server for constructing and analyzing disease-specific transcription factor and miRNA co-regulatory networks. Nucleic acids research. 2015;43:W283-8.

[24] Bisognin A, Sales G, Coppe A, Bortoluzzi S, Romualdi C. MAGIA(2): from miRNA and genes expression data integrative analysis to microRNA-transcription factor mixed regulatory circuits (2012 update). Nucleic acids research. 2012;40:W13-21.

[25] Yan Z, Shah PK, Amin SB, Samur MK, Huang N, Wang X, et al. Integrative analysis of gene and miRNA expression profiles with transcription factor-miRNA feed-forward loops identifies regulators in human cancers. Nucleic acids research. 2012;40:e135.

[26] Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic acids research. 2006;34:D140-4.

[27] Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic acids research. 2015;43:D146-52.

[28] Hsu SD, Tseng YT, Shrestha S, Lin YL, Khaleel A, Chou CH, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic acids research. 2014;42:D78-85.

[29] Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. Nucleic acids research. 2009;37:D105-10.

[30] Ru Y, Kechris KJ, Tabakoff B, Hoffman P, Radcliffe RA, Bowler R, et al. The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. Nucleic acids research. 2014;42:e133.

[31] Chawla K, Tripathi S, Thommesen L, Laegreid A, Kuiper M. TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. Bioinformatics. 2013;29:2519-20.

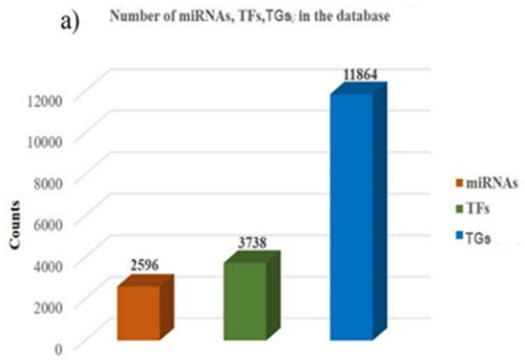
[32] Kummerfeld SK, Teichmann SA. DBD: a transcription factor prediction database. Nucleic acids research. 2006;34:D74-81.

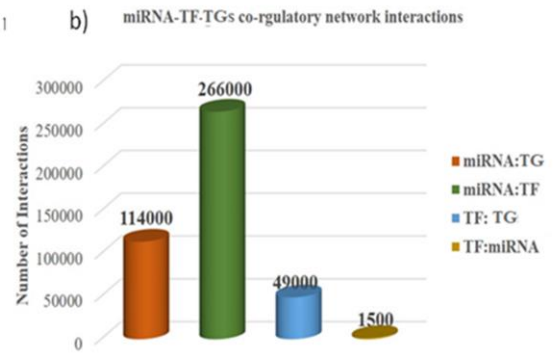
[33] Lamesch P, Li N, Milstein S, Fan C, Hao T, Szabo G, et al. hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. Genomics. 2007;89:307-15.

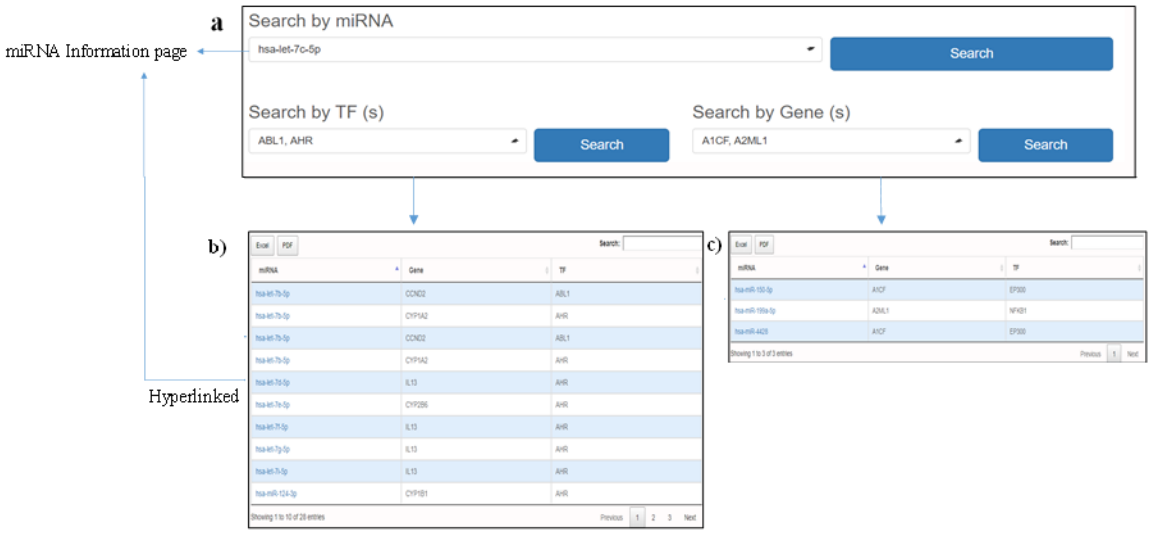
[34] Schmeier S, Alam T, Essack M, Bajic VB. TcoF-DB v2: update of the database of human and mouse transcription co-factors and transcription factor interactions. Nucleic acids research. 2017;45:D145-D50.

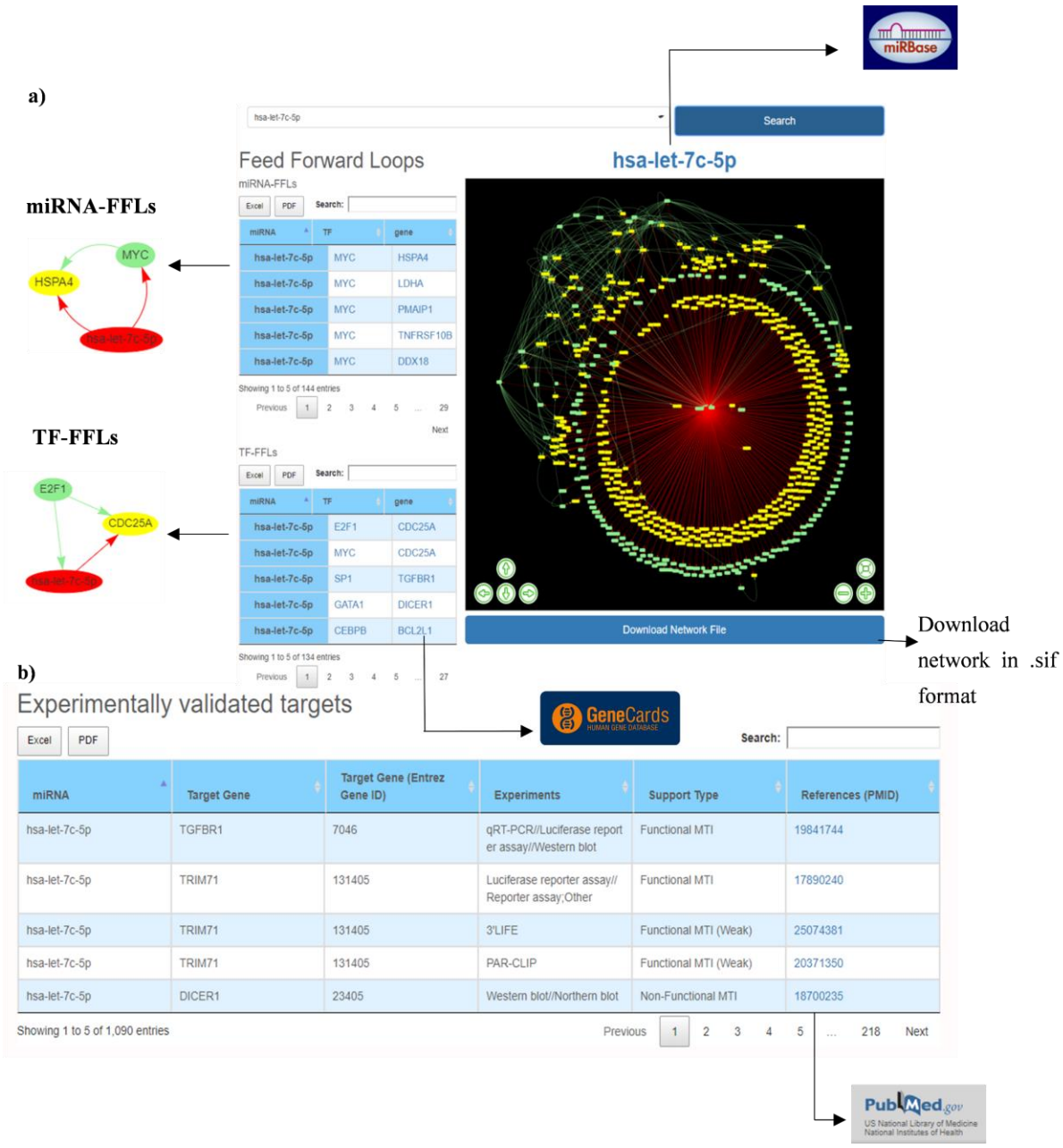
[35] Fulton DL, Sundararajan S, Badis G, Hughes TR, Wasserman WW, Roach JC, et al. TFCat: the curated catalog of mouse and human transcription factors. Genome biology. 2009;10:R29.
[36] Wingender E, Schoeps T, Haubrock M, Donitz J. TFClass: a classification of human transcription factors and their rodent orthologs. Nucleic acids research. 2015;43:D97-102.
[37] Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic acids research. 2003;31:374-8.
[38] Lesurf R, Cotto KC, Wang G, Griffith M, Kasaian K, Jones SJ, et al. ORegAnno 3.0: a community-driven resource for curated regulatory annotation. Nucleic acids research. 2016;44:D126-32.
[39] Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic acids research. 2018;46:D1284.
[40] Tong Z, Cui Q, Wang J, Zhou Y. TransmiR v2.0: an updated transcription factor-microRNA regulation database. Nucleic acids research. 2019;47:D253-D8.
[41] Bandyopadhyay S, Bhattacharyya M. PuTmiR: a database for extracting neighboring transcription factors of human microRNAs. BMC bioinformatics. 2010;11:190.
[42] Rosenthal SB, Len J, Webster M, Gary A, Birmingham A, Fisch KM. Interactive network visualization in Jupyter notebooks: visJS2jupyter. Bioinformatics. 2018;34:126-8.
[43] Schulz WL, Nelson BG, Felker DK, Durant TJS, Torres R. Evaluation of relational and NoSQL database architectures to manage genomic annotations. Journal of biomedical informatics. 2016;64:288-95.
[44] Liu J, Qu Z, Yang M, Sun J, Su S, Zhang L. Jointly integrating VCF-based variants and OWL-based biomedical ontologies in MongoDB. IEEE/ACM transactions on computational biology and bioinformatics. 2019.
[45] Berger SI, Iyengar R, Ma'ayan A. AVIS: AJAX viewer of interactive signaling networks. Bioinformatics. 2007;23:2803-5.
[46] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003;13:2498-504.
[47] Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PloS one. 2014;9:e98679.
[48] Gerasch A, Faber D, Kuntzer J, Niermann P, Kohlbacher O, Lenhof HP, et al. BiNA: a visual analytics tool for biological network data. PloS one. 2014;9:e87397.
[49] Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, et al. GeneCards Version 3: the human gene integrator. Database : the journal of biological databases and curation. 2010;2010:baq020.
[50] Gabay M, Li Y, Felsher DW. MYC activation is a hallmark of cancer initiation and maintenance. Cold Spring Harbor perspectives in medicine. 2014;4.
[51] Shortt J, Johnstone RW. Oncogenes in cell survival and cell death. Cold Spring Harbor perspectives in biology. 2012;4.

[52] Li G, Ross KE, Arighi CN, Peng Y, Wu CH, Vijay-Shanker K. miRTex: A Text Mining System for miRNA-Gene Relation Extraction. PLoS computational biology. 2015;11:e1004391.
[53] Dolezal JM, Dash AP, Prochownik EV. Diagnostic and prognostic implications of ribosomal protein transcript expression patterns in human cancers. BMC cancer. 2018;18:275.
[54] Arthurs C, Murtaza BN, Thomson C, Dickens K, Henrique R, Patel HRH, et al. Expression of ribosomal proteins in normal and cancerous human prostate tissue. PloS one. 2017;12:e0186047.
[55] Nakhoul H, Ke J, Zhou X, Liao W, Zeng SX, Lu H. Ribosomopathies: mechanisms of disease. Clinical medicine insights Blood disorders. 2014;7:7-16.
[56] Shalgi R, Lieber D, Oren M, Pilpel Y. Global and local architecture of the mammalian microRNA-transcription factor regulatory network. PLoS computational biology. 2007;3:e131.
[57] Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic acids research. 2019;47:D155-D62.


a)
Number of miRNAs, TFs,TGs in the database
Counts
12000
10000
8000
6000
4000
2000
0
2596
3738
11864
miRNAs
TFs
TGs


b)
miRNA-TF-TGs co-rgulatory network interactions
Number of Interactions
300000
250000
200000
150000
100000
50000
0
114000
266000
49000
1500
miRNA:TG
miRNA:TF
TF: TG
TF:miRNA


a
Search by miRNA
hsa-let-7c-5p
Search
Search by TF (s)
ABL1, AHR
Search
Search by Gene (s)
A1CF, A2ML1
Search
miRNA Information page
Hyperlinked
b)
c)


a)
miRNA-FFLs
MYC
HSPA4
hsa-let-7c-5p
TF-FFLs
E2F1
CDC25A
hsa-let-7c-5p
hsa-let-7c-5p
Search
Feed Forward Loops
hsa-let-7c-5p
miRNA-FFLs
Excel
PDF
Search:
miRNA TF gene
hsa-let-7c-5p MYC HSPA4
hsa-let-7c-5p MYC LDHA
hsa-let-7c-5p MYC PMAIP1
hsa-let-7c-5p MYC TNFRSF10B
hsa-let-7c-5p MYC DDX18
Showing 1 to 5 of 144 entries
Previous 1 2 3 4 5 ... 29 Next
TF-FFLs
Excel
PDF
Search:
miRNA TF gene
hsa-let-7c-5p E2F1 CDC25A
hsa-let-7c-5p MYC CDC25A
hsa-let-7c-5p SP1 TGFBR1
hsa-let-7c-5p GATA1 DICER1
hsa-let-7c-5p CEBPB BCL2L1
Showing 1 to 5 of 134 entries
Previous 1 2 3 4 5 ... 27
Download Network File
Download network in .sif format
b)
Experimentally validated targets
Excel
PDF
Search:
miRNA Target Gene Target Gene (Entrez Gene ID) Experiments Support Type References (PMID)
hsa-let-7c-5p TGFBR1 7046 qRT-PCR//Luciferase reporter assay//Western blot Functional MTI 19841744
hsa-let-7c-5p TRIM71 131405 Luciferase reporter assay//Reporter assay;Other Functional MTI 17890240
hsa-let-7c-5p TRIM71 131405 3'LIFE Functional MTI (Weak) 25074381
hsa-let-7c-5p TRIM71 131405 PAR-CLIP Functional MTI (Weak) 20371350
hsa-let-7c-5p DICER1 23405 Western blot//Northern blot Non-Functional MTI 18700235
Showing 1 to 5 of 1,090 entries
Previous 1 2 3 4 5 ... 218 Next