

CircuitsDB: a database of mixed microRNA / Transcription Factor feed-forward regulatory circuits in human and mouse

Original

CircuitsDB: a database of mixed microRNA / Transcription Factor feed-forward regulatory circuits in human and mouse / Olivier, Friard; Re, A; Daniela, Taverna; Michele De, Bortoli; Davide, Corà. - In: BMC BIOINFORMATICS. - ISSN 1471-2105. - ELETTRONICO. - 11:AGO(2010), p. 435. [10.1186/1471-2105-11-435]

Availability:

This version is available at: 11583/2970542 since: 2022-08-08T12:21:47Z

Publisher:

BMC Part of Springer Nature

Published

DOI:10.1186/1471-2105-11-435

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

DATABASE

Open Access

CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse

Olivier Friard¹, Angela Re², Daniela Taverna^{1,3,4}, Michele De Bortoli^{1,3}, Davide Corá^{1,5*}

Abstract

Background: Transcription Factors (TFs) and microRNAs (miRNAs) are key players for gene expression regulation in higher eukaryotes. In the last years, a large amount of bioinformatic studies were devoted to the elucidation of transcriptional and post-transcriptional (mostly miRNA-mediated) regulatory interactions, but little is known about the interplay between them.

Description: Here we describe a dynamic web-accessible database, *CircuitsDB*, supporting a genome-wide transcriptional and post-transcriptional regulatory network integration, for the human and mouse genomes, based on a bioinformatic sequence-analysis approach. In particular, *CircuitsDB* is currently focused on the study of mixed miRNA/TF Feed-Forward regulatory Loops (FFLs), i.e. elementary circuits in which a master TF regulates an miRNA and together with it a set of Joint Target protein-coding genes. The database was constructed using an ab-initio oligo analysis procedure for the identification of the transcriptional and post-transcriptional interactions. Several external sources of information were then pooled together to obtain the functional annotation of the proposed interactions. Results for human and mouse genomes are presented in an integrated web tool, that allows users to explore the circuits, investigate their sequence and functional properties and thus suggest possible biological experiments.

Conclusions: We present *CircuitsDB*, a web-server devoted to the study of human and mouse mixed miRNA/TF Feed-Forward regulatory circuits, freely available at: <http://biocluster.di.unito.it/circuits/>

Background

Gene regulation is one of the most important molecular mechanisms occurring in a eukaryotic cell or organism. Control of gene expression is crucial for normal development and maintenance of healthy cells, and alterations from standard coordination programs can lead to severe diseases including cancer. The numerous events going from a DNA gene sequence to the corresponding protein are carefully controlled: from the control of transcription initiation to post-translational modifications that ultimately indicate the fate of the protein product. The primary regulation of gene expression is thought to be performed by Transcription Factors (TFs), proteins that are able to positively or negatively

coordinate gene transcription through the interaction with specific recognition DNA motifs usually located in the gene promoter regions (see [1,2] for recent reviews and perspectives).

In the last years, however, an additional class of gene regulators emerged: the microRNAs (miRNAs). miRNAs are short (~22nt) endogenous non-coding RNAs able to negatively regulate gene expression at the post-transcriptional level, via mRNA cleavage or translational repression. To this purpose, antisense complementary base-pair matching between a mature miRNA and its specific target sequences, located in the 3-UTR of the regulated mRNAs, is usually required (reviewed e.g. in [3]).

As a consequence of the above mentioned discoveries, the study of gene regulation has undergone a deep change of perspective. While past studies usually dealt with individual regulatory interactions, it has become clear that the only way to understand the regulatory activity of a

* Correspondence: davide.cora@ircct.it

¹Center for Molecular Systems Biology, University of Torino, Via Accademia Albertina, 13 - I-10123 Torino, Italy

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

eukaryotic genome is to directly address the complex, combinatorial nature of the whole ensemble of DNA *cis* and *trans* elements involved in such a process. Despite numerous efforts, mechanisms that control gene expression are not fully understood yet. In particular, a lot of methods exist to elucidate TF or miRNA-related regulatory networks, but comparable information to explicitly connect them is still lacking. Given a transcriptional and a post-transcriptional regulatory network, different possible ways to connect them are in principle possible: recently, a strong focus regarding the study of local mixed interactions has emerged in several works [4-10].

In this respect, it is important to notice that the transcription of miRNAs is widely regulated by POLII type promoters [11], and that co-expressed miRNAs are found to be regulated by common TFs [12]. Stemming from these considerations, we previously developed a computational framework for the study of connections between transcriptional and post-transcriptional (miRNA-mediated) regulatory interactions in the human genome [13]. We concentrated our attention on a particular class of local regulatory circuits (i.e. *network motifs*) in which a TF regulates an miRNA and together with it a set of Joint Target protein-coding genes. These circuits, called mixed miRNA/TF Feed-Forward regulatory Loops (FFLs, Figure 1a), were identified through a bioinformatic pipeline, mainly based on an ab-initio sequence analysis of human and mouse genomes. Once equipped with the catalogue of FFLs, we studied different ways to characterize their biological behaviors and implications. In particular, data were used to investigate connections between the mixed regulatory circuits involved in cancer.

Here, we present *CircuitsDB*, a user-friendly web-server that includes and extends our previous work [13].

In addition to the human catalogue of mixed FFLs, the database now contains full datasets relative to the mouse genome as well. Data are stored in a relational database that can be accessed through a dynamic web interface. The interface is composed by series of interactive online forms that allow the users to start with their favourite TF or miRNA or gene, and to follow their mutual connections in regulatory loops. Sequence information as well as functional data are provided (Gene Ontology, genomic annotations and positions of the putative regulatory sites, links to cancer or, more in general, disease information, patterns of tissue expression). Finally, a Wiki is also present, so that users can give their own feedback regarding the proposed interactions.

CircuitsDB is freely available at: <http://biocluster.di.unito.it/circuits/>.

Construction and content

CircuitsDB was constructed using a bioinformatic pipeline mainly based on an ab-initio sequence analysis

applied to regulatory regions of human and mouse genomes. In this section, we first describe the dataset of genomic regions used for the definition of our database. Second, we summarize the methodology originally used in [13] for the identification of the transcriptional and post-transcriptional players of the regulatory networks, in human and mouse and the approach used for their integration in mixed Feed-Forward regulatory Loops (FFLs). Finally, we describe the content and the structure of our web-accessible database.

Definition of the dataset of genes and regulatory regions used to infer the transcriptional and post-transcriptional networks

The promoter regions for protein-coding and microRNA (miRNA) genes as well the 3-UTRs were defined according to [13]. Gene definitions, sequences, and functional annotations were extracted from the Ensembl database [14], release 46 and from miRBase, version 9.2 [15]. The promoter region we selected for protein-coding genes corresponded to (-900/+100) nts around the Transcription Start Site (TSS), being the TSS at position +1. For each protein-coding gene, if more than one transcript was present, we used only the longest one. For miRNA genes, we first grouped pre-miRNAs in the so called Transcriptional Units (TUs) [16] and associated the promoter of the most 5-upstream member to all the pre-miRNAs belonging to the TU itself. Then, based on the fact that the pre-miRNAs were inter- or intra-genic, we defined the following promoters. For inter-genic pre-miRNAs the promoter corresponded to (-900/+100) nts upstream of the TSS of the first pre-miRNA in the TU. The same was true for intra-genic pre-miRNAs which showed opposite orientation with respect to the hosting protein-coding gene. Finally, if the pre-miRNAs were intra-genic but sharing the same orientation of the hosting protein-coding gene, the promoter region was considered coincident with the one defined for the protein-coding host gene.

This procedure was implemented here for both human and mouse genomes. For subsequent analysis we considered only protein-coding and pre-miRNA genes showing at least a direct one-to-one orthology between the two genomes (from [14] and [16]). The final dataset of promoter regions is composed of a collection of 21446 (21316 protein-coding plus 130 pre-miRNA) human and 21944 (21814 protein-coding plus 130 pre-miRNA) mouse regulatory sequences. The 130 pre-miRNAs included in our work encode for 193 mature miRNAs (see Supplementary File S1 of [13]). For protein-coding genes, we then downloaded the 3-UTR regions, considering only the longest transcript in case of multiple alternative isoforms. We ended up with only 17486 human and 15921 mouse sequences, since not all

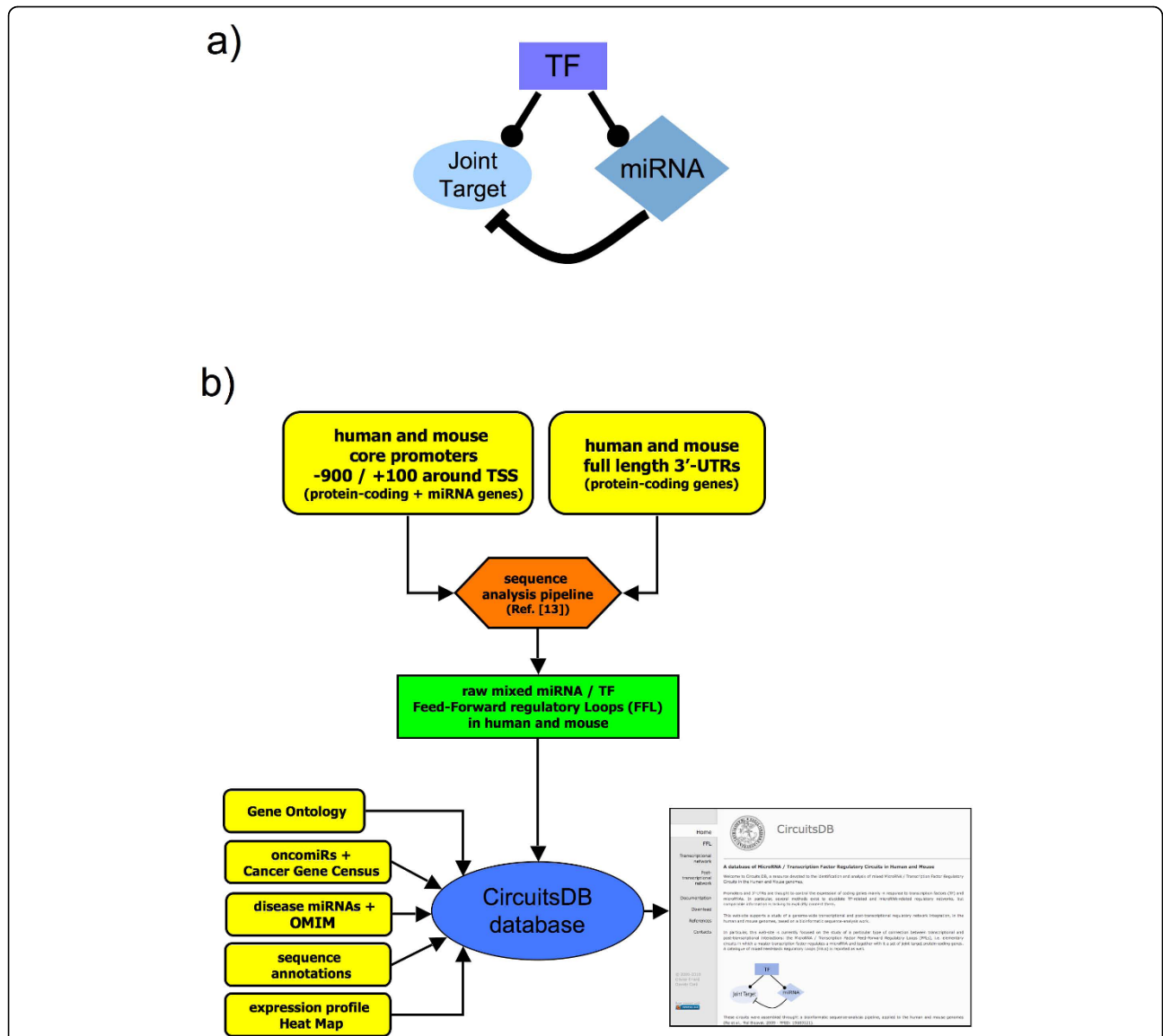


Figure 1 Mixed miRNA/TF Feed-Forward Loops and CircuitsDB construction pipeline. a) Representation of a typical mixed Feed-Forward regulatory Loop (FFL) included in CircuitsDB. Circuits are composed by a master Transcription Factor (TF, square box) that regulates a microRNA (miRNA, diamond-shaped box) and together with it a Joint Target protein-coding gene (round box). Inside each circuit, → indicates transcriptional activation/repression, whilst ⊘ post-transcriptional repression. b) Flow-chart of the strategy used for the construction of CircuitsDB. We first built a genome-wide catalogue of putative core promoter regions for protein-coding and miRNA genes, plus a set of 3'-UTRs for protein-coding genes, in human and mouse. Then we used the analysis pipeline developed in our previous work [13] to infer a dataset of mixed regulatory FFLs, in human and mouse. Finally we integrated different kinds of biological annotations to support the circuits' properties; the CircuitsDB web-site allows a dynamic exploration of such properties.

the genes have a well defined 3-UTR in the Ensembl database. All the sequences were Repeat-Masked using Ensembl default parameters.

Oligo analysis and definition of mixed microRNA/Transcription Factor regulatory Feed-Forward Loops
 Details about the oligo analysis are listed in [13]; here we report our main choices and results (Figure 1b). Briefly, we scanned all the promoter regions and the 3-UTRs for

conserved-overrepresented oligos (6 to 9 nts for promoters; 7 nts for the 3-UTRs) with potential regulatory roles (as Transcription Factor Binding Sites, TFBS, for promoters or miRNA seeds for the 3-UTRs). By doing so, we fixed 0.1 as False Discovery Rate (FDR) in the oligo analysis pipeline. To assess the oligos surviving the motif-finding analysis, we used a catalogue of known TFBS consensus from the Transfac database [17] and from [18] for the oligos located in promoter regions, manually

filtering out those TFs characterized by very long or too degenerate consensus sequences. Similarly, we used a catalogue of known miRNA seeds derived from the mature miRNAs included in our study to identify significant oligos located in the 3-UTRs (see [19] and [20] for additional details concerning the used algorithms). The above analysis was performed here in parallel for human and mouse. In human, for the transcriptional network, we obtained a catalogue of 2031 significant oligos that could be associated to known TFBSs for a total of 115 different TFs. These 2031 oligos targeted 21399 genes (21219 protein-coding and 180 mature miRNAs). For the post-transcriptional network, we ended up with a library of 182 significant oligos, each of them matching with at least one seed present in 140 out of our 193 mature miRNAs and targeting a total of 17266 protein-coding genes. We obtained rather similar results for the mouse: the transcriptional network is composed of 22054 genes (21875 protein-coding and 179 pre-miRNAs) and 115 different TFs targeting the 2031 significant oligos. On the other hand, the post-transcriptional network includes 15755 genes, targeted by 178 significant oligos corresponding to 143 mature miRNAs.

Once we obtained these two regulatory networks, we focused on the integration of the two datasets in order to construct a catalogue of mixed miRNA/TF FFLs (Figure 1a). In human, that integration included 5030 different “single target circuits”, each of them defined by a single TF as master regulator, a single mature miRNA and a single protein-coding Joint Target. From these single target circuits we constructed “merged circuits” grouping together the FFLs sharing the same TF and the same miRNA, thus obtaining 638 merged circuits. These circuits involved a total of 101 TFs, 133 mature miRNAs and 2625 Joint Target genes. In mouse, we found 6684 different “single target circuits”, which could be grouped in 850 “merged circuits”, involving a total of 94 TFs, 142 mature miRNAs and 2968 Joint Target genes. 30 single target circuits were conserved between human and mouse: they share the TF, the miRNA as well as the Joint Target being one-to-one orthologs.

The prediction of reliable miRNA-mediated post-transcriptional regulatory interactions is still an open issue in computational biology and it is well known that different approaches can lead to very different outcomes [21]. To address this problem, we included in *CircuitsDB* conserved miRNA-target predictions obtained from two external resources, namely TargetScan [22] and TargetMiner [23]. We downloaded from the TargetScan website <http://www.targetscan.org/> inferred miRNA targets for the human and mouse genomes and we mapped Entrez Gene symbols and miRNA family names provided on the Ensembl stable identifiers (ids) for protein-coding genes and pre-miRNAs present in our database of FFLs.

A similar procedure was implemented for the human genome-wide predictions obtained from the TargetMiner website http://www.isical.ac.in/~bioinfo_miu/, where human RefSeq ids were again mapped on Ensembl gene stable ids. By TargetScan or TargetMiner we could confirm 1434 out of the 5030 human FFLs included in our study at the post-transcriptional level. Similar results were obtained also for the murine case, resulting in 1107 FFLs having the post-transcriptional link confirmed by TargetScan. These data are reported in Additional File 1 (for human) and Additional File 2 (for mouse).

Furthermore, we also investigated whether in our database connections between TF and miRNAs were present in mixed Feed-Back regulatory Loops (FBLs), i.e. situations in which a master TF regulates an miRNA, being itself the target of the regulated miRNA at the post-transcriptional level. To this end, in order to recognize the miRNA - TF post-transcriptional interactions, we manually prepared a translation table in which Transfac TF ids were associated to known Ensembl gene stable ids, where possible. We ended up with a catalogue of 113 mixed FBLs in human and 38 FBLs in mouse, having the post-transcriptional link confirmed by at least one of the supporting databases included.

Functional annotations, cancer and disease genes

Once equipped with the catalogue of mixed miRNA/TF Feed-Forward loops, we investigated their functional properties with several different criteria with a focus on functional annotations according to the Gene Ontology (GO) database [24] and to their relevance in cancer or other diseases. In the present version of *CircuitsDB*, these three types of biological annotations are included in the on-line web-service.

Gene product GO annotations for the TF and Joint Target protein-coding genes were downloaded from the Ensembl database, version 46.

Regarding the identification of cancer related genes, we obtained a list of oncomiRs from [25,26] and [27] while for the protein-coding target genes we enumerated a list of genes showing mutations in cancer based on the Cancer Gene Census catalogue [28]. We then focused on the annotation of *CircuitsDB* genes and miRNAs in terms of genetic diseases: for protein-coding genes we used the established OMIM[29] catalogue, whilst the HMDD miRNA-disease database [30] was interrogated to annotate the miRNAs present in our FFLs.

Comparison with experimentally supported regulatory interactions for each of the circuit's link

We then compared our predicted human miRNA-target links (miRNA → Joint Target) with two databases of experimentally supported information, Tarbase [31] and the component of miRecords [32] reporting validated

miRNA-target links. 140 miRNAs, used in our post-transcriptional network, were also present in Tarbase or miRecords. On average, 11.5% of our predicted miRNA targets were already validated experimentally. It was of interest to compare the other two computational algorithms for miRNA target predictions, TargetScan and TargetMiner, with our own algorithm, in the same benchmarking setting, in order to compare the relative performances. To this end, we applied a binomial test to assess the proportions of experimentally supported predictions (implemented as the function `prop.test()` in the R [33] statistical environment); for the sake of the comparison, this test was applied separately to each miRNA in common between our post-transcriptional network and TargetScan (33) or TargetMiner (30). Setting the confidence threshold to 0.01, no significant difference was found for 30 out of 33 miRNAs, when using TargetScan, and for 29 out of 30 miRNAs, when using TargetMiner. Therefore, we can conclude that the proportions of true positives recognized separately by each algorithm, evaluated on the miRNA pool present in our database, are substantially comparable (see Additional File 3).

Eventually, we also assessed the reliability of our predictions separately for the other two types of links in a circuit: the TF \rightarrow Joint Target and the TF \rightarrow miRNA regulatory interactions.

A proof of principle of the reliability of the TF \rightarrow Joint Target links was already established in our previous work [13], where we compared our predictions with the experimentally validated results reported in [34], a large-scale study of direct MYC binding target genes in a model of human B lymphoid tumors performed by chromatin immunoprecipitation coupled with pair-end ditag sequencing analysis (ChIP-PET). The intersection between [34] and our predicted MYC interactions proved to be statistically significant ($p = 1.1 \times 10^{-6}$, Fisher's test).

We then proceeded to compare our predicted TF \rightarrow miRNA regulatory interactions links with TransmiR [11], a literature-based database of TF - miRNA links. At present, we were able to safely identify only 16 TF names in common between our dataset and the TransmiR database. 5 TF - miRNA links were found in common out of the 36 identified by TransmiR, involving a transcription factor and an miRNA present in our dataset.

Tissue expression Heat Map

Due to the sequence analysis pipeline that we adopted to identify the FFLs in CircuitsDB, we were not able to recognize if the action of the master TF was activating or repressing its targets and thus if the FFL that we obtained was of the so-called Type I or Type II [5,13]. Moreover, not only the two types of circuits may lead to very different behaviours in terms of expression patterns

of their components [5], but also the understanding of the consequences of miRNA-target interaction is currently challenged [35-37].

Therefore, we decided to simply give the users the possibility to explore the expression values across several tissues for the TF, miRNA and Joint Target belonging to a given FFL by means of a graphical heat map representation. For this purpose, we collected the expression profiling data for 175 miRNAs over 24 human organs from [38] and for more than 40000 human protein-coding transcripts as microarray probe sets over 79 human tissues from [39]. We were able to identify 14 tissue types in common between the two datasets: AdrenalGland, BoneMarrow, Brain, Heart, Kidney, Liver, Lung, Lymph, Pancreas, Placenta, Prostate, Testis, Thymus, Uterus (see also [23]). For the protein-coding genes, we then used the BioMart tool of the Ensembl database to map the original probe sets on the Ensembl gene stable ids used in our CircuitsDB. If more than one probe matched on the same Ensembl id, we retained as expression values for that gene the mean values, for each tissue type. For both miRNAs and protein-coding genes the expression values were then \log_2 -transformed and the Z -score with respect to the mean over all the tissues was evaluated. These values were finally used to create, for each mixed FFL, a heat map, composed by 14 rows (corresponding to the 14 different tissues) and 3 columns, corresponding to the TF, miRNA and Joint Target embedded in the FFL.

Utility

CircuitsDB was built in the PHP script language and as a MySQL relational database system on a Linux server. In the MySQL database pre-compiled transcriptional and post-transcriptional networks, the dataset of mixed FFLs and all functional and biological information, for both human and mouse, are stored. The interactive web interface allows the user to first select an organism of interest, then select a TF, an miRNA id or a protein-coding gene name (or a combination of these three elements) and query the database in order to retrieve a catalogue of mixed miRNA/TF Feed-Forward regulatory Loops in which the search keys are involved (Figure 2a and 2b).

Query's results are displayed in a synthetic tabular-like view, with single target circuits on separate lines. Different color codes for the different circuit elements are used. The view relative to each single-target circuit can be expanded into six related sub-panels showing additional detailed information, similarly to the graphical strategy adopted e.g. in [31].

The additional information is divided into six categories: *Transcription Factor (TF)*, *miRNA*, *Joint Target gene*, *Sequence*, *Tissue Expression Heat Map* and

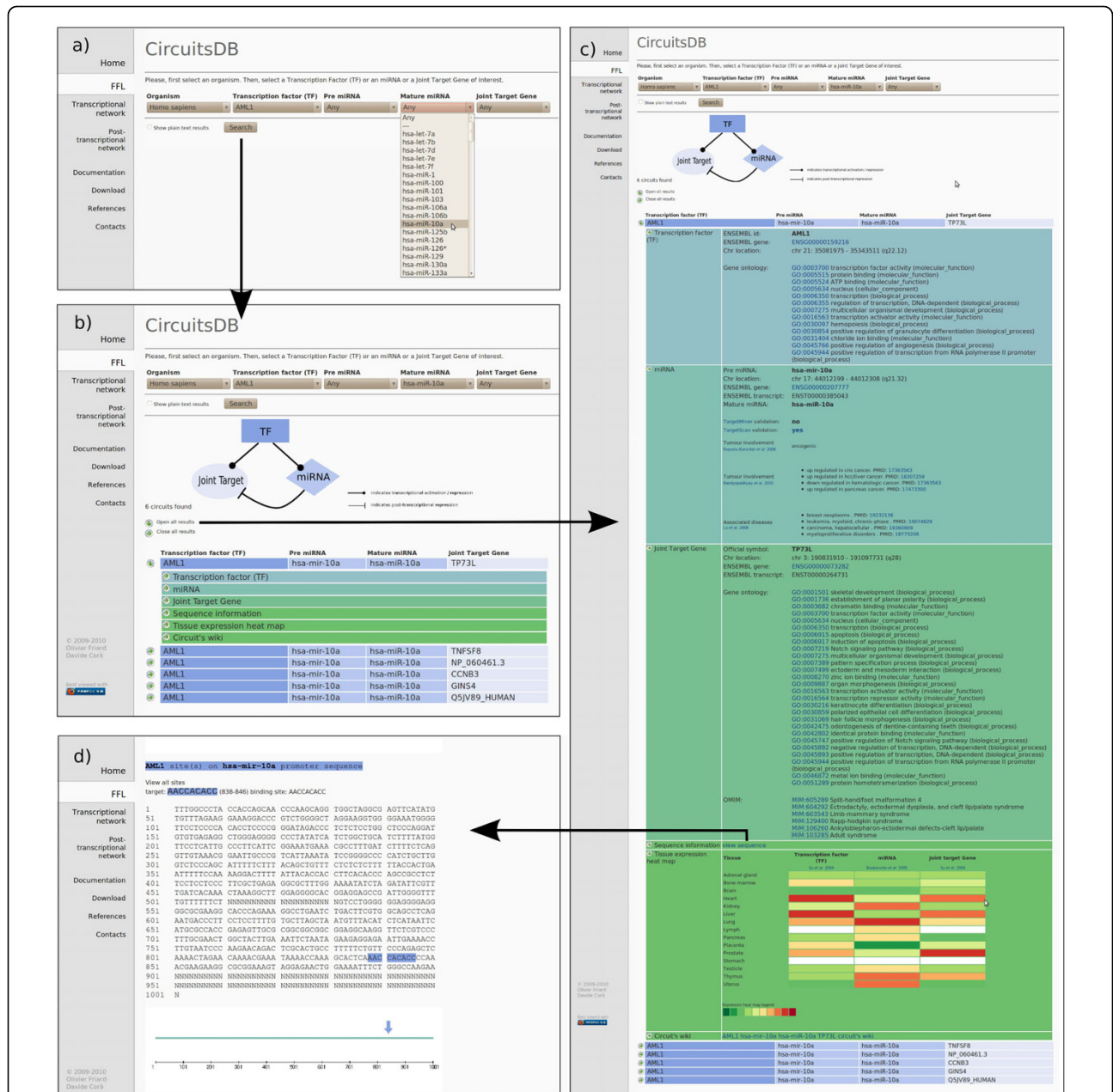


Figure 2 Example of CircuitsDB usage. Graphical snapshots of a typical working session with CircuitsDB. a) In the first step the user selects an organism of interest and at least one search key among the available ones: Transcription Factor (TF) or microRNA (miRNA) or Joint Target gene name. b) In the second panel a list of mixed TF/miRNA regulatory circuits is retrieved. c) In the third panel an example of CircuitsDB output is shown: a single mixed Feed-Forward Loop is identified and its components are reported in a tabular-like fashion. The user can explore TF additional information, as well as miRNA and Joint Target annotations. A sequence analysis panel and a Wiki for community feedback are available. d) Example of sequence analysis panel: TF binding sites are explicitly highlighted on the miRNA core promoter region.

Circuit's Wiki (Figure 2c). The TF information category gathers mostly biological properties of the master TF regulating the circuits: its Ensembl annotations, chromosome location and Gene Ontology annotation, if available. The miRNA information category is composed of the Ensembl annotations of the pre-miRNA composing

the circuit, its chromosome location and corresponding mature miRNA ids. Tumor involvement fields provide annotations of miRNAs known to be involved in cancer. In parallel to that, a disease field reports whether the miRNA is connected to a known disease, if any. The Joint Target gene information panel contains again

Ensembl annotation for the protein-coding Joint Target gene, the gene official name as well as GO annotations. Again, a tumor involvement field shows possible links with cancer. The OMIM fields displays the corresponding mendelian disease associated to the current gene. With the Sequence information link the user is redirected to an additional web-page in which the promoter sequence of the miRNA or the promoter and 3-UTR of the Joint Target present in the circuit are explicitly displayed. TFBS identified with our algorithm and miRNA binding sites are highlighted on the sequences, thus providing direct information for experimental tests (Figure 2d). In the Tissue Expression Heat Map viewer we interactively generated a heat map of tissue specificity for the TF, miRNA and Joint Target belonging to the circuit. Furthermore, we provide CircuitsDB with an embedded Wiki system, that allows interested users to share their knowledge and efforts in annotating circuits: the last category, *Circuit's Wiki*, redirects the user to editable web pages, where a registered user can add personal information in addition to the biological annotations already included in our database.

A Download section provides all the CircuitsDB content as flat files. A Reference section reports a list of the most important data sources used for the construction of our database.

Finally, two additional sections of the database are directly accessible through links in the home page: "Transcriptional" and "Post-Transcriptional Network". In the first one, the user can explore the transcriptional network used for the development of the mixed FFLs catalogue, in human and mouse. Entry points can be a TF of interest, a gene or a DNA oligo. It is worth mentioning that this section allows direct inspection of the subset of our transcriptional regulatory networks involving miRNAs as targets, i.e. the collection of putative TFBS identified on promoter regions for miRNA genes. In the second one, users can explore our post-transcriptional networks, for human and mouse. Entry points can be an miRNA of interest, a gene or again a DNA oligo.

Discussion

CircuitsDB is a database where transcriptional and post-transcriptional (miRNA mediated) network information is fused together in order to propose and recognize non trivial regulatory combinations. Figure 2 shows the main steps that a user should follow during the investigation of the proposed circuits in CircuitsDB web-site: identification of a FFL according to a TF, miRNA or Joint Target gene id (Figure 2a and 2b); exploration of the circuit components and related annotations (Figure 2c); investigation of the regulatory sites comprised in the circuit corresponding to the identified sequences (Figure 2d).

In [13] several examples in which mixed Feed-Forward Loops could exert synergistic biological effects have already been discussed (the MYC/hsa-miR-20a;miR-17-5p circuit, the AREB6/hsa-miR-375 circuit, the MEF-2/hsa-miR-133a circuit, the C-REL/hsa-mir-199a circuit and the HSF2/hsa-let-7f circuit). Here, we will discuss other examples of single-target circuits potentially linked to cancer that we obtained from CircuitsDB.

One circuit involves the master transcription factor Runx1 or Acute Myeloid Leukemia 1 (AML1), miR-10a and the p63 (TP73L), three genes found implicated in leukemia. AML1 is the target of multiple chromosomal translocations in human leukemia and the TEL-AML1 oncogene is the hallmark translocation in Childhood Acute Lymphoblastic Leukemia [40]. miR-10a was proven to be overexpressed and functionally relevant in various tumors, including AML [41] while p63 is a critical transcriptional regulator of cancer cells [42]. This FFL might also coordinate the physiological hematopoiesis during embryonic development since Runx1 is essential for the generation of definitive haematopoietic cells from haemogenic endothelium as shown using a conditional deletion of Runx1 [43]. Nevertheless, p63 has been found to be relevant for tissue development [44,45] and miR-10a is one of the most upregulated miRNA during endodermal differentiation from human embryonic stem cells [46].

Another potentially relevant circuit for tumorigenesis, in particular for ovarian cancer, a leading cause of death from gynecologic malignancies, is composed of HoxA4, miR-125b and ERBB3. HOX genes are transcription factors that control morphogenesis, organogenesis as well as differentiation and that play an important role in ovarian cancer progression by controlling cell migration [47]. miR-125b has been found to be differentially expressed in serous ovarian carcinomas compared with normal ovarian tissues [48]. At the same time, ERBB3 is a tyrosine kinase receptor often activated in ovarian cancer and perturbation of ERBB3-dependent signal transduction by RNA interference resulted in decrease disease progression and prolonged survival in murine models, identifying ERBB3 as a potential therapeutic target in ovarian cancer [49]. The identification of an miRNA, i.e. miR-125b, that could downregulate ERBB3 would be very valuable for ovarian cancer treatments.

Another interesting association with cancer can be observed in the SOX-5/miR-29a/SPARC circuit. Decreased expression of SPARC, an important mediator of cell-matrix interaction, was previously observed in Nasopharyngeal carcinoma (NPC) and in the same system SOX-5 turned out to be upregulated [50]. Considering that SOX-5 overexpression in NPC tumors correlates clinically with poor survival it is essential to understand how SOX-5 regulates tumor progression. It is conceivable that that SOX-5 down-regulates SPARC expression

directly at the level of transcription, while regulating positively miR-29a transcription: this would result in coordinate downregulation of SPARC at the post-transcriptional level. Considering that several evidences connect miR-29a with epithelial tumor invasion and metastasis formation as well as epithelial-mesenchymal transition (EMT) [51] miR-29a could represent a main regulation of SPARC and experimental validations should be carried out.

Some mixed circuits have already been studied from an experimental point of view. For instance, in [52] the authors investigated the interaction between the miR-17-92 cluster, the Myc oncogene and the E2F1 transcription factor, being E2F1 an additional target of Myc that promotes cell-cycle progression (this circuit is predicted also from our analysis and thus present in *CircuitsDB*). In [53] Brosh and colleagues analyzed mixed FFLs in the framework of the mammalian cell proliferation control network. They again concentrated on a network architecture that includes the transcription factor E2F1 and a family of 15 miRNAs, which co-regulate mutual target genes transcriptionally and post-transcriptionally and whose cooperative action reinforces cellular proliferation. Then, this FFL appears to be repressed by p53, possibly by promoting senescence and suppressing cancer progression.

Although in our work we focused on the circuits' properties in relation to cancer biology, by means of the already established biological features of their components, other possible functions might exist for the proposed catalogue of mixed FFLs. In particular, in the original formulation of [4], the wording for mixed circuits referred to an evolutionary perspective: looking at how "canalizing genes", essential for higher organism development, could be influenced by miRNA regulatory networks that act as stabilizers for noise fluctuations in gene expression (type I or incoherent FFLs) [35] and [36] seem to shed new light in supporting this alternative hypothesis for FFLs functioning and in [54] was recently proposed, through stochastic modeling and simulations, a mathematical model for that. It is worth noting that in parallel to this, mixed regulatory circuits are also emerging as key players in regulatory networks of Embryonic Stem Cells (ESC) [55]. Moreover, the interplay between TFs, miRNAs and shared targets is able to influence ESC differentiation and act as defining factors in Induced Pluripotent (iPS) and Cancer Stem Cells (CSC) [56].

Conclusions

We present *CircuitsDB*, a public web application devoted to the study of interactions between transcriptional and post-transcriptional regulatory interactions. *CircuitsDB* is currently mainly focused on the study of mixed miRNA/TF Feed-Forward Loops, i.e. regulatory circuits in which a master TF regulates an miRNA and, together with it, a set of Joint Target protein-coding

genes. These circuits were assembled in our previous experience [13] in the human case, based on a bioinformatic ab-initio analysis. Here, we expanded to the murine case our dataset and provided an integrated web-service to explore and directly investigate such relationships in terms of their sequence and several types of functional annotations. A catalogue of mixed mixed miRNA/TF Feed-Back Loops is also presented.

We consider *CircuitsDB* only the first step of more advanced studies. In particular, we plan to further extend our work to include additional types of mixed miRNA/TF local interactions and other types of post-transcriptional regulators. Prediction methods to infer TF and miRNA regulatory networks also continue to evolve and genome-wide experimental dataset of TF and miRNA interactions will be available in the near future. In subsequent releases, additional bioinformatic methodologies and experimental data for the construction of the database could be easily incorporated in our server.

Availability and requirements

The *CircuitsDB* web-service is freely available at <http://biocluster.di.unito.it/circuits/>. Detailed documentation can be accessed by a link on the left bar in the home page and includes various explanatory applications.

Additional material

Additional file 1: Catalogue of human mixed Feed-Forward regulatory Loops included in *CircuitsDB*. The complete list of mixed Feed-Forward regulatory Loops (FFLs) included in the current release of *CircuitsDB* is reported, for the human genome. Each line corresponds to a single-target closed FFL. The first column includes the FFL id, composed by the Transcription Factor (TF) name and miRNA gene ids (Ensembl stable identifier, pre-miRNA and mature miRNA ids according to the standard nomenclature of miRBase). The second column shows the Joint Target protein-coding gene (Ensembl stable identifier and HGNC standard id). The third and fourth column report the validation of the miRNA \rightarrow Joint Target post-transcriptional interaction according to the TargetScan and TargetMiner databases, respectively.

Additional file 2: Catalogue of mouse mixed Feed-Forward regulatory Loops included in *CircuitsDB*. The complete list of mixed Feed-Forward regulatory Loops (FFLs) included in the current release of *CircuitsDB* is reported, for the mouse genome. Each line corresponds to a single-target closed FFL. The first column includes the FFL id, composed by the Transcription Factor (TF) name and miRNA gene ids (Ensembl stable identifier, pre-miRNA and mature miRNA ids according to the standard nomenclature of miRBase). The second column shows the Joint Target protein-coding gene (Ensembl stable identifier and Mouse gene symbol). The third column reports the validation of the miRNA \rightarrow Joint Target post-transcriptional interaction according to the TargetScan database.

Additional file 3: Results of the comparison with experimentally validated miRNA-gene pairs. The table shows the results of the binomial test used to compare the proportions of experimentally supported miRNA-gene interactions between *CircuitsDB* and TargetScan or TargetMiner. Only miRNAs present in at least one circuit, included by TargetScan or TargetMiner and with a minimum of one validated target were used (0.01 significance threshold).

Abbreviations

TFs: Transcription Factors; miRNAs: microRNAs; FFLs: Feed-Forward regulatory Loops; FBLs: Feed-Back regulatory Loops.

Acknowledgements

We thank Michele Caselle, Ferdinando Di Cunto, Francesca Orso and Paolo Provero for useful suggestions and discussions. We thank Mariama El-Baroudi for assistance with the TarBase and miRecords data. We are grateful to Dr. Qinghua Cui, Peking University, Beijing, China, for the availability of the full HMDD dataset. We are also grateful to the anonymous referees whose comments and suggestions allowed a significant improvement of the work. We acknowledge the Department of Computer Science, University of Torino, for hosting the CircuitsDB web-site.

Author details

¹Center for Molecular Systems Biology, University of Torino, Via Accademia Albertina, 13 - I-10123 Torino, Italy. ²CIBIO-Centre for Integrative Biology, University of Trento, Via delle Regole 101, I-38100 Trento, Italy. ³Department of Oncological Sciences, c/o Institute for Cancer Research and Treatment (IRCC), School of Medicine, University of Torino, Str. Prov. 142 Km. 3.95, I-10060 Candiolo, Italy. ⁴Molecular Biotechnology Center, University of Torino, Via Nizza 52, I-10126 Torino, Italy. ⁵Systems Biology Lab, Institute for Cancer Research and Treatment (IRCC), School of Medicine, University of Torino, Str. Prov. 142 Km. 3.95, I-10060 Candiolo, Italy.

Authors' contributions

OF wrote the software implementation of CircuitsDB. AR and DT participated in data analysis and in their biological assessment. MDB provided funding and resources for the project. DC designed, coordinated the project and participated in data analysis. DC wrote the paper and all the authors read and approved the final version.

Received: 17 December 2009 Accepted: 23 August 2010

Published: 23 August 2010

References

1. Elmitski L, Jin VX, Farnham PJ, Jones SJ: **Locating mammalian transcription factor binding sites: A survey of computational and experimental techniques.** *Genome Res* 2006, **16**(12):1455-64.
2. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM: **A census of human transcription factors: function, expression and evolution.** *Nat Rev Genet* 2009, **10**:252-63.
3. Bartel DP: **MicroRNAs: target recognition and regulatory functions.** *Cell* 2009, **136**:215-33.
4. Hornstein E, Shomron N: **Canalization of development by microRNAs.** *Nat Genet* 2006, **38**(Suppl):S20-4.
5. Tsang J, Zhu J, van Oudenaarden A: **MicroRNA-Mediated Feedback and Feedforward Loops Are Recurrent Network Motifs in Mammals.** *Mol Cell* 2007, **26**(5):753-67.
6. Shalgi R, Lieber D, Oren M, Pilpel Y: **Global and Local Architecture of the Mammalian microRNA-Transcription Factor Regulatory Network.** *PLoS Comput Biol* 2007, **3**(7):e131.
7. Zhou Y, Ferguson J, Chang JT, Kluger Y: **Inter- and intra-combinatorial regulation by transcription factors and microRNAs.** *BMC Genomics* 2007, **8**:396.
8. Yu X, Lin J, Zack DJ, Mendell JT, J Q: **Analysis of regulatory network topology reveals functionally distinct classes of microRNAs.** *Nucleic Acids Res* 2008, **36**:6494-503.
9. Sinha AU, Kaimal V, Chen J, Jegga AG: **Dissecting microregulation of a master regulatory network.** *BMC Genomics* 2008, **8**:88.
10. Martinez NJ, Walhout AJ: **The interplay between transcription factors and microRNAs in genome-scale regulatory networks.** *Bioessays* 2009, **31**:435-45.
11. Wang J, Lu M, Qiu C, Cui Q: **TransmiR: a transcription factor-microRNA regulation database.** *Nucleic Acids Res* 2009, **37** Database: D119-22.
12. Bandyopadhyay S, Bhattacharyya M: **Analyzing miRNA co-expression networks to explore TF-miRNA regulation.** *BMC Bioinformatics* 2009, **10**:163.
13. Re A, Corá D, Taverna D, Caselle M: **Genome-wide survey of microRNA-transcription factor feed-forward regulatory circuits in human.** *Mol Biosyst* 2009, **5**:854-67.
14. Hubbard T, Aken B, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, Down T, Dyer S, Fitzgerald S, Fernandez-Banet J, Graf S, Haider S, Hammond M, Herrero J, Holland R, Howe K: **Ensembl 2007.** *Nucleic Acids Res* 2007, **35** Database: D610-7.
15. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature.** *Nucleic Acids Res* 2006, **34** Database: D140-4.
16. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeifer S, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, Hermida L, Fulci V, Chiaretti S, Foá R, Schliwka J, Fuchs U, Novosel A, Muller RU, Schermer B, Bissels U, Inman J, Phan Q, Chien M, Weir DB, Choksi R, De Vita G, Frezzetti D, Trompeter HI, Hornung V, Teng G, Hartmann G, Palkovits M, Di Lauro R, Wernet P, Macino G, Rogler CE, Nagle JW, Ju J, Papavasiliou FN, Benzing T, Lichter P, Tam W, Brownstein MJ, Bosio A, Borkhardt A, Russo JJ, Sander C, Zavolan M, Tuschl T: **A mammalian microRNA expression atlas based on small RNA library sequencing.** *Cell* 2007, **129**(7):1401-14.
17. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel A, E W: **TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes.** *Nucleic Acids Res* 2006, **34** Database: D108-110.
18. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M: **Systematic discovery of regulatory motifs in human promoters and 3 UTRs by comparison of several mammals.** *Nature* 2005, **434**(7031):338-345.
19. Corà D, Herrmann C, Dieterich C, Di Cunto F, Provero P, Caselle M: **Ab initio identification of putative human transcription factor binding sites by comparative genomics.** *BMC Bioinformatics* 2005, **6**:110.
20. Corà D, Di Cunto F, Caselle M, Provero P: **Identification of candidate regulatory sequences in mammalian 3 UTRs by statistical analysis of oligonucleotide distributions.** *BMC Bioinformatics* 2007, **8**:174.
21. Sethupathy P, Megraw M, Hatzigeorgiou AG: **A guide through present computational approaches for the identification of mammalian microRNA targets.** *Nat Methods* 2006, **3**:881-6.
22. Friedman RC, Farh KK, Burge CB, P BD: **Most mammalian mRNAs are conserved targets of microRNAs.** *Genome Res* 2009, **19**:92-105.
23. Bandyopadhyay S, Mitra R: **TargetMiner: microRNA target prediction with systematic identification of tissue-specific negative examples.** *Bioinformatics* 2009, **25**:2625-31.
24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene Ontology: tool for the unification of biology.** *Nature Genet* 2000, **25**:25-9.
25. Esquela-Kerscher A, Slack FJ: **Oncomirs - microRNAs with a role in cancer.** *Nat Rev Cancer* 2006, **6**(4):259-69.
26. Zhang B, Pan X, Cobb GP, Anderson TA: **microRNAs as oncogenes and tumor suppressors.** *Dev Biol* 2007, **302**:1-12.
27. Bandyopadhyay S, Mitra R, Maulik U, Zhang MQ: **Development of the human cancer microRNA network.** *Silence* 2010, **1**:6.
28. **The Cancer Gene Census.** [http://www.sanger.ac.uk/genetics/CGP/Census/].
29. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD): **Online Mendelian Inheritance in Man, OMIM (TM).** [http://www.ncbi.nlm.nih.gov/omim/].
30. Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, Cui Q: **An analysis of human microRNA and disease associations.** *PLoS One* 2008, **3**:e3420.
31. Papadopoulos GL, Reczko M, Simossis VA, Sethupathy P, Hatzigeorgiou AG: **The database of experimentally supported targets: a functional update of TarBase.** *Nucleic Acids Res* 2009, **37** Database: D155-158.
32. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T: **miRecords: an integrated resource for microRNA-target interactions.** *Nucleic Acids Res* 2009, **37** Database: D105-10.
33. **The R Project for Statistical Computing.** [http://www.r-project.org/].
34. Zeller K, Zhao X, Lee C, Chiu K, Yao F, Yustein J, Ooi H, Orlov Y, Shahab A, Yong H, Fu Y, Weng Z, Kuznetsov V, Sung W, Ruan Y, Dang C, Wei C: **Global mapping of c-Myc binding sites and target gene networks in human B cells.** *PNAS* 2006, **103**(47):17834-39.
35. Wu CI, Shen Y, Tang T: **Evolution under canalization and the dual roles of microRNAs: a hypothesis.** *Genome Res* 2009, **19**:734-43.

36. Li X, Cassidy JJ, Reinke CA, Fischboeck S, Carthew RW: **A microRNA imparts robustness against environmental fluctuation during development.** *Cell* 2009, **137**:273-82.
37. Arvey A, Larsson E, Sander C, Leslie CS, Marks DS: **Target mRNA abundance dilutes microRNA and siRNA activity.** *Mol Syst Biol* 2010, **6**:363.
38. Baskerville S, Bartel DP: **Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes.** *RNA* 2005, **11**:241-7.
39. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB: **A gene atlas of the mouse and human protein-encoding transcriptomes.** *Proc Natl Acad Sci USA* 2004, **101**:6062-7.
40. Schindler JW, Van Buren D, Foudi A, Krejci O, Qin J, Orkin SH, H H: **TEL-AML1 corrupts hematopoietic stem cells to persist in the bone marrow and initiate leukemia.** *Cell Stem Cell* 2009, **5**:5-6.
41. Garzon R, Garofalo M, Martelli MP, Briesewitz R, Wang L, Fernandez-Cymering C, Volinia S, Liu CG, Schnittger S, Haferlach T, Liso A, Diverio D, Mancini M, Meloni G, Foa R, Martelli MF, Mecucci C, Croce CM, Falini B: **Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin.** *Proc Natl Acad Sci USA* 2008, **105**:3945-50.
42. Trink B, Osada M, Ratovitski E, Sidransky D: **p63 transcriptional regulation of epithelial integrity and cancer.** *Cell Cycle* 2007, **6**:240-5.
43. Lancrin C, Sroczynska P, Stephenson C, Allen T, Kouskoff V, Lacaud G: **The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage.** *Nature* 2009, **457**:892-5.
44. Candi E, Rufini A, Terrinoni A, Giamboni-Miraglia A, Lena AM, Mantovani R, Knight R, Melino G: **DeltaNp63 regulates thymic development through enhanced expression of FgfR2 and Jag2.** *Proc Natl Acad Sci USA* 2007, **104**:11999-2004.
45. Carroll DK, Carroll JS, Leong CO, Cheng F, Brown M, Mills AA, Brugge JS, W EL: **p63 regulates an adhesion programme and cell survival in epithelial cells.** *Nat Cell Biol* 2006, **8**:551-61.
46. Tzur G, Levy A, Meiri E, Barad O, Spector Y, Bentwich Z, Mizrahi L, Katzenellenbogen M, Ben-Shushan E, Reubinoff BE, E G: **MicroRNA expression patterns and function in endodermal differentiation of human embryonic stem cells.** *PLoS One* 2008, **3**:e3726.
47. Klausen C, Leung PC, Auersperg N: **Cell motility and spreading are suppressed by HOXA4 in ovarian cancer cells: possible involvement of beta1 integrin.** *Mol Cancer Res* 2009, **7**:1425-37.
48. Nam EJ, Yoon H, Kim SW, Kim H, Kim YT, Kim JH, Kim JW, Kim S: **MicroRNA expression profiles in serous ovarian carcinoma.** *Clin Cancer Res* 2008, **14**:2690-5.
49. Sheng Q, Liu X, Fleming E, Yuan K, Piao H, Chen J, Moustafa Z, Thomas RK, Greulich H, Schinzel A, Zaghlu S, Batt D, Ettenberg S, Meyerson M, Schoeberl B, Kung AL, Hahn WC, Drapkin R, Livingston DM, Liu JF: **An activated ErbB3/NRG1 autocrine loop supports in vivo proliferation in ovarian cancer cells.** *Cancer Cell* 2010, **17**:298-310.
50. Huang DY, Lin YT, Jan PS, Hwang YC, Liang ST, Peng Y, Huang CY, Wu HC, Lin CT: **Transcription factor SOX-5 enhances nasopharyngeal carcinoma progression by down-regulating SPARC gene expression.** *J Pathol* 2008, **214**:445-55.
51. Gebeshuber CA, Zatloukal K, Martinez J: **miR-29a suppresses tristetruprolin, which is a regulator of epithelial polarity and metastasis.** *EMBO Rep* 2009, **10**:400-5.
52. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT: **c-Myc-regulated microRNAs modulate E2F1 expression.** *Nature* 2005, **435**:839-43.
53. Brosh R, Shalgi R, Liran A, Landan G, Korotayev K, Nguyen GH, Enerly E, Johnsen H, Buganim Y, Solomon H, Goldstein I, Madar S, Goldfinger N, Borresen-Dale AL, Ginsberg D, Harris CC, Pilpel Y, Oren M, Rotter V: **p53-Repressed miRNAs are involved with E2F in a feed-forward loop promoting proliferation.** *Mol Syst Biol* 2008, **4**:231.
54. Osella M, Bosia C, Corá D, Caselle M: **The role of incoherent microRNA-mediated feedforward loops in noise buffering.** *arXiv.1004.0336v1* .
55. Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J, Calabrese JM, Dennis LM, Volkert TL, Gupta S, Love J, Hannett N, Sharp PA, Bartel DP, Jaenisch R, Young RA: **Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells.** *Cell* 2008, **134**:521-33.
56. Lin CH, Jackson AL, Guo J, Linsley PS, Eisenman RN: **Myc-regulated microRNAs attenuate embryonic stem cell differentiation.** *EMBO J* 2009, **28**:3157-70.

doi:10.1186/1471-2105-11-435

Cite this article as: Friard et al.: CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. *BMC Bioinformatics* 2010 **11**:435.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

