W416–W423 *Nucleic Acids Research, 2011, Vol. 39, Web Server issue* *Published online 10 May 2011*
doi:10.1093/nar/gkr276

mirConnX: condition-specific mRNA-microRNA network integrator

Grace T. Huang[1,2,3], Charalambos Athanassiou[2] and Panayiotis V. Benos[1,2,3,*]

[1]Joint CMU-Pitt PhD Program in Computational Biology, [2]Department of Computational and Systems Biology, and [3]Clinical and Translational Science Institute, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Received February 27, 2011; Revised April 4, 2011; Accepted April 12, 2011

ABSTRACT

mirConnX is a user-friendly web interface for inferring, displaying and parsing mRNA and microRNA (miRNA) gene regulatory networks. mirConnX combines sequence information with gene expression data analysis to create a disease-specific, genome-wide regulatory network. A prior, static network has been constructed for all human and mouse genes. It consists of computationally predicted transcription factor (TF)-gene associations and miRNA target predictions. The prior network is supplemented with known interactions from the literature. Dynamic TF- and miRNA-gene associations are inferred from user-provided expression data using an association measure of choice. The static and dynamic networks are then combined using an integration function with user-specified weights. Visualization of the network and subsequent analysis are provided via a very responsive graphic user interface. Two organisms are currently supported: *Homo sapiens* and *Mus musculus*. The intuitive user interface and large database make mirConnX a useful tool for clinical scientists for hypothesis generation and explorations. mirConnX is freely available for academic use at http://www.benoslab.pitt.edu/mirconnx.

INTRODUCTION

Since its discovery two decades ago, it has become increasingly clear that microRNAs (miRNAs) play a crucial role in modulating gene expression at the post-transcriptional level. The small, 22-nt long RNA molecules fine-tune gene expression by base pairing to target messenger RNAs, resulting in its degradation or causing translational repression. As Pandit *et al.* (1) has shown, deregulation of even a single miRNA may cause complex human diseases. Regulatory network reconstruction methods have traditionally involved transcriptional regulation only. Incorporating miRNAs thus becomes the next natural step. Only few tools have explored ways to associate mRNA and miRNA expression to infer regulations. MMIA (2) and MAGIA (3), for example, utilize association metrics such as correlation and mutual information. In a different context, Huang *et al.* (4) employed a Bayesian model to identify miRNA targets from sequence features and expression data. However, there are several limitations to these tools. MMIA only examines a subset of miRNAs that are significantly up- or downregulated, and omits those that could potentially be significantly correlated with their targets if they are not considered to be differentially expressed, based on the specific threshold. This only limits the data to those with a control/disease contrast, excluding possible use of time-series data. GenMir++ (4) is a more sophisticated algorithm, but it becomes computationally inefficient when a large number of genes is considered. Furthermore, it does not take into account other supporting information such as the transcriptional regulation. In fact, none of these tools incorporates the full set of transcription factors (TFs) in the global network construction. Additionally, network motifs such as feed-back and feed-forward loops that are known to have an important role in cancer development and other diseases are usually not identified as part of the routine analyses of the currently available tools.

To this end, we developed mirConnX to attempt to address some of the above concerns. mirConnX (http://www.benoslab.pitt.edu/mirconnx) takes advantage of prior knowledge (from sequence data), and incorporates evidence from gene expression data to create condition-specific genome-wide regulatory networks. mirConnX also aims to identify gene network motifs, involving transcription factors and miRNAs, that are associated with

*To whom correspondence should be addressed. Tel: +1 412 6483315; Fax: +1 412 6483163; Email: benos@pitt.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

the corresponding diseases, pathogenesis or phenotype of interest.

TOOL DESCRIPTION

General framework

mirConnX aims to provide an integrated environment that allows the user to infer genome-wide transcriptional (TF-gene/miRNA) and post-transcriptional (miRNA-gene/TF) regulatory networks for a particular disease or condition. We consider mRNA and miRNA expression data measured under the same set of conditions, or at the same time points, or from the same corresponding diseased or normal samples (matching samples). The mRNA and miRNA expression data are pre-processed to remove genes that are lowly expressed with limited variance overall. Then, we connect TFs and miRNAs to genes using a statistical association measure. The *association network* that is constructed reflects the disease status or the condition of interest. This network is an undirected graph, in which an edge exists between two nodes (genes) if an interaction has been detected. Note that such association networks do not discriminate between direct and indirect interactions. This network is then superimposed to a pre-compiled, species-specific *prior network*, which is derived from TF motif scanning and binding, miRNA target predictions and the literature evidence. The prior network is a directed, weighted graph, in which an edge between a TF or miRNA and a gene exists if the former is predicted to regulate the latter. All connections in the prior network correspond to direct, predicted or verified interactions. Superimposing the two networks via an integration function results in a directed network, which is expected to contain significantly fewer indirect interactions (depending on the weight the user assigns on the prior network). mirConnX web tool allows easy visualization and exploration of the network, and identifies network motifs. In the following sections, we describe the construction of the context-dependent (dynamic) association network, the construction of the prior (static) network and their integration. Figure 1 presents an overview of the mirConnX pipeline.

Input formats

mirConnX accepts normalized mRNA and miRNA expression data in tab-delimited files, where the first row contains sample IDs and the first column contains mRNA or miRNA IDs. mirConnX supports gene symbols, Ensembl Gene ID, Ensembl Transcript ID, Entrez Gene ID, RefSeq DNA ID and Unigene ID as mRNA identifiers and miRBase miRNA ID and Accession numbers as miRNA identifiers. An example of the matching mRNA–miRNA data sets can be found and pre-loaded on the front page. Note that the sample IDs for mRNA and miRNA data should match. Any unmatched samples are discarded. mirConnX allows multiple columns with the same header in case of biological or technical replicates. The input data sets are stored only during a user's session and are used to construct the association network. If no miRNA data file is included, the resulting network will show only TF-gene interactions. We currently support two organisms: human (*Homo sapiens*) and mouse (*Mus musculus*), as genome annotation and prior information is most abundant for these species.

Constructing the prior network from sequence and literature data

The prior network is constructed by combining all predictions of TF to gene and TF to miRNA interactions and all miRNA target predictions. The network is then enhanced by literature evidence that confirms the existence of an edge. This results in a directed network that represents the collection of prior knowledge on regulatory potentials between genes.

TF to gene/miRNA regulations. We define the binding potential (Rg_{TF}) of a promoter sequence for a given gene/miRNA as the maximum score between literature evidence ($S_{\mathrm{LIT}} \in \{0,1\}$) and TF binding score ($S_{\mathrm{TF}}$), which is calculated using a sliding window method (5) on the promoters of genes and miRNAs. The JASPAR (6) and TRANSFAC (7) position weight matrices (PWMs) are used for the scanning. A subsequence is considered as a binding site for a TF if its PWM score is on the top 1% of all scores for this PWM. In addition, UCSC Regulation Track Conserved TFBS Scores (S_{CONS}) are added to enhance confidence. The sum of S_{TF} and S_{CONS} is normalized to a score between 0 and 1. Finally, if an experimentally verified binding motif for a given TF is available for this promoter (e.g. in TRANSFAC), then S_{LIT} becomes 1, and so does Rg_{TF}:

$$Rg_{\mathrm{TF}} = \max\left\{(|S_{\mathrm{TF}}|+|S_{\mathrm{CONS}}|), S_{\mathrm{LIT}}\right\}$$

Regular gene promoters were defined as the region 5 kb upstream of TSS. Promoters are obtained from Database of Transcription Start Sites (DBTSS) (8), The Eukaryotic Promoter Database (EPD) (9) and UCSC genome browser Regulation–Transcription track (Eponine and SwitchGear TSS). miRNA TSSs are defined using a combination of predictions and experiments from CoreBoost_HM (10), Marson *et al.* (11) and Corcoran *et al.* (12). Human (NCBI36/Hg18) and mouse (NCBI37/mm9) sequence data were downloaded from UCSC genome browser (13).

miRNA to gene/TF regulations. miRNA target prediction algorithms generally do not agree very well. Thus, we used a combination of five target prediction algorithms that take into account the seed sequence, flanking sequences and context, binding energy and conservation. These algorithms are: PITA (14), miRANDA (15), TargetScan 5.0 (16), RNAhybrid (17) and Pictar (18). We define the regulatory potential (Rg_{miR}) of an miRNA for a gene as the proportion of the target prediction algorithms predicting the gene to contain at least one miRNA target site. If predictions for corresponding genome versions are not available, we ran the algorithms using default parameters and cutoffs. In addition, if the 3′-UTR of a gene contains an experimentally verified site from TarBase (19) or miRecords (20), then S_{TarBase} or $S_{\mathrm{miRecords}}$

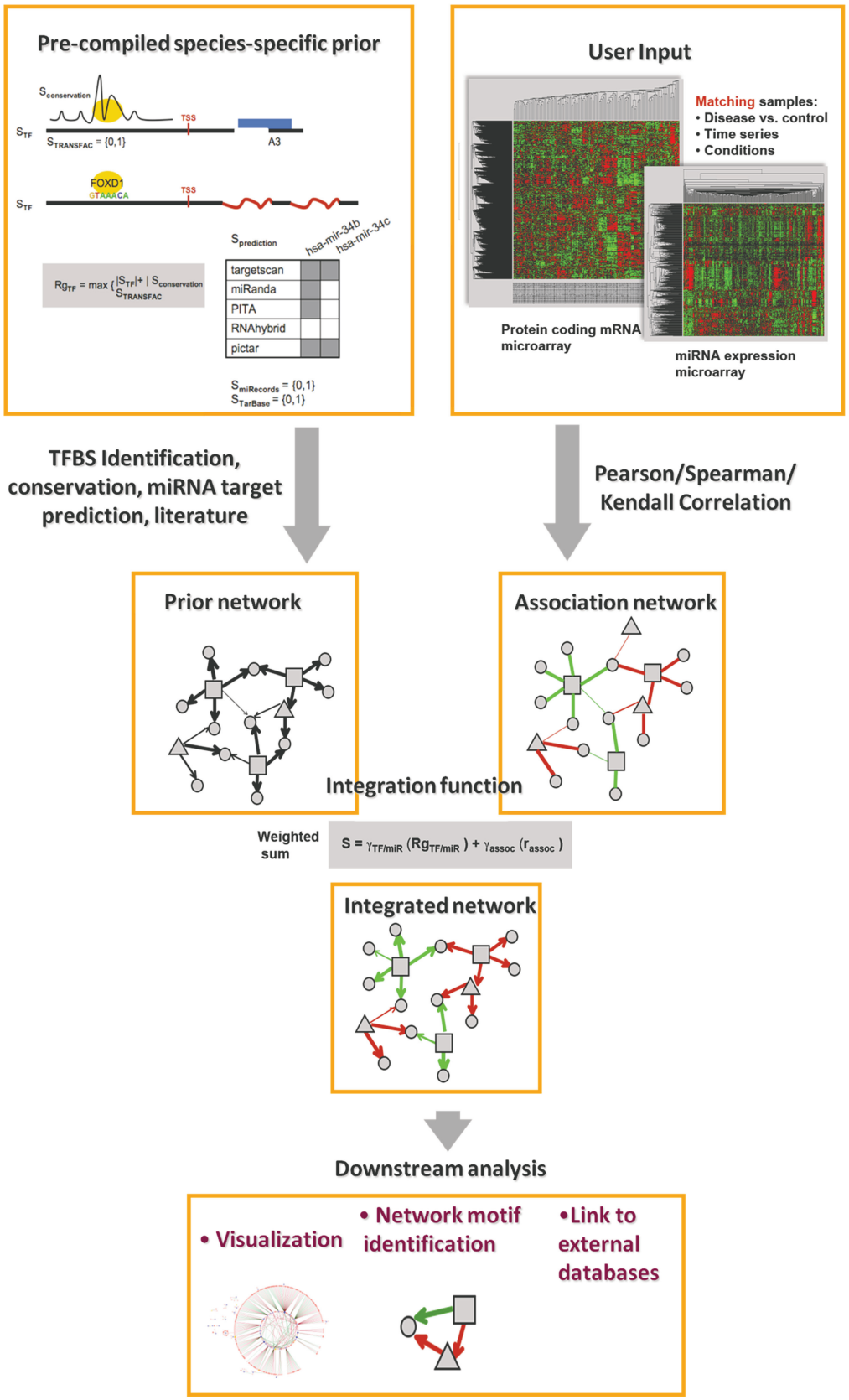

Figure 1. Overview of the integrated analysis in mirConnX. Visualization is achieved with the use of Cytoscape Web v0.7.2 (25). Feed-forward loops are displayed on a separate tab. Links to external databases provided for every coding gene or miRNA. Statistical analysis on GO pathway (45) overrepresentation in a user-selected set of genes will be added soon.

score becomes 1 (otherwise is 0), and so does the regulatory potential of the gene for a given miRNA:

$$Rg_{miR} = \max\{|S_{prediction}|, S_{TarBase}, S_{miRecords}\}$$

Human and mouse 3′-UTRs were downloaded from UCSC genome browser. The list of mature and complement miRNAs, as well as their sequences, were obtained from miRBase v.14 (21).

Gene expression pre-processing

Standard gene symbol and miRNA ID are used as our primary identifier. Genes and miRNAs with multiple probes on the array, or those converted to the same gene symbol/miRNA ID, are collapsed into a single medium value. The normalized mRNA and miRNA expression data are pre-processed using three filters for low (i) absolute expression, (ii) variance and (iii) entropy. A cutoff of 5% is used for mRNA and miRNA expression data individually to remove data that are not likely to be important for the network. A list of the genes filtered and excluded from the analysis is available for user to download. Finally, all matching conditions or samples between the mRNA and miRNA data matrices are retained for analysis. In case of multiple replicates for the same condition, the median value between replicates is used.

Constructing the association network from gene expression data

We construct an association network from the user-supplied expression data by measuring the strength of all pair-wise interactions between TFs, miRNAs and genes across the samples/replicates. A number of parametric and non-parametric association metrics is available to the user for defining these interactions. Correlation coefficient is one of the most intuitive, and most well received. The different flavors of correlation (Pearson, Spearman and Kendall) have been used successfully in the past and achieved different levels of success, for example in the WGCNA R package (22). Pearson correlation coefficient is often used when a linear dependence between the variables exists. In contrast, Spearman ρ correlation coefficient applies Pearson formula on the ranks of the values of the two variables and can detect similarities even if non-linear (but monotonic) association exists. Kendall τ rank correlation coefficient also operates on the ranks, but it calculates the probability of concordance or discordance of any pair of observations. In general, Spearman and Kendall give similar results, but they differ on the magnitude [for more details on correlation measures, see Ref. (23)]. We implemented these three correlation measures and applied them on pairs of gene, TF or miRNA expression values across matching conditions. The absolute magnitude reflects the level of correlation, and the sign suggests positive or negative interaction. Mutual information is a non-parametric test that has been implemented in algorithms such as ARACNE (24) as the measure of an association for genome-wide pairwise interactions. Mutual information is non-negative, and as such it does not provide information about the sign of interaction. Furthermore, it is generally computationally intensive and sample-size sensitive, since it requires an estimation of marginal and joint probabilities of the variables. For all these reasons, we are not currently implementing the mutual information in miRconnX. The degree of association, r_{assoc}, is defined as the probability that two genes are correlated. We used the inverse of correlation coefficient significance $(1-p)$ as the probability of non-random association. The use of significance, instead of the coefficient itself, takes into account of the sample size and allows a fair comparison between networks generated by different sizes of data.

Network integration

Integration of the prior confidence of association (based on sequence data, literature evidence and predictions) and the correlation network (based on the gene expression data) is currently done via a simple weighted sum function (S).

$$S = \gamma_{\text{prior}}(Rg_{\text{TF}}, Rg_{\text{miR}}) + \gamma_{\text{assoc}}(r_{\text{assoc}})$$

In this equation, γ_{prior} is a user-defined parameter between 0 and 1, and $\gamma_{\text{assoc}} = 1-\gamma_{\text{prior}}$. The default for γ_{prior} is 0.3 (i.e. 30%) and a value <0.5 is recommended for the prior information. The user can also define a cutoff for the combined regulation score, S. This is also a number between 0.0 and 1.0. The higher the S the fewer connections will be reported. A value of 0.7−0.99 is recommended. Finally, for practical purposes, we cap the number of interactions to be displayed on screen at 3000, as beyond that the network becomes too large to be efficiently visualized.

Submission and wait time

Depending on the size of the files (number of genes analyzed) and types of analysis chosen, the analysis could take anywhere from minutes to up to an hour. As an example, for 20 000 genes and 500 miRNA, the computing time is roughly 15 min using Pearson correlation. While the job is running, an execution log will be displayed. The user can close the browser window. When the job finishes, the user will receive an email notification and retrieve the results from the link provided.

mirConnX output

Following the link to the result, a visualization of the network is displayed, as shown in Figure 2. Cytoscape Web v 0.7.2 (25) is used for network display. The rendering time for a network with 1500 nodes and 2500 connections is about 15 s. Once uploaded, browsing the various areas of the network is instantaneous. Users can use the tools at the bottom right corner to zoom in/out and edit node placements on the visualization page, and output the visualization as graphics. The list of interactions is also displayed with links to external databases such as miRBase and Entrez Gene (26) for annotation, PubFocus (27), EBIMed (28), miR2Disease (29) and miRo (30) to facilitate clinical research by sifting through a large body of literature and records, as well as Gene Ontology (31) terms

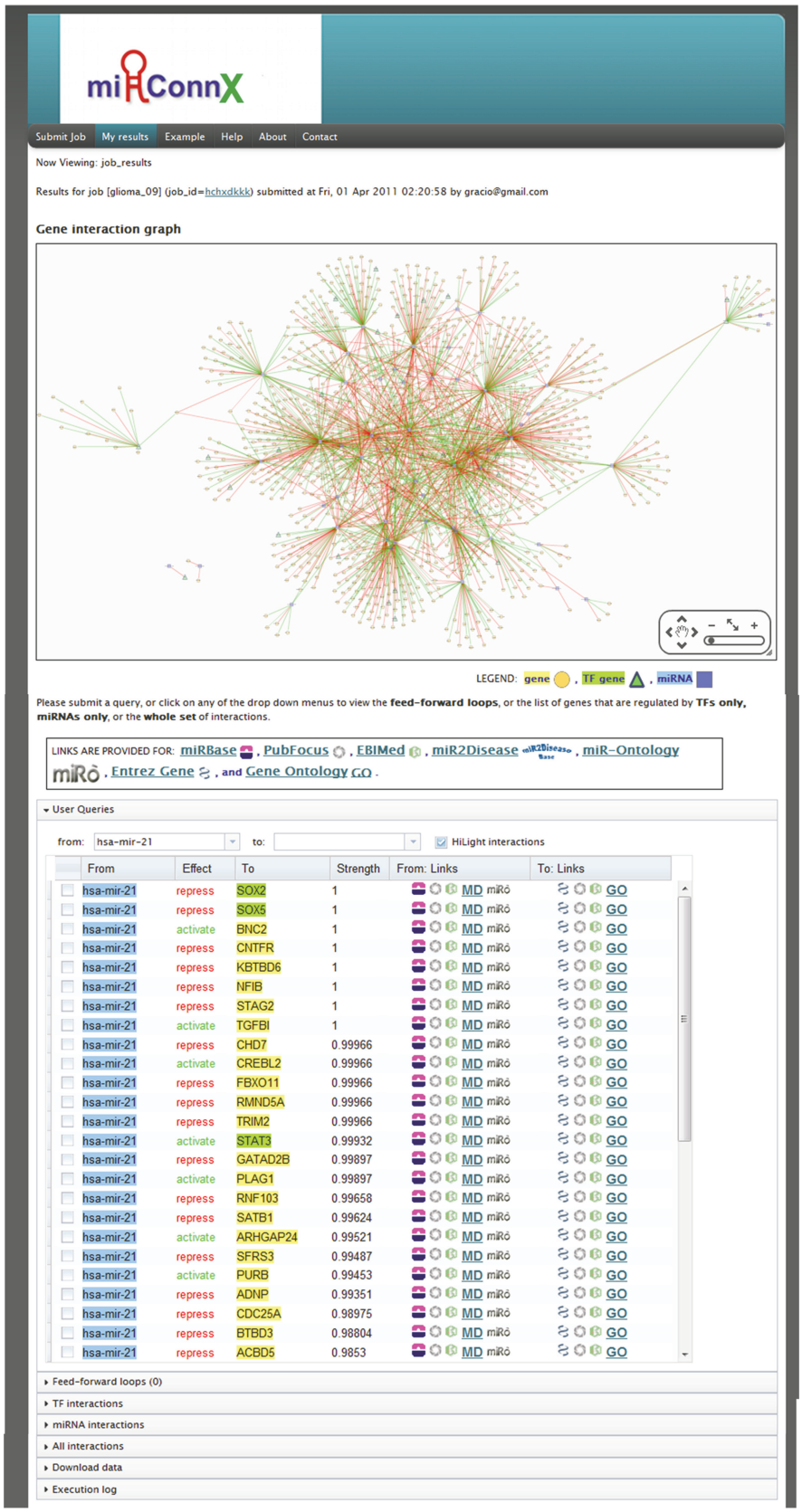

Figure 2. Snapshot of the glioblastoma case study output. An example search for the downstream targets of miR-21, a key player in glioblastoma development, is shown in the middle.

for each gene. We also make available for download: (i) list of interactions above the user-defined display cutoff, ranked by regulatory score in tab-delimited text file; (ii) list of nodes, ranked by degree centrality in text-delimited text file; and (iii) the network in pdf or GRAPHML graph formats compatible with Cytoscape for further exploration. The user can (iv) search for a particular node and its targets/regulators (through the 'List of gene interactions: filtered' drop down menu), a set of particular interaction and highlight or select the corresponding nodes and edges on the graph display. Finally, we display all (v) feed-forward loops and their neighbors at the given threshold. In addition, a summary of statistics, including the actual number of TFs, miRNAs and genes can be retrieved under 'execution log'.

CASE STUDY

The main idea behind mirConnX was first used to analyze lung epithelial gene expression data a few years ago (1,12). In that study, we were able to identify a feed-forward loop that included SMAD TFs, let-7 d and HMGA2 gene, which was central in the regulation of epithelial to mesenchymal transition (EMT). Furthermore, we later found that knocking down of let-7 d in the trachea of mice can cause lung fibrosis few days later (1).

Here, we present a case study that demonstrates the utility of mirConnX. We downloaded a set of publicly available mRNA and miRNA expression profiles from The Cancer Genome Atlas (TCGA) pilot project (http://cancergenome.nih.gov/), where a large compendium of tumor and normal glioblastoma multiforme (GBM; primary brain tumor) expression data is available. The choice of this disease is two-fold: in this repository, GBM is one of the two diseases with both tumor and normal cells. Furthermore, recent studies have revealed distinct patterns of miRNA expression in tumor compared to normal brain (32) and several miRNA targets have in fact been experimentally verified (33–35). The disease samples are characterized by rapid proliferation and stem-cell like behavior that is possibly caused by malfunctioning of characteristic pathways (36). Mutations in miRNA and miRNA targets have been postulated to be involved in tumorigenesis, but have not been specifically identified in GBM.

The expression profiles downloaded consist of a total of 58 matched mRNA and miRNA samples from the Agilent 244 k aCGH platform at data level 3. We used the following parameters on mirConnX: Gene Symbols, miRBase ID, Pearson correlation with a prior weight of 0.3 and 0.9 as the display cutoff threshold. A total of 56 miRNAs, 29 TFs and 1180 genes form a network with a total of 1851 connections. Of these interactions, 43 are miR-TF regulations, 34 TF-gene connections and 1774 miRNA–gene connections.

Among the top interactions, we were able to identify four hubs: miR-21, miR-326, miR-34, and miR-137, which have been verified to be miRNAs involved in Gliobastoma. These four miRNAs are also hubs with some of the highest degree centrality (37), sharing many targets and TFs with other hubs. Among them, miR-21 has been found to be one of the most highly expressed miRNAs in many cancer types, and it has been shown that miR-21 acts as an oncogen in glioblastoma by suppressing apoptosis (38). Among the highest ranking targets we predict for miR-21, SOX2 (39) and TGFB pathway (40) were shown to be regulated by the miRNA. RECK and PDCD4 have been experimentally verified, *in vivo* and *in vitro*, to be involved in proliferation (34,41). In addition, PELI1 and CDC25A have been shown in other cancer types to play a role in apoptosis (38,42,43). Similarly, miR-137 has been shown to be involved in proliferation and neuronal differentiation *in vitro* (44). Indeed, both CDK6 and MITF, the experimentally verified targets from the study were also predicted in our network.

A thorough literature search on all of the predicted interactions for Gliobastoma is not possible here, but we demonstrated that mirConnX is useful for identifying hub genes, their regulators, and their targets involved in diseases, the pathways involved and could potentially be a powerful tool for clinical scientists to create a list of top candidate genes and forming hypotheses.

CONCLUSIONS

In recent years, with the availability of condition-specific high-throughput mRNA and miRNA expression data, there is an increasing need of an integrated environment that combines data analyses and visualization to construct hypothesized networks. While many methods exist for network generation using only expression data, only binding affinity experiments such as ChIP-chip, or even manually curated data from expert knowledge databases, an integrated network that maximally exploits information in both domains is lacking. Additionally, there has not been many attempts to incorporate both TF and miRNA regulations, yet it has become increasingly clear that miRNAs play a crucial role in human diseases. mirConnX is a novel web tool developed specifically to fill the niche. The utility of mirConnX lies in its ability to integrate user-supplied data with pre-compiled information of miRNA targeting and TF binding, and generate a network that reflects characteristics specific to the data guided by some prior beliefs. The user-friendly display of interaction networks and other downstream analyses also provides an integrated environment for clinical researchers to perform further investigation and exploration.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Naftali Kaminski and his group for beta-testing the web site and for helpful suggestions. We also like to thank the people who developed Cytoscape Web and made it available. Finally, the contribution of two anonymous reviewers in the improvement of the manuscript and the web site is gratefully acknowledged.

FUNDING

National Institutes of Health R01 (grants LM007994 and LM009657); National Institutes of Health UL1, Clinical and Translational Science Institute, University of Pittsburgh (grant RR024153 to G.T.H.). Funding for open access charge: National Institutes of Health R01 (grants LM007994, LM009657 and UL1 grant RR024153).

Conflict of interest statement. None declared.

REFERENCES

1. Pandit,K.V., Corcoran,D., Yousef,H., Yarlagadda,M., Tzouvelekis,A., Gibson,K.F., Konishi,K., Yousem,S.A., Singh,M., Handley,D. *et al*. (2010) Inhibition and role of let-7 d in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.*, **182**, 220–229.
2. Nam,S., Li,M., Choi,K., Balch,C., Kim,S. and Nephew,K.P. (2009) MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression. *Nucleic Acids Res.*, **37**, W356–W362.
3. Sales,G., Coppe,A., Bisognin,A., Biasiolo,M., Bortoluzzi,S. and Romualdi,C. (2010) MAGIA, a web-based tool for miRNA and genes integrated analysis. *Nucleic Acids Res.*, **38**, W352–W359.
4. Huang,J.C., Morris,Q.D. and Frey,B.J. (2007) Bayesian inference of microRNA targets from sequence and expression data. *J. Comput. Biol.*, **14**, 550–563.
5. Stormo,G.D. (2000) DNA binding sites: representation and discovery. *Bioinformatics*, **16**, 16–23.
6. Portales-Casamar,E., Thongjuea,S., Kwon,A.T., Arenillas,D., Zhao,X., Valen,E., Yusuf,D., Lenhard,B., Wasserman,W.W. and Sandelin,A. (2010) JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **38**, D105–D110.
7. Wingender,E. (2004) TRANSFAC, TRANSPATH and CYTOMER as starting points for an ontology of regulatory networks. *In Silico Biol.*, **4**, 55–61.
8. Yamashita,R., Wakaguri,H., Sugano,S., Suzuki,Y. and Nakai,K. (2010) DBTSS provides a tissue specific dynamic view of transcription start sites. *Nucleic Acids Res.*, **38**, D98–D104.
9. Schmid,C.D., Perier,R., Praz,V. and Bucher,P. (2006) EPD in its twentieth year: towards complete promoter coverage of selected model organisms. *Nucleic Acids Res.*, **34**, D82–D85.
10. Wang,X., Xuan,Z., Zhao,X., Li,Y. and Zhang,M.Q. (2009) High-resolution human core-promoter prediction with CoreBoost_HM. *Genome Res.*, **19**, 266–275.
11. Marson,A., Levine,S.S., Cole,M.F., Frampton,G.M., Brambrink,T., Johnstone,S., Guenther,M.G., Johnston,W.K., Wernig,M., Newman,J. *et al*. (2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell*, **134**, 521–533.
12. Corcoran,D.L., Pandit,K.V., Gordon,B., Bhattacharjee,A., Kaminski,N. and Benos,P.V. (2009) Features of mammalian microRNA promoters emerge from polymerase II chromatin immunoprecipitation data. *PLoS ONE*, **4**, e5279.
13. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
14. Kertesz,M., Iovino,N., Unnerstall,U., Gaul,U. and Segal,E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
15. Enright,A.J., John,B., Gaul,U., Tuschl,T., Sander,C. and Marks,D.S. (2003) MicroRNA targets in Drosophila. *Genome Biol.*, **5**, R1.
16. Friedman,R.C., Farh,K.K., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
17. Kruger,J. and Rehmsmeier,M. (2006) RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.*, **34**, W451–W454.
18. Krek,A., Grun,D., Poy,M.N., Wolf,R., Rosenberg,L., Epstein,E.J., MacMenamin,P., da Piedade,I., Gunsalus,K.C., Stoffel,M. *et al*. (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.
19. Papadopoulos,G.L., Reczko,M., Simossis,V.A., Sethupathy,P. and Hatzigeorgiou,A.G. (2009) The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Res.*, **37**, D155–D158.
20. Xiao,F., Zuo,Z., Cai,G., Kang,S., Gao,X. and Li,T. (2009) miRecords: an integrated resource for microRNA–target interactions. *Nucleic Acids Res.*, **37**, D105–D110.
21. Griffiths-Jones,S., Saini,H.K., van Dongen,S. and Enright,A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.
22. Langfelder,P. and Horvath,S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.
23. Sokal,R.R. and Rohlf,F.J. (1995) *Biometry: the Principles and Practice of Statistics in Biological Research*, 3rd edn. W. H. Freeman and Co., New York.
24. Margolin,A.A., Nemenman,I., Basso,K., Wiggins,C., Stolovitzky,G., Dalla Favera,R. and Califano,A. (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, **7(Suppl. 1)**, S7.
25. Lopes,C.T., Franz,M., Kazi,F., Donaldson,S.L., Morris,Q. and Bader,G.D. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.
26. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2011) Entrez gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **39**, D52–D57.
27. Plikus,M.V., Zhang,Z. and Chuong,C.M. (2006) PubFocus: semantic MEDLINE/PubMed citations analytics through integration of controlled biomedical dictionaries and ranking algorithm. *BMC Bioinformatics*, **7**, 424.
28. Rebholz-Schuhmann,D., Kirsch,H., Arregui,M., Gaudan,S., Riethoven,M. and Stoehr,P. (2007) EBIMed—text crunching to gather facts for proteins from Medline. *Bioinformatics*, **23**, e237–e244.
29. Jiang,Q., Wang,Y., Hao,Y., Juan,L., Teng,M., Zhang,X., Li,M., Wang,G. and Liu,Y. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, **37**, D98–D104.
30. Lagana,A., Forte,S., Giudice,A., Arena,M.R., Puglisi,P.L., Giugno,R., Pulvirenti,A., Shasha,D. and Ferro,A. (2009) miRo: a miRNA knowledge base. *Database*, **2009**, bap008.
31. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al*. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
32. Dong,H., Siu,H., Luo,L., Fang,X., Jin,L. and Xiong,M. (2010) Investigation gene and microRNA expression in glioblastoma. *BMC Genomics*, **11(Suppl. 3)**, S16.
33. Corsten,M.F., Miranda,R., Kasmieh,R., Krichevsky,A.M., Weissleder,R. and Shah,K. (2007) MicroRNA-21 knockdown disrupts glioma growth *in vivo* and displays synergistic cytotoxicity with neural precursor cell delivered S-TRAIL in human gliomas. *Cancer Res.*, **67**, 8994–9000.
34. Gabriely,G., Wurdinger,T., Kesari,S., Esau,C.C., Burchard,J., Linsley,P.S. and Krichevsky,A.M. (2008) MicroRNA 21 promotes glioma invasion by targeting matrix metalloproteinase regulators. *Mol. Cell Biol.*, **28**, 5369–5380.
35. Li,Y., Guessous,F., Zhang,Y., Dipierro,C., Kefas,B., Johnson,E., Marcinkiewicz,L., Jiang,J., Yang,Y., Schmittgen,T.D. *et al*. (2009) MicroRNA-34 a inhibits glioblastoma growth by targeting multiple oncogenes. *Cancer Res.*, **69**, 7569–7576.
36. Furnari,F.B., Fenton,T., Bachoo,R.M., Mukasa,A., Stommel,J.M., Stegh,A., Hahn,W.C., Ligon,K.L., Louis,D.N., Brennan,C. *et al*. (2007) Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes Dev.*, **21**, 2683–2710.
37. Ulrik,B. and Thomas,E. (2005) *Network Analysis*. Springer-Verlag Berlin Heidelberg, Germany.

38. Chan,J.A., Krichevsky,A.M. and Kosik,K.S. (2005) MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res.*, **65**, 6029–6033.
39. Fang,X., Yoon,J.G., Li,L., Yu,W., Shao,J., Hua,D., Zheng,S., Hood,L., Goodlett,D.R., Foltz,G. *et al.* (2011) The SOX2 response program in glioblastoma multiforme: an integrated ChIP-seq, expression microarray, and microRNA analysis. *BMC Genomics*, **12**, 11.
40. Papagiannakopoulos,T., Shapiro,A. and Kosik,K.S. (2008) MicroRNA-21 targets a network of key tumor-suppressive pathways in glioblastoma cells. *Cancer Res.*, **68**, 8164–8172.
41. Chen,Y., Liu,W., Chao,T., Zhang,Y., Yan,X., Gong,Y., Qiang,B., Yuan,J., Sun,M. and Peng,X. (2008) MicroRNA-21 down-regulates the expression of tumor suppressor PDCD4 in human glioblastoma cell T98G. *Cancer Lett.*, **272**, 197–205.
42. Marquez,R.T., Wendlandt,E., Galle,C.S., Keck,K. and McCaffrey,A.P. (2010) MicroRNA-21 is upregulated during the proliferative phase of liver regeneration, targets Pellino-1, and inhibits NF-kappaB signaling. *Am. J. Physiol. Gastrointest. Liver Physiol.*, **298**, G535–G541.
43. Wang,P., Zou,F., Zhang,X., Li,H., Dulak,A., Tomko,R.J. Jr, Lazo,J.S., Wang,Z., Zhang,L. and Yu,J. (2009) microRNA-21 negatively regulates Cdc25A and cell cycle progression in colon cancer cells. *Cancer Res.*, **69**, 8157–8165.
44. Silber,J., Lim,D.A., Petritsch,C., Persson,A.I., Maunakea,A.K., Yu,M., Vandenberg,S.R., Ginzinger,D.G., James,C.D., Costello,J.F. *et al.* (2008) miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells. *BMC Med.*, **6**, 14.
45. Harris,M.A., Clark,J., Ireland,A., Lomax,J., Ashburner,M., Foulger,R., Eilbeck,K., Lewis,S., Marshall,B., Mungall,C. *et al.* (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.*, **32**, D258–D261.