

# Human microRNAs co-silence in well-separated groups and have different predicted essentialities

Gábor Boross<sup>1,2</sup>, Katalin Orosz<sup>1,2</sup> and Illés J. Farkas<sup>2,\*</sup>

<sup>1</sup>Department of Biological Physics, Eötvös Loránd University,

<sup>2</sup>Statistical and Biological Physics Research Group and CellCom RET at the Hung. Acad. of Sci.,  
Pázmány P. stny. 1A, H-1117 Budapest, Hungary

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXXX

## ABSTRACT

**Background:** Short regulating RNAs guide many cellular processes. Compared to transcription factor proteins they appear to provide more specialised control and their deletions are less frequently lethal.

**Results:** We find large differences between computationally predicted lists of human microRNA-target pairs. Instead of integrating these lists we use the two most accurate of them. Next, we construct the co-regulation network of human microRNAs (miRNAs) as nodes by computing the correlation (link weight) between the gene silencing scores of individual miRNAs. In this network we locate groups of tightly co-regulating nodes (modules). Despite explicitly allowing overlaps the co-regulation modules of miRNAs are well separated. We use the modules and miRNA co-expression data to define and compute miRNA essentiality. Instead of focusing on particular biological functions we identify a miRNA as essential, if it has a low co-expression with the miRNAs in its module. This may be thought of as having many workers performing the same tasks together in one place (non-essential miRNAs) as opposed to a single worker performing those tasks alone (essential miRNA).

**Conclusions:** On the system level we quantitatively confirm previous findings about the specialised control provided by miRNAs. For knock-out tests we list the groups of our predicted most and least essential miRNAs. In addition, we provide possible explanations for (i) the low number of individually essential miRNAs in *C. elegans* (Miska *et al.*, 2007) and (ii) the high number of ubiquitous miRNAs influencing cell and tissue-specific miRNA expression patterns (Landgraf *et al.*, 2007) in mouse and human.

**Contact:** fij@elte.hu

## 1 INTRODUCTION

RNA silencing was first observed in modified strains of petunia where over-expressed enzymes did not accelerate but rather decelerated the pathways they catalysed (van der Krol *et al.*, 1990; Napoli *et al.*, 1990). Subsequent studies, in both uni- and multicellular organisms, have identified a number of 20–22bp long regulating RNAs with diverse origins (Bartel *et al.*, 2004). The precursors of these non-coding RNAs can be, for example, transcribed in separate 1–2 kbp units, extracted from eukaryotic introns or inserted into the host cell by viruses. Based on their biogenesis and

in vivo observations, short regulating RNAs have been categorised as microRNAs (miRNAs), short interfering RNAs (siRNAs), Piwi-interacting RNAs (piRNAs), etc. Here we investigate the translational silencing interactions of human miRNAs.

In metazoans miRNAs have a widespread effect on mRNA and protein levels: at least hundreds of them control differentiation during development as well as other tissue-specific functions (Farh *et al.*, 2005). The bulk of this impact is due to the destabilisation and cleavage of messenger RNAs and translational repression. It is estimated that in vertebrates RNAs coding for proteins account for as little as 1.5% of all transcribed RNA, a huge drop compared to over 75% in prokaryotes (Mattick and Gagen, 2001). The largely expanded warehouse of non-coding RNAs has resulted in organisms of previously unseen complexity (Heimberg *et al.*, 2008). To describe biological and other systems of such complexity, networks (graphs) have proven to be a highly useful tool: they map interactors, e.g., miRNAs or proteins, and their interactions to nodes and links. We map a protein, its mRNA and its gene to a node, and represent a silencing interaction as a directed link from a miRNA to a target gene.

A central goal of large-scale biomolecular interaction studies is to connect the static structure of the interaction map(s) with dynamical properties of the underlying biological system (Barabási and Oltvai, 2004; Warner *et al.*, 2006). Well-known examples for this connection include graph motifs, often performing basic signal processing tasks (Alon, 2007), and functional modules (Hartwell *et al.*, 1999). Each of these groups (motifs, modules) is a set of molecules densely linked to each other by interactions and carrying out specific biological tasks together. Overlaps between such groups show that multiple processes use the same molecules (the shared nodes), while well-separated groups indicate that jobs are managed independently. Here we search for modules in the network of co-regulating miRNAs. In this network two regulators (nodes) are connected, if they share at least one target and the strength of a co-regulation link is computed from the silencing efficiencies of the two connected miRNAs. miRNAs participating in the same module of the co-regulation network co-silence many of their targets.

After comparing the lists of human miRNA - target gene interactions from four computational databases we find that even though the four interaction lists are highly different, their similarity (relative overlap) moderately increases with growing sample size. Experimental benchmarks sufficient for calibrating the prediction quality

\*to whom correspondence should be addressed

of each data set are not yet available. Two recent studies, however, found that mRNA and protein level fold changes due to miRNA control are best predicted by TargetScan and PicTar (Baek *et al.*, 2008; Selbach *et al.*, 2008). Based on a list of unified miRNA - target gene interaction scores derived from TargetScan data we compute the co-regulation network of miRNAs. Next, we search for modules – groups of tightly co-regulating miRNAs – in this network. Despite explicitly allowing for overlaps between the modules we find that most nodes belong to a single module. We conclude that the co-silencing groups of human miRNAs work with little cross-talk. Following the functional analysis of two selected modules we define and compute the essentiality of miRNAs. A miRNA co-expressed with the miRNAs of the same module can be easily replaced by another member of the same module in most of its regulatory roles. On the other hand, miRNAs not co-expressed with their module partners may not be easily replaced and are therefore more essential.

## 2 METHODS

### 2.1 miRNA-target gene interaction scores

First we compared human miRNA - human target interactions from one manually curated and four computational data sources. Then, for our analyses we used data from TargetScan and – as a control – from PicTar (see Figs. 4 and 5 and the Supplementary Material for controls). The five data sources were TarBase (as provided in a filtered form under “known targets” by miRBase in June 2008) (Sethupathy *et al.*, 2006), miRBase (version 5) (Griffiths-Jones *et al.*, 2008), PicTar (vertebrates: “conservation in mammals”, Dec. 2007) (Lall *et al.*, 2006), PITA (top: “3-15”, Nov. 2007) (Kertesz *et al.*, 2007) and TargetScan v4.1 (conserved and non-conserved sites) (Lewis *et al.*, 2003). TarBase provides a manually collected list of experimentally verified interactions, while the four computational data sets (i) provide a score for each predicted miRNA - target (transcript, protein or gene) link quantifying the efficiency of silencing and (ii) apply a lower cutoff score (a threshold) below which they discard all links.

We consider only human miRNAs and omit the *hsa*- prefix from each name. In PicTar and PITA target transcripts are identified by RefSeq mRNA IDs, miRBase contains Ensembl transcript IDs, while TargetScan and TarBase contain gene/protein names. We mapped all target names to Ensembl gene IDs. In each of the four computationally predicted lists and for each miRNA - target gene pair ( $M, G$ ) we calculated a single unified interaction score,  $w_{M,G}$ . See the Supplementary Material for details.

### 2.2 The co-regulation network of miRNAs and its modules

In the co-regulation network of miRNAs two miRNAs (nodes) are connected, if they share at least one target gene. The weight (score) of each link is calculated as the similarity of the regulation patterns of the two miRNAs. To compute the score of each link, we first listed for each miRNA,  $M$ , its unified interaction score with all genes in the genome,  $\vec{v}_M = (w_{M,G_1}, w_{M,G_2}, \dots)$ , based on TargetScan data. We set the silencing score to zero for any non-interacting regulator-target pair and computed the co-regulation score (link weight) of two miRNAs (network nodes) as the correlation of their  $\vec{v}_M$  vectors.

To find modules in the co-regulation network of miRNAs, we discarded co-regulation links with weights below a fixed threshold,  $W$ . With the Clique Percolation Method, Palla *et al.* (2005), implemented by CFinder, we simultaneously selected the optimal link weight threshold,  $W$ , and computed the modules of the network. The Clique Percolation Method (CPM) finds groups of nodes (modules) in the network such that the density of links is higher inside the groups than between them. At the same time the CPM selects an optimal link weight threshold,  $W$ . Note that the CPM explicitly allows

for overlaps between the identified network modules. For details please see the Supplementary Material.

### 2.3 Co-expression of miRNAs and miRNA essentiality

We downloaded the correlations between the expression profiles of human miRNAs across several tissue types and experimental conditions from a recent compendium (Supplementary Table 20. of Landgraf *et al.*, 2007). For each miRNA,  $M$ , we computed its average expression correlation with (i) all other miRNAs ( $A_{\text{all}}$ ) and with (ii) miRNAs co-regulating with  $M$  more strongly than the co-regulation score threshold,  $W$  ( $A_{\text{top}}$ ). We defined the essentiality of each miRNA, relative to other miRNAs, as the ratio of strengths of its two roles:  $E = (1 + A_{\text{all}})/(1 + A_{\text{top}})$ . We note that the co-expression data set was filtered: miRNAs not present in at least ten tissues with an overall frequency of at least five clones were discarded. Thus, a total of 4 005 co-expression links were listed among 90 miRNAs. Out of these miRNA - miRNA pairs 3 081, connecting 79 miRNAs, were also present in the list of co-regulation links computed from TargetScan data.

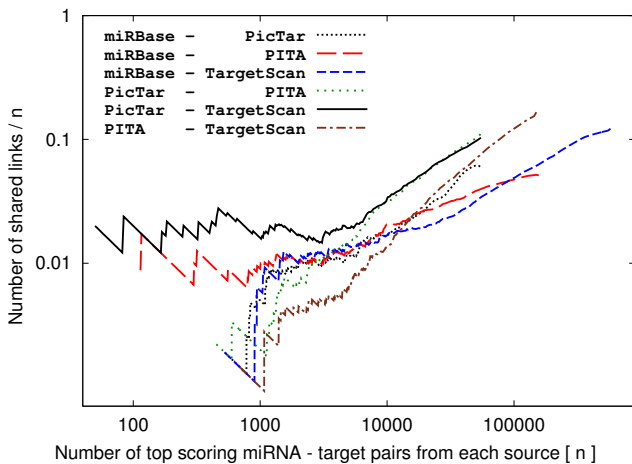
## 3 RESULTS

### 3.1 Computationally predicted lists of miRNA-target pairs differ strongly

Many biochemical and thermodynamical factors influence miRNA-target pairing and the description of these factors is still an ongoing process (Grimson *et al.*, 2007; Filipowicz *et al.*, 2008). Currently, computationally predicted miRNA - target pairs (interactions, links) by far outnumber experimentally verified ones. For example, miRBase (version 5) lists 584 403 predicted miRNA - target gene pairs in human, while, according to TarBase, direct experimental evidence is available only for a total of 61 pairs so far. Moreover, when comparing the human gene silencing interactions predicted by miRBase, PicTar, PITA and TargetScan, we find that only 0.17% of all predicted miRNA - target gene pairs are listed by all four sources and only 12.4% by at least two (see Table 1 for details). We have checked that these ratios are similar in other metazoans. As an example for the large differences between prediction algorithms, we compare two well-known silencing interactions from *C. elegans*.

Database	Number of miRNAs / target genes	Number of interactions	Experimentally verified interactions according to TarBase
<b>Experimental</b>			
TarBase	33 / 54	61	
<b>Computationally predicted</b>			
miRBase	711 / 22 474	584 403	16 (0.003%)
PicTar	171 / 6 885	54 947	31 (0.06%)
PITA	470 / 8 720	152 040	23 (0.02%)
TargetScan	455 / 15 878	955 644	44 (0.004%)

**Table 1.** Statistics of experimentally verified and computationally predicted lists of human miRNA - target gene pairs (silencing interactions). The four computationally predicted interaction lists – from miRBase, PicTar, PITA and TargetScan – contain a total of 1 529 836 miRNA - target gene pairs between 816 miRNAs and 22 968 target genes. Out of these pairs only 2 565 (0.17%) are contained by all four sources, 190 480 (12.4%) by at least two and 48 (0.003%) have been experimentally verified.



**Fig. 1.** Similarity of the lists containing the top scoring  $n = 1, 2, \dots$  human miRNA - human target gene pairs (links) from two computational databases, e.g., miRBase - PicTar or PITA - TargetScan. As larger numbers of top scoring links are chosen from any two of the four databases, the relative size of their intersection (number of pairs contained by both sources divided by  $n$ ) grows. For large  $n$  we used larger step sizes.

Scoring schemes vary across databases, and therefore, scores from different databases cannot be directly compared. However, scores within each database can be ordered and one can easily see whether two selected interactions have the same strength order in two different databases. In *C. elegans* two thoroughly studied miRNA - target interactions are the silencing of *lin-14* by *lin-4* (Lee *et al.*, 1993; Wightman *et al.*, 1993) and the silencing of *lin-41* by *let-7* (Slack *et al.*, 2000). PicTar lists both interactions, with scores 11.20 and 4.18, respectively and TargetScan lists only the first, *lin-4*  $\dashv$  *lin-14*, but not the second. (We use the sign  $\dashv$  to denote a silencing interaction.) Thus, PicTar and TargetScan predict that the first interaction is significantly stronger than the second. On the other hand, miRBase and PITA contain only the second interaction, *let-7*  $\dashv$  *lin-41*, but not the first, indicating that the second is significantly stronger. In summary, even for well-studied miRNA - target pairs, computational algorithms may provide very different predictions. This is likely due to the different biological and thermodynamical effects that each algorithm emphasises as well as the quantification of these effects.

### 3.2 Computationally predicted interaction lists become increasingly similar with growing sample size

Despite the large differences that may occur between computational predictions for a single miRNA - target pair, computationally predicted interaction lists do indeed converge as more pairs are considered. To analyse how sample size affects the similarity of predictions, we selected a representative sample, the  $n = 1, 2, \dots$  interactions (links) with the top scores, from each of the four investigated computational sources. We found that for any two, e.g., miRBase-PicTar or miRBase-PITA, the relative intersection size of the two selected lists, i.e., the number of shared links divided by  $n$ , grows with  $n$  (see Fig. 1). This implies that as sample size grows predictions of computational algorithms become increasingly similar.

Even though the four predicted interaction lists become more similar with increasing sample size, their qualities significantly differ (Sethupathy *et al.*, 2006), thus, for further analyses it is necessary to select a list of high-confidence miRNA - target gene interactions. One option for selecting high-quality interactions is (i) to apply benchmarks for computing a confidence score for each data source and then (ii) to use these scores to a single interaction score. This method is frequently applied for integrating protein-protein association and interaction (PPI) data and relies on the availability of benchmarks with a sufficiently high coverage of the full interactome, see, e.g., Ref. (Lee *et al.*, 2004). For RNA silencing however, experimental data sets have not yet reached the coverage necessary for quantifying the precision of computational predictions. Based on two recent studies (Baek *et al.*, 2008; Selbach *et al.*, 2008), we have decided to restrict ourselves to the interaction scores from TargetScan and use PicTar scores as a control.

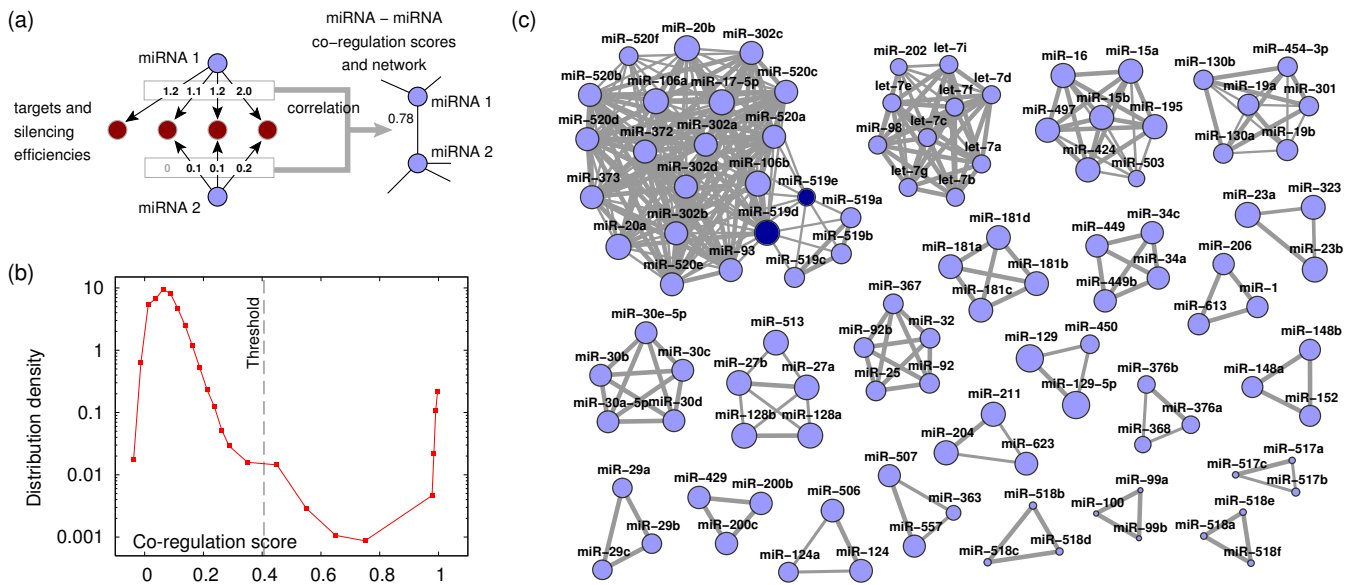
### 3.3 Modules of co-regulating miRNAs are well separated

A simple format for listing the overall silencing effect of a given miRNA,  $M$ , is a vector  $\vec{v}_M$ , in which the  $i$ th component,  $v_{M,i}$ , is the score of the silencing interaction between  $M$  and the  $i$ th gene. We quantified the similarity,  $w_{M,N}$ , of the silencing functions of two miRNAs,  $M$  and  $N$ , as the Pearson correlation of their silencing vectors,  $\vec{v}_M$  and  $\vec{v}_N$ , and applied these  $w_{M,N}$  similarity scores as link weights in the co-regulation network of miRNAs (see Fig. 2a).

For non-normally distributed variables rank correlation measures, e.g., the Spearman rank correlation, are often preferred over the Pearson correlation. After computing both correlation measures for all miRNA-miRNA pairs, we found that in the current case they provide almost identical top scoring pairs, thus, neither of them is significantly better than the other (see the Supplementary Material for details).

Compared to other similarity measures, the Pearson correlation efficiently finds small numbers of significantly co-regulated targets and at the same time strongly suppresses those silencing scores,  $v_{M,i}$ , of a miRNA that are small (and may contain a higher relative error) compared to other silencing scores of the same miRNA. Next, to illustrate why the Pearson correlation of silencing scores is biologically more relevant than the number of shared targets of the two miRNAs, we discuss two examples. First consider two hypothetical miRNAs, both controlling all genes of the human genome and each silencing all of its targets with the same large  $w_{M,G}$  score plus a random error. Despite the large number of shared target genes, the correlation of the  $\vec{v}_M$  vectors these two miRNAs is 0, i.e., there is no co-regulation at all. On the other hand, two miRNAs silencing strongly the same ten target genes with equal strength, have maximally correlated  $\vec{v}_M$  vectors, and thus, a co-regulation score of 1. Interestingly, in Fig. 2b we found a strongly bimodal distribution of co-regulation link weights: the score of a miRNA - miRNA co-regulation link is almost always below  $\approx 0.4$ , while for a small fraction of the links the score is above 0.9. In other words, it is meaningful to filter the co-regulation links by discarding those below the co-regulation score threshold,  $W$  (see Methods and the Supplementary Material for details on how the precise value of  $W$  was selected).

We proceed now to computing the modules of co-regulating miRNAs. A network module (also called cluster or community) in



**Fig. 2.** Co-regulation scores of human miRNAs and co-regulating modules. (a) The co-regulation score between two miRNAs is defined as the Pearson correlation of their silencing scores over all target genes. This score is used as the weight of the link connecting the two miRNAs in the co-regulation network. (b) The distribution of miRNA co-regulation scores is bimodal. There are very few values between 0.4 and 0.9. For finding modules, i.e., densely internally linked groups of nodes in the miRNA co-regulation network, we used only co-regulation scores above the threshold value. Note that the vertical scale is logarithmic. The line connecting the data points is a guide to the eye. (c) The modules of co-regulating human miRNAs identified by CFinder. Despite explicitly allowing for overlaps between the modules, they are well-separated. Only two modules share nodes, these are shown with dark colour. Most of the miRNA modules are (close to) fully connected with co-regulation links of weight 0.95 or above, thus, in one module all miRNAs silence roughly the same group of targets. The area of a node is proportional to the number of target genes of the given miRNA; the width of a link is proportional to the co-regulation score of the two miRNAs. The functions of two selected modules are discussed in Sec. 3.4. The layout of this network was produced with Cytoscape (Shannon *et al.*, 2003). While this figure displays each miRNA separately, Supplementary Figure 5. shows the same modules by merging all miRNAs from the same family (as defined by miRBase).

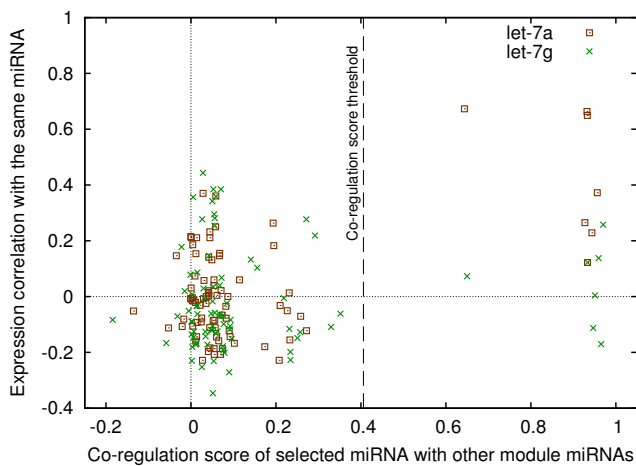
the miRNA co-regulation graph corresponds to a group of miRNAs silencing together a group of target genes: there are many regulator-target pairs between the two groups, and both groups (regulators and targets) have few further silencing interactions. Overlaps among the identified miRNA modules would indicate that the shared miRNAs strongly participate in more than one regulatory task. Most clustering algorithms, including commonly used hierarchical clustering methods, by their definition prohibit overlaps between the identified clusters and break down the network (or any other data set) into non-overlapping clusters. As our goal was not only to identify groups (modules) of co-silencing miRNAs, but also to determine whether these modules overlap, we applied CFinder (Adamcsek *et al.*, 2006), a fast algorithm explicitly allowing for overlaps between the detected network modules.

Despite allowing overlaps between the modules of co-regulating miRNAs, we found very few. Figure 2c shows the groups of miRNAs densely internally linked by co-regulation links, as identified by CFinder. We conclude that, according to current experimental data, groups of human miRNAs silence well-separated groups of target genes. In other words, the scarcity of overlaps between co-silencing groups shows that in human most translational silencing tasks are managed by well-defined groups of miRNAs. Interestingly, not only specialised miRNAs work in the modules, but also two ubiquitously expressed miRNAs, e.g., *let-7a* and *miR-30c* (Sempere *et al.*, 2004), may silence well-separated groups of target genes.

### 3.4 Seed sequences and functions in selected modules of co-regulating miRNAs

In mammals perfect base pairing between residues 2 to 8, the seed, of a miRNA and a transcript was shown to be a better predictor of the silencing interaction than perfect pairing on any other heptamer of the miRNA (Lewis *et al.*, 2003). Observe in the main panel of Fig. 2c that the miRNAs *miR-15a*, *miR-15b*, *miR-16*, *miR-195*, *miR-424* and *miR-497* (all from the same highly conserved family) are in the same module and have the seed AGCAGCA. Replacing the last residue, A, in this seed by the other purine, G, allows a miRNA to be in the same module: *miR-503* (from a different conserved family), which has the seed AGCAGCG. On the other hand, replacing the 5th residue, G, by a pyrimidine, C, separates miRNAs from this group: *miR-29a,b,c* all have the seed AGCACCA, and they form a co-regulating module isolated from that of *miR-15a*.

Even though the two modules are separated, they control similar processes. Both modules contain experimentally verified repressors of anti-apoptotic members from the *Bcl-2* protein family and thereby induce apoptosis: *miR-15a* and *miR-16* downregulate *Bcl2* (Cimmino *et al.*, 2005), while *miR-29* represses *Mcl1* (Mott *et al.*, 2007). In addition to this similarity, the two modules can also control cell fate through biochemically different mechanisms. Members of the *miR-16* module cause cell cycle arrest by directly silencing targets enriched with cell cycle functions (Linsley *et al.*, 2007; Liu



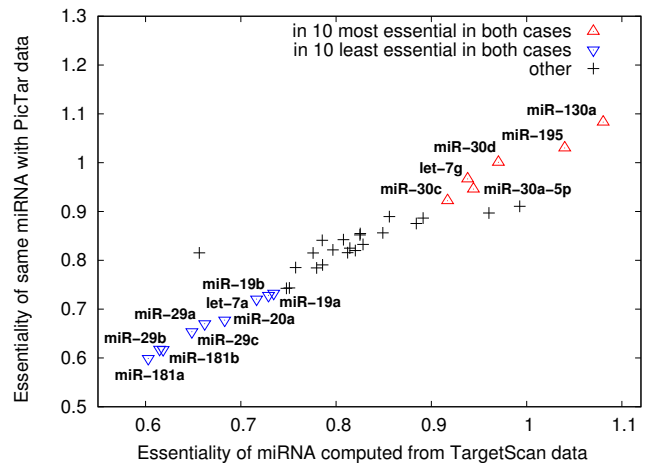
**Fig. 3.** Co-expression and co-regulation scores between two selected miRNAs and other miRNAs. We call two miRNAs “strongly co-regulating”, if their co-regulation score is above the threshold value. The average co-expression of *let-7a* with its module member miRNAs is  $A_{top} = 0.42$ . With all miRNAs of the modules this average is  $A_{all} = 0.027$ . Thus, the essentiality of *let-7a*, relative to other miRNAs, is  $E = (1 + A_{all}) / (1 + A_{top}) = 0.72$ . While *let-7a* is co-expressed with its strongly co-regulating partners, *let-7g* is not, and therefore, it is more essential:  $E = 0.94$ .

*et al.*, 2008), while *miR-29* was found to keep the methylation pattern of DNA favourable for the expression of tumour-suppressor genes *FHIT* and *WWOX* by targeting DNA methyl transferases *DNMT3A/B* (Fabbri *et al.*, 2007).

From the above examples one may reach the conclusion that in general a purine-purine difference in one (e.g., the last) residue of the seed can be present in the same co-regulation module of miRNAs, but a purine-pyrimidine swap inside the seed can already distinguish between co-regulation modules. This result is somewhat refined by the module in the top left corner of Fig. 2c, where the miRNAs *miR-520a,b,c,d,e* all have the same seed, AAGUGCU, while *miR-520f* and *miR-520g* have different seed regions, but contain the same AAGUGCU sequence at positions 1 to 7 and 4 to 10, respectively.

### 3.5 Predicted essentiality of human miRNAs

MiRNAs fit into the full regulatory web of the cell through upstream and downstream connections: they control well-separated groups of target transcripts and their expression is also controlled, often in groups. The co-expression of two miRNAs may be achieved through different sets of controlling molecules, nevertheless, it usually reflects a common biological cause. Co-expressing miRNAs that also co-regulate allows the cell to defend itself against random failures. Each of these regulators can be replaced with several others in most of their regulatory roles. Again, one may think of a group of workers performing similar tasks together. Removing a single worker (a co-expressed miRNA) will not halt production. One may also imagine individual workers (non-co-expressed miRNAs) performing their tasks in different locations. Remove one of these workers and production halts immediately. Such diversified expression of co-regulating miRNAs provides the cell with the opportunity



**Fig. 4.** Predicted essentiality of human miRNAs with TargetScan and PicTar data. MiRNAs that are among the 10 most (least) essential in both cases are indicated with upward (downward blue) triangles. See Supplementary Figure 4. for a randomised control.

to integrate signals and silence the same target group under a number of different conditions. To determine the participation of each miRNA in these two mechanisms, error tolerance and signal integration, we have computed how strongly (on average) each miRNA is co-expressed with (i) all others and (ii) those in its module. Observe that if a miRNA is co-expressed with those miRNAs that it strongly co-regulates with, then in case of a knock-out most of its regulatory functions can be replaced. On the other hand, the disruption of a miRNA that is expressed differently from its co-regulating partners is more likely to strongly reduce the viability of the cell. Consequently, we define the essentiality of each miRNA, relative to others, as the ratio between how strongly it participates in replacement and signal integration (see Methods for details). The full list of predicted miRNA essentiality values is available in the Supplementary Material.

Certainly, each miRNA may contribute both to error tolerance and signal integration to some extent, but since both tasks are based on the collective action of miRNAs, only module members can contribute significantly. In Fig. 3 we take – as an example – two miRNAs from the same co-regulation module and show that one of them, *let-7a*, is better co-expressed with its co-regulating partners, than the other, *let-7g*. Thus, *let-7g* is more essential than *let-7a*. Below we show that the set of predicted most and least essential miRNAs changes little, if TargetScan is replaced with PicTar as a data source. Indeed, six of the predicted ten most essential miRNAs are identical in both cases (see Fig. 4): *miR-130a*, *miR-195*, *miR-30d*, *let-7g*, *miR-30a-5p* and *miR-30c*. Out of the ten least essential miRNAs identified with TargetScan and PicTar data, nine are identical: *miR-181a*, *miR-29b*, *miR-181b*, *miR-29c*, *miR-29a*, *miR-20a*, *let-7a*, *miR-19b* and *miR-19a*. Many of the miRNAs on these two lists have been experimentally linked with some form of tumour, as well as other functions. For example, *miR-195* and *miR-130a* were connected with cardiac growth and angiogenesis, respectively, and *miR-181* with hematopoietic lineage differentiation (Chen *et al.*, 2004; van Rooij *et al.*, 2006; Chen and Gorski, 2008). As we could not identify marked functional differences between the predicted most and

least essential miRNAs, we conclude that essentiality defined in the above way – low co-expression with strongly co-regulating partners – cannot be linked with a particular function. Rather, we suggest that the essentiality of a miRNA is determined merely by its position in the co-regulation network relative to other members of the same co-regulating module.

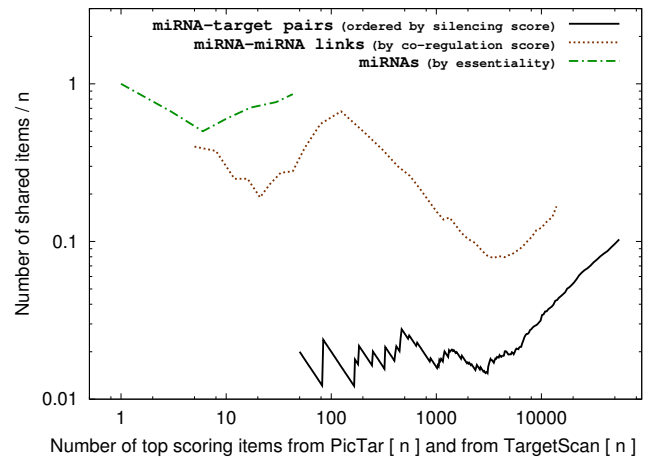
### 3.6 MiRNA co-regulation scores, modules and miRNA essentialities extract high-quality information

The four computational databases compared in Table 1 share only a small portion of their predicted interactions and up to this point we have used data only from TargetScan. To validate our results, we have also performed all analyses above with PicTar data. Similar to Fig. 1, where the top scoring miRNA - target gene interactions of, e.g., PicTar and TargetScan were compared, in Fig. 5 we compare miRNA - miRNA co-regulation scores and miRNA essentiality levels computed from PicTar and TargetScan data. The agreement between the strongest co-regulation scores is much better than between the top scoring miRNA - target interaction scores. As a consequence, the co-regulation modules are also well preserved when TargetScan is replaced by PicTar as a source of primary data. Note that usually not all miRNAs fall into the co-regulation modules. If TargetScan interactions are used, then 111 of the 455 miRNAs are in the co-regulation modules. Next, consider PicTar as a data source. All miRNAs listed by PicTar are listed by TargetScan as well. If we use PicTar data, 70 miRNAs are module members, out of which 55 were also module members with TargetScan data. As a randomised control the chance for at least 55 miRNAs to be in both results is, according to the hypergeometric distribution,  $3.62 \times 10^{-26}$ . Thus, the co-regulating modules of miRNAs computed with PicTar and TargetScan are indeed highly similar. In the Supplementary Material we show that the co-regulation scores, modules and miRNA essentialities computed from miRBase and PITA are also highly similar to TargetScan results.

## 4 DISCUSSION AND OUTLOOK

Considering the impact of human miRNAs on messenger RNA and protein levels, direct experimental evidence on the underlying miRNA - transcript interactions is scarce. Due to this absence it is not yet possible to measure the precision and coverage of computationally predicted miRNA - target interaction lists nor to integrate these lists. Two recent experimental studies (Baek *et al.*, 2008; Selbach *et al.*, 2008) focusing on small numbers of human miRNAs found two of the computational databases to be the most precise predictors of mRNA and protein level changes upon miRNA removal and insertion. We used miRNA - target silencing scores from one of these two sources, TargetScan, and applied data from the other, PicTar, as a control.

After defining and computing miRNA co-regulation scores we found that human miRNAs work in well-separated co-regulating modules. For a possible biochemical explanation consider the following. In transcription regulation transcription factor (TF) proteins need to recognise a section of a double helical nucleic acid, while in RNA silencing template formation between sections of two nucleic acids is necessary. Thus, a short RNA regulator is constrained to be more specific in the space of transcript (seed) sequences than a TF protein in the space of DNA sequence motifs. In other words,



**Fig. 5.** Comparing TargetScan and PicTar through silencing scores between miRNA - target gene pairs (solid line, this curve is repeated from Fig. 1), miRNA - miRNA co-regulation scores (dotted line) and predicted essentialities of miRNAs (dash-dot). When changing between TargetScan and PicTar as primary data sources the list of top scoring miRNA co-regulation links and the list of predicted most essential miRNAs are clearly much better preserved than the list of top scoring miRNA - target gene pairs. The end points of the two upper curves are higher than the end point of the lower curve, thus, the improvement in data quality cannot be accounted to the reduced number of items ( $n$ ).

compared to TF proteins miRNAs are more strongly forced to be organised into tightly co-regulating modules. One such module is a group of miRNAs all of which silence almost the same group of targets and regulate few others.

The examples in Sec. 3.4 show that even though the seeds of miRNAs in the same co-regulation module are mostly similar (often identical), there is no one-to-one mapping between modules and seeds. Rather, each module seems to correspond to a small group of similar seeds that may be even shifted from the usual position (residues 2 – 8) to, e.g., the positions 1 – 7 or 4 – 10. We found that the groups of miRNAs defined by these small groups of seeds share almost none of their members.

The setup outlined here may be called collective targeting or “group-to-group control”. We anticipate that in other animal species translational silencing operates in a similarly specialised, group-to-group manner. This may partly explain the recent experimental result that only a very small portion of single miRNAs is essential for development or viability in *C. elegans* (Miska *et al.*, 2007). Landgraf *et al.* (Landgraf *et al.*, 2007) have reported that in human and mouse a small set of ubiquitously expressed miRNAs account for the bulk of cell lineage- and tissue-specific miRNA expression. Seeking an explanation for this result we argue that – compared to TFs – among miRNAs ubiquitous expression may be a stronger indicator of essentiality which may lead to the preservation of a number of ubiquitously expressed miRNAs over evolutionary time scales. We identified a miRNA as more essential, if it is expressed differently from the miRNAs it co-regulates with. If such a miRNA is turned off, there are few others that can take over its silencing tasks precisely without side effects. For example, assume that a group of target genes is silenced by the miRNA  $M_1$  that is expressed in many tissues and the miRNAs  $M_2, M_3, \dots$  that are

all expressed in the same single tissue only. We liken this to many workers performing the same task together (non-essential miRNAs) as opposed to a single worker performing the task alone (essential miRNAs). In this case the removal of  $M_1$  will leave the group of target genes unsilenced under many conditions, but the removal of, e.g.,  $M_2$  can be compensated for easily. As discussed above, due to the mechanism of RNA silencing miRNAs often have to regulate their targets together with closely co-silencing partners. Within such co-regulation modules the most broadly expressed miRNAs are likely to be the most essential.

Recently, Huang *et al.* (Huang *et al.*, 2007) have combined miRNA and mRNA expression profiles with predicted miRNA - mRNA silencing scores to improve the silencing scores. Here we built the co-regulation network of mRNAs first – using only silencing scores – and then compared miRNA-miRNA co-regulation with miRNA-miRNA coexpression. The method of the current paper is probably less applicable to improving the predicted silencing scores of single miRNA-mRNA pairs. However, silencing scores have currently little direct experimental support and are still often rather ambiguous (see Table 1 and Fig. 5). The co-regulation modules of miRNAs and the essentiality scores obtained in the current paper extract biological information that is robust to this ambiguity (Fig. 5).

In the current paper we have defined the predicted essentialities of miRNAs through expression correlations. An interesting, though conceptually different, definition may focus on single tissues (or conditions) and compare the expression level of a miRNA in a single tissue to its expression levels in all other tissues. We note also that two recent *in silico* studies have revealed pathway components collectively targeted by miRNAs (Gusev *et al.*, 2007) and regulatory modules comprising miRNAs and target genes (Yoon and De Micheli, 2005). However, neither of these two studies discussed the separation of the co-regulating modules and the improvement in data quality that can be achieved by data filtering steps. We provide the experimentally testable hypothesis that after individual deletions of the predicted six most essential miRNAs, the percentage of inviable strains will be higher as compared with deletions of the predicted nine least essential.

## ACKNOWLEDGEMENTS

We thank T. Vicsek for discussions and suggesting the concept of “group-to-group” silencing and E. Segal for comments; S. van Dongen, G. Bell and M. Kertesz for advice on database issues; N. Blumm for careful reading of the manuscript.

*Funding:* This work was supported by the Hungarian National Science Foundation (K68669, K75334) and the Hungarian National Office for Research and Technology (CellCom).

*Author contributions:* G.B., O.K., and I.J.F. performed research; I.J.F. and G.B. designed research; I.J.F. wrote the paper.

*Conflict of interest:* none declared.

## REFERENCES

Adamcsek, B. *et al.* (2006) CFinder: Locating cliques and overlapping modules in biological networks, *Bioinformatics*, **22**, 1021-1023.

Alon, U. (2007) Network motifs: theory and experimental approaches, *Nat. Rev. Genet.*, **8**, 450-461.

Baek, D. *et al.* (2008) The impact of microRNAs on protein output, *Nature*, **455**, 64-71.

Bartel, D.P. *et al.* (2004) Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat. Rev. Genet.*, **5**, 396-400.

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

Chen, Y., Gorski, D.H. (2008) Regulation of angiogenesis through a microRNA (miR-130a) that down-regulates antiangiogenic homeobox genes GAX and HOXA5, *Blood*, **111**, 1217-1226.

Chen, C.-Z. *et al.* (2004) MicroRNAs Modulate Hematopoietic Lineage Differentiation, *Science*, **303**, 83-86.

Cimmino, A. *et al.* (2005) miR-15 and miR-16 induce apoptosis by targeting BCL2, *Proc. Natl. Acad. Sci. U S A*, **102**, 13944-13949.

Fabbri, M. *et al.* (2007) MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B, *Proc. Natl. Acad. Sci. U S A*, **104**, 15805-15810.

Farh, K.K. *et al.* (2005) The Widespread Impact of Mammalian MicroRNAs on mRNA Repression and Evolution. *Science*, **310**, 1817-1821.

Filipowicz, W. *et al.* (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?, *Nat. Rev. Genet.*, **9**, 102-114

Griffiths-Jones, S. *et al.* (2008) miRBase: tools for microRNA genomics. *Nucl. Acids Res.*, **36** (Database issue), D154-D158.

Grimson, A. *et al.* (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing, *Mol. Cell.*, **27**, 91-105.

Gusev, Y. *et al.* (2007) Computational analysis of biological functions and pathways collectively targeted by co-expressed microRNAs in cancer, *BMC Bioinformatics*, **8**, S16

Hartwell, L.H. *et al.* (1999) From molecular to modular cell biology, *Nature*, **402**, C47-52.

Heimberg, A.M. *et al.* (2008) MicroRNAs and the advent of vertebrate morphological complexity, *Proc. Natl. Acad. of Sci. U S A*, **105**, 2946-2950.

Huang, J.C. *et al.* (2007) Using expression profiling data to identify human microRNA targets. *Nat. Meth.*, **4**(12), 1045-1049.

Kertesz, M. *et al.* (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**(10), 1278-1284.

van der Krol, A.R. *et al.* (1990) Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell.*, **2**(4), 291-299.

Lall, S. *et al.* (2006) A Genome-Wide Map of Conserved MicroRNA Targets in *C. elegans*. *Curr. Biol.*, **16**, 460-471

Landgraf, P. *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401-1414.

Lee, I. *et al.* (2004) A Probabilistic functional network of yeast genes, *Science*, **306**, 1555-1558.

Lee, R.C. *et al.* (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, **75**(5), 843-54.

Lewis, B.P. *et al.* (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787-798.

Linsley, P.S. *et al.* (2007) Transcripts Targeted by the MicroRNA-16 Family Cooperatively Regulate Cell Cycle Progression, *Mol. Cell. Biol.*, **27**, 2240-2252.

- Liu, Q. et al. (2008) miR-16 family induces cell cycle arrest by regulating multiple cell cycle genes, *Nucl. Acids Res.*, **36** 5391-5404.
- Mattick, J.S., Gagen, M.J. (2001) The evolution of controlled multi-tasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms, *Mol. Biol. Evol.*, **18**, 1611-1630.
- Miska, E.A. et al. (2007) Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability, *PLoS Genet.*, **3**, e215.
- Mott, J.L. et al. (2007) mir-29 regulates Mcl-1 protein expression and apoptosis, *Oncogene*, **26**, 6133-6140.
- Napoli, C. et al. (1990) Introduction of a Chimeric Chalcone Synthase Gene into *Petunia* Results in Reversible Co-Suppression of Homologous Genes in trans, *Plant Cell.*, **2**(4), 279-289.
- Palla, G. et al. (2005) Uncovering the overlapping community structure of complex networks in nature and society, *Nature*, **435**, 814-818.
- van Rooij, E. et al. (2006) A signature pattern of stress-responsive microRNAs that can evoke cardiac hypertrophy and heart failure, *Proc. Natl. Sci. Acad. U S A*, **103** 18255-18260.
- Selbach, M. et al. (2008) Widespread changes in protein synthesis induced by microRNAs, *Nature*, **455**, 57-63.
- Sempere, L.F. et al. (2004) Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation, *Genome Biol.*, **5**, R13.
- Sethupathy, P. et al. (2006) A guide through present computational approaches for the identification of mammalian microRNA targets, *Nat. Meth.*, **3**, 881 - 886.
- Sethupathy, P. et al. (2006) TarBase: A comprehensive database of experimentally supported animal microRNA targets, *RNA*, **12**, 192-197.
- Shannon, P. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Research*, **13**, 2498-2504.
- Slack, F.J. et al. (2000) The lin-41 RBCC gene acts in the *C. elegans* heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor, *Mol. Cell*, **5**, 659-669.
- Warner, G.J., Adeleye, A.Y., Ideker, T. (2006) Interactome networks: the state of the science, *Genome Biol.*, **7**, 301.
- Wightman, B. et al. (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*, *Cell*, **75**(5), 855-62.
- Yoon, S., De Micheli, G. (2005) Prediction of regulatory modules comprising microRNAs and target genes, *Bioinformatics*, **21** Suppl.2, ii93-100.