BIOINFORMATICS ORIGINAL PAPER

Vol. 28 no. 7 2012, pages 983–990
doi:10.1093/bioinformatics/bts077

Systems biology

Advance Access publication February 10, 2012

Susceptibility of the human pathways graphs to fragmentation by small sets of microRNAs

Guy Naamati[1,†], Yitzhak Friedman[2,†], Ohad Balaga[1] and Michal Linial[2,*]

[1]School of Computer Science and Engineering and [2]Department of Biological Chemistry, Institute of Life Sciences, Sudarsky Center for Computational Biology, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Associate Editor: Ivo Hofacker

ABSTRACT

Motivation: MicroRNAs (miRNAs) are short sequences that negatively regulate gene expression. The current understanding of miRNA and their corresponding mRNA targets is primarily based on prediction programs. This study addresses the potential of a coordinated action of miRNAs to manipulate the human pathways. Specifically, we investigate the effectiveness of disrupting the topology of human pathway graphs through a regulation by miRNAs.

Results: From a set of miRNA candidates that is associated with a pathway, an exhaustive search for all possible doubles and triplets (coined miR-Duo, miR-Trios) is performed. The impact of each miR-combination on the connectivity of the pathway graph was quantified. About 170 human pathways were tested, and the miR-Duos and miR-Trios were scored for their ability to disrupt these pathway graphs. We show that 75% of all pathways are effectively disconnected by a small number of pathway-specific miR-Trios. Only 15% of the human pathways are resistant to fragmentation by miR-Duos or miR-Trios. Significantly, the combination of the most effective miR-Trios is unique. Thus, a specific regulation of a pathway within the cell is guaranteed. The impact of the selected miR-Duo/Trios on various diseases is discussed.

Conclusions: The methodology presented shows that the synthesis of the topology of a network with a detailed understanding of the miRNAs' regulation is useful in exposing critical nodes of the network. We propose the miR-Trio approach as a basis for rationally designed perturbation experiments.

Contact: michall@cc.huji.ac.il

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on October 10, 2011; revised on January 27, 2012; accepted on February 7, 2012

1 INTRODUCTION

Non-coding microRNAs (miRNAs) play a role in stem cell differentiation, cell division, immunological cell function, stress response and cells' identity [reviewed in (Ambros, 2004; Wienholds and Plasterk, 2005)]. Furthermore, miRNAs are involved in the human pathogenesis including cancer (Ventura and Jacks, 2009) and viral infection (Cullen, 2009). Recent studies confirmed the presence of hundreds of miRNAs in healthy and diseased tissues (Landgraf *et al.*, 2007). A growing number of tools and algorithms were developed for predicting miRNA-target pairs (Maziere and Enright, 2007; Mendes *et al.*, 2009). Currently, miRBase (Griffiths-Jones *et al.*, 2006) is the most comprehensive collection of miRNAs with over 1000 miRNAs from human and 700 from mouse (Bentwich, 2008). The miRNAs are estimated to target over a third of the genes in human (John *et al.*, 2004). More than dozen miRNA-target predictors assigned potential targets for individual miRNAs. However, the consistency among the principal miRNA-target predictors is relatively small (Mendes *et al.*, 2009; Thomas *et al.*, 2010). A direct pairing of miRNAs and mRNAs was revealed using deep sequencing technology from Argonaute bound sequences (Chi *et al.*, 2009). Currently, the TarBase (Papadopoulos *et al.*, 2009b) and the miRecords databases (Xiao *et al.*, 2009) are used as sound sources for validated miRNAs and their targets.

The presence of multiple miRNAs' binding sites, on the same transcript, augments its regulation level (Brodersen and Voinnet, 2009; Hon and Zhang, 2007; Peter, 2010). The parallel over-expression of three miRNAs resulted in a synergistic effect on the transcription level of the candidate genes (Krek *et al.*, 2005; Tu *et al.*, 2009). For example, the regulation of the tumor suppressor Fus1 in cancer cells depends on the coordinated action of three miRNAs (Du *et al.*, 2009). Importantly, introducing several miRNAs not only affected gene candidates but also impacted regulatory pathways (Ivanovska and Cleary, 2008; Mu *et al.*, 2009). Considering pathways and diseases as the predictions' objective reduces the noisy predictions of miRNA-gene targets, as demonstrated by miRNApath (Chiromatzo *et al.*, 2007) and DIANA-miRPath (Papadopoulos *et al.*, 2009a, b).

This study raises the possibility that a cooperative action of a small, selected group of miRNA can alter the expression of gene-targets that belong to some pathways, without scarifying specificity. More importantly, the concept that is investigated argues that disruption of the pathway's topology has the potential to alter the outcome of the targeted pathway. A support for this notion stems from the following: (i) A number of miRNAs can act cooperatively; (ii) Multiple miRNAs disrupt disease pathways; (iii) Genes that comprise the regulatory and signaling pathways are amenable to miRNA regulation; (iv) The quantities of miRNAs in cells are strongly regulated; (v) Diseases and developmental stages are often specified by altered expression of a number of miRNAs.

Herein, we prioritize each pathway according to its susceptibility to a regulation by a small group of miRNAs. We included the most reliable resources for human pathways from pathway interaction database (PID) (hosted by the National Cancer Institute, NCI) (Schaefer *et al.*, 2009) and KEGG (Ogata *et al.*, 1999). Many of the

*To whom correspondence should be addressed.

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

pathways were previously analyzed in view of their modular nature, redundancy and robustness against perturbations (Ideker and Sharan, 2008; Stelling *et al.* 2004). While recent studies have acknowledged the effectiveness of miRNAs on regulation pathways (Cui *et al.*, 2006; Gusev *et al.*, 2007), the connectivity of the genes in the pathway was not taken into account.

First, we represent all high quality human pathways as undirected graphs and present a pathway disconnecting score. Then, we define a set of miRNA candidates based on the likelihood to regulate the genes of the subject pathway. The miRNA set is then exhaustively searched for miRNA-triplets (i.e. miR-Trios) that maximize the disruption of a given pathway accordingly to various measures for network graph fragmentation. This procedure was repeated for miRNA doublets (coined miR-Duos).

We present a method for prioritizing a small subset of miRNAs with maximal impact on pathway connectivity. A quantitative change in a set of miRNAs is expected to alter the pathway outcome. This procedure implements the updated view on cellular miRNAs that 'work together' (Peter, 2010; Salmena *et al.*, 2011). Our pathway-specific approach complements a direct manipulation of genes via RNAi with the use of siRNA and shRNA. We demonstrate the strength and weakness of the approach by analyzing the top scoring miRNA sets for different signaling pathways, especially in the context of diseases.

2 METHODS

2.1 Databases

The database of miRror (Friedman *et al*., 2010) is based on the resources from the principal miRNA-target prediction tools. The resources used are: (i) TargetScan (Lewis *et al.*, 2003); (ii) microCosm that is based on miRanda algorithm (John *et al.*, 2004); (iii) PicTar (Krek *et al.*, 2005); (iv) DIANA–MicroT (Maragkakis *et al.*, 2009); (v) PITA (Kertesz *et al.*, 2007); (vi) mirZ (Hausser *et al.*, 2009); (vii) microRNA.org that is based on miRanda algorithm (Betel *et al.*, 2008); (viii) miRDB resource (Wang, 2008); (ix) TargetRank (Nielsen *et al.*, 2007) and several smaller databases such as RNA22.

Regulatory pathways resources: (i) The PID (Schaefer *et al.*, 2009) compiles 79 human regulatory pathways that describe accepted interactions. (ii) KEGG (Kanehisa and Goto, 2000) is a comprehensive collection of manually drawn pathway maps. These maps represent the information on the molecular interaction and reaction networks for 93 maps.

2.2 miRror application for cooperative regulation

The miRror platform (Friedman *et al.*, 2010) was used to match a selected gene list to the minimal number of preferred miRNAs sets. The statistical method of miRror is the miRtegrate algorithm. It calculates the probability of matches between a gene list and the genes that are reported by each of the miRNA-targets resources. Calculating the *P*-value for genes as input was performed according to the hypergeometric distribution (Friedman *et al.*, 2010). The *P*-value thresholds that were used in this study are 0.01, unless otherwise stated.

2.3 ID conversion for pathways

miRror platform is used to determine the match between a gene list in a pathway and its most likely set of miRNAs that target these genes, collectively. miRror uses the RefSeq identifiers as its central entry. We applied an internal ID matching protocol to convert the KEGG 'official gene name' into RefSeq identifiers. Conversion was performed using BioMart and UniProtKB ID mapping. The match of RefSeq to the KEGG reaches 97% coverage for human.

2.4 Constructing a pathway graph

A procedure by which the pathway is transformed into a number of genes is by converting the pathway into a graph unit. The pathways from KEGG (signaling and regulatory) are stored in XML format. We obtained the list of nodes and edges from the XML format. In KEGG, nodes were entries of type 'gene'. Edges were connected between two nodes if a relation existed between them. Entries of type 'group' were defined as a node that includes all the genes linked to it.

For each input of genes that were extracted from the pathway resource, the miRror output set of miRNAs was mapped to these genes. Each miRNA is mapped to a number of genes, and the set of three miRNAs (without repetition) is referred to as a miR-Trio. In cases that the miRror output consisted of >60 miRNAs, the most significant 60 miRNAs with the lowest *P*-values were used for the analysis. Only representative miRNAs were included according to miRBase family assignment (Griffiths-Jones *et al.*, 2006).

2.5 Disconnection scores

2.5.1 Partition square score Targeting a gene is translated to its elimination together with its adjacent edges. The partition to a maximum number of small graphs provides a maximal score. Formally, for each miR-Trio, cc_del is the CC after gene deletion. CC is the connected components before deletion.

$$\mathrm{PAS} = 1 - \frac{\sum(\mathrm{cc\_del})^2}{\sum(\mathrm{cc})^2}$$

2.5.2 Betweenness score Betweenness is a centrality measure of a node within a graph. The betweenness of a particular node is determined by how often it appears on the shortest paths between pairs of remaining nodes. Nodes that appear on many shortest paths have higher betweenness. Formally, del_betweenness is the betweenness for each vertex after gene deletion. Betweenness is the betweenness score (BES) of the graph before deletion.

$$\mathrm{BES} = 1 - \frac{\sum(\mathrm{del\_betweenness})}{\sum(\mathrm{betweenness})}$$

2.5.3 Fraction of edges score A miR-Trio that removes many edges will hold a high score. Formally, denote del_edges as the edges after gene deletion and edges as the edges before gene deletion.

$$\mathrm{FES} = 1 - \frac{\sum(\mathrm{del\_edges})}{\sum(\mathrm{edges})}$$

In order to attenuate the score for miR-Trios that eliminate all genes in the network, a normalization factor was included. This factor is defined as the square root of the number of remaining vertices after vertex elimination divided by the total number of vertices in the graph. The factor is then multiplied (before getting the root) by (number of vertices in the graph/number of vertices in the graph −1) to provide a perfect score of 1 (if only one gene has been targeted). Notice that 'number of vertices in the graph' is canceled out.

$$\mathrm{Normalized\_score} = \mathrm{score} \cdot \sqrt{\frac{\sum(\mathrm{del\_vertices})}{\sum(\mathrm{vertices} - 1)}}$$

Each score (as above) ranges between 0 and 1, before and after normalization. For each miR-Trio we define the maximal disintegration score (DIS) as the maximal score obtained by any of the three scoring methods described above.

2.6 miR-Trio exhaustive search

The number of miRNAs that are assigned by miRror was used to determine the number of miR-Trios tests. Only pathways that were associated with at least three miRNAs were analyzed. miRror was activated by using all

possible gene-targets (ignoring tissues/cell-line filters) (Friedman *et al.*, 2010). The analysis is limited to 60 miRNAs per pathway. Therefore, the maximal search is the number of different possible triplets (60 choose 3 = 34 220). Each miR-Trio was tested individually in view of the graph topology and the different scoring methods (see above). The miR-Trios that achieved maximal DIS were defined as 'top scoring'. We repeated the protocol to test the concept of miR-Duos (pairs of miRNAs). All miR-Trios or miR-Duos that are within the top 5% of the maximal DIS for each pathway are defined as 'top 5%'.

3 RESULTS

We compiled the manually curated pathways from KEGG and PID (NCI human pathways) resources, covering 93 and 79 human pathways, respectively. The conversion relies on a simplification of the KEGG pathway by excluding non-protein components and localization information. The conversion creates pathway graphs that are used throughout this study. About half of the pathway graphs are composed of only one or two connected components (CCs, Fig. 1A). The quantitative properties of KEGG pathways are listed in Supplementary Table S1. The conversion of the KEGG pathway to a graph is illustrated for 'Insulin signaling pathway' (Fig. 1B).

3.1 Selecting the miRNA candidates for a pathway

Analysis of a pathway via the gene-target collection by any of the miRNA prediction tools usually results in tens and hundreds of miRNAs. For example, considering the miRNA predictions for some of the 54 genes participating in 'Axonal Guidance pathway' (KEGG:hsa04360) suggests thousands of predictions by the miRanda algorithm (John *et al.*, 2006). Specifically, the EPHA7 (Ephrin receptor A7) is mapped to 159 miRNAs, PPP3CA (protein phosphatase 3) to 184 miRNAs with several other proteins including RASA1, RAC1, CDC42 associated with ~100 miRNAs each. Other miRNA-target predictors connect these genes with even larger sets of putative miRNAs (data not shown).

In order to determine the number of miRNA candidates for a given human pathway, the number of genes within the pathway is considered. This number for human KEGG pathways ranges from 12 to 247 (Supplementary Table S1). Analysis of a pathway via the gene-targets set by individual miRNA prediction databases results in hundreds of possible miRNAs. Therefore, matching a small number of miRNAs to a particular pathway is virtually impossible with the current predictors. To address this challenge, we selected the candidate miRNAs by the miRror approach (Friedman *et al.*, 2010).

miRror is a context-dependent statistically based tool that determines the most likely set of miRNAs that regulates the input genes. Thus, the output of miRror for any pathway is a short list of miRNAs that collectively target the pathway's genes. The list of possible miRNAs is ranked according to a predetermined statistical threshold using the miRNA-target predictors (total 12 databases).

Using the same example of the 'Axon guidance' pathway (KEGG: hsa04360, 54 genes), miRror reports on only 37 miRNAs for a predetermined threshold of *P*-value −0.01. We limited the results for cases that are supported by at least two miRNA-target predicting databases.

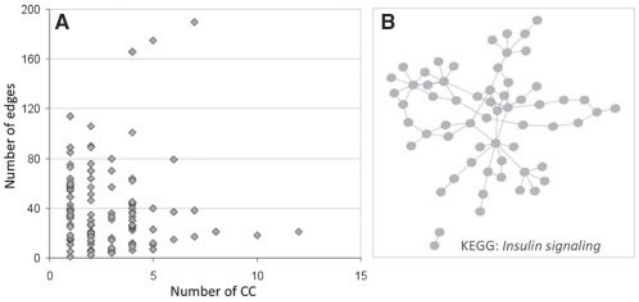

Fig. 1. Most KEGG pathways composed of <5 CCs (**A**). The undirected graph of the KEGG pathway 'Insulin signaling' is composed of two CCs (**B**). Note that only few genes are responsible for the high connectivity of the pathway graph.

Fig. 2. Identifying miRNA triplets according to the pathway perturbation scores. From left to right: (i) Converting KEGG pathways to undirected graphs. (ii) miRror application for the input of all genes (G1–G6) in the pathway resulting in a minimal set of miRNAs (M1–M7). (iii) A collection of all possible miR-Trios based on the miRror's output list. (iv) Each gene targeted by the miR-Trio is 'eliminated' from the graph. The effect on graph connectivity is measured by alternative scoring methods (marked as Score 1, Score 2).

3.2 The procedure for identifying the miRNA pathway perturbations

The goal of our strategy is to propose a small number of miRNAs that are expected to maximally affect the cell function, by a perturbation of some pathway. We quantified the impact of miRNA triplets on any selected pathway. Obviously, there is nothing extraordinary in choosing triplets for the analyses. We considered triplets as possible sets from a computational perspective. We examined, for each pathway, all possible combinations of triplet miRNAs (referred to as miR-Trios) based on the initial list of the miRNA candidates. A schematic representation of the procedure for miR-Trios selection is shown in Figure 2. For example, for 40 miRNAs 9880 possible miR-Trios were tested and this number is doubled for 50 miRNAs (see Section 2).

3.3 The disconnecting score is indicative of the pathway's fragility

There are several approaches that quantify and evaluate the importance of an individual component in a complex biological pathway. In seeking miR-Trios with a maximal potential to affect a given pathway, the pathway graph topology and connectivity need to be incorporated into the scoring scheme (Fig. 2).

Since no one scoring method captures the impact on disconnecting the pathway graph, we applied several complementary scoring systems.

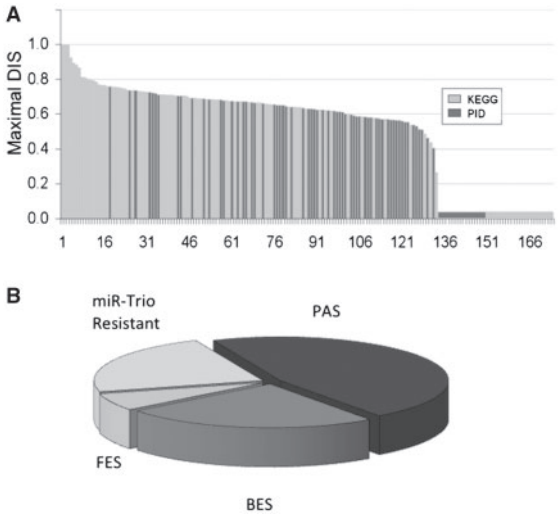

Fig. 3. Maximal DIS for KEGG and PID human pathways. (**A**) One hundred and seventy-two pathways from KEGG and PID resources ranked according to their maximal DIS. Pathways are marked by their source. (**B**) The fraction of each of the scoring modes among the maximal DIS values. miR-Trio resistant refers to the pathways that fail miR-Trios regulation. A unified pathways' list from KEGG and PID is shown in Supplementary Table S2.

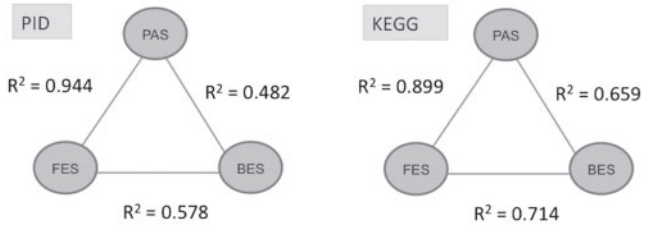

Fig. 4. Pairwise correlations between the top scoring results (BES, PAS or FES) for human PID and KEGG pathways. The correlations between all pairs of the miR-Trio scoring methods are indicated.

- Fraction of edges score (FES). This score simply gives a higher score to miR-Trios that remove the largest percentage of edges.
- Betweenness score (BES). This score rewards the miR-Trios that eliminate vertices with high centrality.
- Partition square score (PAS). This score gives high scores to miR-Trios that breakdown the pathway into smaller CCs.

Disconnecting score (DIS) is the highest maximal score from all three scoring modes (each score ranges from 0 to 1).

Figure 3 shows the results for the top scoring miR-Trios for each KEGG or PID pathway. Two observations are evident from the analysis of the two resources for human pathways: (i) Maximal DIS values are associated with the pathways from the two resources. Still the highest scores are associated with KEGG pathways. (ii) Each of the three scoring modes contributes to the maximal DIS (Fig. 3B).

Figure 4 is a summary of the correlation that is observed between any pairs of the scoring modes. PAS correlates with BES (KEGG and PID at the $R^2 = 0.659$ and 0.482, respectively).

Table 1. Samples of KEGG pathways and their quantitative properties

KEGG pathway	KEGG map	Nodes graphs	Number of miRs $P < 0.01$	Number of miR-Trios $P < 0.01$	Number of miR-Trio maximal DIS
Colorectal cancer	hsa05210	27	55	26 235	25
ECM-receptor interaction	hsa04512	40	7	35	2
Non-small cell lung cancer	hsa05223	29	57	29 260	2
Pathways in cancer	hsa05200	140	56	27 720	2
Renal cell carcinoma	hsa05211	28	54	24 804	1616
Type II diabetes	hsa04930	21	9	54	19

The PAS-FES and BES-FES pairs are also correlated. For most pathways, the difference in the best scores that are obtained from the different methods is significant. For example, the best miR-Trio for 'Neurotrophin signaling' is ranked sixth for the PAS and 61 for the BES score. Correlation coefficients were calculated for all pathways (Fig. 4).

3.4 Only few miR-Trios assigned with most pathways

Most pathways that are subjected to miRNAs regulation are associated with only a few top-scoring miR-Trios. This is best reflected by the compression factor for the selected miR-Trios. The compression factor denotes the fraction of the original number of all combinations of miR-Trios tested for that pathway and the actual top scoring miR-Trios. For example, for pathways that have >5000 potential miR-Trios each, the compression is 6000–10 000-folds.

Table 1 suggests that each pathway is unique in its trend toward the number of candidate miR-Trios. Among the 93 KEGG pathways, there are 21 pathways with >10 000 miR-Trios candidates and 20 pathways with a search space of <100 candidates. We verified that the correlation coefficient between the number of the highest scoring miR-Trios and the total number of the searched miR-Trios is negligible ($R^2 = 0.0306$). Similarly, there is no correlation between the miR-Trios search space and the DIS value for a pathway.

3.5 Selecting the miRNA candidates for a pathway

Figure 5 shows the pathway graph of 'Insulin signaling pathway'. While it is associated with 47 miRror-based miRNAs, only two miR-Trios have maximal DIS. The disconnection of the network by miR-Trios is evident. In this example, fragmentation occurs from the original two CCs (Fig. 1) to 39 CCs, most of them containing a single gene. At such a level of partitioning, some genes (six genes) become isolated as a byproduct of neighboring edge removal (Fig. 5B, isolated circles).

3.6 Maximal specificity by the top scoring miR-Trios

Over 4200 nodes are included in the analysis of human pathway graphs. Naturally, multiple pathways share many of the genes. For example, MAPK1 appears in a quarter of the tested pathways (46 pathways). The genes that are associated with >20 pathways are GRB2, EGFR, BAD, EGFR, AKT3, SRC and more. Not surprisingly, these are proteins that often serve as connectors in

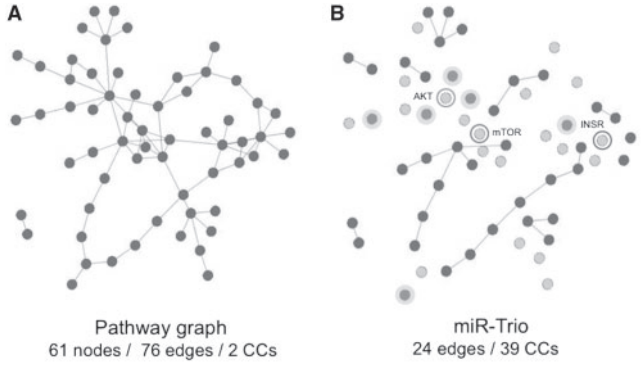

Fig. 5. The impact of miR-Trio on the 'Insulin signaling' pathway. (**A**) The pathway graph is composed of two CCs. (**B**) The top scored miR-Trio led to disconnecting of the graph to 39 CCs. Maximal DIS = 0.76. The key genes AKT3, mTOR, INSR are targeted (circled).

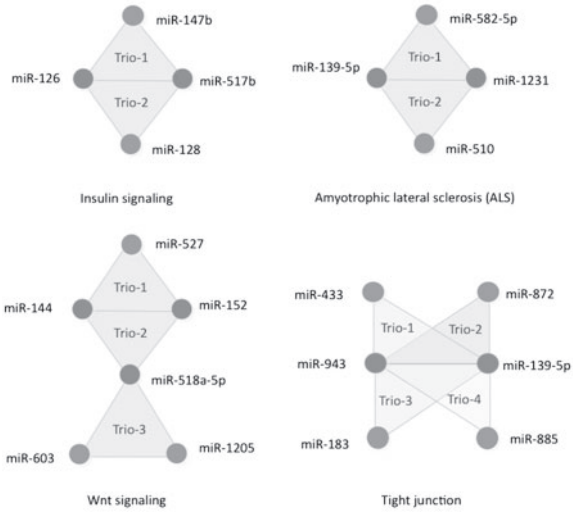

Fig. 6. Examples of top scored miR-Trios and their associated miRNAs. These examples refer to pathways that are marked by—three to four top scoring miR-Trios. Pathways that are covered by six miRNAs are associated with three miR-Trios (Wnt signaling, data not shown) or four miR-Trios (tight junction). For a list of pathways and their miR-Trios, see Supplementary Table S4.

the pathway graphs. Thus, targeting these signaling genes will lead to the 'collapse' of the pathway.

We showed that these proteins are indeed almost always disconnected from the pathway graphs by miR-Trios (Fig. 5B, circled). We thus tested the possibility that only a few miRNAs are repeatedly included among the top scoring miR-Trios, specifically those that target the key signaling proteins. However, this scenario has not been manifested. We noted that only a few miRNAs participate in more than one pathway. From a collection of 214 different miRNAs, 150 miRNAs are involved in only a single pathway.

The miRNAs that are involved in the top scoring miR-Trios for multiple pathways are miR-218 (six), miR-518d-3p (five), miR-487b and miR-147b (four pathways, each). This unexpected diversity argues that the critical factor in selecting the top scored miR-Trio strongly dependent of the graph topology rather than on the identity of the individual nodes in the pathway. Inspecting the overlap among all 133 pathways (having DIS > 0) indicates that no two pathways share an identical miR-Trio.

3.7 Minimal miRNA combinations are associated with maximal graph fragmentation

From thousands of miR-Trios combinations, only few miR-Trios led to maximal DIS for a given pathway (Table 1). The number of miRNAs involved in these miR-Trios is often limited to a few combinations. Analyzing all pathways revealed that on average, only—four to five miRNAs are associated with the set of selected top scoring miR-Trios.

A similar analysis was repeated using pairs of miRNA as seed for the regulation of the pathway graphs. We refer to this analysis as miR-Duos. The overall results for the human pathways (93 KEGG and 79 from PID) resemble that obtained for the miR-Trios and for 40% of the results, the DIS values of the miR-Duo were identical to those obtained from the miR-Trio analysis. The majority of the cases are explained by the 'redundancy' in selection of miRNAs in the sets of miR-Trio. For example, in the case of 'ErbB signaling' (Supplementary Table S4), the top scoring miR-Trio includes miR-518b, miR-518d-3p and miR-551b. The miR-Duo considers only one miR-518 as a representative. In most cases, the overlap between the miR-Trios and miR-Duos is explained by the selection of the repeatedly used miRNAs among the top scoring miR-Trios (examples in Fig. 6). For example, only five top scoring miR-Trios are associated with the KEGG pathway 'Vascular smooth muscle contraction' (Supplementary Table S4). These miR-Trios combine only five miRNAs (from a maximal of 15 combinations). This finding further substantiates the dominant role of some miRNAs of the pairs or triplets. Moreover, while 39/172 pathways failed to receive positive DIS values, only 27 of these pathways were listed for the miR-Duo analysis. For 12 of the pathways the positive DIS value was satisfied by only two miRNAs (for example Cell cycle: hsa04110, Oocyte meiosis: hsa04114, PPAR signaling: hsa03320, Supplementary Table S3).

Figure 6 shows examples of the highest miR-Trios and their miRNAs. These examples are the pathways associated with—two to five miR-Trios. While there is no overlap in miRNAs among different pathways, often the same miRNA repeats in the miR-Trios of the same pathway.

For some of the examples, the same miRNA is included in three top scoring miR-Trios. Recall that while the score of top miR-Trios is identical, the actual number and identity of the targeted genes is not necessarily identical.

3.8 miR-Trio regulation of Alzheimer's disease and Wnt signaling pathways

The biological information that is encoded in each of the 172 tested pathways is extensive. The representation of the pathway as an undirected graph is clearly an over-simplification. Still, we show that even a complex pathway related to a disease state is suitable for the miR-Trio approach. The pathway of 'Alzheimer's disease' (AD) depicts one of the most studied degenerative disease pathway. The AD pathway focuses on parallel processes that involve

the mitochondrion, the endoplasmic reticulum (ER) stress, the Apolipoprotein E (APOE), the Tau phosphorylation and the amyloid precursor protein (APP) products. This is reflected by the pathway graph that is composed of eight CCs (Supplementary Table S1). Surprisingly, only one dominating miR-Trio was assigned with a maximal impact on the disconnection of the pathway graph (miR-487a, miR-452 and miR-625). The fragmentation of the pathway by this miR-Trio covers all known AD related sub-processes.

The Wnt signaling pathway represents another complex pathway that is subjected to miR-Trio manipulation. Wnt signaling operates through the activation of membrane receptors by the Wnt soluble ligands. Three redundant processes are known: the canonical pathway through β-catenin, the non-canonical through Rho- and Jnk-associated kinases and the activation of intracellular calcium sensitive kinases (PKC and CAMKII). While over 22 000 miR-Trios were tested for maximal DIS, only three miR-Trios were selected that cover six miRNAs (Supplementary Tables S3 and S4).

Further support for the susceptibility of the Wnt signaling and AD pathways to miR-Trios comes from testing the mouse KEGG pathways. Recall that while the proteins of the pathways are mostly close homologs, the regulation by miRNAs is not necessarily equivalent as the 3′-untranslated region of human and mouse homologs differs significantly. However, although the miR-Trios are not identical for the mouse and human similar pathways, in both organisms these are among the most sensitive pathways. Specifically, the mouse Wnt signaling and AD pathways were ranked 4 and 5 whereas in the human they were ranked 7 and 10, respectively.

Recall that in most instances, there is no obvious combination of miRNAs that are effective toward a specific pathway. A selection of the top scoring miR-Trios is often sensitive to the parameters of the search. Specifically, Table 2 shows the application of miRror for the Wnt signaling pathway using several statistical thresholds and the selection of different set of prediction resources (DBs). Note that the parameters used drastically change the search space, but the number of top miRNAs among the top-Trios remains limited (3–7). We mark the miRNAs that are shared in the resulting top miR-Trios (named dominating miRs). Note that only miR-487b appears as dominating miRNA after changing the parameters of the analysis (Table 2). It is shown that reducing the search space by almost 40-folds (from 20 835 to 560) led to a marginal reduction in DIS (from 0.753 to 0.721). We conclude that the space of solutions for disconnecting a pathway graph is rich enough to select numerous comparable miR-Trios.

4 DISCUSSION

In this study, we combine the importance of regulatory pathway graph topology with the concept of coordinated regulation by miRNAs. We test the notion that a small set of a combinatorial miRNA action constitutes an effective seed for pathway's manipulations. The protocol applied in this study introduces several sequential filters for the selection of few miRNAs that are potentially effective toward a specific pathway.

As a byproduct of this study, we identified human pathways that are not susceptible for miRNA regulation and others that are sensitive. The later serve as an unbiased resource to test the rebalance of miRNA combination in health and disease.

4.1 miRror algorithm drastically reduces the noise of miRNA-target prediction

Matching a pathway with relevant miRNAs presents a challenging problem. It is estimated that between 30% and 50% of the genes in mammalian genomes are candidates for miRNAs regulation (Friedman *et al.*, 2009; Lewis *et al.*, 2003). A high level of false predictions by miRNA-gene target predicting tools, and the low coherence among them (Mendes *et al.*, 2009) lead to an uncertainty in assigning miRNAs to a specific pathway. Often each miRNA is associated with hundreds to thousands of predicted targets (Shomron *et al.*, 2009). Consequently, looking up on miRNA-targets pairs is impractical. Applying a strict statistical significance threshold such as the one provided by the miRror is critical in reducing false predictions (Friedman *et al.*, 2010).

In recent years several databases and tools were developed with the notion that miRNAs regulate diseases and pathways (Papadopoulos *et al.*, 2009a). However, most of these tools list hundreds of miRNA predictions. For example, the number of the miRNAs associated with 'Wnt signaling' and 'Insulin signaling' based on miRWalk (www.ma.uni-heidelberg.de/apps/zmf/mirwalk) is 572 and 657, respectively. In contrast, our final step of the procedure associated these pathways with six and four miRNAs, respectively, (Fig. 6).

Our systematic search presents the most likely miRNA combinations according to their ability to severely alter the integrity of the target pathway. The apparent ranking of disconnecting human pathways by miRNAs supports the notion that the pathway topology is the prime determinant.

4.2 Some human pathways are resistant to fragmentation by small sets of miRNAs

A substantial fraction of the pathways (39/172) is not suitable for the miR-Trio approach. We propose that many of these pathways can still be subjected to a miRNA-based regulation. Indeed, when miRror was applied with a relaxed threshold (P-value <0.05), $\sim$50% of the pathways that initially were not suitable for our approach appeared legitimate. In accord with the observation seen in Table 2, these pathways are associated with small number of top scoring miR-Trios. The properties of these pathways are not different from the other human pathways. We conclude that a small set of miR-Trios can lead to fragmentation of most (87%) of the known human pathways.

The susceptibility of human pathway to miRNA regulation was proposed based on the enrichment in the candidate genes function (Cui *et al.*, 2006). We confirm that most pathways are indeed susceptible to miRNA regulation. When a search is performed in view of the capacity of the miR-Trios to cause a fragmentation of specific pathways, only a small set of combinations is retrieved. Recent technological advances for introducing several miRNAs in a controllable manner were presented (Brown and Naldini, 2009).

4.3 The gap between the miR-Trio list and network perturbation

We will not discuss the experimental implication that follows the rational choice of miR-Trios. The endogenous level of miRNA expression and the expression levels of the cognate endogenous targets are an important determinant for miRNAs regulation (Arvey *et al.*, 2010; Salmena *et al.*, 2011). A quantitative view on

Table 2. miR-Trios selection for KEGG Wnt signaling pathway

miRror *P*-val	Number of DBs	Number of tested miRs	Number of miR-Trios tested	Top mode	Top score	Number of top scoring Trios	Number of miRs in top Trios	Dominant miR
0.05	15	51	20 825	PAS	0.753	2	5	485-3p
0.05	7	38	8436	PAS	0.785	4	5	219-5p, 487b
0.01	15	28	3276	PAS	0.774	3	5	511, 144
0.01	7	16	560	PAS	0.721	11	7	330-5p
0.001	15	4	4	PAS	0.700	2	4	16, 487b
0.001	7	3	1	PAS	0.663	1	3	

miRNAs *in vivo* indicates a large difference in miRNAs occurrence (e.g. Krichevsky *et al.*, 2003). The estimated quantities of the level of miRNA expression are reported in miRBase following RPT normalization (reads per thousand) (Griffiths-Jones *et al.*, 2006). Importantly, in our analysis we are not considering the endogenous expression of the miRNAs in different tissues or physiological states. We consider the miR-Trios as a pure exogenous set of miRNAs and as a guideline for molecular manipulations (using the preferred combinations of designed miRNAs, siRNAs or other shRNAs).

The outcome from the exhaustive search for miR-Trios is the list of the most sensitive nodes in the pathway graphs. The stability of short RNAs is dependent on the hybridization stability with the target gene. In this view, miRNAs that are characterized by a partial hybridization are often more stable in the cell with respect to siRNAs that are designed to perfectly match the target gene (Pasquinelli, 2010). High specificity in manipulating a desired pathway should be considered through a siRNAs protocol (Nielsen *et al.*, 2007). Under such design, one can expect a minimal off-target effect. However, understanding the effectiveness of RNAi technology remains an active research topic. Recently, it was suggested that the lifetime of mRNAs is an important determinant in the effectiveness and failure in manipulation by short RNAs (Larsson *et al.*, 2010).

For this approach, the selected miRNAs from the top scoring miR-Trios should be considered in a cellular context. In practical terms, there are cases where the subjected cells do not express any of the miR-Trios under consideration. In addition, recent observations regarding miRNA as part of a larger network of interaction cannot be ignored [see discussion in (Brown and Naldini, 2009; Ivanovska and Cleary, 2008; Krek *et al.*, 2005)]. The main components that are sensitive to molecular and cellular manipulations include: the miRNA stability, off-target (van Dongen *et al.*, 2008), dilution effects (Arvey *et al.*, 2010; Linsley *et al.*, 2007), sponge and competition among targets (Peter, 2010; Salmena *et al.*, 2011), cross-mapping (Guo *et al.*, 2011) and negative design (Arvey *et al.*, 2010). Argonaute immunopurification experiments reveal that endogenous Ago2 is occupied with tens of miRNAs, some of which are at exceptionally high amounts. Thus, compartalization must play a key role in determine specificity (Azuma-Mukai *et al.*, 2008).

4.4 The search space of miR-Duos and miR-Trios

Figure 6 and Supplementary Tables S3–S4 list the top miR-Trios but also the expansion of choosing miR-Trio among the top 5% of the results. For about 50 pathways, such relaxation expands the selection of miRNAs substantially (by $\geq$2-folds). For example in 'Chronic myeloid leukemia' (KEGG, hsa05220), only two top scored miR-Trios were proposed, but nine miR-Trio combinations are suggested among the top 5% of the results. As shown in Table 2, the selection of miR-Trios and miR-Duos is sensitive to the parameters used in the protocol. A relaxed inspection of the top 5% of the results is important to (i) extend the list of valid combinations, and (ii) identifying instances for which the relaxation of the parameter (5% top scoring) had no effect. Interestingly, for 34% (DIS $>$ 0) the relaxation for 5% had no effect for miR-Trios or miR-Duos (KEGG pathways, 27/79 and 24/71 for miR-Duos and miR-Trios, respectively). However, on average, the ratio of number of miR-combinations with the highest DIS relative to their number in the top 5% is 2.7 and 3.5 for miR-Duos and miR-Trios, respectively. In a few instances, the expansion was dramatic. For example, the 'JAK-STAT signaling pathway' reported seven top DIS miR-Duos and 839 when 5% top DIS is considered (18 and 17 000 for miR-Trios, respectively). We calculated a high correlation between the expanded lists for the miR-Duos and miR-Trios (*t*-test *P*-value = 0.0061).

We started our study by an exhaustive search for miR-Trios. However it was apparent that for a substantial number of instances, reducing the effective number of miRNAs to pairs or even to apparently a single miRNA is logical. For example, four miR-Trios are listed for Circadian rhythm-mammal (Supplementary Table S4). Among the four miRNAs involved, three belong to the miRNA has-mir-515 family (hsa-miR-519c-3p, hsa-miR-519a, hsa-miR-519b-3p). Applying a miRNA family compression scheme reduces the number of miR-Trios and the miRNAs that need to be considered in analyzing the impact of miRNAs regulation on the pathway.

Finally, we explored the relatively high susceptibility of most pathways to a small set of miR-Trios. We propose that this is not a reflection of targeting only a few commonly used signaling genes. Instead, the topological property of the pathway graph representation is a major factor in the selection of top scoring miR-Trios. Incorporation of a network consideration to the miRNA regulation is discussed in view of the gene interaction and epistasis (Phillips, 2008). Our study adds an additional layer of combinatorial view on small sets of miRNAs in view of the topology robustness of human pathways.

ACKNOWLEDGEMENTS

We thank Nati Linial for his advice and suggestions throughout the project. G.N. and O.B. are recipients of the SCCB fellowship, the Sudarsky Center for Computational Biology.

Funding: The work is partially supported by the ISF 592/07 and the BSF 2007219.

Conflict of Interest: none declared.

REFERENCES

Ambros,V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
Arvey,A. *et al.* (2010) Target mRNA abundance dilutes microRNA and siRNA activity. *Mol. Syst. Biol.*, **6**, 363.
Azuma-Mukai,A. *et al.* (2008) Characterization of endogenous human argonautes and their miRNA partners in RNA silencing. *Proc. Natl Acad. Sci. USA*, **105**, 7964–7969.
Bentwich,I. (2008) Identifying human microRNAs. *Curr. Top. Microbiol. Immunol.*, **320**, 257–269.
Betel,D. *et al.* (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res.*, **36**, D149–D153.
Brodersen,P. and Voinnet,O. (2009) Revisiting the principles of microRNA target recognition and mode of action. *Nat. Rev. Mol. Cell. Biol.*, **10**, 141–148.
Brown,B.D. and Naldini,L. (2009) Exploiting and antagonizing microRNA regulation for therapeutic and experimental applications. *Nat. Rev. Genet.*, **10**, 578–585.
Chi,S.W. *et al.* (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486.
Chiromatzo,A.O. *et al.* (2007) miRNApath: a database of miRNAs, target genes and metabolic pathways. *Genet. Mol. Res.*, **6**, 859–865.
Cui,Q. *et al.* (2006) Principles of microRNA regulation of a human cellular signaling network. *Mol. Syst. Biol.*, **2**, 46.
Cullen,B.R. (2009) Viral and cellular messenger RNA targets of viral microRNAs. *Nature*, **457**, 421–425.
Du,L. *et al.* (2009) miR-93, miR-98, and miR-197 regulate expression of tumor suppressor gene FUS1. *Mol. Cancer Res.*, **7**, 1234–1243.
Friedman,L.M. *et al.* (2009) MicroRNAs are essential for development and function of inner ear hair cells in vertebrates. *Proc. Natl Acad. Sci. USA*, **106**, 7915–7920.
Friedman,Y. *et al.* (2010) MiRror: a combinatorial analysis web tool for ensembles of microRNAs and their targets. *Bioinformatics*, **26**, 1920–1921.
Griffiths-Jones,S. *et al.* (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
Guo, L. *et al.* (2011) Cross-mapping events in miRNAs reveal potential miRNA-mimics and evolutionary implications. *PLoS One*, **6**, e20517.
Gusev,Y. *et al.* (2007) Computational analysis of biological functions and pathways collectively targeted by co-expressed microRNAs in cancer. *BMC Bioinformatics*, **8** (Suppl. 7), S16.
Hausser,J. *et al.* (2009) MirZ: an integrated microRNA expression atlas and target prediction resource. *Nucleic Acids Res.*, **37**, W266–W272.
Hon,L.S. and Zhang,Z. (2007) The roles of binding site arrangement and combinatorial targeting in microRNA repression of gene expression. *Genome Biol.*, **8**, R166.
Ideker,T. and Sharan,R. (2008) Protein networks in disease. *Genome Res.*, **18**, 644–652.
Ivanovska,I. and Cleary,M.A. (2008) Combinatorial microRNAs working together to make a difference. *Cell Cycle*, **7**, 3137–3142.
John,B. *et al.* (2004) Human microRNA targets. *PLoS Biol.*, **2**, e363.
John,B. *et al.* (2006) Prediction of human microRNA targets. *Methods Mol. Biol.*, **342**, 101–113.
Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
Kertesz,M. *et al.* (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
Krek,A. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.
Krichevsky,A.M. *et al.* (2003) A microRNA array reveals extensive regulation of microRNAs during brain development. *RNA*, **9**, 1274–1281.
Landgraf,P. *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401–1414.
Larsson,E. *et al.* (2010) mRNA turnover rate limits siRNA and microRNA efficacy. *Mol. Syst. Biol.*, **6**, 433.
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.
Maragkakis,M. *et al.* (2009) DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acids Res.*, **37**, W273–W276.
Maziere,P. and Enright,A.J. (2007) Prediction of microRNA targets. *Drug Discov. Today*, **12**, 452–458.
Mendes,N.D. *et al.* (2009) Current tools for the identification of miRNA genes and their targets. *Nucleic Acids Res.*, **37**, 2419–2433.
Mu,P. *et al.* (2009) Genetic dissection of the miR-17∼92 cluster of microRNAs in Myc-induced B-cell lymphomas. *Gene. Dev.*, **23**, 2806–2811.
Nielsen,C.B. *et al.* (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA*, **13**, 1894–1910.
Ogata,H. *et al.* (1999) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **27**, 29–34.
Papadopoulos,G.L. *et al.* (2009a) DIANA-mirPath: integrating human and mouse microRNAs in pathways. *Bioinformatics*, **25**, 1991–1993.
Papadopoulos,G.L. *et al.* (2009b) The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Res.*, **37**, D155–D158.
Pasquinelli,A.E. (2010) Molecular biology. Paring miRNAs through pairing. *Science*, **328**, 1494–1495.
Peter,M.E. (2010) Targeting of mRNAs by multiple miRNAs: the next step. *Oncogene*, **29**, 2161–2164.
Phillips,P.C. (2008) Epistasis–the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.*, **9**, 855–867.
Salmena,L. *et al.* (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*, **146**, 353–358.
Schaefer,C.F. *et al.* (2009) PID: the pathway interaction database. *Nucleic Acids Res.*, **37**, D674–D679.
Shomron,N. *et al.* (2009) An evolutionary perspective of animal microRNAs and their targets. *J. Biomed. Biotechnol.*, **2009**, 594738.
Stelling,J. *et al.* (2004) Robustness of cellular functions. *Cell*, **118**, 675–685.
Thomas,M. *et al.* (2010) Desperately seeking microRNA targets. *Nat. Struct. Mol. Biol.*, **17**, 1169–1174.
Tu,K. *et al.* (2009) Combinatorial network of primary and secondary microRNA-driven regulatory mechanisms. *Nucleic Acids Res.*, **37**, 5969–5980.
van Dongen,S. *et al.* (2008) Detecting microRNA binding and siRNA off-target effects from expression data. *Nat. Methods*, **5**, 1023–1025.
Ventura,A. and Jacks,T. (2009) MicroRNAs and cancer: short RNAs go a long way. *Cell*, **136**, 586–591.
Wang,X. (2008) miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA*, **14**, 1012–1017.
Wienholds,E. and Plasterk,R.H. (2005) MicroRNA function in animal development, *FEBS Lett.*, **579**, 5911–5922.
Xiao,F. *et al.* (2009) miRecords: an integrated resource for microRNA-target interactions, *Nucleic Acids Res.*, **37**, D105–D110.

APPENDIX

An interactive website with the application miRrorNet provides an upload for any of the 172 human pathways collected from KEGG and PID resources. The website allows using any of the listed KEGG pathways to retrieve the top scoring miR-Trios. The user can select the default or change the parameters of miRrorNet.

Supplementary Table S1: KEGG human pathways (93) and their pathway graphs quantitative information.

Supplementary Table S2: Results of miRror candidate miRNAs using P-value <0.01 for all KEGG human pathways and for all three scoring methods.

Supplementary Table S3: A unified analyzed list of 172 human pathways from of KEGG and PID.

Supplementary Table S4: A list of top scoring miR-Trios.

miRrorNet: protonet.cs.huji.ac.il/mirror/mirnet.php