

RESEARCH

Open Access

# A structural-based statistical approach suggests a cooperative activity of PUM1 and miR-410 in human 3'-untranslated regions

Limor Leibovich<sup>1</sup>, Yael Mandel-Gutfreund<sup>2\*</sup>, Zohar Yakhini<sup>1,3\*</sup>

## Abstract

**Background:** Micro (mi)RNAs comprise a large family of small non-coding RNAs that are thought to regulate a large fraction of protein-coding genes. Generally, miRNAs downregulate messenger (m)RNA expression by binding to the 3' untranslated regions (UTRs) of the RNA molecules. An important factor for binding specificity is the matching in the seed region. In addition, target site accessibility is thought to be crucial for efficient repression of miRNA targets. Several recent studies indicated that miRNA repression can be facilitated by RNA-binding proteins. In this study, we examine the conjecture that RNA-binding proteins are involved in ushering miRNAs to bind targets that are initially less accessible.

**Results:** We analyzed human 3'-UTR sequences containing potential binding sites of 153 conserved miRNA families, and ranked sequences around the sites according to their miRNA accessibility. By applying a rank-based motif search tool to these miRNA targets, we found motifs that are enriched among less accessible targets. As expected from our ranking method, most of the significant motifs were GC-rich. However, one AU-rich motif was found to be enriched among miR-410 less accessible targets. This motif resembles the Pumilio homolog 1 (PUM1) consensus binding site. We observed a stronger enrichment of the PUM1 motif in conserved targets than in non-conserved targets; moreover, the enrichment of this motif was found to be conserved in a subset of placental mammals. Further, we analyzed publicly available gene expression data, and found that the mutual expression of PUM1 and miR-410 has a greater negative influence on the expression of low accessibility targets than on other targets, an effect that was stronger than when considering both miR-410 and PUM1 separately.

**Conclusions:** Taken together, our findings suggest a cooperative relationship between miR-410 and PUM1 in regulating human highly structured 3'-UTRs. This kind of cooperation can allow a second level of regulation of such targets. Considering cases in which miRNAs bind low accessibility targets may help to improve current miRNA prediction tools and to obtain a better understanding of the mechanisms underlying miRNA regulation activity.

## Background

Micro (mi)RNAs are small RNA molecules (approximately 22 nucleotides) participating in a large variety of cellular processes in animals, plants and viruses [1-3]. miRNAs act by binding to the 3'-untranslated region (3'-UTR) of messenger (m)RNAs, forming hybrids that consist of the binding site in the 3'-UTR and of the

miRNA seed region (positions 2-8 in the miRNA) [4,5]. miRNAs regulate mRNAs through two main mechanisms: mRNA degradation and inhibition of mRNA translation [6]. It has been shown that the match between the mRNA and the miRNA seed region is important for target recognition [1,7]. However, the number of nucleotide matches in the seed is not the only factor that determines site functionality, and other factors such as site accessibility influence the target recognition [8].

Several recent reports have demonstrated that miRNA repression can be reversed or modulated by RNA-binding proteins (RBPs) interacting with the 3'-UTR of target mRNAs. It was reported that the RNA-binding

\* Correspondence: yaelm@tx.technion.ac.il; zohar\_yakhini@agilent.com

<sup>1</sup>Department of Computer Sciences, Technion - Israel Institute of Technology, Technion City, Haifa 32000, Israel

<sup>2</sup>Faculty of Biology, Technion - Israel Institute of Technology, Technion City, Haifa 32000, Israel

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

protein ELAVL1 (embryonic lethal abnormal vision-like protein 1; also known as HuR) reverses miR-122 repression of *SLC7A1* (also known as *CAT-1*) mRNA in human hepatocarcinoma cells subjected to stress [9]. This effect on *SLC7A1* is mediated by ELAVL1 translocation from the nucleus to the cytoplasm upon stress, and is accompanied by *SLC7A1* release from processing (P) bodies, structures involved in RNA metabolism, leading to active translation of the message. The process requires the association of ELAVL1 with AU-rich sequences in the 3'-UTR region of the *SLC7A1* mRNA, through an as yet unknown mechanism. RBP modulation of miRNA-mediated repression has also been reported for dead end homolog 1 (DND1). In zebrafish, it alleviates miR-430 repression of *nanos1* and *tdrd7* in primordial germ cells [10,11]. DND1 can also relieve the repression of cyclin-dependent kinase inhibitor 1B (*CDKN1B*) mediated by miR-221 and the repression of *LATS2* by miR-372 in HEK293T cells [10]. In zebrafish and in humans, DND1 counteracts miRNA-mediated repression by binding to uridine-rich regions located near the miRNA binding sites within the 3'-UTR of the message. DND1 binding to these sequences might interfere with miRNA-mRNA interaction. Another indication of the functional relationship between miRNAs and RBPs was found in the rat hippocampal neurons, for which treatment with brain-derived neurotrophic factor (BDNF) was shown to partially relieve miR-134 mediated repression of *Limk1* [12]. When miRNAs regulate mRNAs, they are assembled into ribonucleoprotein complexes known as the miRNA-induced silencing complex (miRISC) [2]. In contrast to ELAVL1, DND1 and BDNF, which relieve the miRNA repressive function, the TRIM-NHL protein family (NHL-2 in *Caenorhabditis elegans* [13] and TRIM32 in mouse [14]) increase the activity of specific miRNAs, including let-7, by binding to miRISC components.

Pumilio family (PUF) proteins constitute a highly conserved family of RNA-binding proteins that regulate target mRNAs via binding to their 3'-UTRs [15]. PUF proteins are vital in developmental processes, including stem cell maintenance [16-18]. They are also required for long-term memory, and control neuron excitability and development [19-21]. PUF proteins bind specific RNA sequences in 3'-UTRs that contain a core 'UGUR' tetranucleotide followed by sequences that vary between members of this family. mRNA-PUF protein complexes are thought to trigger translational repression or promote mRNA degradation [22-24]. PUF proteins have been recently shown to be associated with miRNAs. It was observed that predicted miRNA binding sites are enriched among validated PUF targets near PUF-binding motifs in humans [25]. In *C. elegans*, the Pumilio homolog PUF-9 is suggested to cooperate with let-7 family

members to repress *hbl-1* in the hypodermis and the ventral nerve cord [26]. This repression requires a region of the *hbl-1* 3'-UTR that contains binding sites for PUF and let-7.

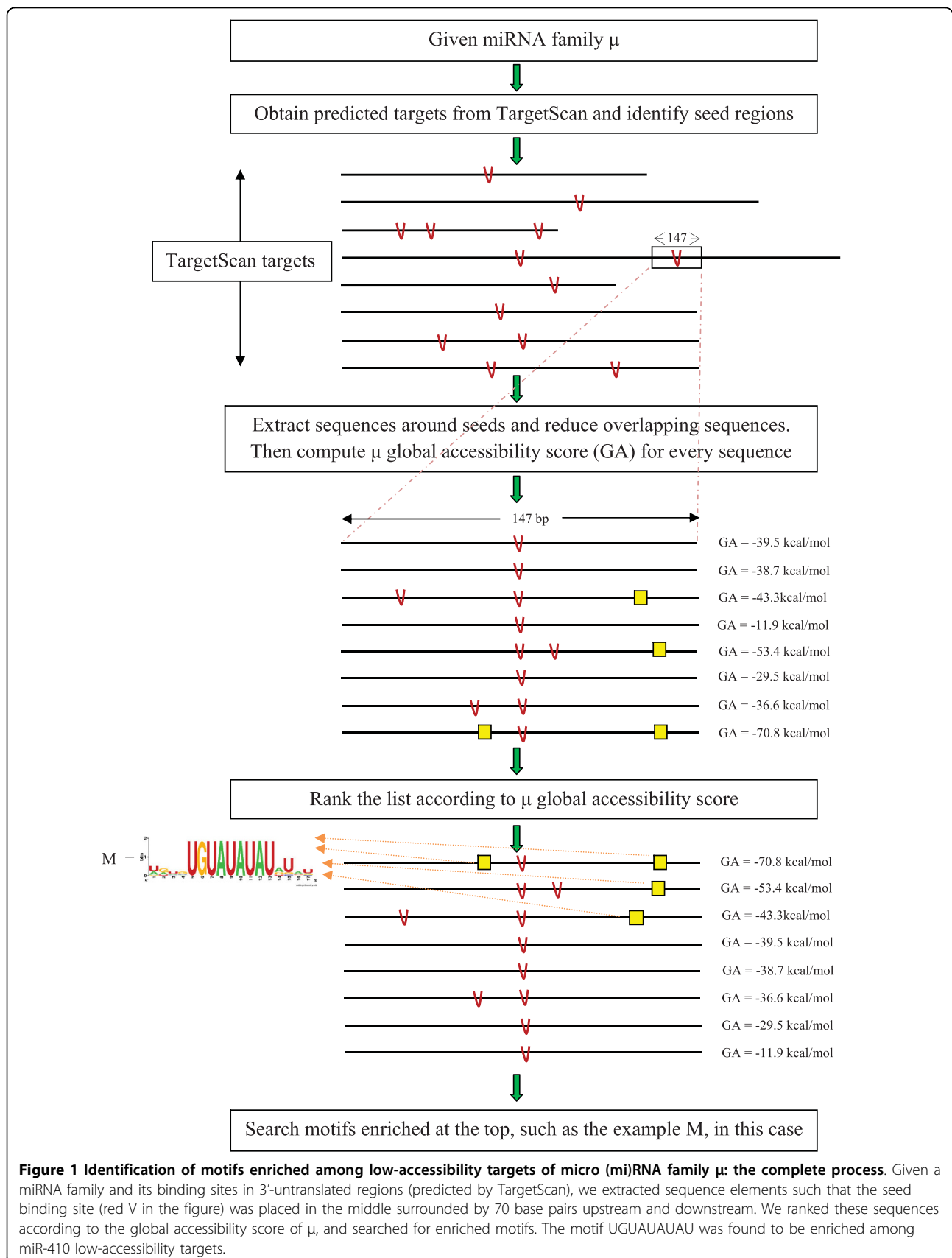
Overall, many examples suggest extensive crosstalk between RBPs and miRNAs. It is likely that additional cases of RBPs modulating miRNA interactions exist. Because efficient repression of miRNA targets is strongly dependent on site accessibility [8], RNA-binding proteins might function as ushers that mediate the opening of the structure, thereby allowing interaction between miRNA and its low-accessibility targets. In this study, we describe a computational approach to seeking evidence for such a mechanism. The approach makes use of a statistical process that includes thermodynamics-based ranking. We highlight one of the cases for which we found significant evidence. Our findings suggest a cooperative mechanism associating the RNA-binding protein Pumilio homolog 1 (PUM1) with miR-410 targeting of low-accessibility target sites in human 3'-UTRs. We found enrichment of the PUM1 binding motif in less accessible miR-410 targets. This association between miR-410 and PUM1 in the context of low-accessibility targets was also significant in other placental mammals (chimpanzee and horse). Furthermore, as a sequence-independent test, we analyzed publicly available gene expression data. We found an inverse relationship between the mutual expression profile of PUM1 and miR-410, and between the expression profiles of the least accessible targets. This inverse relationship was significantly stronger for the combination of PUM1 and miR-410 than for each of them separately. To summarize, by demonstrating a significant association between PUM1 binding sites and highly structured miR-410 targets, our findings suggest that this pair of RBP and miRNA may work together to repress low accessibility targets. Further experiments will be needed to prove this suggested mechanism.

## Results

Our conjecture in this work was that RNA-binding proteins might assist miRNAs in their repression of low-accessibility miRNA targets, with the RBP binding mediating the opening of the secondary structure, thus allowing the miRNA to access the mRNA. This kind of cooperation between RBP and miRNA requires a region of the target 3'-UTR to contain binding sites for both RBP and miRNA. The approach we developed for the exploration of such cooperative pairs of RBP and miRNA is described below.

## Approach description

To seek cooperative mechanisms for miRNA activity from sequence data, we implemented a computational process as follows (described schematically in Figure 1):



**Figure 1 Identification of motifs enriched among low-accessibility targets of micro (mi)RNA family  $\mu$ : the complete process.** Given a miRNA family and its binding sites in 3'-untranslated regions (predicted by TargetScan), we extracted sequence elements such that the seed binding site (red V in the figure) was placed in the middle surrounded by 70 base pairs upstream and downstream. We ranked these sequences according to the global accessibility score of  $\mu$ , and searched for enriched motifs. The motif UGUUAUUAU was found to be enriched among miR-410 low-accessibility targets.

1) For a given miRNA family  $\mu$ , we obtained all the conserved predicted targets of  $\mu$ , including positions of the seed in the 3'-UTR. The predictions were taken from TargetScan [27-29].

2) For every seed occurrence (red V in Figure 1), we considered 70 base pairs on each side.

At the end of this step, we obtained a collection of sequences of length 147 bp with the binding site for  $\mu$  in the middle, denoted as  $S_{1,\dots,S_{L(\mu)}}$ .

3) To reduce overlap of sequences, we reduced the collection containing  $S_{1,\dots,S_{L(\mu)}}$  using a maximal independent set algorithm (see Methods).

We performed this process for every conserved miRNA family, defined according to [29]. Consequently, for every conserved miRNA family we obtained a set of minimally overlapping conserved sequences containing the predicted target site at the centre of the sequence. The sequences contained the miRNA seed binding site surrounded by 140 bases (70 upstream and 70 downstream), which is sufficient for a reliable prediction of local secondary structures [8].

To detect sequence elements playing a role in structure-driven cooperation with the miRNA of interest given a target sequence  $S_i$ , we defined a criterion that reflects miRNA accessibility to  $S_i$ . Accessibility is reflected by the global accessibility score of each miRNA family to each  $S_i$ . The score takes into consideration both the accessibility of the entire target sequence and the local accessibility of the miRNA binding site (for more details, see Methods). We ranked the target sequences of the miRNA family according to miRNA accessibility, with the least accessible targets located at the top of the list and the more accessible targets ranked lower in the list. Furthermore, to avoid motifs that overlap with the miRNA binding site or with its reverse complement, we masked the miRNA binding site (located at the centre of the sequence) and the nucleotides that are predicted to form base pairs with it. We next sought motifs enriched among the least accessible targets in comparison with the accessible targets. This was done using DRIM [30], which was adapted for finding RNA motifs. To reduce the number of false positive results, we concentrated on relatively long motifs of length 9 bp.

#### The Pumilio binding motif is enriched in miR-410 low accessibility targets

To search for RNA-binding motifs that may be associated with low accessibility target sites, we repeated the process described above for the 153 conserved miRNA families known for humans [29]. We found 163 enriched motifs for the 153 miRNA families that passed the minimum hypergeometric (mHG) score threshold of  $10^{-4}$  (see Methods). Every motif is associated with a miRNA

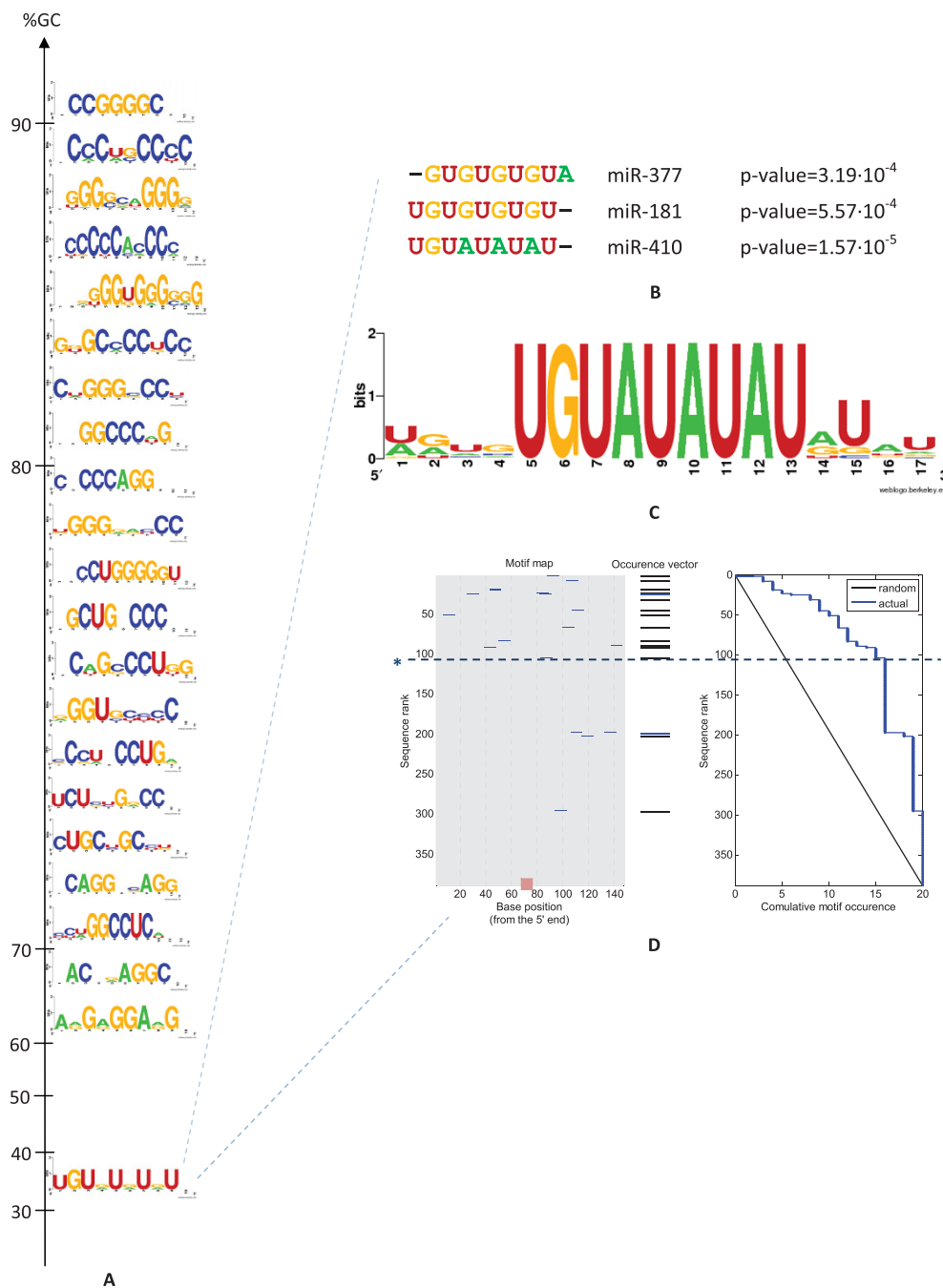
family  $\mu$ , such that it is enriched among the least accessible targets of  $\mu$ . As expected from our accessibility score and ranking approach, sequences appearing at the top of the ranked list tended to have greater GC content than did those located at the bottom of the list (see Additional file 1, Figure S1). Therefore, we expected to find enrichment of GC-rich motifs in low-accessibility targets.

Furthermore, we clustered the identified motifs into groups based on sequence similarity, considering only clusters containing  $\geq 3$  motifs (Figure 2A). As control, we conducted the process described above for *Saccharomyces cerevisiae* 3'-UTRs, working with human miRNAs and using a similar number of targets as in humans. Because the miRNA machinery is not known to exist in *S. cerevisiae*, we considered the results we obtained here as being random (or as negative control). In the *S. cerevisiae* control, we found 33 enriched motifs versus 163 motifs in humans (at  $mHG \leq 10^{-4}$ ). Clustering the *S. cerevisiae* motifs yielded clusters containing only one or two motifs (see Additional file 1, Figure S2 for the mHG scores of human motifs versus *S. cerevisiae* motifs; for more details on this control, see Methods).

Among the results for humans, we detected one exceptional cluster having a relatively low GC content (the multiple sequence alignment for this cluster is shown in Figure 2B). In this cluster, the motif found for miR-410 (UGUAUUAU) contained only one G (11.1% GC). Interestingly, no motif having such low GC content was found in the *S. cerevisiae* control (at  $mHG \leq 10^{-4}$ ). This motif perfectly contains the consensus binding site of PUM1 and PUM2, which is UGUAHAUA [25,31], suggesting an association between the RNA-binding proteins of the Pumilio family and the miR-410 low accessibility targets (the motif is shown in Figure 2C, its enrichment among miR-410 low accessibility targets is illustrated in Figure 2D, and the low accessibility targets containing the motif are listed in Table 1; see Additional file 1, Figure S3 for their structures).

#### Controls

To further study the proposed association, we conducted a list of control experiments, described below. Previous studies have shown that miRNA binding sites are occasionally found in multiple copies in 3'-UTRs [32,33]. Multiplicity of miRNA binding sites has been suggested to be correlated with degree of repression. Because the presence of multiple binding sites of miR-410 in the 3'-UTRs could influence our results, we checked whether binding site multiplicity was correlated with our ranking for miR-410 targets. We found that miR-410 binding site multiplicity did not correlate with the accessibility ranking of miR-410 predicted targets



**Figure 2 Clustered motifs found to be enriched among micro (mi)RNA low-accessibility targets in humans. (A)** The motifs found to be enriched among miRNAs low-accessibility targets were clustered according to sequence similarity. Clusters containing at least three motifs are shown. For each cluster, we calculated the average GC content over all motifs in the cluster. The clusters were sorted according to their GC content. Note that the GC content axis is variable. **(B)** The multiple sequence alignment of the motifs belonging to the cluster with the lowest GC content is shown. For each member of the cluster, the miRNA family for which it was found and the enrichment *P* value are mentioned. **(C)** The logo of the motif enriched among miR-410 least accessible targets is shown. This motif holds the lowest GC content found. **(D)** On the left, the occurrences of UGUUAUUAU among miR-410 predicted targets are shown. The targets were ranked according to the miR-410 global accessibility score. The location of the miRNA binding site is marked with a pink rectangle on the x axis. In the middle, the occurrences vector plot illustrates the number of motif occurrences in each sequence (black for one occurrence, blue for two occurrences). On the right, the actual motif accumulated occurrences versus the expected motif accumulated occurrences in a random dataset are shown, highlighting the observed enrichment. The dashed line (and asterisk) indicates the minimum hypergeometric (mHG) cutoff used for partitioning the sequences into two subsets (such that the motif is enriched in the upper subset).

**Table 1 MiR-410 low accessibility targets containing the motif UGUUAUUAU**

Gene symbol	Gene name	PANTHER molecular function
<i>FAM120C</i>	Family with sequence similarity 120C	Unclassified
<i>C14orf102</i>	Chromosome 14 open reading frame 102	Unclassified
<i>NRIP1</i>	B-cell CLL/lymphoma 11B (zinc finger protein)	Zinc finger transcription factor Nucleic acid binding
<i>DGKG</i>	Diacylglycerol kinase, gamma 90 kDa	Kinase
<i>RAI1</i>	Retinoic acid induced 1	Transcription factor
<i>MCAM</i>	Melanoma cell adhesion molecule	CAM family adhesion molecule
<i>STARD13</i>	START domain containing 13	Other G-protein modulator
<i>CELSR1</i>	Cadherin, EGF LAG seven-pass G-type receptor 1 (flamingo homolog, <i>Drosophila</i> )	G-protein coupled receptor Cadherin
<i>ZC3H7B</i>	Zinc finger CCCH-type containing 7B	Nucleic acid binding
<i>NDST1</i>	<i>N</i> -deacetylase/ <i>N</i> -sulfotransferase (heparan glucosaminyl) 1	Other transferase
<i>FZD5</i>	Frizzled homolog 5 ( <i>Drosophila</i> )	Molecular function unclassified
<i>HACE1</i>	HECT domain and ankyrin repeat containing, E3 ubiquitin protein ligase 1	Ubiquitin-protein ligase
<i>TEC</i>	Tec protein tyrosine kinase	Non-receptor tyrosine protein kinase
<i>PRRX1</i>	Paired related homeobox 1	Homeobox transcription factor Other DNA-binding protein

Fourteen genes at the subset of top 105 low accessibility targets contain the motif UGUUAUUAU. Their gene symbols, gene names and PANTHER [49] molecular functions are shown.

(Pearson correlation coefficient = 0.062), thereby ruling out multiplicity as a driver for the aforementioned enrichment (for more details, see Methods).

An alternative explanation for the observed enrichment of the Pumilio motif among low accessibility miR-410 targets is that Pumilio binding sites are generally more prevalent in highly folded targets. To test this, we calculated the enrichment of the Pumilio consensus motif [25,31] in UTR sequences ranked according to  $\Delta G$  calculated from the predicted secondary structure. Interestingly, we did not observe any enrichment of the Pumilio motif in highly structured UTRs ( $P = 0.24$  for the best enrichment among 100 repetitions; for more details, see Methods). These results reinforce the association between miRNA accessibility and the observed enrichment of the Pumilio motif in the least accessible miR-410 targets.

To examine whether Pumilio binding sites are generally enriched among GC-rich sequences, we ranked the predicted targets of each miRNA family in our list of conserved miRNA families according to their GC content, and recalculated the enrichment of the Pumilio consensus [25,31]. Among the 153 conserved miRNA families (including miR-410), the best enrichment found for the Pumilio motif had a  $P$  value of 0.21, thus excluding this explanation.

Furthermore, as a control for our ranking approach, we took the 100 least accessible targets and 100 most accessible targets (of miR-410) and searched for enriched motifs in each subset, using MEME [34]. We masked the miRNA binding site and its complement to avoid enrichment of motifs derived from the miRNA

binding site. In the subset of least accessible targets, the most enriched motif found was an AU-rich motif that is similar to the Pumilio motif (its regular expression is AU[AG][UC]AUAUAUAUAUA; e-value =  $1.4 \times 10^{-30}$ ). This motif (or any similar motif) was not found for the subset of accessible targets. Moreover, the best e-value per motif in the latter subset was  $7.3 \times 10^{-6}$ .

To assess whether the observed association between the Pumilio motif and the predicted, least accessible miR-410 sites could reflect a functional relationship between the RBP and the miRNA, we generated a subset of predicted miR-410 targets that are evolutionarily conserved and thus more likely to be functional miR-410 sites [35]. We then compared the enrichment of the Pumilio motif among predicted targets of human miR-410 taken from three datasets: conserved targets, targets with no restriction on conservation, and non-conserved targets. We found that Pumilio was most significantly enriched in the conserved dataset ( $P = 1.57 \times 10^{-5}$ ), whereas it was less enriched in both the miR-410 predicted targets with no restriction on conservation ( $P = 2.15 \times 10^{-3}$ ) and in the non-conserved dataset ( $P = 4.2 \times 10^{-2}$ ; for more details, see Methods). The higher enrichment of Pumilio binding sites among the conserved, least accessible miR-410 targets may indicate that this association is related to miRNA function.

To further validate that the Pumilio motif is functionally related to miR-410, we used TargetScan to predict a set of miR-410 pseudo-targets (computed based on sequence match with the miRNA seed) in the 3'-UTRs of organisms lacking miR-410 (*C. elegans*) or miRNAs in general (*S. cerevisiae*). We then calculated the

enrichment of the motif among low accessibility targets for each organism (using broadly the same process described above). As expected, the Pumilio motif was not found to be enriched in any of these organisms (*S. cerevisiae*:  $P = 0.35$ , *C. elegans*:  $P = 0.76$ ). Next, we applied the latter test to placental mammals in which miR-410 is conserved [29]. For this analysis, we used the dataset of conserved targets predicted by TargetScan for chimpanzee, horse and dog. Here, again, the Pumilio motif was found to be enriched among miR-410 low-accessibility targets, specifically in chimpanzee ( $P = 5.28 \times 10^{-4}$ ) and horse ( $P = 1.4 \times 10^{-4}$ ), but it was only weakly enriched in dog ( $P = 3.3 \times 10^{-2}$ ). It is important to note that these organisms were chosen because the sizes of their datasets are very similar to that of the human dataset, and thus the differences in  $P$  values cannot be due to the size of the datasets.

#### Additional controls using validated Pumilio targets in humans

To further investigate the relationship observed between miR-410 and the Pumilio family and to show its dependence on target accessibility, we used experimental Pumilio binding data available for humans [25]. We obtained 3'-UTR sequences of validated Pumilio targets and ranked them according to their corresponding affinity to Pumilio, as reported previously [25]. We then calculated the enrichment of the miR-410 binding site among validated Pumilio targets. The miR-410 binding site was not found to be enriched among Pumilio targets (PUM1:  $P = 0.86$ , PUM2:  $P = 0.48$ ). As a control, we calculated the enrichment of the Pumilio recognition motif (UGUAHAUA [25,31]) among validated Pumilio targets, and found it to be strongly enriched among PUM1 targets ( $P = 2.47 \times 10^{-9}$ ) and weakly enriched among PUM2 targets ( $P = 2.7 \times 10^{-2}$ ). To summarize, the analysis presented here demonstrates that the miR-410 binding motif is not generally enriched in Pumilio targets (see Additional file 1, Figure S4). These findings indicate that the observed association of Pumilio and miR-410 cannot be explained by a general association between them. They therefore support a more specific role of Pumilio, closely associated to low-accessibility miR-410 targets.

Next, we evaluated the association of Pumilio proteins with other miRNAs in the context of low-accessibility targets using the knowledge of experimentally validated Pumilio targets [25]. We used mHG statistics to calculate the enrichment of validated Pumilio targets in the list of miRNA targets ranked according to global accessibility (see Methods). We did this for all 153 human miRNA families. Of the 153 experiments, the most significant enrichment was found for miR-410 ( $P = 1.4 \times 10^{-2}$ ; 14 out of 388 miR-410 targets are validated

Pumilio targets and all are in the list of the top 251 least accessible targets).

#### PUM1 cooperation with miR-410 based on gene expression analysis

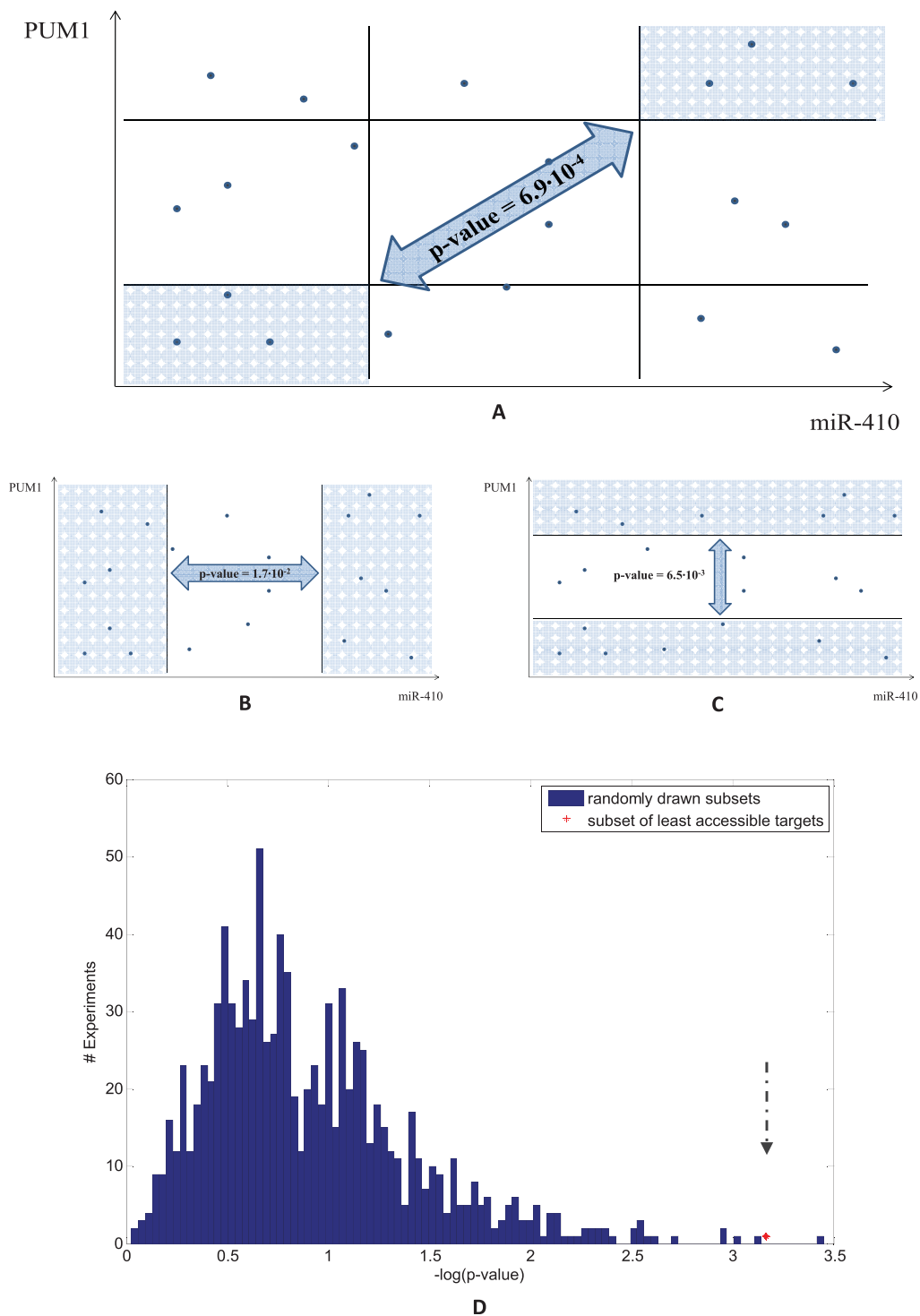
The aforementioned experiments suggest the involvement of the Pumilio family in modulating miR-410 repression of low-accessibility targets. In this section, we used gene expression data (on the NCI60 panel [36,37]) to test PUM1 cooperation with miR-410.

According to the suggested mechanism, two main observations should hold. First, in the presence of high levels of PUM1 and miR-410, there should be a stronger repression effect on the set of least accessible targets than in the presence of low levels of PUM1 and miR-410. The set of least accessible targets (see Methods) is hereby denoted by  $\Phi$ . The difference in the extent of repression comparing the two states (defined according to PUM1 and/or miR-410 expression levels) is hereafter termed as 'differential repression'. Differential repression is expected to be more dramatic for  $\Phi$  than it is for other (more accessible) miR-410 targets. Indeed, the significance of differential repression of the least accessible targets compared with the rest of the targets was  $6.9 \times 10^{-4}$  (for more details, see Methods). The second expected observation is that the differential repression of the least accessible targets,  $\Phi$ , versus the rest of the targets should be less significant than when partitioning samples according to only miR-410 or only PUM1, owing to the suggested cooperative interaction. Indeed, the mutual differential repression was ~25-fold more significant than when considering only miR-410, and ~10-fold more significant than when considering only PUM1 (Figure 3, A-C).

To validate that the mutual repressive influence of PUM1 and miR-410 is related to the low accessibility of the miRNA targets, we repeated this process 1,000 times, each time for a randomly drawn subset of genes,  $\Phi'$ , taken from the pool of miR-410 predicted targets (each  $\Phi'$  contains the same number of elements as  $\Phi$ ; each  $\Phi'$  is disjoint from  $\Phi$ ). As shown in Figure 3D, of the 1,000 experiments, only one result was better than  $6.9 \times 10^{-4}$  ( $P = 3.55 \times 10^{-4}$ ).

#### Discussion

The extent to which miRNAs interact with low accessibility targets is not clear, but if such binding takes place there could be a molecular mechanism allowing the miRNA to bind highly structured targets, possibly by involving RNA-binding proteins. To date, there is increasing evidence for cooperation between miRNAs and RBPs [9,10,12-14,26]; however, the possible role of RBPs in facilitating miRNA binding to inaccessible sites has yet not been examined carefully. In this study, we



**Figure 3 Cooperative repressive influence of PUM1 and miR-410 on least accessible targets. (A)** In this test, we compared samples having high levels of PUM1 and miR-410 expression with samples having low expression, and calculated the degree of differential repression for all targets. We found that the set of least accessible targets was significantly differentially repressed compared with other targets ( $P = 6.9 \times 10^{-4}$ ). Note that the expression values of PUM1 and miR-410 are drawn schematically. **(B,C)** We repeated the test described in (A) but now ignored PUM1 and miR-410, respectively. In each of these tests, the differential repression of the least accessible targets compared with other targets was weaker than in (A) ( $P$  values obtained are indicated in the figure). **(D)** We repeated the test described in (A) for 1,000 randomly drawn subsets instead of the subset of least accessible targets. Each random subset contained the same number of targets as the subset of least accessible targets and was disjoint from it. The differential repression of 999 subsets was less significant than  $6.9 \times 10^{-4}$ .



sought sequence motifs that are enriched among miRNA low-accessibility targets, because these motifs may point to RNA-binding proteins that take part in this kind of mechanism. We clustered the motifs found into groups based on sequence similarity (see Methods). The cluster containing the motif UGUUAUUAU that was found for miR-410 (Figure 2B) was exceptional because of its low average GC content (33%). The above motif was especially interesting because it is the well-characterized Pumilio recognition motif [25,31] and was unusually AU-rich. The other motifs in the cluster comprised UG runs known as binding motifs of the heterogenous RNP protein trans-activation-response DNA binding protein (TARDBP; also known as TDP-43) [38].

Testing the enrichment of the Pumilio motif among miR-410 low-accessibility targets using various controls strongly supported the conjecture that this association is related to the miRNA accessibility of the targets. In addition, the accessibility-dependent association between PUM1 and miR-410 was found to be conserved in humans, chimpanzee and horse, and it was more significant in the human conserved dataset than in the human non-conserved dataset, implying functionality. Moreover, analyzing publicly available gene expression data revealed that mutual expression of PUM1 and miR-410 has a greater negative influence on low-accessibility targets than on other targets. Interestingly, mutual expression of PUM1 and miR-410 had a greater negative influence than did their individual influences separately, supporting our conjecture that miR-410 and PUM1 act together, possibly by facilitating the repression of low-accessibility miR-410 targets.

In a previous study by Fiore *et al.* [21], it was demonstrated that the miRNA cluster miR379-410 (containing miR-410) is transcriptionally activated upon activation of mouse cortical neurons, and that one component of this cluster, miR-134, takes part in triggering activation-dependent dendritogenesis. The latter study further suggests that *Pum2* is a miR-134 target in this process. Based on the verified cooperation between Pumilio and the miR379-410 cluster in dendritogenesis, we speculate that the cooperation between miR-410 and Pumilio suggested in this study might be part of the process regulating dendritogenesis. Clearly, further experimental assays should be carried out to validate this observed association of miR-410 and PUM1 in human 3'-UTRs, and to explain the related mechanisms.

## Conclusions

An association between miRNAs and Pumilio was suggested previously [25], based on the observation that predicted miRNA binding sites are enriched among validated Pumilio targets near Pumilio binding motifs. In this work, we did in fact observe an association, specific

to the context of low accessibility, between Pumilio binding sites and miR-410 targets. We suggest that PUM1 and miR-410 may cooperate in repressing highly structured targets, allowing a second level of regulation of these targets. We hypothesized a mechanism in which PUM1 plays a role in ushering miR-410 to highly structured targets. It is likely that additional pairs of miRNAs and RBPs, as yet undiscovered, cooperate in a similar way. Taking into account the possibility that miRNAs can bind low-accessibility targets with the assistance of RNA-binding proteins may help in improving the accuracy of miRNA target prediction tools and in identifying novel regulatory mechanisms.

## Methods

### NCI60 dataset

This dataset comprised a panel of 60 cancer-derived cell lines. For each of the 60 cell lines, we obtained mRNAs and miRNA expression data from the literature [36,37].

### Human validated Pumilio targets dataset

Lists of PUM1 and PUM2 target mRNAs in HeLa S3 cells were obtained [25]. The dataset included gene information and numerical data related to the measured affinity between Pumilio proteins and the potential targets.

### Maximal independent set

Given a set of sequence elements  $S_1, \dots, S_L$ , which are substrings of a gene 3'-UTR, there can be overlap between sequences. Each  $S_i$  is associated with two coordinates that define its start and its end positions in the 3'-UTR. To reduce overlap between sequences, given  $S_i$  and  $S_j$ , we required that  $\text{start}(S_i) + 100$  be less than or equal to  $\text{start}(S_j)$  (that is, that the distance between the starts of every two sequences in the set must be greater than 100 nucleotides). Consider the interval graph [39]  $G = (V, E)$ , whose set of vertices is  $V = \{S_1, \dots, S_L\}$  and set of edges is  $E = \{(S_i, S_j) \mid i < j \text{ and } \text{start}(S_i) + 100 > \text{start}(S_j)\}$ . Without loss of generality, the sequences are sorted according to their starting positions, such that the starting point of  $S_1$  is minimal. To find a maximal set of minimally overlapping sequences, we need to find a maximal independent set in  $G$ . Because  $G$  is an interval graph, the optimal solution can be found efficiently [39].

### TargetScan

TargetScan predicts the biological targets of miRNAs by searching for the presence of conserved eight-mer and seven-mer sites that match the seed region of each miRNA [27-29]. As an option, non-conserved sites are also predicted. Sites with mismatches in the seed region that are compensated for by conserved 3'-pairing are also identified [29]. We used TargetScanHuman [40],

which considers matches to annotated human UTRs and their orthologs.

TargetScan provides a code that enables making custom predictions of miRNA binding sites for any arbitrary given set of syntactically valid sequences. We used it to obtain predictions in a variety of organisms (such as *S. cerevisiae*, even if they lack miRNA activity).

Conserved miRNA families were defined according to previously published data [29]. They included broadly conserved families (which are conserved across most vertebrates, usually as far as zebrafish) and conserved families (which are conserved across most mammals, but usually not beyond placental mammals).

#### RNAfold and miRNA global accessibility score

RNAfold is a software application that predicts the secondary structures of single-stranded RNA or DNA sequences [41].

Given a miRNA family  $\mu$  and a target sequence  $S_i$ , we defined the miRNA global accessibility criterion for  $S_i$  as follows:

- We calculated the free energy of the entire sequence, denoted as  $\Delta G_{\text{all}}(S_i)$ .
- We calculated the free energy of  $S_i$  when the area surrounding the seed binding site is forced to be unpaired, denoted by  $\Delta G_{\text{masked}}(S_i)$ .
- The free energy lost in opening the structure at the binding site of  $\mu$  in  $S_i$  was then defined as:

$$\Delta G_{\text{open}}(S_i) = \Delta G_{\text{all}}(S_i) - \Delta G_{\text{masked}}(S_i).$$

This number reflects the local accessibility of the miRNA binding site; the more negative this value, the greater the energy required for opening the target site secondary structure.

- The global accessibility of  $S_i$  to  $\mu$  binding is represented by:

$$\text{GA}(S_i) = \Delta G_{\text{open}}(S_i) + \Delta G_{\text{all}}(S_i) = 2\Delta G_{\text{all}}(S_i) - \Delta G_{\text{masked}}(S_i)$$

It should be noted that to calculate  $\Delta G_{\text{masked}}(S_i)$ , given a sequence  $S_i$ , we masked the area surrounding the target site in  $S_i$  (25 bases in total, with the miRNA binding site located in the middle) and calculated the free energy of the modified sequence using RNAfold. Technically, putting the letter N in any region in a sequence leads RNAfold to avoid base pairing in that region (see Additional file 1, Figure S5) and thereby masks the region.

#### DRIM

DRIM (discovery of rank imbalanced motifs) is a software application that identifies sequence motifs in lists of ranked DNA sequences [30], and it has also been adapted for RNA sequences [42]. DRIM employs a flexible

threshold statistical approach [30,43] to discover motifs that are significantly enriched at the top of a ranked list of sequences compared with the rest of the list.

The motifs returned by DRIM are seed motifs that are sufficiently significant to be passed as input to the heuristic search mechanism of DRIM. The default threshold, which is also used here, is mHG score =  $10^{-4}$  (motifs with greater scores are discarded).

#### *S. cerevisiae* control

We downloaded 3'-UTR sequences of *S. cerevisiae* from the UCSC Table Browser [44] and predicted the (imaginary) binding sites of human miRNA families within these UTRs using TargetScan script. To allow the same statistical power as in the human dataset, we used a similar number of targets per miRNA family for the tests described in this paper (if there turned out to be more sequences than needed, we would then filter the sequences in the middle of the ranked list). The number of *S. cerevisiae* targets per miRNA is the same as in humans, except for sporadic cases in which there are fewer targets in *S. cerevisiae* (on average, the difference between the number of targets in humans and the number of targets in *S. cerevisiae* per miRNA is 12).

#### Procedure for clustering motifs

Given two sequences  $s = s_1, \dots, s_n$  and  $t = t_1, \dots, t_n$ , the  $i$ -level distance between  $s$  and  $t$  is defined as  $i$  plus the number of mismatches between the sequences  $s_1, \dots, s_{n-i}$  and  $t_{i+1}, \dots, t_n$ .

Given two sequences  $s = s_1, \dots, s_n$  and  $t = t_1, \dots, t_n$ , the distance between  $s$  and  $t$  is defined as the minimal  $i$ -level distance for  $i = 0, \dots, n-1$ .

Given two sets of sequences  $C_1$  and  $C_2$ , we defined the distance between  $C_1$  and  $C_2$  as the average of distances between every two sequences  $s$  and  $t$ , such that  $s \in C_1$  and  $t \in C_2$ .

The clustering procedure we applied here is given as its input a set of sequences to cluster and a parameter denoted as diameter. It begins with clusters that are singletons; each sequence is a single cluster. It recursively merges the pair of closest clusters, and halts when the distance between the closest clusters is greater than the diameter given as a parameter. This is a variant of the nearest neighbour hierarchical clustering approach.

To produce the motif Shannon logo for a cluster, we calculated the multiple sequence alignment for the members of the cluster using ClustalW2 [45] and drew the motif using WebLogo [46].

#### Multiplicity test

To test whether multiplicity is correlated with our ranking, we counted the number of miRNA binding sites within each predicted target sequence by counting the

number of matches of the miRNA seed in the sequence. In addition to Watson-Crick base pairing, we allowed wobble base pairs (G-U). We also allowed one mismatch between the miRNA seed and its binding site and counted overlapping binding sites.

### Testing possible association between Pumilio and $\Delta G$

To calculate the enrichment of Pumilio in highly folded targets, we generated 100 sets of UTR sequences and ranked them according to their  $\Delta G$ . The length of the sequences and the size of each set were the same as for the miR-410 dataset to allow for the same statistical power. The UTR sequences in each set were chosen randomly.

### Conservation control

In this control, we used three datasets of miR-410 predicted targets in humans: conserved targets, non-conserved targets, and targets with no constraint on conservation. The conserved predicted targets dataset and the non-conserved predicted targets dataset were obtained from TargetScan. For the third dataset, we downloaded human 3'-UTR sequences from the UCSC Table Browser [44], and calculated miR-410 targets within these sequences using TargetScan script. In this control, as in the *S. cerevisiae* control, we used the same number of targets for the three datasets. We calculated the enrichment of the Pumilio motif among the target sequences of each set.

### Enrichment analysis (using mHG statistics)

An approach has been previously described [30] to identify the enrichment of a set of genes, A, in a ranked list of genes using mHG statistics. Given a total number of genes N, with B of these genes belonging to A, and n of these genes being in the target set (for example, differentially expressed genes), the probability that b or more genes from the target set are also in A is given by the hypergeometric tail (HGT):

$$P(X \geq b) = HGT(b; N, B, n) = \sum_{i=b}^{\min(n, B)} \frac{\binom{n}{i} \binom{N-n}{B-i}}{\binom{N}{B}}$$

If a ranked genes list  $g_1, \dots, g_N$  is provided in place of a target set, we define a label vector  $\lambda = \lambda_1, \dots, \lambda_N \in \{0, 1\}^N$  according to the association of the ranked genes to A, that is,  $\lambda_i = 1$  if and only if  $g_i$  is in A. The mHG score is then defined as:

$$mHG(\lambda) = \min_{1 \leq n < N} HGT(b_n(\lambda); N, B, n)$$

where

$$b_n(\lambda) = \sum_{i=1}^n \lambda_i$$

In other words, the mHG score is the optimal HGT probability that is found over all possible partitions induced by the ranking. As such, this score must be corrected for multiple testing. A dynamic programming algorithm for computing the exact P value of a given mHG score has been described previously [30]. More specifically, given a ranked list of genes, a subset A, and a corresponding mHG score s, the mHG P value tells us the exact probability of observing an mHG score  $s' \leq s$  under the null assumption that all occurrence configurations of A in the ranked list are equiprobable.

### Enrichment of Pumilio experimentally validated targets among low accessibility miRNA targets

The set of validated Pumilio targets contains 1,482 genes (comprising both PUM1 and PUM2 targets). Given a miRNA family denoted as  $\mu$ , using the mHG statistics we calculated the enrichment of Pumilio targets in the list of  $\mu$  targets ranked according to their global accessibility scores (such that least accessible targets are at the top) as follows.

1) We used the list of conserved predicted targets,  $S_1, \dots, S_N$ , ranked according to global accessibility,  $GA(S_i)$  of  $\mu$  (the calculation of  $GA(S_i)$  is described above).

2) We produced a binary vector  $\lambda(\mu)$  as follows:

For each sequence  $S_i$  in the list of predicted targets for  $\mu$ ,  $\lambda_i(\mu) = 1$  if and only if  $S_i$  was reported as a Pumilio validated target and also contained the Pumilio consensus motif (UGUAHAUA [25,31]). We required  $S_i$  to contain the Pumilio motif in addition to being a validated target because Pumilio target genes have been reported [25] without an indication of where the binding sites reside in the 3'-UTR. In addition, we required that the Pumilio binding site did not overlap with the predicted miRNA binding site or with its complement.

3) We used the mHG statistics on  $\lambda(\mu)$  to calculate the enrichment of validated Pumilio targets in low-accessibility targets of  $\mu$ .

### Differential expression

To compute the differential expression, we used the threshold number of misclassifications (TNoM) score and P value previously described [47].

### Samples classification algorithm (used in the gene expression analysis)

The expression analysis requires having two disjoint subsets (denoted as A and B) such that A contains the

samples in which PUM1 and miR-410 are highly expressed, whereas B contains the samples in which PUM1 and miR-410 are expressed at low levels. To identify an appropriate partition for this purpose, we performed a class discovery process. Let the vector  $x = x_1, \dots, x_N$  be the expression profile of miR-410 in samples 1, ..., N and let the vector  $y = y_1, \dots, y_N$  be the expression profile of PUM1 in samples 1, ..., N. Because  $(x_i, y_i)$  is the expression of miR-410 and PUM1 in the  $i^{\text{th}}$  sample, respectively, we can define the set  $P \subseteq \mathbf{R}^2$  that contains  $N$  points representing the expression of miR-410 and PUM1 in the  $N$  samples.

We sorted  $P$  by constructing a two-dimensional  $k$ -dimensional (kd)-tree ([48]) that would follow the sorting order. At the root, we split the set  $P$  with a vertical line into two subsets of roughly equal size.  $P_{\text{left}}$ , the subset of points to the left or on the splitting line was stored in the left subtree, and  $P_{\text{right}}$ , the subset to the right of the splitting line, was stored in the right subtree. At the left child of the root, we split  $P_{\text{left}}$  into two subsets with a horizontal line: the points below or on it were stored in the left subtree of the left child and the points above it were stored in the right subtree. Similarly, the subset  $P_{\text{right}}$  was split with a horizontal line into two subsets that were stored in the left and right subtrees of the right child. At the grandchildren of the root, we again split with a vertical line. In general, we split with a vertical line at nodes whose depth was even, and with a horizontal line at nodes whose depth was odd. The algorithm terminates when all the subtrees are leaves. Scanning the leaves from left to right produces the sorted list.

The construction of the kd-tree uses  $O(N)$  storage and takes  $O(N \log N)$  time.

Having the sorted list of points  $p_1, \dots, p_N$ , such that  $p_1 < p_2 < \dots < p_N$ , we defined the following configurations:

$$S_k = \{ (A_k, B_k) \mid A_k = \{p_{N-k+1}, \dots, p_N\} \text{ and } B_k = \{p_1, \dots, p_k\} \}, k = 1, \dots, \left\lfloor \frac{N}{2} \right\rfloor.$$

Given a figure of merit, we calculated the figure of merit for all the configurations and took the configuration  $S_k = \{A_k, B_k\}$  that held the optimal value.

We now explain the figure of merit used in our process. Consider a set  $\Phi$  of mRNAs of interest (for example, miR-410 least accessible targets). We evaluated the differential expression of  $\Phi$  due to miR-410 and PUM1 expression levels as follows.

1) Consider a configuration  $A, B$ .

2) For every miR-410 predicted target  $\tau$  we calculated the TNoM  $P$  value, measuring whether its expression in  $A$  was lower than its expression in  $B$  (termed as the differential repression of  $\tau$ ).

3) We ranked miR-410 targets according to their TNoM  $P$  values in increasing order.

4) We calculated the enrichment of the genes of  $\Phi$  at the top of this ranked list.

5) Finally, we selected the best configuration  $A, B$  using the enrichment calculated in the previous step as the figure of merit.

#### Definition of least accessible targets

Given a list of miRNA targets ranked according to miRNA accessibility, the set of least accessible targets (denoted as  $\Phi$ ) was defined as the top 20% targets of the ranked list.

#### Additional material

**Additional file 1: Supplemental Figures S1-S5.** Supplemental Figure **S1** - Correlation between rank and GC content calculated for the targets of every micro (mi)RNA family. Supplemental Figure **S2** - Minimum hypergeometric (mHG) scores of human motifs versus mHG scores of *Saccharomyces cerevisiae* motifs. Supplemental Figure **S3** - Predicted structures of selected mRNA targets. Supplemental Figure **S4** - The enrichment of the Pumilio binding motif versus the enrichment of the miR-410 binding motif in validated Pumilio targets. Supplemental Figure **S5** - Illustration of masking a region in a sequence.

#### Acknowledgements

We thank Yoav Arava for his valuable review of the manuscript. We thank Israel Steinfeld for his useful comments and for providing access to several software implementations and statistical enrichment tools. We thank Eran Eden, Gal Shaul and Naama Hazan for providing access to a motif visualization tool. We thank the anonymous reviewer for suggesting some controls added after the review. LL was partially supported through a Sherman Fellowship.

#### Author details

<sup>1</sup>Department of Computer Sciences, Technion - Israel Institute of Technology, Technion City, Haifa 32000, Israel. <sup>2</sup>Faculty of Biology, Technion - Israel Institute of Technology, Technion City, Haifa 32000, Israel. <sup>3</sup>Agilent Laboratories Israel, Azorim Business Park, 94 Em Hamoshavot Road, 49527 Petach-Tikva, Israel.

#### Authors' contributions

LL, YMG and ZY designed the study. YMG proposed the research question. LL and ZY developed the methodology and the algorithmic approaches. LL developed the software and performed the data analysis. All authors contributed to and read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

Received: 24 June 2010 Accepted: 22 September 2010

Published: 22 September 2010

#### References

1. Bushati N, Cohen SM: **microRNA functions.** *Annu Rev Cell Dev Biol* 2007, **23**:175-205.
2. Williams AE: **Functional aspects of animal microRNAs.** *Cell Mol Life Sci* 2008, **65**:545-562.
3. Pfeffer S, Zavolan M, Grässer FA, Chien M, Russo JJ, Ju J, John B, Enright AJ, Marks D, Sander C, Tuschl T: **Identification of virus-encoded microRNAs.** *Science* 2004, **304**:734-736.
4. Eulalio A, Huntzinger E, Izaurralde E: **Getting to the root of miRNA-mediated gene silencing.** *Cell* 2008, **132**:9-14.
5. Filipowicz W, Bhattacharyya SN, Sonenberg N: **Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?** *Nat Rev Genet* 2008, **9**:102-114.

6. Chekulaeva M, Filipowicz W: **Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells.** *Curr Opin Cell Biol* 2009, **21**:452-460.
7. Bartel DP: **MicroRNAs: target recognition and regulatory functions.** *Cell* 2009, **136**:215-233.
8. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E: **The role of site accessibility in microRNA target recognition.** *Nat Genet* 2007, **39**:1278-1284.
9. Bhattacharyya SN, Habermacher R, Martine U, Closs EI, Filipowicz W: **Relief of microRNA-mediated translational repression in human cells subjected to stress.** *Cell* 2006, **125**:1111-1124.
10. Kedde M, Strasser MJ, Boldajipour B, Oude Vrielink JA, Slanchev K, le Sage C, Nagel R, Voorhoeve PM, van Duijse J, Ørom UA, Lund AH, Perrakis A, Raz E, Agami R: **RNA-binding protein Dnd1 inhibits microRNA access to target mRNA.** *Cell* 2007, **131**:1273-1286.
11. Ketting RF: **A dead end for microRNAs.** *Cell* 2007, **131**:1226-1227.
12. Schratz GM, Tuebing F, Nigh EA, Kane CG, Sabatini ME, Kiebler M, Greenberg ME: **A brain-specific microRNA regulates dendritic spine development.** *Nature* 2006, **439**:283-289.
13. Hammell CM, Lubin I, Boag PR, Blackwell TK, Ambros V: **nhl-2 modulates miRNA activity in *Caenorhabditis elegans*.** *Cell* 2009, **136**:926-938.
14. Schwamborn JC, Berezikov E, Knoblich JA: **The TRIM-NHL protein TRIM32 activates microRNAs and prevents self-renewal in mouse neural progenitors.** *Cell* 2009, **136**:913-925.
15. Wickens M, Bernstein DS, Kimble J, Parker R: **A PUF family portrait: 3'UTR regulation as a way of life.** *Trends Genet* 2002, **18**:150-157.
16. Crittenden SL, Bernstein DS, Bachorik JL, Thompson BE, Gallegos M, Petcherski AG, Moulder G, Barstead R, Wickens M, Kimble J: **A conserved RNA-binding protein controls germline stem cells in *Caenorhabditis elegans*.** *Nature* 2002, **417**:660-663.
17. Moore FL, Jaruzelska J, Fox MS, Urano J, Firpo MT, Turek PJ, Dorfman DM, Pera RA: **Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (deleted in AZoospermia) and DAZ-like proteins.** *Proc Natl Acad Sci USA* 2003, **100**:538-543.
18. Salvetti A, Rossi L, Lena A, Batistoni R, Deri P, Rainaldi G, Locci MT, Evangelista M, Gremigni V: **DjPum, a homologue of *Drosophila* Pumilio, is essential to planarian stem cell maintenance.** *Development* 2005, **132**:1863-1874.
19. Schweers BA, Walters KJ, Stern M: **The *Drosophila melanogaster* translational repressor pumilio regulates neuronal excitability.** *Genetics* 2002, **161**:1177-1185.
20. Dubnau J, Chiang AS, Grady L, Barditch J, Gossweiler S, McNeil J, Smith P, Buldoc F, Scott R, Certa U, Broger C, Tully T: **The *staufen/pumilio* pathway is involved in *Drosophila* long-term memory.** *Curr Biol* 2003, **13**:286-296.
21. Fiore R, Khudayberdiev S, Christensen M, Siegel G, Flavell SW, Kim TK, Greenberg ME, Schratz G: **Mef2-mediated transcription of the miR379-410 cluster regulates activity-dependent dendritogenesis by fine-tuning Pumilio2 protein levels.** *EMBO J* 2009, **28**:697-710.
22. Olivas W, Parker R: **The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast.** *EMBO J* 2000, **19**:6602-6611.
23. Goldstrohm AC, Hook BA, Seay DJ, Wickens M: **PUF proteins bind Pop2p to regulate messenger RNAs.** *Nat Struct Mol Biol* 2006, **13**:533-539.
24. Kadyrova LY, Habara Y, Lee TH, Wharton RP: **Translational control of maternal cyclin B mRNA by Nanos in the *Drosophila* germline.** *Development* 2007, **134**:1519-1527.
25. Galgano A, Forrer M, Jaskiewicz L, Kanitz A, Zavolan M, Gerber AP: **Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system.** *PLoS One* 2008, **3**:e3164.
26. Nolde MJ, Saka N, Reinert KL, Slack FJ: **The *Caenorhabditis elegans* pumilio homolog, puf-9, is required for the 3'UTR-mediated repression of the let-7 microRNA target gene, hbl-1.** *Dev Biol* 2007, **305**:551-563.
27. Lewis BP, Burge CB, Bartel DP: **Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.** *Cell* 2005, **120**:15-20.
28. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP: **MicroRNA targeting specificity in mammals: determinants beyond seed pairing.** *Mol Cell* 2007, **27**:91-105.
29. Friedman RC, Farh KK, Burge CB, Bartel DP: **Most mammalian mRNAs are conserved targets of microRNAs.** *Genome Res* 2009, **19**:92-105.
30. Eden E, Lipson D, Yogev S, Yakhini Z: **Discovering motifs in ranked lists of DNA sequences.** *PLoS Comput Biol* 2007, **3**:e39.
31. Morris AR, Mukherjee N, Keene JD: **Ribonomic analysis of human Pum1 reveals cis-trans conservation across species despite evolution of diverse mRNA target sets.** *Mol Cell Biol* 2008, **28**:4093-4103.
32. Pillai RS, Artus CG, Filipowicz W: **Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis.** *RNA* 2004, **10**:1518-1525.
33. Hon LS, Zhang Z: **The roles of binding site arrangement and combinatorial targeting in microRNA repression of gene expression.** *Genome Biol* 2007, **8**:R166.
34. MEME. [[http://meme.sdsc.edu/meme4\\_4\\_0/cgi-bin/meme.cgi](http://meme.sdsc.edu/meme4_4_0/cgi-bin/meme.cgi)].
35. Chen K, Rajewsky N: **Natural selection on human microRNA binding sites inferred from SNP data.** *Nat Genet* 2006, **38**:1452-1456.
36. Shankavaram UT, Reinhold WC, Nishizuka S, Major S, Morita D, Chary KK, Reimers MA, Scherf U, Kahn A, Dolginow D, Cossman J, Kaldjian EP, Scudiero DA, Petricoin E, Liotta L, Lee JK, Weinstein JN: **Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study.** *Mol Cancer Ther* 2007, **6**:820-832.
37. Blower PE, Verducci JS, Lin S, Zhou J, Chung JH, Dai Z, Liu CG, Reinhold W, Lorenzi PL, Kaldjian EP, Croce CM, Weinstein JN, Sadee W: **MicroRNA expression profiles for the NCI-60 cancer cell panel.** *Mol Cancer Ther* 2007, **6**:1483-1491.
38. Ayala YM, Pantano S, D'Ambrogio A, Buratti E, Brindisi A, Marchetti C, Romano M, Baralle FE: **Human, *Drosophila*, and *C.elegans* TDP43: nucleic acid binding properties and splicing regulatory function.** *J Mol Biol* 2005, **348**:575-588.
39. Leung JYT: **Fast algorithms for generating all maximal independent sets of interval, circular-arc and chordal graphs.** *Journal of Algorithms* 1984, **5**:22-35.
40. TargetScanHuman. [<http://www.targetscan.org>].
41. Hofacker IL, Stadler PF: **Memory efficient folding algorithms for circular RNA secondary structures.** *Bioinformatics* 2006, **22**:1172-1176.
42. DRIM. [<http://bioinfo.cs.technion.ac.il/drim>].
43. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z: **GORilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists.** *BMC Bioinformatics* 2009, **10**:48.
44. UCSC Table Browser. [<http://genome.ucsc.edu/cgi-bin/hgTables>].
45. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23**:2947-2948.
46. WebLogo. [<http://weblogo.berkeley.edu/logo.cgi>].
47. Ben-Dor A, Bruhn L, Friedman N, Nachman I, Schummer M, Yakhini Z: **Tissue classification with gene expression profiles.** *J Comput Biol* 2000, **7**:559-583.
48. de Berg M, Cheong O, van Kreveld M, Overmars M: **Kd-trees.** *Computational Geometry - Algorithms and Applications* Springer, 3 2008, 99-101.
49. Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, Guo N, Muruganujan A, Doremioux O, Campbell MJ, Kitano H, Thomas PD: **The PANTHER database of protein families, subfamilies, functions and pathways.** *Nucleic Acids Res* 2005, **33**:D284-288.

doi:10.1186/1758-907X-1-17

**Cite this article as:** Leibovich et al: **A structural-based statistical approach suggests a cooperative activity of PUM1 and miR-410 in human 3'-untranslated regions.** *Silence* 2010 **1**:17.