Cell, Vol. 120, 15–20, January 14, 2005, Copyright ©2005 by Elsevier Inc. DOI 10.1016/j.cell.2004.12.035

Letter to the Editor

Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets

We predict regulatory targets of vertebrate microRNAs (miRNAs) by identifying mRNAs with conserved complementarity to the seed (nucleotides 2–7) of the miRNA. An overrepresentation of conserved adenosines flanking the seed complementary sites in mRNAs indicates that primary sequence determinants can supplement base pairing to specify miRNA target recognition. In a four-genome analysis of 3′ UTRs, approximately 13,000 regulatory relationships were detected above the estimate of false-positive predictions, thereby implicating as miRNA targets more than 5300 human genes, which represented 30% of our gene set. Targeting was also detected in open reading frames. In sum, well over one third of human genes appear to be conserved miRNA targets.

MicroRNAs pair to the messages of protein-coding genes to target these genes for posttranscriptional repression (Bartel, 2004). Several computational methods have been developed to predict mRNAs targeted by miRNAs in animals (Enright et al., 2003; Lewis et al., 2003; Stark et al., 2003; John et al., 2004; Kiriakidou et al., 2004; Rajewsky and Socci, 2004). One such method is TargetScan, which was applied to predict miRNA target sites conserved among orthologous 3′ UTRs of vertebrates (Lewis et al., 2003). When using 3′ UTR regions of human, mouse, and rat (Human Sequencing Consortium, 2001; Mouse Sequencing Consortium, 2002; Rat Sequencing Consortium, 2004), 451 human targets were predicted by TargetScan using score and rank cutoffs that gave an estimated 140 false-positive predictions, for a signal:noise ratio of 3.2:1 (Lewis et al., 2003).

With the recent availability of the chicken (Chicken Sequencing Consortium, 2004) and dog genome assemblies, together with updated annotations of the human, mouse, and rat genomes, we revisited miRNA target predictions. Requiring target-site conservation in all five genomes (human, mouse, rat, dog, and chicken) reduced the noise (estimated number of false-positive predictions) such that the TargetScan score and rank cutoffs could be dramatically relaxed, or even eliminated. Moreover, the requirement of a 7-nt match to the seed region of the miRNA (nucleotides 2–8) could be relaxed to require a 6-nt match to a reduced seed comprising nucleotides 2–7 of the miRNA while still retaining modest specificity. Running TargetScan in this way without cutoffs amounted to predicting a target simply by virtue of the presence of at least one 6-nt seed match to the miRNA in orthologous UTRs of each of the five genomes. Signal:noise was improved when the seed match was required to occur at corresponding positions in a multiple alignment of the orthologous UTRs. Therefore, the availability of newly sequenced genomes, improved annotations, and whole-genome alignments (Karolchik et al., 2003; Blanchette et al., 2004) allowed use of a greatly simplified method: miRNA targets could be predicted by finding perfect Watson-Crick (W-C) seed matches that are conserved in the UTR regions of whole-genome alignments, as exemplified by the miR-23a-*HIC* seed pairing (Figure 1A).

Starting with the UTRs corresponding to mRNAs annotated in the UC Santa Cruz Genome Browser database (Karolchik et al., 2003), and a set of 62 unique seed matches that represented 148 human miRNA genes and defined the families of known miRNAs conserved in the five genomes (Supplemental Table S1 at http://www.cell.com/cgi/content/full/120/1/15/DC1), we identified 14301 instances of conserved seed matches within the 3′ UTRs, thereby predicting 14,301 unique target sites. Because some UTRs had multiple conserved target sites for the same miRNA seed, this analysis implicated 12,839 unique miRNA-target regulatory relationships (Figure 1B, left graph). Because many UTRs had conserved target sites for different miRNA seeds, which often could enable combinatorial control of these messages, these 12,839 predictions involved the UTRs of 3,227 unique human genes. Supplemental Table S2 (also available at the TargetScan web site http://genes.mit.edu/targetscan/) lists genes and target sites, with annotation of additional specificity determinants, which are described below.

To estimate the number of false positives, we picked for each authentic seed match at least five hexamers of comparable abundance in the UTR dataset and repeated the analysis with these control sequences, averaging the results for each set of control sequences (Supplemental Data). Summing these averages yielded 5,817 target sites corresponding to 5,386 unique false-positive predictions, for a signal:noise ratio of 2.4:1 (Figure 1B, left graph). The number of genes targeted above the noise was estimated by removing 5,817 randomly chosen hits from the set of 14,301, leaving 8,484 hits that involved the messages of 2,767 human genes. Thus the five-genome analysis implicated 25% of our set of 10,938 orthologous vertebrate genes as conserved targets of the miRNAs.

We next looked at the sequence flanking the seed matches for conserved positions that might contribute specificity to miRNA:target interactions (Figure 1C). The position immediately upstream of the seed match was highly conserved and had a high propensity to be a conserved W-C match to the eighth nucleotide of the miRNA. (We designate these target and miRNA positions t8 and m8, respectively, and use "M" to designate W-C matches between corresponding target and miRNA positions.) Requiring a conserved match at this position markedly increased specificity, improving the signal:noise to 3.8 (Figure 1B, SeedM + m8M). However, the sensitivity, calculated as signal above noise, decreased substantially, implying that more than 3500 authentic target sites lack m8 matches.

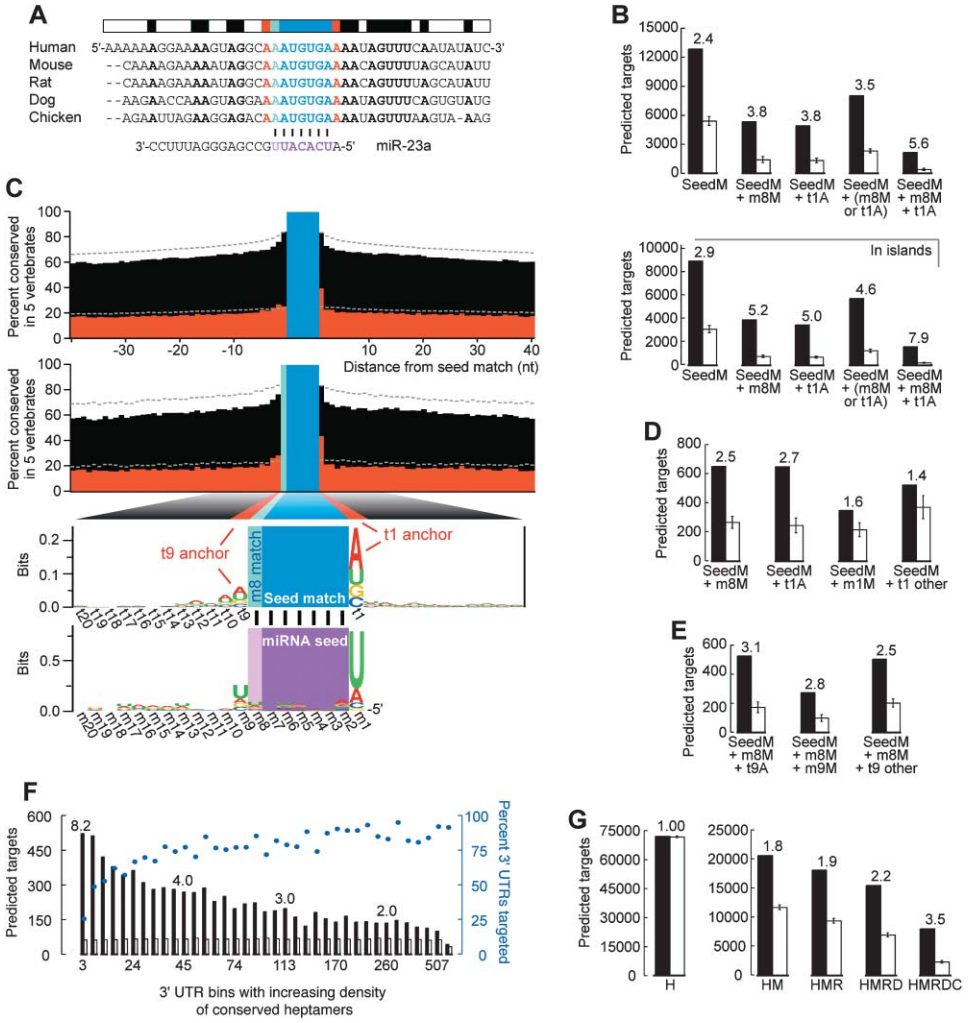

Figure 1. Conserved Seed Matches, Often with Anchoring A's, Predict miRNA Targets

(A) Alignment of orthologous segments of the *HIC* UTR, showing the conserved match to the miR-23a seed. Residues of the seed (purple), seed matches (dark blue), m8 (light purple), m8 matches (light blue), and anchoring A's (red) are indicated.

(B) The number of miRNA–target relationships predicted (solid bars), with estimates of the number of false positives (open bars), for searches based on the indicated criteria. In this and subsequent panels, error bars indicate one standard deviation, based on analyses of control cohorts. Standard error on these values was much smaller (not visible if shown as error bars) because each estimate of the number of false positives was calculated using many control sequences. The numbers above each graph indicate the value for signal divided by that of the noise. Also graphed are the subsets of predictions in which seed matches fell within islands of conservation (in islands).

(C) Overall conservation and sequence identity flanking conserved seed matches and miRNA seeds. Related seeds arising from 5′-end heterogeneity within a miRNA family were excluded from this analysis (Supplemental Table S1). For each position flanking the conserved seed match, the percentage of seed matches in which that position was conserved in all five vertebrates is shown (top panel), with the height of the black bar indicating conservation of any of the four possibilities, and that of the red indicating conservation of adenosine. The gray dashes indicate the same analysis for conserved matches to control sequences. The second panel shows the same analysis for sites that have both a conserved seed match and a conserved m8 match. The third panel shows the sequence identity immediately flanking the seed matches, with the height of the letters corresponding to the information content, measured using the relative entropy relative to the background base composition of 3′ UTRs (Gorodkin et al., 1997). The bottom panel shows the analogous representation of the sequence identity at the first 20 positions of the miRNAs, giving equal weight to each miRNA family used in our analysis (Supplemental Table S1).

(D) The utility of a t1 A anchor for predicting targets when the miRNA does not begin with a U. For this set of miRNAs, the signal:noise ratio in the basic SeedM analysis (before requiring additional conserved pairing or nucleotides) was 1.8:1, which was lower than that for miRNAs that either begin with U or have paralogs that begin with U.

(E) The utility of a t9 A anchor for predicting targets when the miRNA does not have a U at position 9. The set of 36 miRNAs used in this analysis yields a signal:noise of 2.1:1 in a seed-only analysis.

(F) Increased accuracy of target prediction for UTRs with a lower density of conservation. Of the 10,968 UTRs in our dataset, 4,887 had at least one conserved heptamer. These were ranked by their density of conserved heptamers, then binned such that each bin had enough UTRs to contain 8,000 conserved heptamers. For each bin, predictions for the real miRNA seeds (black) are compared to averages for the control cohorts (open). The value for signal divided by that of the noise is shown above representative bins, with the number of conserved heptamers per kb shown below. Also plotted are the percentage of UTRs in each bin that are predicted to be miRNA targets (blue circles, right axis).

(G) Analysis with one to five genomes (H, human; M, mouse; R, rat; D, dog; C, chicken) using the set of 10,968 genes aligned in the five genomes.

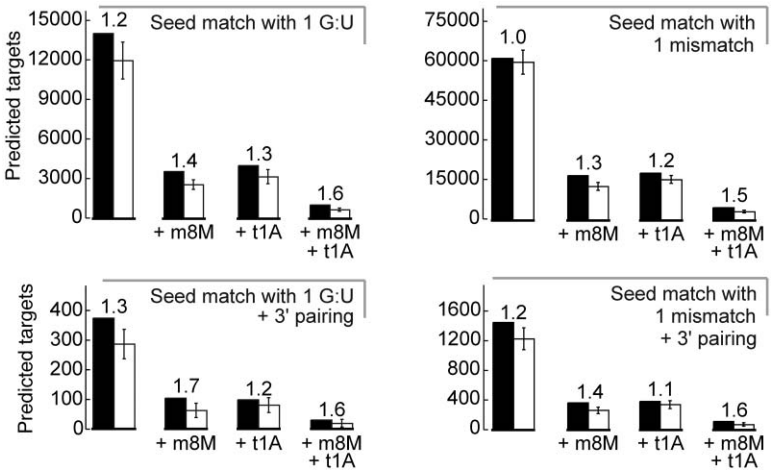

Figure 2. Signal (Black) and Estimates of False Positives (Open) When Searching for miRNA Targets that Have a Conserved G:U Pair (Left Panels) or Mismatch (Right Panels) Disrupting the Seed Match

The effects on signal and noise when requiring pairing to the 3′ portion of the miRNA (six contiguous pairs allowing one G:U wobble) are also shown (bottom panels).

High conservation was also observed at the first position downstream of the seed match. This nucleotide was often a conserved A, which could pair to the first nucleotide of a miRNA whose first nucleotide is U, a class which includes the majority of miRNAs (Figure 1C). However, a conserved A was also observed next to seed matches for miRNAs that do not begin with a U. For miRNAs that begin with A, C, or G (and which do not have any known or predicted paralogs that begin with U), the nucleotide immediately downstream of the conserved seed + m8 matches was twice as frequently a conserved A than any other conserved nucleotide, including the nucleotide that could form a W-C match to the first nucleotide of the miRNA.

The discovery that an A appears to anchor the very 3′ terminus of the miRNA complementary site suggested that requiring a 6-nt W-C seed match followed by this "A anchor" would increase the specificity of target prediction. Indeed, searching for this type of 7-nt composite match increased signal:noise to 3.8:1 in the five-genome analysis (Figure 1B, SeedM + t1A). This improved signal:noise was accompanied by a 51% loss in sensitivity. When focusing on the subset of our set of miRNAs that begin with A, C, or G, none of this drop in sensitivity was attributed to the loss of matches that involved conserved W-C pairing to the first nucleotide of the miRNA. For these nine representative miRNAs that do not begin with a U and do not share a common seed sequence with a related microRNA that starts with a U (Supplemental Table S1), demanding the W-C seed match followed by the A anchor gave 625 predictions (Figure 1D, SeedM + t1A), whereas demanding that the seed match be followed by a conserved W-C match to the miRNA gave 348 predictions, barely above the estimate of the false positives (Figure 1D, SeedM + m1M) and with signal:noise not significantly better than when requiring conservation of a non-A mismatch at this position (Figure 1D, SeedM + t1 other). We conclude that not all of the specificity of metazoan miRNA-target recognition can be explained by base-pairing to the message—a component of this specificity lies at the level of mRNA primary sequence. We speculate that a protein of the silencing complex recognizes this A in a manner that allows simultaneous or sequential interaction between the A and the first nucleotide of the miRNA, thereby explaining the strong bias toward a U at the first nucleotide of miRNAs, which has been a curiosity since it was first observed (Lau et al., 2001).

Requiring both the m8 match and the t1 anchor improved the specificity, with signal:noise of 5.6:1 in the five-genome analysis (Figure 1B). However, most of the conserved seed matches had only one of these specificity determinants, such that requiring one or the other yielded 8,012 predicted targets with signal:noise of 3.5:1. Calculating as before the number of unique genes predicted above the noise yielded 2,421 unique human genes as miRNA targets, or 22% of our set of 10,968 orthologous genes.

The ability to predict thousands of targets with confidence that most are authentic built on two features of our previous TargetScan analysis not shared by other miRNA target prediction algorithms: a requirement for perfect W-C seed pairing and the use of rigorous control cohorts to assess the utility of algorithmic refinements (Lewis et al., 2003). However, our new analysis differed by starting with whole-genome alignments, thereby requiring that the conserved seed matches be at conserved positions within the UTRs, and by focusing only on an 8-nt segment of the UTR centered on the seed match—without consideration of other criteria, such as predicted thermodynamic stability of pairing, pairing outside the immediate vicinity of the seed, or presence of multiple complementary sites per UTR, all of which were previously considered by TargetScan and by most other target-prediction algorithms. We call the refined (and considerably simplified) algorithm "TargetScanS" because of its emphasis on pairing to a 6-nt miRNA seed. The standard TargetScanS algorithm predicts targets that have a conserved 6-nt seed match flanked by either a m8 match or a t1A anchor.

Surprisingly little conservation was detected beyond the residues immediately flanking the conserved seed matches, even though our analysis was restricted to the miRNA families that are highly conserved in the five genomes, each of which has a member with no more than one substitution separating the human and chicken orthologs. Conservation was slightly elevated at t9, particularly when restricting the analysis to sites with m8

matches (Figure 1C). As seen for t1, there was again enrichment for an A at t9. This bias could not be explained by the nucleotide composition of the miRNAs, even though there is a marked preference for a U at position 9 of the miRNA (Figure 1C). When looking more closely at the conserved matches for miRNAs that do not have a U at position 9, we found an overabundance of a conserved A forming a mismatch to this nucleotide. When predicting targets for these miRNAs, requiring a conserved t9A mismatch provided substantially more specificity gain than did requiring a conserved W-C match or conserved non-A mismatches (Figure 1E).

Beyond this modestly conserved t9 anchor, conservation upstream of the seed match, where the 3′segment of the miRNA would be expected to pair, was no greater than that downstream of the seed match (Figure 1C). The same was true when restricting the analysis to sites predicted with greater specificity because they have either m8 matches or t1 anchors (Figure 1C; data not shown). The gradual downward slope in conservation observed when going in either direction from the seed match paralleled that of the background expectation and was a consequence of starting at positions that were confidently aligned in the five genomes (Figure 1C). The lack of conservation upstream of the t9 anchor suggests that thousands of vertebrate miRNA–target interactions are mediated primarily by seed matches, supplemented with either a t1A anchor or an m8 match, but with little, if any, role for pairing to the 3′ portion of the miRNA.

The observation that miRNA target sites are often not conserved beyond the 8-nt site centered on the seed match suggested that the specificity of miRNA target prediction might actually be improved by excluding those seed matches that occur in the context of more extensive conservation. Incorporating the criterion that seed matches must fall in short "islands" of conservation surrounded by the expected background level of divergence substantially increased the signal:noise ratios (Figure 1B, island row of histograms). The somewhat counterintuitive use of excess flanking conservation as a contrary indicator for target prediction improved specificity by reducing the frequency of false positives, thereby increasing the signal:noise ratio. To further explore this phenomenon, we binned the UTRs based on their density of conserved heptamers and then calculated the signal:noise ratio of TargetScanS separately for each bin (Figure 1F). The bins with a low density of conserved heptamers had high signal:noise values (greater than 8:1), whereas those with high-density heptamer conservation had poor signal:noise values (less than 2:1). In other words, as conservation in the UTRs increases, a smaller fraction of the conservation can be explained by pairing to miRNAs. For this reason, the 30 UTRs with the highest density of conserved heptamers were excluded from the analyses reported in this paper (other than that of Figure 1F). Although these 30 messages are likely to be miRNA targets, it seemed prudent not to include them because of the high likelihood that they would have fortuitous conserved pairing to many other miRNAs that do not regulate them.

For many examples of metazoan miRNA–target interactions with experimental support, recognition appears to involve multiple complementary sites to the same miRNA (Bartel, 2004). However, a number of examples of regulation have been identified that involve what appears to be only a single complementary site for a particular miRNA (Moss et al., 1997; Kiriakidou et al., 2004; Poy et al., 2004; Yekta et al., 2004). The original TargetScan analyses primarily predicted targets with more than one match to the same miRNA, although the cutoffs used for the four-genome analysis (human, mouse, rat, pufferfish) did include some predicted targets with single sites (Lewis et al., 2003). With TargetScanS, targets are predicted without preference for those that have multiple matches. Requiring a second syntenic match to the same miRNA seed increased the signal:noise ratio to 3.2:1 but reduced by 90% the number of predictions (Supplemental Table S2). Thus, demanding more than one conserved match excluded most of the apparently authentic miRNA-target pairs identified in our analyses. Of course, the finding that single conserved matches are sufficient to confidently predict miRNA-target pairs in a comparative genomic analysis is completely compatible with the idea that, within the cell, biochemical specificity is augmented by additional determinants, such as mRNA structure, binding of accessory proteins, and/or the presence of nonconserved or imperfect seed matches at additional sites in the message.

Performing TargetScanS on fewer genomes provided modest gains in sensitivity (Figure 1G), mostly from removing chicken from the analysis, which allows identification of miRNA–target interactions that were lost in the five-genome analyses either because they are specific to the mammalian lineages or because they lie in portions of the chicken genome that are missing or misassembled in the database. When extending the four-genome analysis to include genes aligned among the mammals but not to chicken, 13,044 regulatory interactions were predicted above the estimate of the false-positive predictions—an average of over 200 targets for each of the miRNA families represented (Supplemental Table S3). Calculating as for the five-genome analysis the number of unique genes predicted above the noise yielded 5,300 unique human genes as miRNA targets, or 30% of our set of 17,850 orthologous mammalian genes.

The four-genome mammalian analysis provided a set of predictions suitable for comparing to the results of previous mammalian target-prediction efforts. After accounting for the different starting sets of miRNAs and protein-coding genes, 343 of the 451 predictions in the original three-genome TargetScan analysis remained, and 67% of these overlapped (Lewis et al., 2003; Supplemental Table S3). However, there was less overlap with the results of other mammalian target predictions. After accounting for the differing gene sets, 39% of the remaining 145 human predictions reported by Kiriakidou et al. (2004) as conserved in mouse and only 18% of the remaining 5,028 predicted miRNA:UTR interactions reported by John et al. (2004) as conserved among mammals were present among our four-genome predictions. As described below, TargetScanS misses some targets when demanding perfect seed matching confined to the 3′ UTRs. Nonetheless, the small overlap with the predictions of John et al. (2004) is not necessarily evidence that TargetScanS is missing a large class of authentic targets. Instead, the small overlap could be due to a large number of false positives generated by the

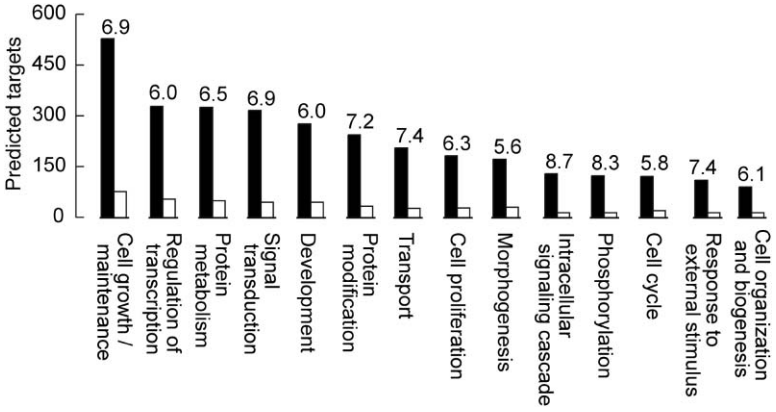

Figure 3. Biological Process Classification of the Vertebrate miRNA Targets Predicted in the Seed + t1A + m8M Analysis

Shown are plots for categories that are well represented by targets and have signal:noise of at least 5.6:1, which was the ratio for the overall analysis (Figure 1B).

method used by John et al. for arriving at their predictions (see Supplemental Data for detailed discussion).

By requiring perfect seed pairing, TargetScan and TargetScanS miss miRNA–target interactions with wobbles or mismatches that disrupt seed pairing, such as the nematode *let-7–lin-41* or vertebrate miR-196–*HoxB8* interactions, both of which have been validated in animals (Reinhart et al., 2000; Yekta et al., 2004). The loss of such interactions from our original TargetScan analysis was tolerated because allowing wobbles or mismatches in the seed pairing would have decreased the signal:noise ratio using rigorous estimates of false positives to essentially 1:1, casting doubt on all such interactions identified with imperfect seed matching.

Revisiting this issue in an analysis including the newly sequenced genomes revealed some signal above noise, with moderate improvement in specificity when requiring the t1 A anchor or m8 match, but the quality of these predictions was still far below that observed for perfect seed matches (Figure 2). The *let-7–lin-41* and miR-196–*HoxB8* interactions both include extensive pairing to the 3′ portion of the miRNA, each involving at least nine contiguous W-C pairs, which might compensate for the imperfect seed pairing and impart specificity (Doench and Sharp, 2004). Requiring conserved 3′ pairing with at least six contiguous pairs (allowing one G:U wobble) yielded little if any increased specificity of target prediction (Figure 2, bottom panels). The existence of a class of conserved sites of this type could explain the observed pattern of sequence conservation of vertebrate miRNAs, which typically extends throughout the miRNA. However, compared to searches requiring perfect seed pairing (Figure 1B), fewer targets were predicted. Overall, it appears that there are relatively few conserved interactions that lack perfect seed pairing. However, additional parameters need to be examined, and it remains possible that many such interactions exist but most of them have not yet been confidently identified by existing algorithms.

If relatively few miRNA interactions lack perfect seed pairing, this could be explained if these types of interactions typically require extensive pairing outside the seed, thereby increasing the total required base-pairing to the message. As a result, such interactions would emerge more rarely and be more difficult to maintain over the course of evolution, perhaps occurring preferentially under circumstances in which regulation by a specific member of a multi-miRNA family is required. For example, if *C. elegans lin-41* were to be repressed by any of the other three *let-7* family members, which have the same seeds (Lim et al., 2003) but are expressed earlier than is *let-7* RNA, then premature downregulation of *lin-41* might cause larval cells to precociously assume adult cell fates (Reinhart et al., 2000). Perhaps to achieve the proper timing of repression, the *lin-41* 3′ UTR has imperfect seed pairing to the entire *let-7* family, which prevents regulation by the other three family members while the extensive pairing to the unique 3′ region of *let-7* RNA enables regulation by *let-7*.

In animals, all known target sites are in 3′ UTRs, whereas in plants they are sometimes in the 3′ UTR but are usually in the ORFs (Jones-Rhoades and Bartel, 2004) and also have been predicted to reside in 5′ UTRs (Sunkar and Zhu, 2004). TargetScanS applied to 5′ UTRs found little or no signal above noise. Because of their high sequence conservation, ORFs were more difficult to analyze by our methods. Nonetheless, a five-genome ORF analysis requiring conserved seed matches flanked by both an m8 match and a t1A yielded 2,371 predicted targets (Supplemental Table S2), which was significantly above the 1,300 estimated false positives. Although this analysis provided evidence that many messages have functional miRNA complementary sites in ORFs, our data are consistent with the idea that most functional mRNA–miRNA pairing resides in the 3′ UTRs, and that miRNA pairing explains a substantial fraction of the conservation observed in metazoan 3′ UTRs.

The plant miRNAs have a strong propensity to target messages of developmental regulators, particularly transcription factors involved in plant development (Rhoades et al., 2002; Jones-Rhoades and Bartel, 2004). Although many of the TargetScanS predictions were annotated as controlling transcription or development, most had other functions (Figure 3), as seen previously for the TargetScan predictions (Lewis et al., 2003). Some miRNAs had a propensity to target genes of a particular category. An interesting example is the miRNAs of the *mir-17-18-19-20* gene cluster, which resides in a region of the genome that is amplified in many lymphomas and solid tumors (Ota et al., 2004). These miRNAs had a striking propensity to target genes with known or suspected roles in growth control, including both oncogenes and genes that repress growth (Supplemental Table S2). Among those with roles in growth arrest were numerous

genes in the TGF-β signaling pathway (including *TGF-β receptor II*, *Smad2*, *Smad6*, *Smad7*, and *SARA*), *SOCS* genes (*SOCS-1*, *SOCS-3*, *SOCS-5*, and *SOCS-6*), *p130*, and *PTEN*.

TargetScanS analyses indicate that a substantial fraction of the mammalian genes are subject to miRNA control and that primary sequence determinants supplement pairing in specifying target recognition. Initial analyses indicate that the same is true in invertebrates. As additional mammalian genomes are added to the whole-genome alignments, the accuracy of predictions will undoubtedly increase, as fortuitous conservation is reduced and flaws in any single assembly can be diluted by requiring conservation in an appropriate quorum of genomes, bringing the conserved roles of miRNAs in mammals into increasingly sharper focus.

Acknowledgments

We thank the Genome Sequencing Center, Washington School of Medicine and the Chicken Mapping Consortium for making available the chicken genome assembly, the Broad Institute of MIT/Harvard and Agencourt Bioscience for making available the dog genome assembly, and Ernest Fraenkel for helpful comments on this manuscript. This work was supported by a pre-doctoral fellowship from the DOE (B.P.L.) and by grants from the Searle Scholars Program (C.B.B.) and from the N.I.H (C.B.B. and D.P.B).

Benjamin P. Lewis,[1,2] Christopher B. Burge,[1,*] and David P. Bartel[1,2,*]
[1]Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
[2]Whitehead Institute for Biomedical Research
9 Cambridge Center
Cambridge, Massachusetts 02142

*Correspondence: cburge@mit.edu (C.B.B.), dbartel@wi.mit.edu (D.P.B.)

Selected Reading

Bartel, D.P. (2004). Cell *116*, 281–297.

Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., et al. (2004). Genome Res. *14*, 708–715.

Chicken Sequencing Consortium. (2004). Nature *432*, 695–716.

Doench, J.G., and Sharp, P.A. (2004). Genes Dev. *18*, 504–511.

Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). Genome Biol. *5*, R1.

Gorodkin, J., Heyer, L.J., Brunak, S., and Stormo, G.D. (1997). Comput. Appl. Biosci. *13*, 583–586.

Human Sequencing Consortium. (2001). Nature *409*, 860–921.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S. (2004). PLoS Biol 2(11): e363 DOI: 10.1371/journal.pbio.0020363.

Jones-Rhoades, M.W., and Bartel, D.P. (2004). Mol. Cell *14*, 787–799.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., et al. (2003). Nucleic Acids Res. *31*, 51–54.

Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., and Hatzigeorgiou, A. (2004). Genes Dev. *18*, 1165–1178.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). Science *294*, 858–862.

Lewis, B.P., Shih, I., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Cell *115*, 787–798.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003). Genes Dev. *17*, 991–1008.

Moss, E.G., Lee, R.C., and Ambros, V. (1997). Cell *88*, 637–646.

Mouse Sequencing Consortium. (2002). Nature *420*, 520–562.

Ota, A., Tagawa, H., Karnan, S., Tsuzuki, S., Karpas, A., Kira, S., Yoshida, Y., and Seto, M. (2004). Cancer Res. *64*, 3087–3095.

Poy, M.N., Eliasson, L., Krutzfeldt, J., Kuwajima, S., Ma, X., Macdonald, P.E., Pfeffer, S., Tuschl, T., Rajewsky, N., Rorsman, P., and Stoffel, M. (2004). Nature *432*, 226–230.

Rajewsky, N., and Socci, N.D. (2004). Dev. Biol. *267*, 529–535.

Rat Sequencing Consortium. (2004). Nature *428*, 493–521.

Reinhart, B.J., Slack, F.J., Basson, M., Bettinger, J.C., Pasquinelli, A.E., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). Nature *403*, 901–906.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Cell *110*, 513–520.

Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. (2003). PLoS Biol 1(3): e60 DOI: 10.1371/journal.pbio.0000060.

Sunkar, R., and Zhu, J.K. (2004). Plant Cell *16*, 2001–2019.

Yekta, S., Shih, I.H., and Bartel, D.P. (2004). Science *304*, 594–596.