nature genetics

Combinatorial microRNA target predictions

Azra Krek[1,2,4], Dominic Grün[1,4], Matthew N Poy[3,4], Rachel Wolf[1], Lauren Rosenberg[1], Eric J Epstein[3], Philip MacMenamin[1], Isabelle da Piedade[1], Kristin C Gunsalus[1], Markus Stoffel[3] & Nikolaus Rajewsky[1]

MicroRNAs are small noncoding RNAs that recognize and bind to partially complementary sites in the 3′ untranslated regions of target genes in animals and, by unknown mechanisms, regulate protein production of the target transcript[1–3]. Different combinations of microRNAs are expressed in different cell types and may coordinately regulate cell-specific target genes. Here, we present PicTar, a computational method for identifying common targets of microRNAs. Statistical tests using genome-wide alignments of eight vertebrate genomes, PicTar's ability to specifically recover published microRNA targets, and experimental validation of seven predicted targets suggest that PicTar has an excellent success rate in predicting targets for single microRNAs and for combinations of microRNAs. We find that vertebrate microRNAs target, on average, roughly 200 transcripts each. Furthermore, our results suggest widespread coordinate control executed by microRNAs. In particular, we experimentally validate common regulation of *Mtpn* by *miR-375*, *miR-124* and *let-7b* and thus provide evidence for coordinate microRNA control in mammals.

Expression profiling of microRNAs by cloning and sequencing, northern blotting or microarrays[4] has shown that a small number of microRNAs (typically one to ten) are often expressed in specific tissues and developmental stages. Many known target genes of microRNAs

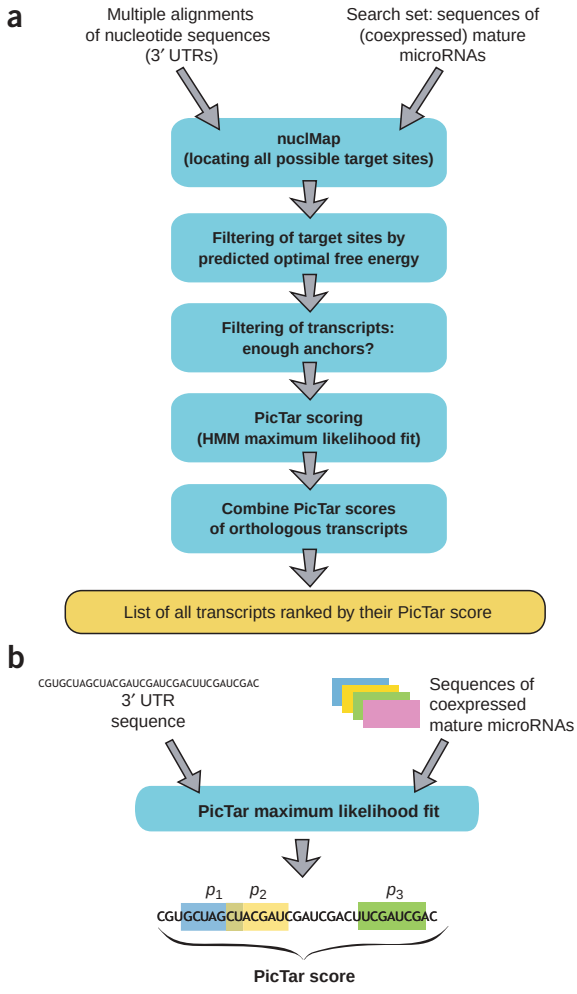

Figure 1 The PicTar algorithm. (**a**) Schematic overview. Input to PicTar consists of multiple alignments of RNA sequences (typically 3′ UTRs) and a search set of mature (coexpressed) microRNA sequences. The program nuclMap locates all perfect nuclei (length 7, starting at position 1 or 2 of the 5′ end of the microRNA) and imperfect nuclei in 3′ UTR sequences. Highly probable nuclei that survive the optimal free energy filter and fall into overlapping positions in the alignments for all species under consideration are called anchors. If a 3′ UTR multiple alignment has a minimal (user-defined) number of anchors, each UTR in the alignment will be scored by the central PicTar maximum likelihood procedure (**b**). Scores for individual UTRs in an alignment are combined to obtain the final PicTar score, which can be used to obtain a ranked list of all sets of orthologous transcripts. (**b**) PicTar scoring of a single 3′ UTR sequence. PicTar tallies all segmentations of the RNA sequence (3′ UTR) into binding sites and background sequences[15]. PicTar computes the maximum likelihood score (PicTar score) that the RNA sequence is targeted by combinations of microRNAs from the search set when compared to background and the individual probability p_i for each subsequence of the RNA sequence to be bound by a microRNA (only the nuclei for the binding sites are depicted). These posterior probabilities are different from the probability that a single subsequence is a microRNA binding site and reflect the competition of microRNAs and background for binding in the UTR.

[1]Center for Comparative Functional Genomics, Department of Biology, New York University, 100 Washington Square East, New York, New York 10003, USA. [2]Department of Physics, New York University, New York, New York 10003, USA. [3]Laboratory of Metabolic Diseases, The Rockefeller University, New York, New York 10021, USA. [4]These authors contributed equally to this work. Correspondence should be addressed to N.R. (nikolaus.rajewsky@nyu.edu).

Published online 3 April 2005; doi:10.1038/ng1536

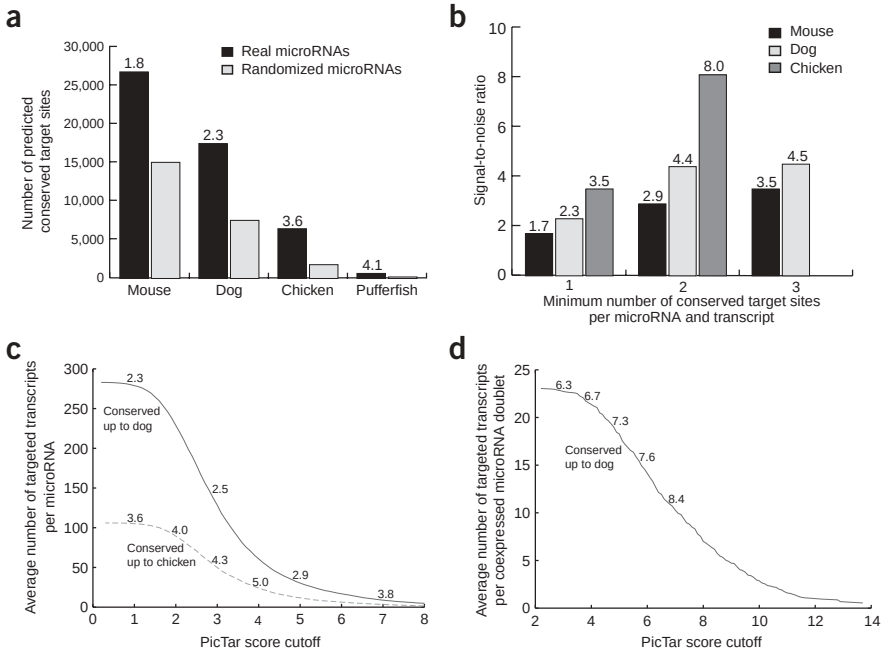

Figure 2 Signal-to-noise ratio for vertebrate microRNA target site predictions. (**a**) Signal-to-noise ratio for predicted single target sites. The number of predicted conserved target sites (anchors) for the set of 58 unique conserved human microRNAs versus the corresponding number for randomized microRNAs, requiring conservation of anchor sites between human, chimpanzee, mouse (first column), rat and dog (second column), chicken (third column) and pufferfish (last column). Inclusion of more distantly related species substantially boosts signal-to-noise ratio (indicated above black bars). For human, chimpanzee, mouse, rat and dog, we predict 17,542 conserved target sites with a signal-to-noise ratio of 2.3 and therefore ~10,000 true target sites. (**b**) Multiplicity of target sites boosts the signal-to-noise ratio. The ratio of the number of transcripts with at least *n* anchor sites per microRNA for real versus random microRNAs provides an estimate of the signal-to-noise ratio for sites conserved in human, chimpanzee, mouse (black bars), rat, dog (light gray bars) and chicken (dark gray bars). The multiplicity of target sites (scored by PicTar) helps to raise the signal-to-noise ratio. (**c**) PicTar score–dependent sensitivity and specificity of single microRNA target site predictions. The average number of predicted targets of a single microRNA with at least one anchor site per transcript in human, chimpanzee, mouse, rat and dog (upper curve) or in human, chimpanzee, mouse, rat, dog and chicken (lower curve) is plotted as a function of a PicTar score cutoff (discarding transcripts with a score below cutoff). Signal-to-noise ratios are indicated above each curve. (**d**) PicTar score–dependent sensitivity and specificity of target site predictions for four sets of coexpressed microRNAs[4] (**Supplementary Table 4** online) and corresponding sets of randomized microRNAs, requiring two anchor sites for different microRNAs in human, chimpanzee, mouse, rat and dog. The plot shows the average number of targets per pair of microRNAs as a function of the PicTar score cutoff (signal-to-noise ratios above the curve).

contain several microRNA binding sites, and the degree of translational repression may increase exponentially with the number of microRNA binding sites in the 3′ untranslated region (UTR)[5]. Thus, as in transcriptional regulation, the concentrations of the *trans*-acting microRNAs in a cell may be read out by *cis*-regulatory sites and used to fine-tune gene expression[6]. Hence, to understand biological microRNA function, it may be important to search for combinations of microRNA binding sites for sets of coexpressed microRNAs. Previously developed computational algorithms can identify targets for single microRNAs[7–14] but have not so far been used to score common targets of several microRNAs. Furthermore, they typically have relatively high false-positive rates when the number of binding sites for a given microRNA in a 3′ UTR is small[3]. Our method, probabilistic identification of combinations of target sites (PicTar), overcomes these problems by generalizing previous methods and allows the identification of targets for both single microRNAs and combinations of microRNAs.

Input to PicTar (**Fig. 1a**) is a fixed search set of microRNAs and multiple alignments of orthologous nucleotide sequences (3′ UTRs). Output are scores that rank genes by their likelihood of being a common target of members (subsets) of the search set and probabilities for the predicted binding sites in each UTR. The algorithm follows the general logic of Ahab, a validated probabilistic algorithm for the identification of combinations of transcription factor binding sites[15,16]. PicTar tallies all segmentations of a sequence into binding sites and background and computes the maximum likelihood score that the sequence is bound by combinations of microRNAs (**Fig. 1b** and **Supplementary Note** online). In this probabilistic model, microRNAs compete with each other and background for binding. The model accounts for synergistic effects of multiple binding sites of one microRNA or several microRNAs acting together, as well as for the appropriate scoring of overlapping sites. The probabilities assigned to a single site were modeled in accordance with experimental[7,8,12,17] and computational[7–14] results. Cross-species comparisons are crucial for filtering out false positives: candidate target genes are defined as UTRs with a minimal (user-defined) number of evolutionarily conserved putative binding sites. PicTar then scores the candidate sequences for each species separately. The resulting scores are combined to obtain the final PicTar score for a gene. Future insights into microRNA target site recognition and repression efficacy can easily be incorporated into the model.

In *Caenorhabditis elegans*, the sequential stage-specific expression of the microRNAs *lin-4* and *let-7* coordinates developmental timing[18]. To test PicTar, we applied it to search our genome-wide set of 10,607 *C. elegans* and *Caenorhabditis briggsae* 3′ UTR sequences (**Supplementary Methods** online) for targets of *lin-4* or *let-7*. The known targets *lin-14*, *hbl-1*, *daf-12* and *lin-28* were ranked first, second, fourth and seventh, respectively, and only one known target gene (*lin-41*) was not recovered, suggesting that PicTar has excellent specificity and sensitivity. In accordance with previous studies[18,19], PicTar predicts *lin-14* and *lin-28* to be targeted by both *lin-4* and *let-7*. Notably, PicTar found only a few genes with sites for both *lin-4* and *let-7*, suggesting that the number of *lin-4*–*let-7* common targets is relatively small. We further tested PicTar by computing predictions for each microRNA separately over all *C. elegans* 3′ UTRs without any cross-species comparisons. Randomization tests (**Supplementary Methods** online) indicated that a highly significant (>10 s.d.) fraction of predicted sites is evolutionarily conserved, strengthening confidence in PicTar.

For target predictions in vertebrates, we constructed multiple alignments of 20,254 annotated human 3′ UTRs to genomic sequences from seven other vertebrates, chimpanzee, mouse, rat, dog, chicken,

Table 1 Target validation for *miR-124* and *miR-375* by immunoblotting and luciferase reporter assays

microRNA	RefSeq gene ID	Gene functional annotation	Immunoblotting	Luciferase
miR-375	NM_013464.2	*Mus musculus* aryl-hydrocarbon receptor (*Ahr*)	–	ND
miR-375	NM_010847.1	*M. musculus* Max interacting protein 1 (*Mxi1*)	–	ND
miR-375	NM_016889.1	*M. musculus* insulinoma-associated 1 (*Insm1*)	–	ND
miR-375	**NM_008413.1**	***M. musculus* Janus kinase 2 (*Jak2*)**	**ND**	**+**
miR-375	**NM_007573.1**	***M. musculus* complement component 1, q subcomponent binding protein (*C1qbp*)**	**ND**	**+**
miR-375	**NM_146144.1**	***M. musculus* ubiquitin specific protease 1 (*Usp1*)**	**ND**	**+**
miR-375	**NM_008098.2**	***M. musculus* myotrophin (*Mtpn*)**	**+**	**+**
miR-375	**NM_197985**	***M. musculus* adiponectin receptor 2 (*Adipor2*)**	**+**	**+**
miR-124	NM_009498.3	*M. musculus* vesicle-associated membrane protein 3 (*Vamp3*)	–	ND
miR-124	NM_013464.2	*M. musculus* aryl-hydrocarbon receptor (*Ahr*)	–	ND
miR-124	**NM_011951.1**	***M. musculus* mitogen activated protein kinase 14 (*Mapk14*)**	**+**	**ND**
miR-124	**NM_008098.2**	***M. musculus* myotrophin (*Mtpn*)**	**+**	**+**
miR-124	NM_197985	*M. musculus* adiponectin receptor 2 (*Adipor2*)	–	–

The genes tested were selected from the single microRNA target prediction lists requiring different levels of conservation (human, chimpanzee and mouse or human, chimpanzee, mouse, rat and dog). Genes indicated in bold were validated by immunoblotting or luciferase reporter assay (+). –, no detectable decrease of endogenous target protein levels or luciferase reporter activity following microRNA overexpression; ND, not determined. Predictions to be tested were not selected by their PicTar score. The average score of the predicted targets for *miR-375* and *miR-124* is 2.66 and 2.04, respectively, which is low. Therefore, our number of false positives is comparable to the predicted signal-to-noise ratio (**Fig. 2**).

pufferfish and zebrafish, using the University of California at Santa Cruz (UCSC) database[20]. Of these alignments, 92% covered all mammalian species, 55% included chicken sequences and 21% spanned all eight vertebrates. Comparing human-mouse sequence pairings of the alignments to pairings independently defined through a gene orthology table (**Supplementary Methods** online) yielded a low error rate of ~3%.

To estimate false positive rates for vertebrate microRNA target site predictions, we recorded all perfectly binding conserved target sites ('anchors') for 58 unique human microRNAs conserved in human, chimpanzee, mouse, rat, dog and chicken (**Supplementary Table 1** online) and for randomized microRNAs[11,13]. **Figure 2a** shows the ratio of the total number of anchor sites for real and randomized microRNAs at different degrees of conservation. The inclusion of the dog and chicken genomes substantially improved the signal-to-noise ratio from 1.8 for human, chimpanzee and mouse to 2.3 and 3.6, respectively. Overall, our results suggest that on average, each microRNA targets ~200 transcripts above noise, similar to other predictions[21]. We recorded signal-to-noise ratios for the number of transcripts with at least n conserved anchors for each microRNA separately (**Fig. 2b**). The multiplicity of sites in a UTR, which is scored by PicTar, also leads to a substantial increase of signal-to-noise[11]. Encouraged by these results, we used PicTar to make ranked target predictions for all currently available, conserved microRNAs (**Supplementary Tables 2** and **3** online). Specificity and sensitivity strongly correlated with the PicTar score (**Fig. 2c**). The specificity as a function of PicTar score for sets of coexpressed microRNAs used for the tissue-specific target predictions in four mammalian tissues (**Supplementary Table 4** online) is shown in **Figure 2d**. In addition, we validated 7 of 13 predicted targets of *miR-124* and *miR-375* by either western blotting or luciferase reporter assays, consistent with our false positive estimates (**Table 1**).

It has been proposed that microRNAs can target genomic DNA in animals and induce transcriptional silencing of genes through chromatin modification[1]. To test this hypothesis, we ran PicTar on genome-wide sets of nontranscribed upstream sequences (**Supplementary Methods** online). We found, in contrast to our results for 3′ UTRs, neither correlation of binding site positions with evolutionary conservation nor substantial differences in the conservation of putative target sites for real or randomized microRNAs (data not shown). Our data suggest that most animal microRNAs recognize targets in genomic DNA by mechanisms not captured by our algorithm, target sequences other than proximal upstream sequences or do not target genomic DNA to a considerable extent.

To provide a crude estimate of the number of microRNAs that may coordinately regulate target genes, we counted how many sets of three microRNAs have anchor sites in common transcripts. We plotted the probability that a fixed triplet of microRNAs has an anchor site for

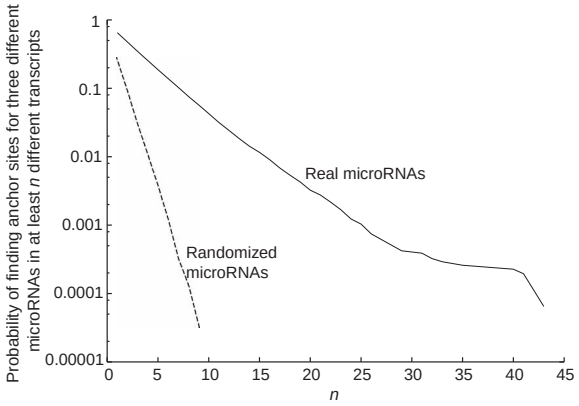

Figure 3 Estimate of the number of coordinately regulated targets for sets of three microRNAs. The probability $p(n)$ that a set of three microRNAs hits at least n transcripts is plotted on a log scale for real (upper curve) and random (lower curve) microRNAs as a function of n. The criterion for a 'hit' was the presence of at least one anchor site, conserved in human, chimpanzee, mouse, rat and dog, for each microRNA in the triplet. The probability of obtaining not a single hit for a triple of real microRNAs is $1 - p(1) = 0.35$. $p(n)$ drops off exponentially and much more steeply for random microRNAs, indicating that PicTar runs with random microRNAs will typically yield a vastly reduced number of predictions. $p(25)$ is ~0.001 for real microRNAs, thereby indicating that only ~30 of ~30,000 possible sets of microRNA triples (sampled from our set of 58 microRNAs) are candidates to coordinately regulate at least 25 different transcripts each.

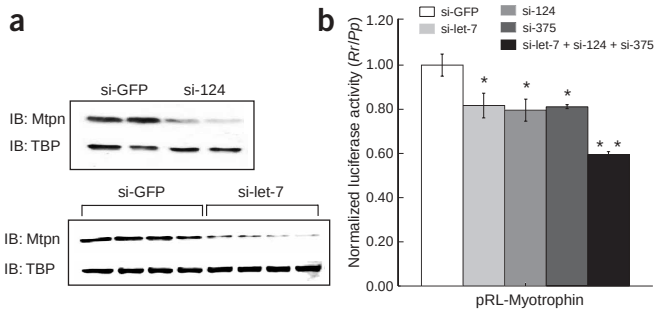

Figure 4 Regulation of *Mtpn* by *miR-375*, *miR-124* and *let-7b*. (**a**) Immunoblotting (IB). N2A cells were transiently transfected with siRNAs designed against eGFP (si-GFP), *miR-124* (si-124) or *let-7b* (si-let-7b) and lysed after 48 h. *Mtpn* expression was assessed after SDS-PAGE and immunoblotting with antibodies to Mtpn. TBP (TATA-binding protein) expression was analyzed as a loading control. (**b**) Dual luciferase assay of N2A cells transfected with a *Renilla reniformis* luciferase (*Rr*-luc) construct containing the full length 3′ UTR of *Mtpn* and si-GFP, si-124, si-let-7b, si-375 or all three siRNAs for 48 h and lysed. A *Photinus pyralis* luciferase (*Pp*-luc) served as an internal transfection control. The ratios of *Rr*-luc to *Pp*-luc expression were normalized to the si-GFP transfections. Error bars represent the standard error (s.e.) from three independent experiments. $^{*}P \leq 0.05$; $^{**}P \leq 0.01$.

each microRNA in at least *n* different transcripts (**Fig. 3**). This probability decays exponentially with *n*. Roughly two thirds of all possible microRNA triplets could coordinately regulate transcripts. The probability that a fixed set of three microRNAs hit more than 10 or 25 targets was 0.04 or 0.001, respectively. Roughly 600 of all triples drawn from the 58 microRNAs could coordinately regulate more than ten targets each with at least two of the three anchor sites above noise. We also computed the same statistics for randomized microRNA sequences (**Fig. 3**), which showed that running PicTar with sets of randomized microRNAs resulted in markedly fewer predicted targets.

We hypothesized that the three most highly expressed microRNAs in the murine pancreatic cell line MIN6 (ref. 22), *miR-124* (11.5% of the total microRNA profile), *miR-375* (6.5%) and *let-7b* (6.5%), may act together on a target gene. We used them as input for PicTar, requiring at least one anchor site to be conserved between human, chimpanzee, mouse, rat and dog for each microRNA. We examined the results for *Mtpn*, a known target of *miR-375* (ref. 22). When we searched ~18,500 3′ UTR alignments with *miR-375* only, *Mtpn* was at rank 102, with one predicted binding site with a nucleus 3,135 nt downstream of the stop codon and a probability of 0.74. We previously validated the functionality of this site[22]. Searching for *miR-124* targets put *Mtpn* at rank 727. Using both *miR-124* and *miR-375* as the search set boosted the rank to 14. Finally, searching for targets of *miR-124*, *miR-375* and *let-7b* placed *Mtpn* at rank 4.

To validate these predictions experimentally, we tested *Mtpn* regulation by *miR-124* and *let-7b* using two methods. First, we transfected neuroblastoma N2A cells with short interfering RNA (siRNA) duplexes that are homologous in sequence to *miR-124* (si-124) or *let-7b* (si-let-7b). We observed in both cases a decrease in endogenous *Mtpn* expression by western blotting (**Fig. 4a**). Second, to test whether *Mtpn* is a target of *miR-124* or *let-7b*, we subcloned the *Mtpn* 3′ UTR downstream of a luciferase reporter gene. Cotransfection of N2A cells with this reporter construct and with si-124 or si-let-7b substantially decreased luciferase activity compared with cotransfection with a control siRNA targeting eGFP (si-GFP; **Fig. 4b**). Downregulation by si-124 and si-let-7b was similar to that caused by si-375. Furthermore, cotransfection of N2A cells with the luciferase reporter and a pool of si-124, si-let-7b and si-375 resulted in normalized luciferase activity that was substantially less than the activity in any of the other cotransfections, suggesting that *Mtpn* is regulated by the coordinate action of all three microRNAs. Together, our results provide evidence for a direct and microRNA concentration–dependent regulation of *Mtpn* by *miR-375*, *miR-124* and *let-7b* and thus establish that *Mtpn* expression is coordinately regulated by three highly expressed pancreatic microRNAs.

In summary, we developed a computational approach that successfully identifies not only microRNA target genes for single microRNAs but also targets that are likely to be regulated by microRNAs that are coexpressed or act in a common pathway. We showed that massive sequence comparisons using previously unavailable genome-wide alignments across eight vertebrate species strongly decreased the false positive rates of microRNA target predictions, allowing PicTar to predict (above noise), on average, ~200 targeted transcripts per microRNA. PicTar's combinatorial microRNA target predictions led to the experimental validation of *Mtpn* as the first mammalian gene shown to be regulated coordinately by three microRNAs. Our results thus provide a computational and experimental model for studying translational gene regulation by multiple microRNAs and a first glimpse at the complexity of translational gene regulation executed by microRNAs.

METHODS

Vertebrate 3′ UTR sequences and alignments. We extracted genome-wide multiple alignments of eight vertebrates from the UCSC Genome Database. These alignments were built from the following genome assemblies: human, May 2004 (hg17); chimpanzee, November 2003 (panTro1); mouse, May 2004 (mm5); rat, June 2003 (rn3); dog, July 2004 (canFam1); chicken, February 2004 (galGal2); pufferfish, August 2002 (fr1); and zebrafish, November 2003 (danRer1). We used the UCSC mappings of the human RefSeq mRNA data[23] (Release 6, 5 July 2004) to the human genome to define multiple alignments of 3′ UTRs. These alignments cover 19,971 sequences for human and chimpanzee; 19,289 for human, chimpanzee and mouse; 18,717 for human, chimpanzee, mouse and rat; 18,567 for human, chimpanzee, mouse, rat and dog; 11,190 for human, chimpanzee, mouse, rat, dog and chicken; 6,136 for human, chimpanzee, mouse, rat, dog, chicken and pufferfish; and 4,355 for human, chimpanzee, mouse, rat, dog, chicken, pufferfish and zebrafish. The sequence space of each species in the alignment is given by the respective number of nucleotides (**Table 2**). The multiple alignments cover human, chimpanzee, mouse, rat and dog for 90% of all human 3′ UTR sequence nucleotides. Sequences of all eight species are aligned for 21% of all human 3′ UTRs. The coverage for human, chimpanzee, mouse, rat, dog and chicken (55%) is consistent with the estimated number of orthologous human-chicken genes[24].

Table 2 Total number of nucleotides per species in the multiple alignment

	Human	Chimpanzee	Mouse	Rat	Dog	Chicken	Pufferfish	Zebrafish
1	19,253,481	18,720,159	15,610,779	15,071,221	17,356,774	5,485,265	1,334,211	1,688,879
2	14,575,934	14,224,691	13,144,375	12,699,682	13,873,555	4,398,114	1,136,336	1,430,061

Row 1 enumerates the total number of nucleotides in our raw 3′ UTR multiple alignments; row 2 lists the same quantities for unique and repeat masked sequences.

For generating our statistics, we produced a final data set of 3′ UTRs by restricting human 3′ UTR sequences to unique sequences and by masking repeats using the UCSC repeat masks (**Table 2**).

Vertebrate promoter sequences and alignments. Similarly, we used UCSC mappings of human RefSeq mRNA sequences to define 500 bp upstream of transcription start sites ('promoters'). To exclude possible overlaps with 3′ UTRs, we did not include sequences that overlapped any transcript, arriving at a total of 17,883 human sequences. In some cases, however, our promoters will overlap with 5′ UTRs, because transcription start sites are often not known. We constructed multiple alignments for promoters across vertebrates as described above.

PicTar algorithm: identification of single microRNA target sites. The 'nucleus' (or 'seed'), typically a perfectly Watson-Crick–base-paired stretch of ~7 nt in the microRNA:mRNA duplex, has a key role in both target site recognition and repression of the target transcript. The nucleus is usually located in the 5′ end of the microRNA starting at the first or second position[3]. Its free energy correlates with the ability of the microRNA:mRNA duplex to repress translation of the targeted transcript[17]. We used these and other experimental results[12] to define probabilities for a mRNA sequence to be a binding site for a given microRNA. More precisely, we defined a 'perfect nucleus' as a perfectly Watson-Crick–base-paired stretch of 7 nt starting at either the first or the second base of the microRNA (counted from the 5′ end). Insertions or mutations in the mRNA sequence of a perfect nucleus are allowed as long as its free energy of binding, determined by standard RNA secondary structure prediction software, does not increase and does not contain G:U base-pairings. These mutated nuclei are called imperfect nuclei. In accordance with previous studies[7–14], we also require that the free energy of the entire microRNA:mRNA duplex be below a cutoff value. For sites with perfect nuclei, this value is set to 33% of the optimal free energy of the entire mature microRNA binding to a perfectly complementary target site. This filter discarded, on average, only ~5% of all perfect nuclei but increased our signal-to-noise ratio. At present we use a much more stringent filter (66% of the optimal free energy) for sites with imperfect nuclei to safeguard against false positives. A perfect nucleus that survives the filtering is assigned a probability p to be a binding site for the microRNA. The probability for imperfect nuclei is $1 - p$ divided by the total number of imperfect nuclei (typically in the range of 2–20). We worked with a high p ($p \approx 0.8$) because most of the known target sites do not have imperfect nuclei but checked that rankings of UTRs with PicTar scores were not sensitive to particular settings of reasonably high values of p. The current settings strongly disfavor contributions from imperfect sites. More sophisticated ways to assign probabilities to microRNA binding sites will be possible once more targets are validated.

PicTar algorithm: scoring combinations of target sites. PicTar computes a maximum likelihood score that a given RNA sequence (typically a 3′ UTR) is targeted by a fixed set of microRNAs (**Supplementary Note** online). Once the probabilities for each subsequence of the RNA sequence to be a binding site for a microRNA are fixed, the scoring of PicTar is similar to the Ahab algorithm as described[15] with the following five implementation details. First, PicTar sets the length of putative microRNA binding sites to the length of the corresponding nuclei. This captures the experimental result that overlapping binding sites seem to act independently as long as their nuclei are not overlapping[17]. Second, a short 3′ UTR (<300 bp) cannot be used to reliably estimate its own background nucleotide frequencies. In these cases, we take the linear combination of the background nucleotide frequencies estimated from the UTR and background frequencies estimated from all UTRs for the same species in our data set. Third, we use the Baum-Welch algorithm[25] to compute maximum likelihoods. Convergence of the logarithm of the partition sum is checked up to a precision of 0.0005. Fourth, we use the optimized prior for background when computing the partition sum for background only. Fifth, the order of the model for background sequence is set to 0.

PicTar algorithm: genome-wide PicTar runs and cross-species comparisons. We first precomputed the positions of all possible microRNA nuclei in all UTR sequences with the program nuclMap. We checked whether nuclei for the same microRNA fall into overlapping alignment positions for all species under consideration. If nuclei are conserved by these criteria, we checked whether the optimal free energy of their predicted microRNA:mRNA duplexes passed our filtering criteria. Perfect nuclei that survived these steps are called anchors. The number of anchors in a UTR determines whether a transcript will be scored by PicTar. If so, the optimal free energy of all sites with perfect or imperfect nuclei in each UTR sequence is used to filter out improbable target sites. The remaining sites for each UTR are input to PicTar to compute a score for each UTR in the multiple alignment. To obtain a final score that reflects the probability that the UTR is regulated by the given set of microRNAs, we averaged the scores for all species that were used to define anchor sites. This average should reflect the different evolutionary distances between species. We averaged the human and chimpanzee scores and the mouse and rat scores independently to obtain a primate score and a rodent score. These scores were then averaged with the dog score to obtain a score reflecting conservation in all mammals. Similarly, we averaged this mammalian score, the chicken score and the averaged fish scores for an overall score, as appropriate. Running an entire analysis on a standard PC with 2 GB of memory took ~15 min when searching for targets of one to six microRNAs.

Optimal free energy estimates of RNA:RNA duplexes. We calculated free energies of RNA:RNA duplexes using RNAhybrid[14] with options -s3utr_human for vertebrate sequences, -s3utr_worm for nematode sequences and default settings otherwise.

Data sets of known and randomized mature microRNA sequences. We downloaded mature microRNA sequences from Rfam[26] (Release 5.0) and added nine microRNAs[22]. We extracted a subset of microRNAs conserved between human, chimpanzee, mouse, rat, dog and chicken using Rfam annotations of mature vertebrate microRNAs homologous to a human microRNA. Whenever no annotation was available, we used stringent criteria to check conservation of the precursor and for the mature microRNA. We constructed a set of unique microRNAs by lumping together microRNAs with identical bases at positions 1–7 or 2–8 (starting at the 5′ end). We obtained 58 unique microRNAs that are conserved in human, chimpanzee, mouse, rat, dog and chicken. Similar to a previously described method[11], we generated cohorts of unique randomized microRNAs by extracting 8-mers with approximately the same abundance (± 15%) of the 7-mer starting at positions 1 and 2 and the corresponding 7-mer of the considered microRNA in all human 3′ UTRs. Experimenting with numerous other randomization schemes led to comparable signal-to-noise ratios. We attached the 3′ end of each microRNA to the corresponding random 8-mer.

Assay of luciferase activity. We excised the wild-type mouse myotrophin 3′ UTR from IMAGE clone 6839739 and subcloned it downstream of the stop codon in pRL-TK (Promega). We transfected N2A cells with 0.1 μg of the pRL-TK reporter vector encoding *Rr*-luc and 0.1 μg of the pGL3 control vector encoding *Pp*-luc (Promega). We transfected cells with 200 ng of siRNAs; in cases where pools were compared with single siRNAs, the difference was made up using si-GFP. We collected and assayed cells 30–36 h after transfection.

siRNAs. Synthetic microRNAs and siRNAs were synthesized by Dharmacon Research, Inc. We transfected N2A cells with vectors and siRNAs using Lipofectamine 2000 (Invitrogen) in accordance with the manufacturer's instructions.

Cell culture and western blotting. We cultured N2A cells in Dulbecco's modified Eagle medium containing 25 mM glucose and 10% fetal bovine serum. We used a polyclonal antibody to myotrophin at 1:1,000 dilution for western blotting[22].

URLs. PicTar results will be available at http://pictar.bio.nyu.edu. The UCSC Genome Browser is available at http://www.genome.ucsc.edu/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank V. Miljkovic and S. Pueblas for preparing figures for the manuscript. N. Rajewsky thanks T. Tuschl, P. Macino and F. Piano for discussions. This project was funded in part by a grant from the US National Institutes of Health

(to M.S.). D.G. acknowledges a scholarship by the German Academic Exchange Service. K.C.G. and P.M. were supported by grants from the US National Institutes of Health (to F. Píaro) and the US National Science Foundation (to K.C.G.). This research was supported in part by the Howard Hughes Medical Institute grant through the Undergraduate Biological Sciences Education Program to New York University.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 4 January; accepted 23 February 2005
Published online at http://www.nature.com/naturegenetics/

1. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
2. Cullen, B.R. Transcription and processing of human microRNA precursors. *Mol. Cell* **16**, 861–865 (2004).
3. Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
4. Barad, O. *et al.* MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. *Genome Res.* **14**, 2486–2494 (2004).
5. Doench, J.G. & Sharp, P.A. SiRNAs can function as miRNAs. *Genes Dev.* **17**, 438–442 (2003).
6. Hobert, O. Common logic of transcription factor and microRNA action. *Trends Biochem. Sci.* **29**, 426–428 (2004).
7. Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B. & Cohen, S.M. Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**, 25–36 (2003).
8. Stark, A., Brennecke, J., Russell, R.B. & Cohen, S.M. Identification of Drosophila microRNA targets. *PLoS Biol.* **1**, E60 (2003).
9. Rajewsky, N. & Socci, N.D. Computational identification of microRNA targets. *Dev. Biol.* **267**, 529–535 (2004).
10. Enright, A.J. *et al.* MicroRNA targets in Drosophila. *Genome Biol.* **5**, R1 (2003).
11. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. & Burge, C.B. Prediction of mammalian microRNA targets. *Cell* **26**, 787–798 (2003).
12. Kiriakidou, M. *et al.* A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.* **18**, 1165–1178 (2004).
13. John, B. *et al.* Human MicroRNA targets. *PLoS Biol.* **2**, e363 (2004).
14. Rehmsmeier, M., Steffen, P., Hochsmann, M. & Giegerich, R. Fast and effective prediction of microRNA/target duplexes. *RNA* **10**, 1507–1517 (2004).
15. Rajewsky, N., Vergassola, M., Gaul, U. & Siggia, E.D. Computational detection of genomic cis-regulatory modules applied to body patterning in the early Drosophila embryo. *BMC Bioinformatics* **3**, 30 (2002).
16. Schroeder, M.D. *et al.* Transcriptional control in the segmentation gene network of Drosophila. *PloS Biol.* **2**, E271 (2004).
17. Doench, J.G. & Sharp, P.A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
18. Banerjee, D. & Slack, F. Control of developmental timing by small temporal RNAs: a paradigm for RNA-mediated regulation of gene expression. *Bioessays* **24**, 119–129 (2002).
19. Reinhart, B.J. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **24**, 901–906 (2000).
20. Karolchik, D. *et al.* The UCSC Genome Browser Database. *Nucleic Acids Res.* **31**, 51–54 (2003).
21. Lewis, B.J., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
22. Poy, M.N. *et al.* A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* **432**, 226–230 (2004).
23. Pruitt, K.D., Tatusova, T. & Maglott, D. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts, and proteins. *Nucleic Acids Res.* **33**, D501–D504 (2005).
24. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
25. Sinha, S., van Nimwegen, E. & Siggia, E.D. A probabilistic method to detect regulatory modules. *Bioinformatics* **19**, i292–i301 (2003).
26. Griffiths-Jones, S. The microRNA Registry. *Nucleic Acids Res.* **32**, D109–D111 (2004).