Nucleic Acids Research, 2005, Vol. 33, No. 8 ***2697–2706***
doi:10.1093/nar/gki567

Clustering and conservation patterns of human microRNAs

Yael Altuvia, Pablo Landgraf[1], Gila Lithwick, Naama Elefant, Sébastien Pfeffer[1], Alexei Aravin[1], Michael J. Brownstein[2], Thomas Tuschl[1] and Hanah Margalit*

Department of Molecular Genetics and Biotechnology, Faculty of Medicine, The Hebrew University, PO Box 12272, Jerusalem 91120, Israel, [1]Laboratory of RNA Molecular Biology, The Rockefeller University, 1230 York Avenue, Box 186, New York, NY 10021, USA and [2]Laboratory of Genetics NIMH/NHGRI, National Institutes of Health, Building 36, Room 3D06 Bethesda, MD 20892, USA

Received February 2, 2005; Revised and Accepted April 21, 2005

ABSTRACT

MicroRNAs (miRNAs) are ~22 nt-long non-coding RNA molecules, believed to play important roles in gene regulation. We present a comprehensive analysis of the conservation and clustering patterns of known miRNAs in human. We show that human miRNA gene clustering is significantly higher than expected at random. A total of 37% of the known human miRNA genes analyzed in this study appear in clusters of two or more with pairwise chromosomal distances of at most 3000 nt. Comparison of the miRNA sequences with their homologs in four other organisms reveals a typical conservation pattern, persistent throughout the clusters. Furthermore, we show enrichment in the typical conservation patterns and other miRNA-like properties in the vicinity of known miRNA genes, compared with random genomic regions. This may imply that additional, yet unknown, miRNAs reside in these regions, consistent with the current recognition that there are overlooked miRNAs. Indeed, by comparing our predictions with cloning results and with identified miRNA genes in other mammals, we corroborate the predictions of 18 additional human miRNA genes in the vicinity of the previously known ones. Our study raises the proportion of clustered human miRNAs that are <3000 nt apart to 42%. This suggests that the clustering of miRNA genes is higher than currently acknowledged, alluding to its evolutionary and functional implications.

INTRODUCTION

MicroRNAs (miRNAs) are very small (~22 nt) single-stranded non-coding RNA (ncRNA) molecules, processed from hairpin precursors of ~70 nt (pre-miRNA), extracted, in turn, from primary transcripts (pri-miRNA) (1–3). MiRNAs have lately gained much interest, as recent genome-wide studies have shown that they are widespread in a variety of organisms and are conserved in evolution. In fact, they are now considered one of the largest gene families, and a growing number of biological processes involving miRNAs are continuously discovered [for review see (4–6)].

MiRNAs in animals are found in diverse genomic locations. Most miRNAs are encoded in intergenic regions, but there are also many miRNAs that are hosted within the introns of pre-mRNAs or encoded within ncRNA genes (7,8). Interestingly, it was observed that there are miRNA genes, both hosted and non-hosted, which are clustered. In fact, the clustering propensity of miRNA genes was noticed from the first days of the massive direct cloning of short ncRNA molecules (9–12). Usually, there are between two to three miRNA genes in a cluster. However, larger clusters were also identified, such as the human hsa-miR-17 cluster composed of six members (10,11), which is also conserved in other mammals (13), or the *Drosophila melanogaster* cluster of eight miRNA genes (12). Recently, Seitz *et al*. (14) predicted a huge cluster of 40 miRNA genes located in the ~1 Mb human imprinted 14q32 domain, several of which were shown to be expressed. Clustered miRNA genes may show high similarity in sequence, but they can also differ [(12), and reviewed in (5)]. Accumulating evidence suggests that clustered miRNAs are transcribed as polycistrons and have similar expression patterns [(1,9,12,14–18), and reviewed in (2,5)].

Previous studies estimated the fraction of clustered miRNA genes in *D.melanogaster* to be ~50%, noting that the miRNA genes in human are less clustered [reviewed in (5)]. Here, we conduct a comprehensive analysis of the genomic organization and conservation patterns of miRNA genes in the human genome. We show that 37% of the known human miRNA genes included in this study are highly clustered and have typical conservation patterns, persistent along the cluster.

*To whom correspondence should be addressed. Tel: +972 2 6758614; Fax: +972 2 6757308; Email: hanah@md.huji.ac.il

Furthermore, we show that conserved stem–loop structures with miRNA-like properties appear preferentially in the vicinity of previously known miRNA genes, suggesting that these may be putative miRNA genes clustered with the previously known ones. For 18 of our predictions, we found corroborations either by cloning results or by sequence similarity to experimentally verified miRNAs in other mammals. Our results indicate that the clustering of miRNA genes in the human genome is higher than previously recognized. Such clustering may be related to the evolutionary mechanisms responsible for the spreading of miRNA genes in the human genome. It may also suggest that the expression of miRNAs in a cluster is co-regulated, and that they play a role in a common molecular process.

MATERIALS AND METHODS

Database organization

The human miRNA precursors used in the analysis (207 in total), were derived from the miRNA registry release 4.0 (http://www.sanger.ac.uk/Software/Rfam/mirna/) (19). Around 55% of these entries are supported by experimental evidence and ~45% are considered human miRNAs by sequence similarity to known miRNAs in other mammals. For the general clustering analysis, we used all 207 human miRNA precursors. For all other analyses, we divided the human miRNA precursors into two separate sets based on their genomic locations. The genomic locations were derived from the UCSC July 2003 human genome assembly build 34, hg16 (20,21) (http://genome.ucsc.edu). The first set included 71 miRNA genes that reside within the introns or untranslated regions of same-strand protein-coding genes. These regions were extracted from the refGene and knownGene tables of the UCSC annotation database (21). This set was termed MH (pre-mRNA-hosted). The second set included 132 miRNA genes in intergenic regions that did not overlap any same-strand protein-coding gene. This set was termed NMH (non-pre-mRNA-hosted). Three miRNAs (hsa-miR-205, hsa-miR-125a and hsa-miR-133a-2) with ambiguous location definitions were excluded. In addition, hsa-miR-22 was excluded due to an ambiguous annotation at the time of the analysis.

Evaluation of the general clustering of miRNA genes

Nearest neighbor distances between consecutive miRNA genes were computed and the average distance was calculated across all chromosomes. To evaluate the statistical significance of the miRNA clustering, the average of the nearest neighbor distances was compared to an equivalent average of the nearest neighbor distances of random positions. On each chromosome, we selected random positions whose number was equal to the number of miRNA genes. We computed the nearest neighbor distances between consecutive random points and the average as described above, and checked whether the average nearest neighbor distance of the random group is lower than that of the miRNAs. This procedure was repeated 1000 times. The fraction of times for which the random averages were smaller than the miRNAs average provides the statistical significance for the miRNA clustering.

Extraction of the miRNAs with their flanking regions

For the miRNA precursors in the NMH set, we extracted a fragment that included the miRNA precursor flanked by 3000 nt on each side, or up to the nearest same-strand protein-coding gene if the latter was closer than 3000 nt. In cases where the distance between consecutive miRNAs was <3000 nt, they were merged into a single fragment. Merging all neighboring miRNAs resulted in a total of 103 regions with an average length of 5911 nt.

As the MH miRNAs reside within a defined genomic entity and since most of the hosted clusters of miRNAs tend to reside in the same genomic entity (7), we extracted for MH miRNAs the whole intron, rather than a flanking region of defined length. For alternative splicing variants, the largest annotated intron was used, and non-UTR (untranslated region) introns were preferred over UTR introns. One miRNA (hsa-miR-198) resided on a UTR/CDS exon. Only the UTR-exon region was extracted for this miRNA. The final MH set included 55 fragments with an average length of 16 122 nt and a median length of 4770 nt.

Generation of random sets

Pool of random fragments for comparison with the NMH miRNAs. Same-direction protein-coding gene coordinates were extracted from the refGene and knownGene tables of the UCSC annotation database (21). Alternatively spliced variants and overlapping same-direction genes were merged into a single consecutive region, defining the boundaries of the intergenic regions. Two sets of intergenic regions were extracted, one for each strand. Regions that were annotated as gaps and the NMH set fragments themselves were excluded. We also excluded all the repeats and low complexity regions of length $\geqslant$200 nt.

Pool of random fragments for comparison with the MH miRNAs. Intronic regions were used to generate random datasets, one for each strand. These were based on the refGene and knownGene tables (21). Alternatively spliced and overlapping same-direction gene introns were merged, and intronic regions that overlapped same-direction exons were excluded.

Twenty-five sets of 103 random sequences for the NMH set and 25 sets of 55 sequences for the MH set were generated. The random sequences used in the analysis maintained the same length distribution as the MH and NMH regions. Chromosomal distribution was not maintained, except for the exclusion of chromosome Y, which has no reported miRNA genes.

Conservation score assignments and extraction of conserved regions

For the extraction of the conserved regions, we used the UCSC phastCons conservation scores, which are based on the multi-species alignments of Human (hg16), Chimp (pnTro1), Mouse (mm3), Rat (rn3) and Chicken (galGal2) (22,23) (http://genome.ucsc.edu). Regions containing at least 40 successive positions with a conservation score of at least 0.3 were defined as conserved. Within these regions, occurrences of up to two consecutive positions with conservation values <0.3 were also allowed.

RNA folding procedure

The entire conserved region was folded using the RNAfold program (with the −d0 option) (24). Sub-regions that folded into a stem–loop with at least 20 bp were then refolded independently to ensure that they maintain their stem–loop structure.

Statistical evaluation of the enrichment of miRNA-like sequences in the miRNA-flanking regions

To test whether miRNA-flanking regions are enriched in sequences with miRNA-like properties, we compared the appearance of such sequences in the vicinity of known miRNAs and in random genomic regions. The statistical significance of the difference between the occurrences in the vicinity of miRNAs and in random sequences was determined by Fischer's exact test (with $\alpha = 0.05$). This comparison was repeated 25 times (for all random sets), and the number of tests that showed statistically significant enrichment was counted. The fraction of statistically significant tests out of the 25 tests was compared to that expected at random (0.05) by a binomial test, and this provided the statistical significance of the whole analysis.

Total RNA isolation, cloning and annotation

Small RNAs were isolated from 100–200 μg of total RNA and cloned as described previously (25–27). The annotation was based on information from GenBank (http://www.ncbi.nih.gov/Genbank/), a dataset of human tRNA sequences (http://lowelab.ucsc.edu/GtRNAdb), a dataset of human sn/snoRNA sequences (http://condor.bcm.tmc.edu/smallRNA/Database, snoRNA-LBME-db at http://www-snorna.biotoul.fr/index.php and NONCODE v1 at http://noncode.bioinfo.org.cn/), the miRNA registry release version 5.1, and the repeat element annotation of version 17 of the human genome assembly from UCSC (http://genome.ucsc.edu).

Cell lines and tissues

Pituitary gland was dissected 2 h postmortem following the written consent of the person's relatives. The identity of the person was obscured for privacy reasons.

The human breast cancer cell lines MCF7 and SkBr3 were gifts of Dr Neal Rosen (Memorial Sloan-Kettering Cancer Center, NY), and were maintained in 1:1 mixture of DME:F12 medium supplemented with 100 U/ml penicillin, 100 μg/ml streptomycin, 4 mM glutamine and 10% heat-inactivated fetal bovine serum, and incubated at 37°C in 5% CO_2. The human neuroblastoma cell line BE(2)-M17 (ATCC:CRL-2267) was maintained in 1:1 mixture of OptiMem:F12 medium supplemented with non-essential amino acids, 10% heat-inactivated fetal bovine serum, and incubated at 37°C in 5% CO_2.

RESULTS

Clustering of miRNA genes in the human genome

Short distances between consecutive genes on the chromosome imply that they are clustered. However, the definition of 'short distance' is not obvious and depends to a great extent on the chromosome length. In the human genome, where the chromosomes consist of millions of nucleotides, distances of thousands of nucleotides may still be considered as relatively short, especially considering the considerable size of some genes. To analyze the distance distribution of miRNA gene pairs, we computed the distances between same-strand consecutive miRNA genes obtained from the miRNA registry release 4.0 (19). The 'same-strand' approach that was performed throughout this study was chosen since same-strand miRNA clustering might imply that they are transcribed together [(1,9,12,14–18) and reviewed in (2,5)]. The cumulative distance distribution of the miRNA gene pairs is presented in Figure 1. As can be seen, 10, 20 and 30% of the miRNA gene pairs are separated by <100, 1000 and 10 000 nt, respectively. Comparison of the miRNA pair distances with random distances revealed that the miRNA pair distances are statistically significantly smaller than expected at random ($P < 0.001$) (see Methods). Interestingly, the shortest distance found between two miRNAs on opposite strands in our database was 38 040. This suggests that indeed 'short-range' clustering is strongly linked to 'same-strand' clustering, which in turn is more likely to be linked to polycistronic transcription.

We compared the clustering of the miRNA genes to the clustering of other genes encoded in the human genome. As shown in Figure 1, at the very short distances the fractions of miRNA genes are higher than those of other non-coding RNA genes (snoRNA, tRNA and snRNA genes). tRNA gene pairs are found preferentially in the distance range of 10^4 to 10^6 nt while most snoRNA genes are highly clustered with a pairwise distance of $<10^4$. As to snRNAs, although it has been noted that at least some of them tend to repeat in tandem [e.g. (28)], they seem to be the least clustered among the four RNA types

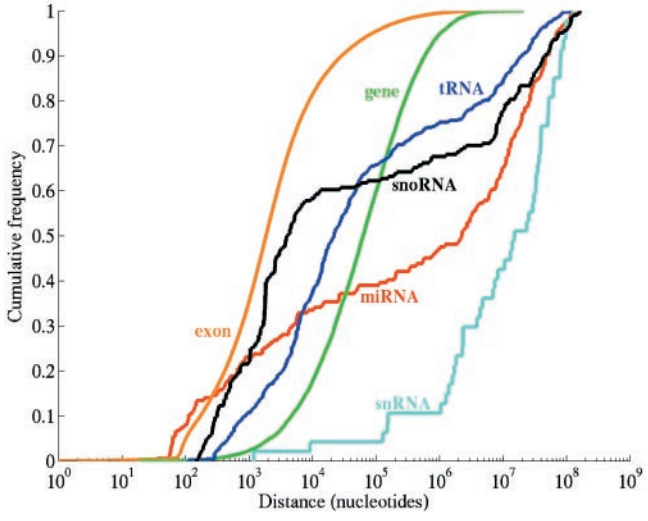

Figure 1. Cumulative distance distribution of miRNA genes and other types of human genomic functional elements. For each of the described elements, the distances (in nucleotides) between every two same-chromosome same-strand successive elements were calculated. Distance is drawn on a logarithmic scale. The different elements are marked: orange (exon of protein-coding genes), green (protein-coding gene), black (snoRNA), blue (tRNA), red (miRNA) and cyan (snRNA). The genomic coordinates were derived from the UCSC July 2003 human genome assembly build 34, hg16 (20,21) (http://genome.ucsc.edu). Protein-coding genes and exons were based on the refGene and knownGene tables. SnoRNA, tRNA and snRNA pseudogenes were excluded.

studied. As expected, the highest clustering is exhibited by protein-coding exons, where 80% of the exon pairs are separated at distances of $<10^4$ nt.

We next wished to examine the number of miRNA genes per cluster. Such an analysis, however, requires a more precise definition of a cluster. As there are miRNA genes that reside in intergenic regions and others that reside within the pre-mRNA introns or untranslated regions, we treated these two groups separately (see Methods). For the 132 miRNA genes in intergenic regions (termed NMH for non-pre-mRNA-hosted), we defined 3000 nt as the maximal distance for two miRNA genes to be considered as clustered (see Methods). The threshold of 3000 was determined because 27% of the pairwise distances were below 3000, and extending the threshold to 10^4 added relatively few miRNA pairs. Furthermore, this rather small distance makes our analysis more stringent and should prevent overestimation of the number of clusters. By this definition, the miRNA genes in intergenic regions were found to be organized in 22 clusters that included 17 pairs, four triplets and one group of five. For the 71 miRNA genes hosted in pre-mRNA (termed MH for pre-mRNA-hosted), we considered those encoded within the same non-coding element (intron or UTR-exon) as clustered, regardless of their pairwise distance. Nine such clusters were identified, including five pairs, three triplets and one group of six (Supplementary Table 1). Interestingly, 24 of the 25 clustered MH miRNAs were also <3000 nt apart from their nearest clustered miRNA, although this distance was not set as a requirement for the determination of MH miRNAs as clustered. In total, 76 human miRNAs, MH and NMH, were found to be included in 31 clusters (Supplementary Table 1), implying that 37.2% of the miRNAs are clustered. Out of these, 75 miRNAs (36.8%) were found to be clustered with pairwise chromosomal distances of at most 3000 nt. In a recent study, Weber reported a total of 37 human miRNA gene clusters (8), which are in a good agreement with the clusters reported here. Eighty-nine percent of the clustered miRNAs detected by the above analysis were also reported by Weber. About 20% of the clustered miRNAs reported by Weber do not follow our more stringent definition of a cluster, mainly due to the strict distance limit we apply. Two additional clusters reported solely in our study comprised eight miRNAs listed in the more recent Rfam version used in our analysis. Furthermore, in addition to the clusters based on the known miRNAs included in Rfam 4.0, we report on the discovery of additional clusters and on the extension of known clusters (see below).

Conservation patterns of miRNAs

Comparing the human miRNA genes to their homologs in the genomes of mouse, rat, chicken and chimpanzee reveals a typical pattern of conservation. It is similar to the conservation pattern of *Drosophila* miRNA genes based on the comparison of *D.melanogaster* and *D.pseudoobscura* (29), and to the conservation pattern that was very recently found for 10 primate species (30). This conservation pattern is especially impressive in the clusters, where typical peaks of high conservation are found in close proximity (Figure 2). The conservation peaks span the miRNA and its precursor. Interestingly, in many cases there is a trough in the middle of the conservation peak, generating a saddle-like shape, which results from the lower conservation of the loop region in the secondary structure of the miRNA precursor.

Examination of the conservation patterns of the miRNA gene sequences along with the conservation of the sequences in their vicinity, revealed a remarkable phenomenon: in quite a few cases, the typical patterns of conservation, each spanning similar length regions, were also observed in the miRNA gene's flanking regions (Figure 2C–E). This may suggest that overlooked miRNA genes reside in these regions, and that the clustering of miRNA genes along the genome is extensive.

MiRNA-like properties of sequences in the vicinity of known miRNA genes

If indeed the conservation patterns in the vicinity of known miRNA genes allude to possible additional miRNA genes encoded in these regions, these sequences should show sequence features typical to miRNAs and their precursors. To this end, we characterized the sequences in the flanking regions of known miRNA genes, and then repeated the same characterization procedure for 25 random sets. For the MH set, sequences were extracted randomly from intronic regions of pre-mRNAs. For the NMH set, sequences were extracted randomly from intergenic regions (see Methods). We then tested whether the regions adjacent to known miRNAs were statistically significantly enriched with miRNA-like properties compared to the random sets. This analysis was applied separately to the two sets of miRNA genes, MH and NMH.

We first tested whether sequences in the flanking regions of miRNA genes show the conservation patterns and the typical stem–loop structure more than expected at random. Conserved subsequences within these regions were identified and we examined their potential to fold into a stem–loop structure (see Methods). Since only one of the miRNA genes in our data overlapped a protein-coding exon on the complementary strand (miR-150), conserved stem–loop regions overlapping exons on the opposite strand were excluded from all sets, including miR-150. As shown in Table 1, the number of conserved stem–loops is >2-fold higher in the miRNA regions than in random sequences.

A total of 113 NMH miRNA genes (86%) and 60 MH miRNA genes (84%) adhere to both the conservation and folding potential criteria. As shown in Table 1, after subtracting the 113 NMH miRNAs and 60 MH miRNAs from the sequences that satisfy the two criteria, there are an additional 140 sequences in the vicinity of NMH miRNAs and 62 sequences in the vicinity of MH miRNAs that show miRNA-like properties. When compared to random sequences, the enrichment of miRNA-like sequences is statistically significant for the NMH group but not for the MH group ($P \leqslant 3.97 \times 10^{-22}$ and $P \leqslant 0.7$, respectively).

The relatively high frequency of conserved stem–loop subsequences in the flanking regions of known NMH miRNA genes supports the conjecture that additional miRNA genes may reside in these regions. However, as was previously shown, conservation and stem–loop folding potential are not sufficient to differentiate miRNAs from random sequences (29,31), and additional features (mainly more refined structural and evolutionary characteristics of the miRNA precursor) are needed [(29,32–36) and reviewed in (5)]. We therefore

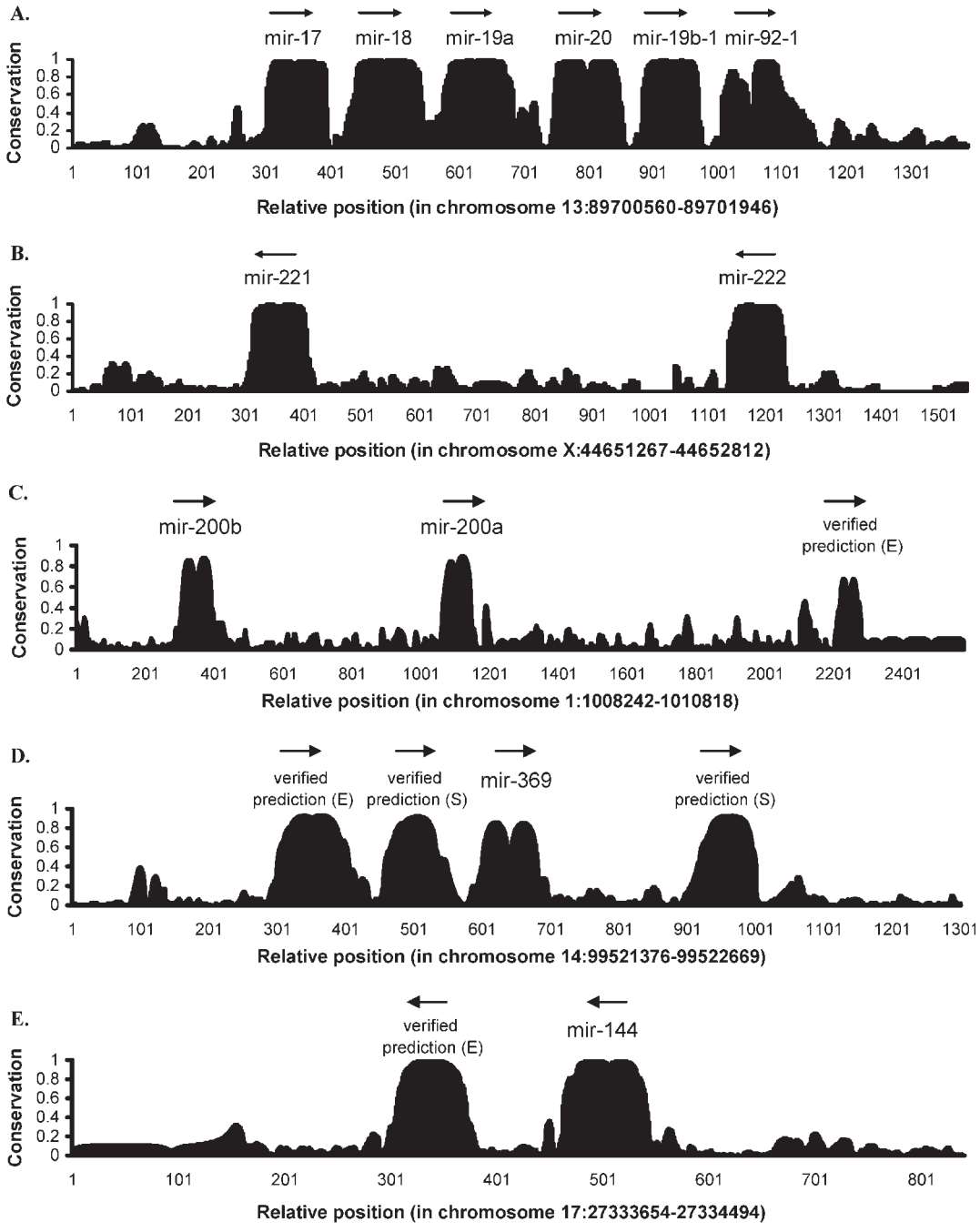

Figure 2. Conservation patterns of known and predicted human miRNAs. The conservation patterns are based on the UCSC phastCons scores (22,23) (http://genome.ucsc.edu). The chromosomal regions of the miRNAs with an additional 3000 nt flanking on both sides are presented. The chromosomal coordinates follow the build 34 assembly (hg16) of the human genome from UCSC (20,21) (http://genome.ucsc.edu/). For simplicity the *x*-axis displays the relative positions. Known miRNAs are designated by their Rfam name omitting the 'hsa' prefix (19). The predicted miRNAs that were verified experimentally fall into two categories: (E)-verified experimentally in this study, and (S)-verified by similarity to a homologous miRNA in another organism. The miRNA orientation is marked by an arrow. (**A**) known large miRNA cluster; (**B**) known miRNA clustered pair; (**C**) example of a miRNA prediction that extends a known pair cluster; (**D**) reveals a new multi-member cluster; and (**E**) reveals a new clustered pair. The plots are not plotted to scale and, therefore, the conserved region width is a function of the length of the presented region; the longer the region, the narrower is the presented profile).

tested whether conserved stem–loops in the vicinity of known miRNAs are indeed enriched in such properties.

We used the 113 NMH miRNAs that passed the conservation and folding criteria as a training set for defining miRNA properties that could potentially filter out non-miRNAs from the subsequences that fulfilled the initial criteria. Three attributes were tested, two are related to the predicted stem–loop structure, and the third is related to the conservation pattern: (i) the free energy of the predicted fold; (ii) the fraction of paired bases out of the total length of the predicted fold; and

Table 1. Comparison of conservation and stem–loop folding potential between miRNA regions and random regions

	NMH[a]	NMH random[b]	MH[c]	MH random[d]
No. of conserved subsequences	525	407	259	147
No. of conserved subsequences with predicted stem–loop[e]	253 (48%)	109 (27%)	122 (47%)	47 (32%)
No. of known miRNAs included in the conserved stem–loop regions[f]	113 (86%)	—	60 (84%)	—
No. of miRNA-neighboring sequences with conserved stem–loops[g]	140	—	62	—

[a]The NMH set includes a total of 103 fragments of 131 miRNA sequences within intergenic regions and their flanking regions.
[b]'NMH random' is comprised of 25 sets, each consisting of 103 random sequences in intergenic regions. This column shows the number of random fragments with a tested property, averaged over the 25 random sets. The percentage is also an average over the 25 sets.
[c]The MH set includes a total of 55 fragments of 71 miRNA gene sequences within pre-mRNA non-coding sequences and their flanking regions.
[d]'MH random' is comprised of 25 sets, each consisting of 55 sequence fragments, chosen randomly from pre-mRNA intronic regions. The column contains information as in 'b'.
[e]As miRNAs are rarely found to overlap exons on the opposite strand, such conserved stem–loop regions were filtered out. The percentage is out of the conserved subsequences.
[f]Percentage is out of the known miRNA genes (131 for NMH and 71 for MH).
[g]Subtraction of the third row from the second.

Table 2. miRNA-related properties in miRNA genes and random data

	Free energy of folding (kcal/mole)	No. of base-paired nucleotides/fold length	Conserved region length (nucleotides)
NMH training set	-41.6 ± 9.9	0.37 ± 0.025	130 ± 81
NMH random set[a]	-23.39 ± 12.6	0.31 ± 0.036	259 ± 213
Applied threshold	$X < -18.8$	$X > 0.31$	$X < 333$

[a]Average number over the 25 random sets.

Table 3. Identification of sequences with miRNA-like properties

	Initial no. of sequences[a]	No. of sequences that passed additional filters[b]
NMH (training set)	113	108 (96%)
MH (test set)	60	60 (100%)
NMH random[c]	109	27 (25%)
MH random[c]	47	12 (25%)
Vicinity of NMH[d]	140	76 (54%)
Vicinity of MH[d]	62	21 (34%)

[a]Number of sequences that show the conservation pattern and folding potential.
[b]Number of sequences that in addition to the conservation and folding potential show three other properties that regard the length of the conserved region and the properties of the folded structure (detailed in Table 2).
[c]Average number and average percentage over the 25 random sets.
[d]All predictions excluding the known miRNAs.

(iii) the length of the conserved region. Based on the differences in these parameters between the miRNA sequences and random sequences, thresholds were set to optimally distinguish between the two sets (Table 2). To test the sensitivity of these thresholds, we applied them to the 60 MH miRNA genes that were left out as a test set (Table 3): 100% of the 60 conserved stem–loops that were related to known MH miRNA genes passed these thresholds. This rate was even higher than the 96% success of the training set (Table 3). Interestingly, applying these thresholds to the random NMH sets filtered out ~75% of the conserved stem–loops. The fraction of paired bases and the length of the conserved region were the attributes that played the major role in the filtering. In the MH group, ~66% of the conserved stem–loops were filtered out. Surprisingly, all the predictions filtered by the conservation attribute were filtered also by the combination of the free energy and fraction of base-pairing attributes. This underlines the importance of evolutionary conservation of miRNA genes on the one hand and of the specific attributes of miRNA precursors on the other hand.

Prediction of novel miRNA genes

We applied the criteria specified in Table 2 to all of the MH and NMH conserved stem–loops that did not coincide with known miRNA genes in our database (Table 3): 54% of the NMH conserved stem–loops fulfilled the selection criteria, more than twice the fraction of such sequences in the random sets ($P \leqslant 2.98 \times 10^{-33}$) (each one of the 25 random sets differed significantly from the NMH set). In the MH set, 34% of the conserved stem–loops fulfilled the selection criteria, 1.36-fold higher than in the random sets. This difference, although statistically significant ($P \leqslant 0.007$), is not very substantial (only 5 of the 25 random sets differed significantly from the MH set).

We compared the predicted sequences to cloning results from human tissues and cell lines, as well as to sequences of experimentally verified miRNAs in other mammals. In applying similarity considerations, we followed Rfam, where >45% of the human entries are supported by similarity to miRNAs in other mammals. Figure 2 demonstrates some of the verified predictions. Figure 2C shows the extension of the cluster of miR-200 to include an additional member that was verified by cloning from human tissues, located ~1000 nt downstream to miR-200a (Tables 4 and 5). Figure 2D demonstrates the identification of three additional miRNA genes in the vicinity of miR-369, one verified by cloning and two supported by sequence similarity to their mouse homolog (Tables 4 and 5). In the last example (Figure 2E), a new miRNA gene, located 200 nt downstream to miR-144, was verified by cloning (Tables 4 and 5), generating a cluster of two genes. In total, in the MH group two predictions could be verified by our cloning data (Tables 4 and 5). In the NMH group, 11 predictions were confirmed by our cloning results, and five predictions were supported by sequence similarity to homologous miRNAs (Tables 4 and 5). Many of the new predictions resided within the Dlk1-Gtl2 imprinted domain of chromosome 14, which is already known to contain clustered miRNAs (14). Our analysis extends these clusters, as well as two other clusters in other chromosomes, and adds three more clusters. All 97 predictions (76 NMH and 21 MH), both verified and

Table 4. Supporting evidence for the predicted miRNA genes in the vicinity of known miRNAs

Coordinates of cluster-founding miRNAs[a]				Predicted miRNA precursor coordinates		Supporting evidence	
Cluster-founding miRNAs	Chromosome[b]	Start	End	Start[c]	End	By cloning (this study)[d]	By similarity[e]
Predicted miRNA genes supported by cloning							
miR-200b, miR-200a	1 (+)	1 008 542	1 009 390	1 010 452	1 010 518	hsa-miR-429 (37)	miR-429 (38)
miR-191 (MH)	3 (−)	49 017 063	49 017 154	49 016 591	49 016 681	hsa-miR-425–3p,5p	Rfam: hsa-miR-425
miR-127,miR-136	14 (+)	99 339 357	99 341 161	99 337 372	99 337 503	**hsa-miR-431**	
				99 338 264	99 338 356	**hsa-miR-433**	
miR-299,miR-323	14 (+)	99 480 172	99 482 195	99 478 434	99 478 519	hsa-miR-379	Rfam: hsa-miR-379
				99 483 163	99 483 242	**hsa-miR-329**	
miR-368	14 (+)	99 496 068	99 496 133	99 497 151	99 497 236	hsa-miR-376a-3p	Rfam:hsa-miR-376a
miR-134	14 (+)	99 511 065	99 511 137	99 510 681	99 510 762	hsa-miR-382	Rfam: hsa-miR-382
				99 512 568	99 512 647	**hsa-miR-453**	
miR-154	14 (+)	99 516 133	99 516 216	99 518 408	99 518 516	hsa-miR-377	Rfam: hsa-miR-377
miR-369	14 (+)	99 521 976	99 522 045	99 521 669	99 521 773	**hsa-miR-409–3p,5p**	Rfam: mmu-miR-409
miR-144	17 (−)	27 334 114	27 334 199	27 333 954	27 334 017	**hsa-miR-451**	cand919 (30)
miR-224 (MH)	X (−)	149 744 663	149 744 743	149 745 713	149 745 797	**hsa-miR-452**	
Predicted miRNA genes supported by similarity							
miR-92,miR-19b,miR-106a	X (−)	132 009 175	132 009 915	132 009 008	132 009 096	—	cand343 (30)
miR-299,miR-323	14 (+)	99 480 172	99 482 195	99 481 385	99 481 464	—	Rfam: hsa-miR-380
miR-368	14 (+)	99 496 068	99 496 133	**99 496 814**	**99 496 913**	—	Rfam: mmu-miR-376b
miR-369	14 (+)	99 521 976	99 522 045	**99 521 825**	**99 521 915**	—	Rfam: mmu-miR-412
				99 522 290	**99 522 369**	—	Rfam: mmu-miR-410

[a]The precursor coordinates are listed. When the predicted miRNA is in the vicinity of a previously known miRNA cluster, the coordinates of the whole cluster are listed, from the initial coordinate of the precursor of the first miRNA to the end coordinate of the precursor of the last miRNA. MiRNAs from the MH group are marked. The cluster-founding miRNA sequences and their precursor sequences are listed in Supplementary Tables 4 and 5, respectively.

[b]The chromosome number, strand and coordinates were taken from the UCSC July 2003 human genome assembly build 34 (hg16) (http://genome.ucsc.edu).

[c]The coordinates of predicted miRNAs are on the same chromosome and strand as the known cluster member/s. Coordinates in bold designate new predictions that were submitted to Rfam since their similarity to a known ortholog was very high. They were named hsa-miR-376b, hsa-miR-412 and hsa-miR-410 respective to their listed order.

[d]Cloned miRNAs were named following Rfam convention. miRNA names in bold designate miRNAs that were submitted to Rfam either as novel miRNAs or as new human orthologs. Hsa-miR-429 was submitted to Rfam by (37) while this paper was submitted. Hsa-miR-376a, hsa-miR-379, hsa-miR-377, hsa-miR-382 and hsa-miR-425 were each listed in Rfam as a miRNA confirmed by similarity. Here, we present experimental evidence for the existence of these miRNAs. miRNAs that were identified from both sides of the precursor stem and matched our predictions were designated with 3p and 5p. The cloned miRNA sequences and their predicted precursor sequences are listed in Supplementary Tables 4 and 5, respectively.

[e]There are three types of 'by similarity' supporting evidence: (i) Similarity to miRNAs in other mammals not recorded in Rfam (ii) Similarity to miRNAs in other mammals where only non-human orthologs are recorded in Rfam. (iii) Similarity to miRNAs in other mammals with a human ('hsa') ortholog recorded in Rfam. All the 'hsa' Rfam entries that are presented here as supporting the predictions were not included in our analysis as they are new entries of the miRNA registry 5.1 (19). All these entries do not have direct experimental evidence in human and they are tagged in Rfam as 'not_experimental'. However they are regarded as human miRNAs 'by similarity'.

unverified, are listed in Supplementary Tables 2 and 3. We compared our predictions with miRNA precursors predicted independently in two recent studies (30,37). Twenty-one of our predictions overlapped miRNA precursors predicted in at least one of these studies: 10 that were verified by cloning, 3 that were verified by similarity and 8 that were not verified. This information is presented in Supplementary Table 2.

It is conceivable that some of the experimentally unverified miRNAs are false positive predictions. Around 25% of the sequences in the random sets passed our filters (Table 3), indicating that there may exist in both our MH and NMH sets sequences with miRNA-like properties which are not actual miRNAs. At the same time, the members of lowly expressed miRNA clusters are cloned rarely and the clone numbers scatter too much to reveal any correlation with respect to co-expression. It is also conceivable that certain miRNA sequences are more difficult to clone due to intrinsic secondary structures or sequence biases that interfere with their experimental detection. However, it is possible to detect low expressed miRNAs by specifically amplifying the candidate miRNAs from miRNA cDNA libraries known to express one of the members of a cluster (33).

DISCUSSION

The clustering propensity of miRNAs and the enrichment of new miRNAs in the vicinity of known miRNA genes have been suggested previously (8,9,12,14,30,34), but have not been comprehensively studied in human. In this study, we systematically evaluated the clustering of miRNA genes in the human genome, exploring different aspects of miRNA clustering. We applied both computational and experimental approaches to rigorously define the clusters on the one hand, and to search for new miRNAs in the vicinity of known ones on the other hand. In using the clustering property for predicting new miRNAs in the vicinity of already known ones, we used the same clustering criteria that were used in the analysis of the known miRNAs.

A pairwise distance analysis of same-strand adjacent miRNAs shows that the distances between the known miRNA genes are smaller than expected at random. Figure 1 shows that for very small distances (up to 100 nt), the fraction of clustered miRNA genes even exceeds that of exons, probably due to their very small size. Especially interesting is the comparison with other RNA genes, where miRNAs are more clustered than all other non-coding RNAs in the very short

Table 5. Cloning frequencies of experimentally verified newly predicted human miRNAs clustered with known human miRNAs[a]

Cluster-founding miRNAs[b]	miRNA[b]	Cell line/Tissue[c] Pituitary gland	MCF7	SkBr3	BE(2)-M17
miR-200b, miR-200a	miR-429	2	2	—	—
	miR-200a	9	7	1	—
	miR-200a*	1	—	—	—
	miR-200b	9	7	1	—
miR-368	miR-368	10	2	—	1
	miR-376a-3p	3	—	—	—
miR-369	miR-409–5p	2	—	—	—
	miR-409–3p	—	—	—	1
	miR-369–5p	1	—	—	—
	miR-369–3p	1	—	—	—
miR-144	miR-451	20	—	—	—
	miR-144	—	—	—	—
miR-224	miR-452	—	—	1	—
	miR-224	—	—	1	—
miR-191	miR-425–5p	1	3	5	1
	miR-425–3p	—	—	3	—
	miR-191	3	5	16	—
	miR-191*	—	—	—	1
miR-127, miR-136	miR-431	—	—	—	3
	miR-433	1	—	—	—
	miR-127	4	—	—	1
	miR-136	2	—	—	31
miR-299, miR-323	miR-329	1	—	—	—
	miR-379	4	1	—	30
	miR-299–3p	—	—	—	1
	miR-299–5p	—	—	—	—
	miR-323	—	—	—	—
miR-134	miR-453	—	—	—	1
	miR-382	—	—	—	1
	miR-134	2	—	—	4
miR-154	miR-377	4	—	—	13
	miR-154	5	—	—	2
Total miRNA clones		1502	767	794	616

[a]Given are the absolute numbers of cloned sequences. The total number of miRNA sequences in the library is indicated at the bottom.
[b]The miRNA sequences and the precursor sequences are listed in Supplementary Tables 4 and 5, respectively.
[c]MCF7 and SkBr3 are human breast cancer cell lines. BE(2)-M17 is a human neuroblastoma cell line.

pairwise distances, but other RNA types, like snoRNA and tRNA exceed them in other distance ranges.

By comparing the frequency of miRNA-like sequences in the vicinity of previously known miRNA genes to their frequency in random regions, we were able to demonstrate that there is a significant enrichment in miRNA-like sequences in the flanking regions of NMH miRNAs. Interestingly, the frequency of miRNA-like conserved stem–loops near MH miRNAs, although statistically significantly higher than in random sequences, seems to be less substantial. We suggest that this difference between the NMH and MH sets results from the different cluster definitions we have used. Unlike the NMH miRNA set, where the search region was limited to at most 3000 nt on each side, the search region for MH miRNAs was defined to be their hosting intron or UTR, which could vary considerably in length. In fact, >40% of the MH regions exceed 10^4 nt. However, except for one MH miRNA, all the known MH miRNAs that were found to be clustered in this study had pairwise distances <3000 nt. This observation still holds for the two MH miRNA predictions that could be verified experimentally: miR-452 that resides 969 nt upstream to miR-224 and miR-425 that resides 381 nt downstream to miR-191 (Tables 4 and 5). This implies that MH clustered miRNAs tend to be closely clustered even within their introns, while the rest of the intron does not show statistically significant enrichment with miRNA-like sequences. Thus, overall, we conclude that NMH and MH miRNAs have similar clustering distance thresholds. The preference for relatively small distances of clustered miRNAs within pre-mRNAs implies that miRNA clustering might be beneficial not only for shared transcription but also for other stages of miRNA processing, such as cleavage or transport.

The analysis scheme that we applied for the study of the miRNA-like properties shares many of its features and methodologies with previously described methods for the prediction of new miRNA genes [reviewed in (5)]. Still, there are several interesting insights that may have implications for the improvement of future algorithms, especially regarding the conservation pattern. As miRNAs are highly conserved across different organisms, all previously reported algorithms for the prediction of animal miRNAs relied on this trait (8,29,30,32–36). Most methods emphasize the importance of a multi-organism comparison [e.g. (29,30,32,33)]. However, this was usually performed by merging the results of several pairwise comparisons. In this study, we used the UCSC phastCons conservation scores (22,23) (http://genome.ucsc.edu), which were based on a multiple alignment of five organisms (human, mouse, rat, chimpanzee and chicken). Since these scores were calculated for each position along the human genome, we could easily derive the conservation patterns of known miRNAs and their proximal regions. Interestingly, miRNAs seem to have a typical conservation pattern which is mainly characterized by its relatively short width and high peak. We have also noticed that there is a symmetric saddle-like pattern, which stems from the more divergent nature of the loop in the structure. This type of pattern was previously observed in a pairwise alignment of two species of *Drosophila* (29), and very recently in a multiple alignment of 10 primates (30). In both cases, however, the compared organisms were closely related. We believe that the derivation of the pattern properties from multiple alignments of distant organisms filters out other conserved regions around the miRNA, which may mask the typical conservation patterns. Indeed, the conservation patterns demonstrated in our study are more pronounced than those observed in the alignments of the 10 primates (30).

Our analysis demonstrated that explicit incorporation of the pattern conservation property as a miRNA-like attribute is very powerful in filtering out random intergenic sequences. This is supported by the findings of Berezikov *et al*. (30). Interestingly, the known miRNAs that we failed to identify by our computational procedure were missed because they did not adhere to the conservation criterion. Among these, half had a longer 'middle' gap than we allowed, suggesting that allowing a longer 'trough' in the conservation pattern may improve the predictions.

In this study, we focused on the characterization of clusters of same-strand miRNA genes that are relatively close (or reside within the same genomic unit), using high resolution conservation scores derived from multiple-organism alignments, and

explicitly including the conservation pattern attributes in the filtering process. Our results show that the extracted miRNA-like features highly succeed in identifying the miRNAs and in filtering noise. We were able to identify 100% of the MH miRNA genes, which were excluded from the training set and kept as a test set. When applied to identification of miRNA sequences in the vicinity of previously known miRNA genes, selecting among the conserved regions the ones with folding potential reduced the predictions by >50% (Table 1), and application of the additional criteria narrowed down these remaining sequences by an additional 50% (Table 3). All in all, the various filters narrowed down the predictions in the vicinity of miRNA genes by ~80%, resulting in 97 predictions: 18 of these predictions could be supported either by cloning results or by sequence similarity. The 18 miRNA predictions were all within <3000 nt apart from their nearest clustered member, raising to 42% the proportion of clustered human miRNA genes with pairwise chromosomal distances of at most 3000 nt. The determination of 3000 nt as the distance threshold may lead to an underestimation of the number of miRNA clusters in human. In a recent study, Baskerville and Bartel (18) demonstrated that proximal pairs of miRNAs tend to coexpress, and that the correlation in expression dropped when the distance between the miRNA pairs exceeded 50 kb. They suggested that clustered miRNAs are expected to be found within this range, which is one order of magnitude larger than our threshold. Indeed, if we set the distance threshold to 10 000 nt, the fraction of clustered miRNA genes rises to 48%. Also, expressed sequence tag evidence indicates that distant miRNAs may reside on the same transcript (e.g. miR-100, let-7 and miR-125). Thus, by adding transcript considerations to the definition of clusters, the number of clusters may further increase. Still, with the stringent definitions used here we demonstrate a strong phenomenon of miRNA clustering.

The validated miRNAs encoded in the vicinity of previously known ones, together with neighboring sequences with miRNA-like properties, revealed new miRNA clusters and increased the number of known members of some of the previously identified clusters. The polycistronic organization of miRNA genes may have important implications for the evolution of miRNA sequences.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank M. Zavolan, R. Sheridan and C. Sander for contributing to identification of miRNAs in the cloning data. M. Chien, J.J. Russo and J. Ju from Columbia Genome Center are thanked for small RNA library sequencing, and R. Hershberg, E. Akiva and Y. Seldin are thanked for their useful comments. This study was supported by grants from the US–Israel Bi-national Science Foundation (H.M. and T.T.), and by a grant from The Israeli Cancer Research Foundation (H.M.). S.P. is supported by the Lehman Brothers Foundation Fellowship through the Leukemia & Lymphoma Society, P.L. is supported by the Dr Mildred Scheel Stiftung für Krebsforschung of the Deutsche Krebshilfe (German Cancer Aid). Funding to pay the Open Access publication charges for this article was provided by The US-Israel Bi-National Science Foundation.

Conflict of interest statement. None declared.

REFERENCES

1. Lee,Y., Jeon,K., Lee,J.T., Kim,S. and Kim,V.N. (2002) MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.*, **21**, 4663–4670.
2. Cullen,B.R. (2004) Transcription and processing of human microRNA precursors. *Mol. Cell*, **16**, 861–865.
3. Tomari,Y. and Zamore,P.D. (2005) MicroRNA biogenesis: drosha can't cut it without a partner. *Curr. Biol.*, **15**, R61–R64.
4. Ambros,V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
5. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
6. He,L. and Hannon,G.J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nature Rev. Genet.*, **5**, 522–531.
7. Rodriguez,A., Griffiths-Jones,S., Ashurst,J.L. and Bradley,A. (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res.*, **14**, 1902–1910.
8. Weber,M.J. (2005) New human and mouse microRNA genes found by homology search. *FEBS J.*, **272**, 59–73.
9. Lau,N.C., Lim,L.P., Weinstein,E.G. and Bartel,D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, **294**, 858–862.
10. Lagos-Quintana,M., Rauhut,R., Lendeckel,W. and Tuschl,T. (2001) Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853–858.
11. Mourelatos,Z., Dostie,J., Paushkin,S., Sharma,A., Charroux,B., Abel,L., Rappsilber,J., Mann,M. and Dreyfuss,G. (2002) miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.*, **16**, 720–728.
12. Aravin,A.A., Lagos-Quintana,M., Yalcin,A., Zavolan,M., Marks,D., Snyder,B., Gaasterland,T., Meyer,J. and Tuschl,T. (2003) The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell*, **5**, 337–350.
13. Tanzer,A. and Stadler,P.F. (2004) Molecular evolution of a microRNA cluster. *J. Mol. Biol.*, **339**, 327–335.
14. Seitz,H., Royo,H., Bortolin,M.L., Lin,S.P., Ferguson-Smith,A.C. and Cavaille,J. (2004) A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Res.*, **14**, 1741–1748.
15. Houbaviy,H.B., Murray,M.F. and Sharp,P.A. (2003) Embryonic stem cell-specific MicroRNAs. *Dev. Cell*, **5**, 351–358.
16. Bashirullah,A., Pasquinelli,A.E., Kiger,A.A., Perrimon,N., Ruvkun,G. and Thummel,C.S. (2003) Coordinate regulation of small temporal RNAs at the onset of *Drosophila* metamorphosis. *Dev. Biol.*, **259**, 1–8.
17. Sempere,L.F., Sokol,N.S., Dubrovsky,E.B., Berger,E.M. and Ambros,V. (2003) Temporal regulation of microRNA expression in *Drosophila melanogaster* mediated by hormonal signals and broad-Complex gene activity. *Dev. Biol.*, **259**, 9–18.
18. Baskerville,S. and Bartel,D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.
19. Griffiths-Jones,S. (2004) The microRNA Registry. *Nucleic Acids Res.*, **32**, D109–D111.
20. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
21. 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) The UCSC Genome Browser Database. *Nucleic Acids Res.*, **31**, 51–54.
22. Siepel,A. and Haussler,D. (2004) Combining phylogenetic and hidden Markov models in biosequence analysis. *J. Comput. Biol.*, **11**, 413–428.
23. Siepel,A. and Haussler,D. Phylogenetic hidden Markov models. In Nielsen,R. (ed.), *Statistical Methods in Molecular Evolution*. Springer (in press).

24. Hofacker,I.L., Fontana,W., Stadler,P.F., Bonhoeffer,L.S., Tacker,M. and Schuster,P. (1994) Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, **125**, 167–188.
25. Meister,G., Landthaler,M., Patkaniowska,A., Dorsett,Y., Teng,G. and Tuschl,T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell*, **15**, 185–197.
26. Pfeffer,S., Lagos-Quintana,M. and Tuschl,T. (2003) Cloning of small RNA molecules. In Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (eds), *Current Protocols in Molecular Biology*. John Wiley and Sons, New York, pp. 26.4.1–26.4.18.
27. Pfeffer,S., Zavolan,M., Grasser,F.A., Chien,M., Russo,J.J., Ju,J., John,B., Enright,A.J., Marks,D., Sander,C. *et al.* (2004) Identification of virus-encoded microRNAs. *Science*, **304**, 734–736.
28. Van Arsdell,S.W. and Weiner,A.M. (1984) Human genes for U2 small nuclear RNA are tandemly repeated. *Mol. Cell. Biol.*, **4**, 492–499.
29. Lai,E.C., Tomancak,P., Williams,R.W. and Rubin,G.M. (2003) Computational identification of *Drosophila* microRNA genes. *Genome Biol.*, **4**, R42.
30. Berezikov,E., Guryev,V., van de Belt,J., Wienholds,E., Plasterk,R.H. and Cuppen,E. (2005) Phylogenetic shadowing and computational identification of human microRNA genes. *Cell*, **120**, 21–24.
31. Ambros,V., Bartel,B., Bartel,D.P., Burge,C.B., Carrington,J.C., Chen,X., Dreyfuss,G., Eddy,S.R., Griffiths-Jones,S., Marshall,M. *et al.* (2003) A uniform system for microRNA annotation. *RNA*, **9**, 277–279.
32. Lim,L.P., Lau,N.C., Weinstein,E.G., Abdelhakim,A., Yekta,S., Rhoades,M.W., Burge,C.B. and Bartel,D.P. (2003) The microRNAs of *Caenorhabditis elegans*. *Genes Dev.*, **17**, 991–1008.
33. Lim,L.P., Glasner,M.E., Yekta,S., Burge,C.B. and Bartel,D.P. (2003) Vertebrate microRNA genes. *Science*, **299**, 1540.
34. Ohler,U., Yekta,S., Lim,L.P., Bartel,D.P. and Burge,C.B. (2004) Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. *RNA*, **10**, 1309–1322.
35. Ambros,V., Lee,R.C., Lavanway,A., Williams,P.T. and Jewell,D. (2003) MicroRNAs and other tiny endogenous RNAs in *C.elegans*. *Curr. Biol.*, **13**, 807–818.
36. Legendre,M., Lambert,A. and Gautheret,D. (2005) Profile-based detection of microRNA precursors in animal genomes. *Bioinformatics*, **21**, 841–845.
37. Xie,X., Lu,J., Kulbokas,E.J., Golub,T.R., Mootha,V., Lindblad-Toh,K., Lander,E.S. and Kellis,M. (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature*, **434**, 338–345.
38. Watanabe,T., Takeda,A., Mise,K., Okuno,T., Suzuki,T., Minami,N. and Imai,H. (2005) Stage-specific expression of microRNAs during *Xenopus* development. *FEBS Lett.*, **579**, 318–324.