20 Zhang, K. *et al*. (2002) A dynamic programming algorithm for haplotype block partitioning. *Proc. Natl. Acad. Sci. U. S. A.* 99, 7335–7339
21 Ideraabdullah, F.Y. *et al*. (2004) Genetic and haplotype diversity among wild-derived mouse inbred strains. *Genome Res.* 14, 1880–1887
22 Yalcin, B. *et al*. (2004) Genetic dissection of a behavioral quantitative trait locus shows that Rgs2 modulates anxiety in mice. *Nat. Genet.* 36, 1197–1202
23 Matsuzaki, H. *et al*. (2004) Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nature Methods* 1, 109–111
24 Twyman, R.M. (2004) SNP discovery and typing technologies for pharmacogenomics. *Curr. Top. Med. Chem.* 4, 1423–1431
25 Doerge, R.W. (2002) Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet.* 3, 43–52
26 Glazier, A.M. *et al*. (2002) Finding genes that underlie complex traits. *Science* 298, 2345–2349

0168-9525/$ - see front matter
doi:10.1016/j.tig.2005.03.010

Genome Analysis

Mammalian microRNAs derived from genomic repeats

Neil R. Smalheiser and Vetle I. Torvik

University of Illinois at Chicago, UIC Psychiatric Institute, MC 912, 1601 W. Taylor Street, Chicago, IL 60612 USA

In this article, we show that a subset of conventional mammalian microRNAs is derived from LINE-2 transposable elements and other genome repeats. These repeat-derived microRNAs arise from conventional precursor hairpins and are distinct from the rasiRNAs, which appear to be processed from long double-stranded RNA precursors. The insertion of transposable elements into new genomic sites appears to be one of the driving-forces that create new microRNAs during mammalian evolution. Two of the LINE-2-derived microRNAs exhibit perfect complementarity to a large family of mRNA and EST transcripts that contain portions of MIR and other LINE-2 elements in their 3′-untranslated regions.

Introduction

MicroRNAs (miRNAs) are small (~18–24 nt) noncoding RNAs that are cleaved from larger (~70 nt) precursors [1,2]. They are thought to elicit mRNA degradation (if they bind in perfect complementarity to the target mRNA) or to arrest mRNA translation (if binding is imperfect). In plants miRNA precursor genes can derive from transposable elements and other genome repeats in both sense and antisense directions [3]; however, the only repeat-associated small RNAs that have been described in fungi and animals are the recently identified rasiRNAs [4–6], which are processed from long double-stranded RNA precursors. In this article, we demonstrate that a subset of conventional mammalian miRNAs also derive from transposable elements. This has implications for the manner in which new miRNA precursors arise during evolution. Moreover, some of these repeat-encoded miRNAs are perfectly complementary to a large family of mRNA and EST transcripts.

Analyzing repeat miRNAs

All human, mouse and rat miRNA precursor sequences from the Sanger miRNA registry (http://www.sanger.ac.uk/Software/Rfam/mirna/; version 4.0; each species has ~220 known miRNAs) were analyzed by the RepeatMasker program (http://www.repeatmasker.org/) to detect well-characterized repeats. Eleven different miRNA precursors contained repeat sequences, including four derived from long interspersed nuclear element 2 (LINE-2) repeats, and others having short interspersed nuclear elements (SINEs), tRNA, mariner DNA repeats, long terminal repeats (LTRs) and simple repeats (Table 1). The majority were highly conserved across human, mouse and rat [miR-28, miR-151/151* (which arise from opposing sides of the same hairpin), miR-321, miR-325 and miR-340], whereas others were restricted to one or two species (human miR-95, rat miR-327, rat miR-333 and rodent miR-341). miR-297 was excluded from further analysis because it lacked an unambiguous precursor, but all of the others arose from classic hairpin precursors and their length distribution resembled the entire set of mammalian miRNAs.

The repeat-containing miRNA precursors were then mapped onto their respective genomes (UCSC Genome Browser, assembly 34, http://genome.ucsc.edu; supplementary file 1 online). The LINE-2-related miRNA precursors were particularly interesting – not only were the hairpins entirely derived from this genomic repeat but also the hairpin foldbacks were formed by the junction of two adjacent LINE-2 segments apposed in opposite orientation (Table 1; Figure 1; supplementary file 1 online). The miRNAs corresponded to two discrete locations near the 3′end of the LINE-2 consensus sequence (Figure 2). By contrast, the other types of genomic repeats were more variably related to the miRNA precursors (supplementary file 1 online).

LINE-2 elements (particularly MIR repeats, which derive from the 3′end of LINE-2) are incorporated into many transcripts [7]. We examined all human, mouse and rat miRNAs and computed the extent of complementarity with all human and mouse mRNAs and ESTs (NCBI RefSeq and EST databases as they appeared in June

Corresponding author: Smalheiser, N.R. (smalheiser@psych.uic.edu).
Available online 27 April 2005

Table 1. Mammalian miRNA precursors that contain genomic repeats[a]

Query sequence	Position in query			Matching repeat	Repeat class (family)
	Begin	End	Dir[b]		
hsa-mir-28	2	35	−	L2	LINE (L2)
hsa-mir-28	43	86	+	L2	LINE (L2)
hsa-mir-95	2	34	+	L2	LINE (L2)
hsa-mir-95	47	81	−	L2	LINE (L2)
hsa-mir-151/151*	2	31	−	L2	LINE (L2)
hsa-mir-151/151*	4	78	+	L2	LINE (L2)
hsa-mir-321	2	47	+	tRNA-Arg-AGG	tRNA
hsa-mir-325	3	42	+	L2	LINE (L2)
hsa-mir-325	52	98	−	L2	LINE (L2)
hsa-mir-340	9	46	−	MARNA	DNA (mariner)
mmu-mir-28	2	35	−	L2	LINE (L2)
mmu-mir-28	43	86	+	L2	LINE (L2)
mmu-mir-151/151*	1	28	−	L2	LINE (L2)
mmu-mir-297–1	1	73	+	RMER12	LTR (ERVK)
mmu-mir-297–2	3	61	+	(TATATG)n	Simple repeat
mmu-mir-321	2	47	+	tRNA-Arg-AGG	tRNA
mmu-mir-325	3	42	+	L2	LINE (L2)
mmu-mir-325	52	98	−	L2	LINE (L2)
mmu-mir-340	12	49	−	MARNA	DNA (mariner)
mmu-mir-341	12	84	+	(CGGT)n	Simple repeat
rno-mir-28	35	86	+		INE (L2)
rno-mir-151/151*	7	36	−	L2	LINE (L2)
rno-mir-151/151*	38	78	+	L2	LINE (L2)
rno-mir-297	1	68	+	(TATG)n	Simple repeat
rno-mir-321	2	47	+	tRNA-Arg-AGG	tRNA
rno-mir-325	3	42	+	L2	LINE (L2)
rno-mir-325	52	98	−	L2	LINE (L2)
rno-mir-327	4	94	−	RodERV21	LTR (ERV1)
rno-mir-333	36	95	+	B2_Rat1	SINE (B2)
rno-mir-340	12	49	−	MARNA	DNA (mariner)
rno-mir-341	12	84	+	(CGGT)n	Simple repeat

[a]Human, mouse and rat miRNA precursor sequences from the Sanger miRNA registry were analyzed by the RepeatMasker program to detect genomic repeats.
[b]Direction (dir) indicates whether the repeat was in sense (+) or antisense (−) orientation with respect to the precursor.

2004). Because there is evidence that G:U base-pairing is tolerated in RNA interference and arrested translation [8], G:U was scored as a potential match. To establish the number of 'hits' that would be expected by chance in the RefSeq databases, each miRNA sequence was scrambled 200 times (100 times permuted randomly, and 100 times permuted to maintain the dinucleotide composition), and their complementarity with human and mouse RefSeq mRNAs was computed. Because there were more hits for miRNAs in EST databases than in RefSeq, it was sufficient to test only 20 scrambled sets (ten random; ten maintaining dinucleotide composition) for each miRNA. EST sequences were assayed in both forward and reverse-complement directions. We scored all cases of perfect complementarity and listed them for each miRNA according to the number of G:U matches involved (up to five; supplementary file 2 online).

Among repeat-derived miRNAs, two (miR-95 and miR-151*) identified multiple transcripts with perfect complementarity and above the level expected by chance. Examination of EST annotation (supplementary file 2) and manual inspection confirmed that the majority of these 'hits' were in the correct (antisense) orientation on the transcripts. MiR-95 had perfect complementarity (with two-to-five G:U matches) to dozens of mRNAs and EST transcripts that were otherwise unrelated except that they contained LINE-2 or MIR repeats in their 3′-UTRs (supplementary file 3). MiR-151* 'hit' LINE-2 repeats within two human mRNAs perfectly (Figure 3), including a hit that involved no G:U matches (GenBank accession no. NM_014400, GPI-anchored metastasis-associated protein homolog C4.4A), and one with two G:U matches (GenBank accession no. NM_005373, myeloproliferative leukemia virus oncogene MPL). Both of these mRNAs were encoded at a different chromosomal site than the miRNA and contained 3′-UTR LINE-2 fragments (Figure 3).

Furthermore, the same analysis was extended to the human genome (ftp://ftp.ncbi.nlm.nih.gov/blast/db/, assembly 34); chromosomes were analyzed in both directions, and controls were performed as described for EST databases. All miRNAs exhibited perfect complementarity (with no G:U matches) to only one-to-three genome sites, with the exception of the tRNA-derived miR-321, which matched five sites. There was no indication of matches to large numbers of genomic sites such as satellite repeats. All of the LINE 2-derived miRNAs stood out from the general miRNA population (and from the other repeat-derived miRNAs) because they were complementary to a significant number of genomic sites with one-to-five G:U matches (supplementary file 2). In each case, the number of genomic sites exceeded the number of sites in the mRNA and EST databases: (i) miR-95 matched 1124 sites perfectly in the genome assembly compared with 159 transcripts in the human databases; (ii) miR-151* matched 27 genomic sites compared with six in the transcript databases; and (iii) miR-28 matched five genomic sites and no sites in the transcript databases.

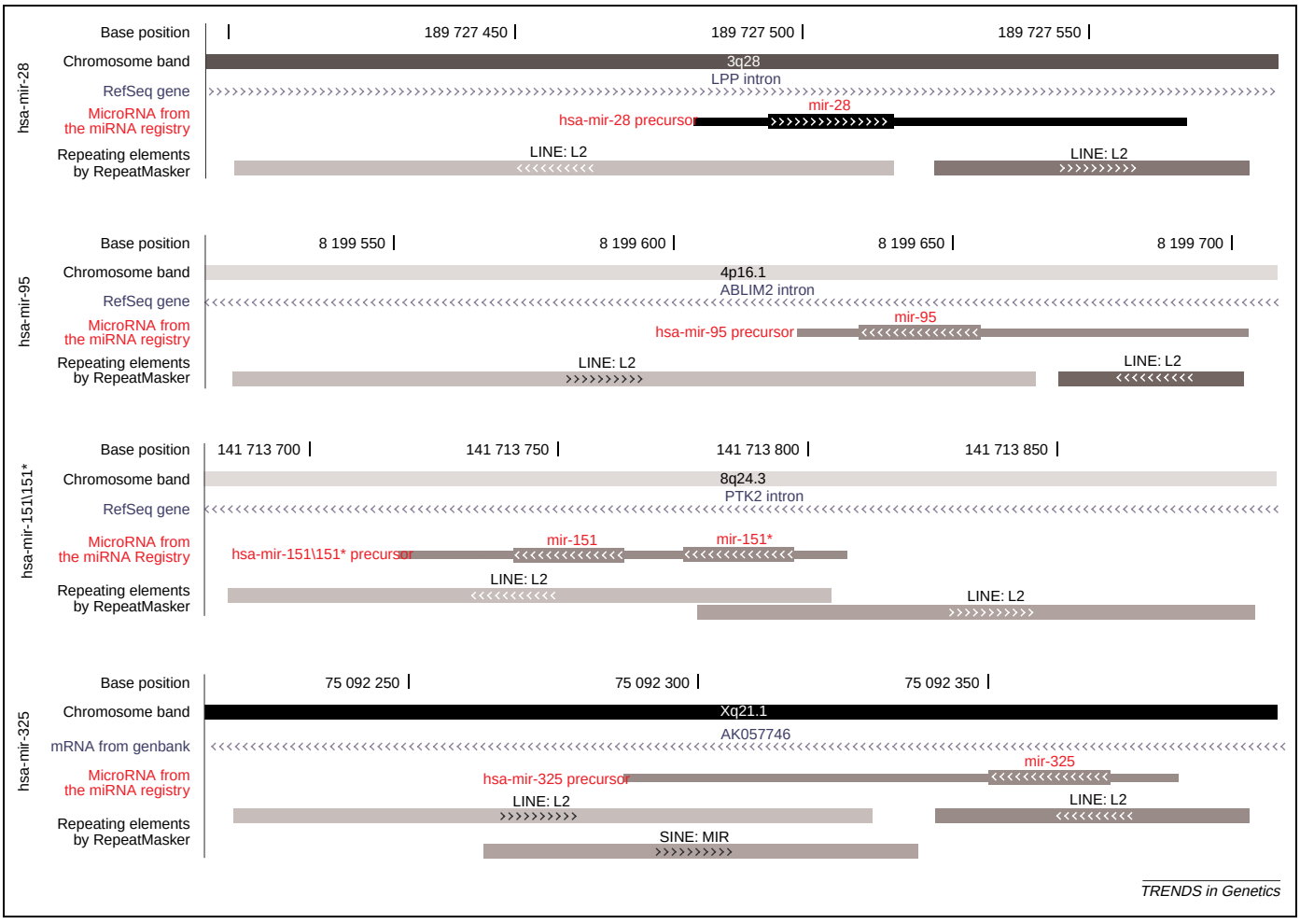

Figure 1. Genomic structure of human LINE-2-derived miRNA precursors. Information was downloaded and edited from the UCSC Genome Browser. Each of the precursors resides within an mRNA intron, and each flanks the junction of two L2 repeats in opposite orientation (darker shading indicates less divergence from the L2 consensus sequence).

Although these raw numbers are not corrected for redundancy, orientation of hits and incomplete database coverage, the perfect chromosomal sites are certain to include many non-transcribed copies of LINE-2 elements.

Finally, Repeatmasker detected no repeat-containing miRNA precursors in chicken or *Drosophila*, and only one in *Caenorhabditis elegans* (cel-mir-69). Thus, the phenomenon described here appears to be primarily associated with the expansion of transposable elements in the mammalian genome.

The significance of mammalian repeat-derived miRNAs

It was previously documented in plants that some miRNAs derive from genome repeats [3]. Our data presented here increases the number of mammalian repeat-derived miRNAs to 13 (including miR-127 and miR-136, which are encoded opposite a

Figure 2. The position of the human LINE-2-derived mature miRNAs near the 3′end of the L2A and L2B consensus sequences. Consensus sequences were obtained from RepBase (http://www.girinst.org/). The five miRNAs are derived from only two locations. Thus, the breakpoints and junctions between apposed L2 genomic repeats that form the miRNA precursors (see Figure 1) occur at specific locations along the L2 consensus sequence.

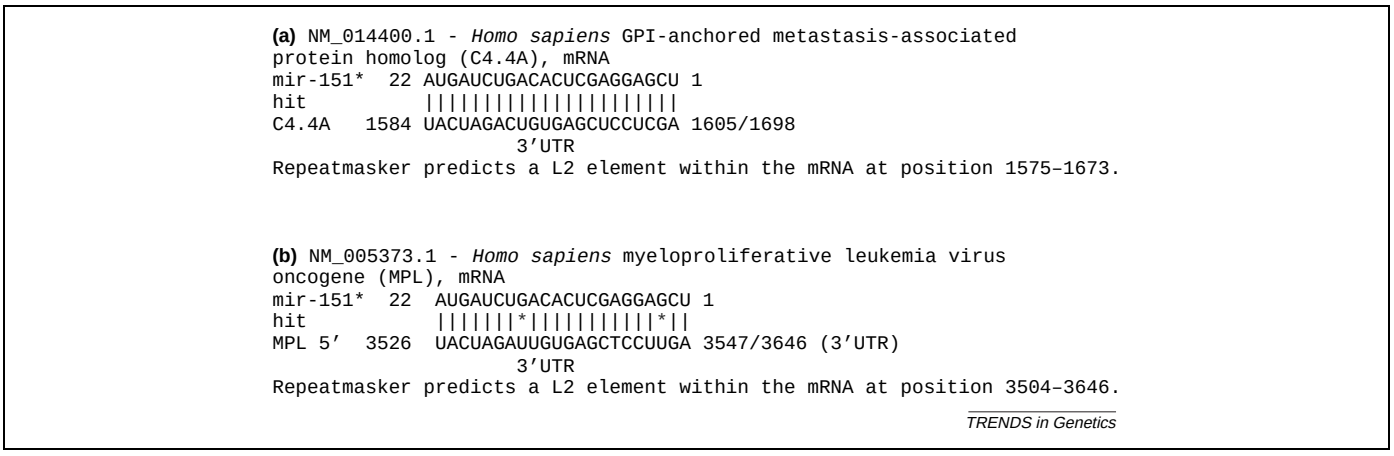

Figure 3. miR-151* matches two human mRNAs perfectly: **(a)** GPI-anchored metastasis-associated protein homolog C4.4A (GenBank accession no. NM_014400); and **(b)** myeloproliferative leukemia virus oncogene MPL (GenBank accession no. NM_005373).

retrotransposon [9,10], and miR-297, which has low-complexity sequence [11]). These repeat-derived miRNAs arise from conventional precursor hairpins and are distinct from rasiRNAs processed from long dsRNA precursors [4–6]. The insertion of transposable elements into new genomic sites thus appears to be one driving-force that creates new miRNAs during mammalian evolution. Although it is not clear whether these elements provide transcriptional start sites for the precursor genes, most if not all of the human repeat-derived miRNAs reside within introns of protein-coding genes, where they could potentially be co-transcribed with the host genes.

Transposable elements can be particularly relevant to generating mammalian miRNAs and miRNA targets that are not broadly conserved across species, presumably because of the timing of the transposition events that generate them. Mir-95 appears to have occurred *de novo* in the human lineage, because no miR-95 precursor can be identified in mouse or rat. Interestingly, although miR-151* is itself conserved across species, its putative perfect target mRNAs (Figure 3) are both human-specific, reflecting the fact that LINE-2 elements became inserted into the 3′-UTR of the human mRNAs but not in their mouse and rat homologs. Rat miR-333 is rodent-specific, not surprisingly because its precursor arises from insertion of a B2 SINE repeat found only in the rodent lineage.

Furthermore, the LINE-2-derived miRNAs miR-95 and miR-151* are perfectly complementary (with zero-to-five G:U matches) to a large family of mRNAs and ESTs that contain MIR LINE-2 elements in their 3′-UTRs. These elements are incorporated into a variety of transcribed genes where they can regulate gene expression [7,12], and can also arise by read-through transcription from LINE-2 elements into neighboring genes. Are LINE-2-bearing transcripts functional targets of LINE-2-derived miRNAs under appropriate conditions? If so it might serve as a mechanism for detecting and neutralizing aberrant transcripts (having read-through transcription from retained introns or neighboring genomic regions) in addition to regulating specific mRNAs. Any computational analysis requires experimental confirmation, but we emphasize that such perfect 'hits' simply cannot be accounted for by chance (even those that have up to five G:U matches).

A variety of non-repeat derived miRNAs also exhibit perfect hits on putative target mRNAs (supplementary file 2; data not shown). One of these examples is HOXB8, which was identified by miR-196 with one G:U match. Since this manuscript was originally submitted for publication, HOXB8 has been confirmed experimentally as a miR-196 target that undergoes mRNA degradation *in vivo* [13]. It is likely that miRNA-target binding involving up to seven G:U matches would also result in mRNA degradation [14]. Thus, perfect complementarity might identify a distinct class of miRNA targets.

Conversely, in a previous bioinformatic analysis, we had predicted that several of the miRNAs (now known to be repeat-derived) hit putative mRNA targets with imperfect complementarity, in a manner that was unlikely to be due to chance [15]. For example, miR-28 and miR-95 are imperfectly complementary to a region in the 3′-UTR of human transcription factor E2F6 mRNA (17 bases in a row including G:U) that contains a LINE-2 fragment. This suggests that repeat-derived miRNAs are also likely to identify biologically relevant targets with imperfect complementarity.

Concluding remarks

Finally, the LINE-2-derived miRNAs also exhibited perfect complementarity (with one-to-five G:U matches) to several non-transcribed chromosomal sequences. Do chromosomal LINE-2 repeats represent direct miRNA targets? This is worth exploring because rasiRNAs can interact directly with chromosomal repeat targets in yeast, *Drosophila* and mammals to regulate heterochromatin [4–6,16,17], and because miRNAs, dicer and the RNA-induced silencing complex (RISC) can enter the nucleus in mammalian cells [18].

Acknowledgements

Our research is supported by NIH grants DA15450, LM07292 and the Human Brain Project.

Supplementary data

Supplementary data associated with this article can be found at doi:10.1016/j.tig.2005.04.008

References

1 Lai, E.C. (2003) MicroRNAs: runts of the genome assert themselves. *Curr. Biol.* 13, R925–R936
2 Carrington, J.C. and Ambros, V. (2003) Role of microRNAs in plant and animal development. *Science* 301, 336–338
3 Llave, C. *et al*. (2002) Endogenous and silencing-associated small RNAs in plants. *Plant Cell* 14, 1605–1619
4 Reinhart, B.J. and Bartel, D.P. (2002) Small RNAs correspond to centromere heterochromatic repeats. *Science* 297, 1831
5 Lippman, Z. *et al*. (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430, 471–476
6 Fukagawa, T. *et al*. (2004) Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nat. Cell Biol.* 6, 784–791
7 Tulko, J.S. *et al*. (1997) MIRs are present in coding regions of human genes. *DNA Seq.* 8, 31–38
8 Pusch, O. *et al*. (2003) Nucleotide sequence homology requirements of HIV-1-specific short hairpin RNA. *Nucleic Acids Res.* 31, 6444–6449
9 Seitz, H. *et al*. (2003) Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat. Genet.* 34, 261–262
10 Smalheiser, N.R. (2003) EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues. *Genome Biol.* 4, 403
11 Houbaviy, H.B. *et al*. (2003) Embryonic stem cell-specific MicroRNAs. *Dev. Cell* 5, 351–358
12 Donnelly, S.R. *et al*. (1999) A conserved nuclear element with a role in mammalian gene regulation. *Hum. Mol. Genet.* 8, 1723–1728
13 Yekta, S. *et al*. (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 304, 594–596
14 Miyagishi, M. *et al*. (2004) Optimization of an siRNA-expression system with an improved hairpin and its significant suppressive effects in mammalian cells. *J. Gene Med.* 6, 715–723
15 Smalheiser, N.R. and Torvik, V.I. (2004) A population-based statistical approach identifies parameters characteristic of human microRNA-mRNA interactions. *BMC Bioinformatics* 5, 139
16 Schramke, V. and Allshire, R. (2004) Those interfering little RNAs! Silencing and eliminating chromatin. *Curr. Opin. Genet. Dev.* 14, 174–180
17 Lehnertz, B. *et al*. (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.* 13, 1192–1200
18 Meister, G. *et al*. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* 15, 185–197

0168-9525/$ - see front matter
doi:10.1016/j.tig.2005.04.008

Genome-wide analysis of coordinate expression and evolution of human *cis*-encoded sense-antisense transcripts

Jianjun Chen[1], Miao Sun[1], Laurence D. Hurst[2], Gordon G. Carmichael[3] and Janet D. Rowley[1]

[1]Department of Medicine, University of Chicago, 5841 S. Maryland Avenue, MC2115, Chicago, IL 60637, USA
[2]Department of Biology and Biochemistry, University of Bath, UK BA2 7AY
[3]Department of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, CT 06030-3301, USA

Is sense-antisense (SA) pairing of transcripts a common mode of gene regulation in the human genome? Although >20% of human genes might form SA pairs, the extent to which they are involved in antisense regulation is unknown. Simultaneous expression of paired sense and antisense genes is an essential step and an important indicator of antisense regulation. In this article, we demonstrate that human SA pairs tend to be co-expressed and/or inversely expressed more frequently than expected by chance. Moreover, co-expressed and inversely expressed SA pairs exhibit a striking pattern of evolutionary conservation. These findings suggest that antisense regulation is a common and important mechanism of gene regulation in the human genome.

Corresponding author: Rowley, J.D. (jrowley@medicine.bsd.uchicago.edu).
Available online 20 April 2005

Introduction

Natural *cis*-encoded antisense RNAs are endogenous transcripts that are transcribed from the opposite strand of the same genomic locus as the sense RNA and have a region of perfect overlap with the sense transcripts [1–4]. In the human genome, although an increasing number of natural antisense transcripts have been predicted or identified [5–8], relatively few of them were shown to have regulatory roles [1–3]. Consequently, it is still not clear whether antisense regulation (i.e. antisense-mediated gene regulation) is a common or an exceptional event in the human genome.

Because of their complementarities, the co-expression (i.e. co-occurrence) of sense and antisense transcripts within the same cell makes it possible to form long, double-strand RNAs (dsRNAs) that can in turn lead to antisense regulation [2,9]. Thus, the simultaneous presence of both sense and antisense transcripts in the same cell or tissue is an essential step and an important indicator of