

# Identification of Mammalian microRNA Host Genes and Transcription Units

Antony Rodriguez, Sam Griffiths-Jones, Jennifer L. Ashurst, and Allan Bradley<sup>1</sup>

Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom

To derive a global perspective on the transcription of microRNAs (miRNAs) in mammals, we annotated the genomic position and context of this class of noncoding RNAs (ncRNAs) in the human and mouse genomes. Of the 232 known mammalian miRNAs, we found that 161 overlap with 123 defined transcription units (TUs). We identified miRNAs within introns of 90 protein-coding genes with a broad spectrum of molecular functions, and in both introns and exons of 66 mRNA-like noncoding RNAs (lncRNAs). In addition, novel families of miRNAs based on host gene identity were identified. The transcription patterns of all miRNA host genes were curated from a variety of sources illustrating spatial, temporal, and physiological regulation of miRNA expression. These findings strongly suggest that miRNAs are transcribed in parallel with their host transcripts, and that the two different transcription classes of miRNAs ('exonic' and 'intronic') identified here may require slightly different mechanisms of biogenesis.

[Supplemental material is available online at [www.genome.org](http://www.genome.org).]

MicroRNAs (miRNAs) are an evolutionarily conserved large class of noncoding RNAs (ncRNAs) 18–22 nucleotides long that mediate posttranscriptional silencing of genes. miRNAs were first discovered in *Caenorhabditis elegans* with the identification of the *lin-4* and *let-7* miRNA genes, which act as posttranscriptional repressors of target genes by antisense binding to their 3' untranslated regions (UTRs; for review, see Ambros 2000). Shortly thereafter, hundreds of other miRNAs were found in worms as well as in flies, plants, and vertebrates (for review, see Carrington and Ambros 2003; Bartel 2004). Rapid progress has begun to unravel the genetic roles of miRNAs in development and other biological processes. For example, in *C. elegans*, *let-7* and *lin-4* miRNAs function as heterochronic genes, and mutations in either disrupt proper specification of cell fates (Ambros 2000). In *Drosophila*, a mutation in *miR-14* leads to a disruption in normal patterns of cell death and also defects in fat metabolism (Xu et al. 2003). In mammals, ~230 miRNAs have been identified from a vast array of tissues and cell types (Lagos-Quintana et al. 2001, 2002, 2003; Mourelatos et al. 2002; Dostie et al. 2003; Houbaviy et al. 2003; Lim et al. 2003; Michael et al. 2003; Kim et al. 2004).

Current models for miRNA biogenesis and maturation suggest that compartmentalized stepwise processing of miRNAs takes place first in the nucleus and then in the cytoplasm. The prevailing view is that primary transcripts of miRNAs (pri-miRNAs) are processed in the nucleus by the RNase III enzyme, Drosha, to stem-loop intermediates known as pre-miRNAs (Lee et al. 2003). These pre-miRNAs are then transported to the cytoplasm for cleavage by Dicer and maturation to their active forms (Lee et al. 2003). Although RNAs of 22-, 75-nt as well as longer pri-miRNA transcripts have been detected by Northern blot analyses for a handful of miRNAs, the transcription units (TUs) or gene hosts that give rise to the vast majority of miRNAs have not been examined in great detail.

In the present study, we set out to identify the modes of transcription for mammalian miRNAs by annotating their positions in the human and mouse genomes. We found that more

than half of all known mammalian miRNAs are within introns of either protein-coding or noncoding TUs, whereas ~10% are encoded by exons of long nonprotein-coding transcripts, also known as mRNA-like noncoding RNAs (lncRNAs; Erdmann et al. 2000). Our annotation illuminates the modes of transcription for miRNAs and has allowed the identification of 'intronic' as well as 'exonic' transcription classes of miRNAs. We curated the expression patterns of well defined microRNA host genes to illustrate likely expression patterns for microRNAs in wide-ranging biological contexts.

## RESULTS AND DISCUSSION

The miRNA Registry contains 232 mammalian miRNAs (<http://www.sanger.ac.uk/Software/Rfam/mirna/>; Griffiths-Jones 2004). Of these we identified 117 miRNAs located in introns of protein-coding genes or long ncRNA transcripts (Table 1, Supplemental Tables A,B). Approximately 40% (90) of all miRNAs are found within introns of protein-coding genes, whereas ~10% (27) are located within introns of long ncRNA transcripts (Table 1). Interestingly, 30 miRNAs overlap with exons of ncRNAs (Supplemental Table C). In some cases (14), miRNAs are located in either an exon or an intron ('mixed') depending on alternative splicing of the host transcript (Table 1, Supplemental Table D). Where clusters of miRNAs overlap with a single host transcript, the vast majority of miRNAs are located in the same intron or exon (Supplemental Tables A–D). Additionally, 32 miRNAs overlap with two or more TUs transcribed on opposite DNA strands (Supplemental Tables A–D). This observation indicates that miRNAs are commonly associated with complex transcriptional loci. The remaining 70 miRNAs are of uncertain transcriptional origin and were not analyzed further. To sum up, a total of 161 miRNAs are linked to the transcription of mRNAs or ncRNAs. Based on these observations, we propose that miRNAs be classified as exonic (exon-derived) miRNAs or intronic (intron-derived) miRNAs.





A partial listing of protein-coding miRNA host genes is shown in Table 2. These miRNA host genes encode proteins with a broad spectrum of biological roles ranging from embryonic development, to the cell cycle, and physiology (Supplemental

<sup>1</sup>Corresponding author.

E-MAIL [abrady@sanger.ac.uk](mailto:abrady@sanger.ac.uk); FAX 44-1223-494714.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.2722704>. Article published online before print in September 2004.

**Table 1.** Distribution of Mammalian miRNAs Found in Introns and Exons of Host Transcripts

	Intron			Exon	'mixed'
	Protein Coding	ncRNA	All	ncRNA	
Human 	82	17	99	24	14
Mouse 	68	16	84	14	11
Both 	66	7	73	8	11
Sum 	90	27	117	30	14 (9*)
					<b>Total</b> 161

The numbers given represent total numbers of miRNAs overlapping with transcripts in the human or mouse genome (see Methods). For comparison purposes, miRNAs in introns of transcripts are separated according to the nature of the host gene (protein coding vs. ncRNA). Nine of the 'mixed' miRNAs are transcribed in introns or exons of alternatively spliced mncRNA hosts the rest are associated with protein-coding host genes as well as mncRNAs of alternatively spliced transcripts. Asterisk (\*) denotes the number of mixed miRNAs derived from mncRNA transcripts only.

Tables A,D). To derive a perspective on the classes of protein-coding genes possibly utilized by miRNAs for their transcription, we surveyed the Gene Ontology (GO) classifications of all miRNA host genes (Gene Ontology Consortium 2001; <http://www.geneontology.org/>). The two most commonly identified 'biological process' classifications for GO annotated miRNA gene hosts are 'metabolism' (GO:0008152; 19 of 90; 21%) followed by 'cellular physiological process' proteins (GO:00050875; 14 of 90; 16%) and the most common GO 'molecular function' are 'purine nucleotide binding' (GO:0017076; 8 of 90; 9%) and 'DNA binding' (GO:0003677; 7 of 90; 8%). Comparison of the most commonly identified GO classifications for miRNA host genes and the entire collection of GO annotated mammalian genes does not reveal disproportionate representation of these classes of genes (data not shown). We also note that several genes involved in human disease are hosts to miRNAs. For instance, nonsense, splicing, frameshift, or deletion mutations in the *Chloride Channel Protein 5 (CLCN5)* gene cause Dent disease and nephrolithiasis, an X-linked recessive disorder in human patients (OMIM: 300008; Yamamoto et al. 2000). A causative role for the *CLCN5*-encoded *miR-188* has not been explored in this disease.

A significant number of intron-based miRNAs reside within related families of protein-coding genes (Table 3, Supplemental Tables A,D,E). This allowed us to group specific sets of miRNAs into novel families based on host gene identity. In some cases, host gene families aid in the identification of related sets of intronic miRNAs (Table 3). For instance, *miR-148b* and *miR-152* are sequence-related and located in two genes encoding subunits of the coatmer transporter (see Table 3, Supplemental Table E). In other examples, we note sequence-unrelated miRNAs that are found within related families of proteins. For example, *miR-105* and *miR-224* are located in two genes encoding subunits of the ligand-gated ion channel  $\gamma$ -aminobutyric-acid receptor- $\alpha$ -3 and - $\epsilon$  respectively, although the mature miRNAs are unrelated by sequence (Table 3).

In addition to the miRNAs in protein coding genes, we also noted 66 miRNAs within TUs that lack a significant protein-

coding potential (Table 1). More detailed analysis of these miRNA host TUs revealed a number of previously classified long ncRNAs (see Tables 4,5 and Supplemental Tables B-D). These include: BIC (Tam 2001), Deleted in Leukemia 2 (DLEU2; Migliazza et al. 2001), Noncoding RNA in Rhabdomyosarcoma (NCRM5; Chan et al. 2002), Synapse-Specific-7H4 (7H4; Velleca et al. 1994), FLJ34037 (Ota et al. 2004), and numerous FANTOM2 long ncRNAs (Okazaki et al. 2002; Numata et al. 2003). Only the BIC RNA has previously been noted as a carrier for a miRNA (Lagos-Quintana et al. 2002). These types of ncRNA transcripts are sometimes referred to as mRNA-like ncRNAs (mlncRNAs; Erdmann et al. 2000) because they share properties with mRNAs such as splicing (e.g., BIC: Tam 2001), polyadenylation (e.g., Air: Sleutels et al. 2002), long transcription units (e.g., Xist: Hong et al. 1999), and possibly also spatio/temporally restricted expression (e.g., Ntab: French et al. 2001). The molecular/genetic function of most non-coding mRNA-like RNAs is unclear but some, such as Xist and Air, have been studied in greater detail.

As well as the previously defined ncRNA TUs noted above, we identified dozens of novel TU hosts for miRs with little protein coding potential (see Methods, Tables 4,5, and Supplemental Tables B-D). For example, EST and cDNA data in ENSEMBL allowed the identification of an mlncRNA TU host containing *miR-100*, *let-7a-2*, and also *miR-125b-1* in an intron (Table 4, Fig. 1A). Sequence alignment of the human/mouse host transcript 1 (ENSESTT00000036713/ENSMUSESTT00000030830; data not shown); and transcript 2 (ENSESTT00000036715/2610203C20Rik) does not reveal significant homology (Fig. 1B). Nevertheless, the *miR-100*, *let-7a-2*, and *miR-125b-1* transcription locus is remarkably similar in human and mouse in both arrangement and structure (Fig. 1A). The possibility exists that some miRNA hosts of this type may be so-called 'inside out' genes analogous to some small nucleolar RNA (snoRNA) hosts (Tycowski et al. 1996), in which transcription of the host gene serves purely to yield the intron-encoded RNAs. It is also plausible that some mlncRNA hosts containing miRNAs in their introns serve as functional RNAs.

A significant number of miRNAs overlap exons of novel noncoding transcripts (Tables 1,5; Supplemental Tables C,D). For example, the *miR-22*-predicted hairpin precursor is contained entirely within exon 2 of a noncoding TU, and the splicing pattern is generally conserved in human and mouse, despite the lack of protein-coding potential (Fig. 2A,B). Significant sequence conservation between the mouse and human *miR-22* TU is observed within the 5' end and within the *miR-22*-predicted hairpin precursor (Fig. 2B). Such high overall conservation in the mammalian *miR-22*/TU may indicate an important role in biogenesis or localization. Although high sequence conservation within the predicted hairpin precursors of mammalian miRNAs is a common feature (Lim et al. 2003), it is unclear whether the extended sequence similarity observed in *miR-22*/TU is a more common phenomenon for the exonic class. An alignment of the mouse and human *miR-155/BIC* does not show such extended conservation beyond the predicted hairpin precursor (data not shown). In addition to *miR-22*, a further 29 other miRNAs are contained within exons of transcripts (Table 5). In the majority of cases, exonic miRNAs were identified within mlncRNA transcripts, but in two rare cases we noted that the miRNA overlapped with the 3' UTR of a protein-coding transcript (Supplemental Table D).

The prevalence of miRNAs within introns or exons of TUs prompted us to ask whether it might be possible to derive reliable miRNA expression patterns from host gene expression data. By analogy to mammalian snoRNAs, intronic miRNAs may be co-transcribed with the host gene by inclusion in introns of their pre-mRNAs and excision from debranched introns by an exonuclease (Tycowski et al. 1993). Similarly to yeast snoRNAs,

**Table 2.** miRNAs in Introns of Protein-Coding Genes

microRNA <sup>a</sup>	Mouse host gene ID <sup>b</sup>	Human host gene ID <sup>b</sup>	Host gene protein description <sup>c</sup>	miRNA position <sup>d</sup>
miR-7-3 (miR-7b)		ENSG00000176840	PITUITARY GLAND SPECIFIC FACTOR 1A. [Source: RefSeq; Acc: NM_174947]	Intron 2
miR-10b	ENSMUSG00000038692	ENSG00000170166	HOMEBOX PROTEIN HOX-D4(HOX-4B) (HOX-5.1) (HHO.C13). {Source: SWISS-PROT (P09016)}	Intron 4
miR-15b, -16-2	ENSMUSG00000034349	ENSG00000113810	STRUCTURAL MAINTENANCE OF CHROMOSOMES 4-LIKE 1 PROTEIN (CHROMOSOME-ASSOCIATED POLYPEPTIDE C) (HCAP-C) (XCAP-C HOMOLOG). SMC4L1 [Source: SWISS-PROT; Acc: Q9NTJ3]	Intron 4
miR-25, -93, -106b	ENSMUSG00000029730	OTTHUMG00000023308/ ENSG00000166508	DNA REPLICATION LICENSING FACTOR MCM7 (CDC47 HOMOLOG) (P1.1-MCM3). [Source: SWISS-PROT; Acc: P33993]	Intron 13
miR-30c-1 and -30e	ENSMUSG00000032897	ENSG000000066136	NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT GAMMA (NF-Y PROTEIN CHAIN C) (NUCLEAR FACTOR YC) (NF-YC) (CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT C) (CBF-C) (TRANSACTIVATOR HSM-1/2). NFYC [Source: SWISS-PROT; Acc: Q13952]	Intron 5
miR-33	ENSMUSG00000022463	OTTHUMG00000030492/ ENSG00000100152	STEROL REGULATORY ELEMENT BINDING PROTEIN-2 (SREBP-2) (STEROL REGULATORY ELEMENT-BINDING TRANSCRIPTION FACTOR 2). SREBF2 [Source: SWISS-PROT; Acc: Q12772]	Intron 15 (16)
miR-101-2 (miR-101b)	ENSMUSG00000024785	OTTHUMG00000019474/ ENSG00000120158	RNA 3'-TERMINAL PHOSPHATE CYCLASE-LIKE PROTEIN (HSPC338), RCL1_HUMAN [Source: SWISS-PROT; Acc: Q9Y2P8]	Intron 8
miR-126, 126*	ENSMUSG00000026921	OTTHUMG00000020938/ ENSG00000172889	EGF-LIKE DOMAIN 7; NEU1 PROTEIN, NOTCH4-LIKE PROTEIN (VASCULAR ENDOTHELIAL ZINC FINGER 1) [Source: RefSeq; Acc: NM_178444]	Intron 6 (7)
miR-128b	ENSMUSG00000032503	ENSG00000076062	CAMP-REGULATED PHOSPHOPROTEIN 21 (ARPP-21) (SWISS-PROT: AP21_HUMAN)	Intron 11 (17)
miR-139	ENSMUSG00000030653	ENSG00000186642	CGMP-DEPENDENT 3',5'-CYCLIC PHOSPHODIESTERASE (EC 3.1.4.17) (CYCLIC GMP STIMULATED PHOSPHODIESTERASE) (CGS-PDE) (CGSPDE), PDE2A [Source: SWISS-PROT; Acc: O00408]	Intron 2 (1)
miR-140	ENSMUSG00000031930	ENSG00000088481	NEDD-4-LIKE UBIQUITIN-PROTEIN LIGASE WWP2 (EC 6.3.2.-) (WW DOMAIN-CONTAINING PROTEIN 2) (ATROPIN-1 INTERACTING PROTEIN 2) (AIP2). WWP_2HUMAN [Source: SWISS-PROT; Acc: O00308]	Intron 15 (16)
miR-149	ENSMUSG00000034220	ENSG00000063660	GLYPICAN-1 PRECURSOR, GPC1 [Source: SWISS-PROT; Acc: P35052]	Intron 1
miR-151	ENSMUSG00000022607	ENSG00000169398	FOCAL ADHESION KINASE 1 (EC 2.7.1.112) (FADK 1) (PP125FAK) (PROTEIN-TYROSINE KINASE 2). FAK1 [Source: SWISS-PROT (Q05397)]	Intron 19 (25)
miR-188	ENSMUSG00000004317	ENSG00000171365	CHLORIDE CHANNEL PROTEIN 5 (CLC-5), CLCN5 [Source: SWISS-PROT; Acc: P51795]	Intron 3 (2)
miR-190	ENSMUSG00000035702	ENSG00000171914	TALIN 2. TLN 2 [Source: SWISS-PROT; Acc: Q9Y4G6]	Intron 51 (27)
miR-207 (human miR-207 not found)	ENSMUSG00000028410		DNAJ HOMOLOG SUBFAMILY A MEMBER 1 (HEAT SHOCK 40 KDA PROTEIN 4) (DNAJ) PROTEIN HOMOLOG 2) (HSJ-2), DNAJ_MOUSE [Source: SWISS-PROT; Acc: P54102]	Intron 1
miR-208	ENSMUSG00000040752	OTTHUMG00000028753/ ENSG00000166094	MYOSIN HEAVY CHAIN, CARDIAC MUSCLE ALPHA ISOFORM (MYHC-ALPHA), MYH6 [Source: SWISS-PROT; Acc: P13533]	Intron 28 (29)
miR-335	ENSMUSG00000051855	ENSG00000106484	MESODERM SPECIFIC TRANSCRIPT ISOFORM A; PATERNALLY EXPRESSED GENE 1. MEST [Source: RefSeq (NM_002402)]	Intron 2 (1)
miR-338	ENSMUSG00000025375	ENSG00000181409	APOPTOSIS-ASSOCIATED TYROSINE KINASE. AATK [Source: RefSeq; Acc: NM_007377]	Intron 5 (7)
miR-339	ENSMUSG00000029533		ARSENITE INDUCIBLE RNA ASSOCIATED PROTEIN. (AIRAP) AA407930 [Source: RefSeq; Acc: NM_133349]	Intron 1
miR-340	ENSMUSG00000020376	ENSG00000113269	RING FINGER PROTEIN 130; GOLIATH PROTEIN; G1-RELATED ZINC FINGER PROTEIN, RNF130 [Source: RefSeq (NM_018434)]	Intron 2 (2)
miR-342	ENSMUSG00000021262	OTTHUMG00000029003/ ENSG00000089465	ENA/VASODILATOR STIMULATED PHOSPHOPROTEIN-LIKE PROTEIN (ENA/VASP-LIKE PROTEIN). EVL [Source: SWISS-PROT (Q9UI08)]	Intron 3
miR-346	14803 NCBI	OTTHUMG00000018650/ ENSG00000182771	SIMILAR TO GLUTAMATE RECEPTOR, IONOTROPIC, DELTA 1. GRID1 [Source: SPTREMBL (Q8IXT3)]	Intron 2 (1)

<sup>a</sup>miRNA names listed in parentheses refer to the mouse name if different from the human.<sup>b</sup>Gene IDs are from ENSEMBL, Vega, or NCBI.<sup>c</sup>Protein descriptions are for the human gene.<sup>d</sup>Intron positions for miRNAs in gene are listed. If the human/mouse positions are different, the mouse location is presented in parentheses.

**Table 3.** Families of miRNAs Found in Introns of Related Protein-Coding Genes

Family	Mouse host gene ID <sup>a</sup>	Human host gene ID <sup>a</sup>	Host gene description <sup>b</sup>	microRNA	Related miRNAs <sup>c</sup>
Coatomer	ENSMUSG00000035994 MBuild30	ENSG00000111481	COATOMER ZETA-1 SUBUNIT (ZETA-1 COAT PROTEIN) (ZETA-1 COP) (CGI-120) (HSPC181), COPZ1 [Source: SWISS-PROT (Q9Y3C3)]	miR-148b	+
	ENSMUSG00000018672	ENSG00000005243	COATOMER ZETA-2 SUBUNIT (ZETA-2 COAT PROTEIN) (ZETA-2 COP). COPZ2 [Source: SWISS-PROT; Acc: Q9P299]	miR-152	+
GABA	ENSMUSG00000031343	ENSG00000011677	GAMMA-AMINOBUTYRIC-ACID RECEPTOR ALPHA-3 SUBUNIT PRECURSOR (GABA(A) RECEPTOR). GABRA3 [Source: SWISS-PROT; Acc: P34903]	miR-105-1, -105-2	—
	ENSMUSG00000031340	ENSG00000102287	GAMMA-AMINOBUTYRIC-ACID RECEPTOR EPSILON SUBUNIT PRECURSOR (GABA(A) RECEPTOR). GABRE [Source: SWISS-PROT; Acc: P78334]	miR-224	—
NIF	ENSMUSG00000038995	ENSG00000144677	NUCLEAR LIM INTERACTOR-INTERACTING FACTOR 1 (NLI-INTERACTING FACTOR 1) (NIF-LIKE PROTEIN) (YA22 PROTEIN) (HYA22), NIF1_HUMAN (C3orf8) [Source: SWISS-PROT; Acc: O15194]	miR-26a-1	+
	ENSMUSG00000040540	ENSG00000175215	NUCLEAR LIM INTERACTOR-INTERACTING FACTOR 2 (NLI-INTERACTING FACTOR 2) (PROTEIN OS-4). NIF2_HUMAN [Source: SWISS-PROT; Acc: O14595]	miR-26a-2	+
	ENSMUSG00000026176	ENSG00000144579	NUCLEAR LIM INTERACTOR-INTERACTING FACTOR 3 (NLI-INTERACTING FACTOR 3) (NLI-IF). NIF3_HUMAN [Source: SWISS-PROT; Acc: Q9GZU7]	miR-26b	+
LIM	ENSMUSG00000033306	ENSG00000145012	LIM DOMAIN CONTAINING PREFERRED TRANSLOCATION PARTNER IN LIPOMA; LIM DOMAIN-CONTAINING PREFERRED TRANSLOCATION PARTNER IN LIPOMA; LIPOMA-PREFERRED-PARTNER GENE. LPP [Source: RefSeq; Acc: NM_005578]	miR-28	—
PANK	ENSMUSG00000033610	ENSG00000163995 OTTHUMG00000018718/ ENSG00000152782	ACTIN BINDING LIM PROTEIN 2. ABLIM2 [Source: RefSeq; Acc: NM_032432]	miR-95 miR-107	— +
	ENSMUSG00000034220	OTTHUMG00000031768/ ENSG00000125779	PANTOTHENATE KINASE 1 (EC 2.7.1.33) (PANTOTHENIC ACID KINASE 1) (HPANK1) (HPANK). [Source: SWISS-PROT; Acc: Q8TE04]	miR-103-2	+
	ENSMUSG00000018846	ENSG00000120137	PANTOTHENATE KINASE 2, MITOCHONDRIAL PRECURSOR (EC 2.7.1.33) (PANTOTHENIC ACID KINASE 2) (HPANK2). [Source: SWISS-PROT; Acc: Q9BZ23]	miR-103-1	+
SLIT	20563 NCBI	ENSG00000145147	PANTOTHENATE KINASE 3 (EC 2.7.1.33) (PANTOTHENIC ACID KINASE 3) (HPANK3). [SOURCE: SWISS-PROT; ACC: Q9H999]	miR-218-1	+
PTPRN	20564 NCBI	ENSG00000184347	SLIT HOMOLOG 2 PROTEIN PRECURSOR (H-SLIT-2). SLIT2 [Source: SWISS-PROT; Acc: Q94813]	miR-218-2	+
		ENSG00000054356	SLIT HOMOLOG 3; SLIT3 [Source: RefSeq; Acc: NM_003062]	miR-153-1	+
	ENSMUSG00000054701	ENSG00000155093	PROTEIN-TYROSINE PHOSPHATASE-LIKE N PRECURSOR (R-PTP-N) (PTP IA-2) (ISLET CELL ANTIGEN 512) (ICA 512) (ISLET CELL AUTOANTIGEN 3), PTPRN [Source: SWISS-PROT; Acc: Q16849]	miR-153-2	+
TRPM	ENSMUSG00000030523	ENSG00000134160	RECEPTOR-TYPE PROTEIN-TYROSINE PHOSPHATASE N2 PRECURSOR (EC 3.1.3.48) (R-PTP-N2) (ISLET CELL AUTOANTIGEN RELATED PROTEIN) (ICAAR) (IAR) (PHOGRIN). PTPRN2 [Source: SWISS-PROT; Acc: Q92932]	miR-211	+
	ENSMUSG00000024763	ENSG00000083067	TRANSIENT RECEPTOR POTENTIAL CATION CHANNEL, SUBFAMILY M, MEMBER 1; MELASTATIN 1. TRPM1 [Source: RefSeq; Acc: NM_002420]	miR-204	+
			LONG TRANSIENT RECEPTOR POTENTIAL CHANNEL 3 (LTRPC3) (FRAGMENT), TRPM3 [Source: SWISS-PROT; Acc: Q9HCF6]		

<sup>a</sup>Gene IDs are from ENSEMBL, NCBI, or AceView.<sup>b</sup>Protein descriptions are for the human gene.<sup>c</sup>miRNAs with related nucleotide sequences are marked with plus signs and those that are not related with a minus sign.

**Table 4.** miRNAs in Introns of Noncoding Transcription Units

microRNA	Chrom <sup>a</sup>	Human host TU ID <sup>b</sup>	Mouse host TU ID <sup>b</sup>	miRNA location <sup>c</sup>	Host transcript description
let-7a-2, miR-100, -125b-1	11 (9)	ENSESTG00000015836	ENSMUSESTG00000011922	Intron 2 (4)	Novel mlncRNA
let-7d	9 (13)	OTTHUMG00000020259		Intron 3	Vega mlncRNA
miR-7-2	15 (7)		ENSMUSESTG00000007393	Intron 1	Novel mlncRNA
miR-15a, -16-1	13 (14)	OTTHUMG00000016927	1810047A16Rik	Intron 4	DLEU2 mlncRNA
miR-30a, -30a*, -30c-2	6 (1)	ENSESTG00000011458		Intron 3	Novel mlncRNA
miR-31	9 (4)	OTTHUMG00000019681		Intron 1	Novel mlncRNA
miR-129-2	11 (2)	11_43541975 AceView		Intron 1	Novel mlncRNA
miR-132, -212	17 (11)		ENSMUSESTG00000011295 MBuild 30	Intron 1	Novel mlncRNA
miR-135a-1	3 (9)	3_52291509 AceView			Novel mlncRNA
miR-135a-2	12 (10)	196475 NCBI	ENSMUSESTG00000002435	Intron 9 (1)	NCRMS mlncRNA
miR-135b	1 (1)	ENSESTG000000002491	ENSMUSESTG00000018175	Intron 1	Novel mlncRNA
miR-141, -200c	12 (6)		ENSMUSESTG00000015291 MBuild30	Intron 1	Novel mlncRNA
miR-154	14 (12)		ENSMUSESTG00000016217	Intron 2	Novel mlncRNA
miR-181a, -181b-2	9 (2)	OTTHUMG00000020657		Intron 1	Vega mlncRNA
miR-181b-1, -213	1 (1)		ENSMUSESTG00000012306	Intron 2	Novel mlncRNA
miR-194-1	1 (1)		2010103J01Rik	Intron 1	Riken mlncRNA
miR-302	4 (3)	4_114030028 AceView		Intron 2	Novel mlncRNA
miR-325	— (X)		ENSMUSESTG00000006551	Intron 1	Novel mlncRNA

<sup>a</sup>The human chromosome is listed first followed by the mouse chromosome in parentheses.<sup>b</sup>Transcript IDs are from ENSEMBL, Vega, FANTOM2, NCBI, or AceView.<sup>c</sup>Intron positions for miRNAs in transcript are listed. If the human/mouse positions are different, the mouse location is presented in parentheses.

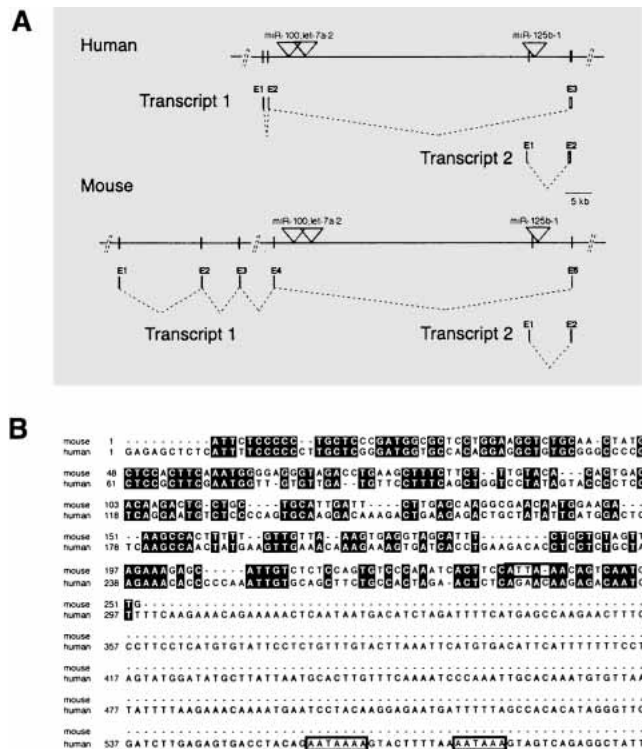
mammalian miRNAs of exonic transcription origin could be derived from ncRNA transcripts by endonucleolytic cleavage within duplex regions (Chanfreau et al. 1998). If parallel transcription of host genes and miRNAs takes place, similar expression patterns for host transcript and miRNAs would be observed. To test this possibility, we analyzed the expression of five predicted miRNA host transcripts by RT-PCR from several organs in adult mice and

compared their expression pattern to that of their corresponding exonic or intronic miRNA (see Methods and Table 6). In four out of five cases, we note an exact correlation between the host gene/transcript expression (Table 6) and the previously reported miRNA expression in mouse adult organs (Sempere et al. 2004). This result suggests that miRNA expression may be reliably derived from host transcript expression information, and strongly

**Table 5.** miRNAs Found in Exons of Nonprotein-Coding Transcription Units

microRNA <sup>a</sup>	Chrom <sup>b</sup>	Human host TU ID <sup>c</sup>	Mouse host TU ID <sup>c</sup>	miRNA location <sup>d</sup>	Host transcript description
let-7b, let-7c-2	22 (15)	OTTHUMG00000030111		Exon 1	Vega mlncRNA
let-7i	12 (10)		D630033A02Rik	Exon 1	Riken mlncRNA
miR-22	17 (11)	ENSESTG00000028158	2210403K04Rik	Exon 2	Riken mlncRNA
miR-29b-2	1 (1)	1_205058566 AceView	C030002C11Rik	Exon 1	Riken mlncRNA
miR-34b, -34c	11 (9)	ENSESTG00000013662		Exon 2	Novel mlncRNA
miR-101-1 (miR-101)	1 (4)		ENSMUSESTG00000022081	Exon 2	Novel mlncRNA
miR-122a	18 (18)	18_54266208 AceView		Exon 2	Novel mlncRNA
miR-133b	6 (1)	6_52060506 AceView		Exon 1	Novel mlncRNA
miR-137	1 (1)	1_97978084 AceView	ENSMUSESTG00000012545	Exon 2 (3)	Novel mlncRNA
miR-143, -145	5 (18)	5_148836826 AceView		Exon 1	Novel mlncRNA
miR-146	5 (11)	ENSESTG00000012459	5830402E17Rik	Exon 2 (1)	Riken mlncRNA
miR-155	21 (16)	114614 NCBI	Bic NCBI	Exon 3 (2)	BIC mlncRNA
miR-192	11 (19)	11_64434198 AceView		Exon 1	Novel mlncRNA
miR-195	17 (11)	LOC284112 AceView		Exon 1	Novel mlncRNA
miR-196b	7 (6)	7_26962152 AceView		Exon 1	Novel mlncRNA
miR-199a-2, -199a*-2	1 (1)	1_169353084 AceView		Exon 1	Novel mlncRNA
miR-202	— (7)		G630041I22Rik	Exon 2	Riken mlncRNA
miR-206	6 (1)		4831414J02Rik	Exon 1	7H4 mlncRNA
miR-214	1 (1)	1_169353084 AceView	D230012O04Rik	Exon 2 (1)	Riken mlncRNA
miR-215	1 (1)		2010103J01Rik	Exon 2	Riken mlncRNA
miR-221	X (X)	X_44651888 AceView		Exon 1	Novel mlncRNA
miR-223	X (X)	X_64102086 AceView	9830169G21Rik	Exon 3 (2)	Riken mlncRNA
miR-296	20 (2)	20_58078195 AceView	D330038P10Rik	Exon 1	Riken mlncRNA
miR-324-5p, -324-3p	17 (11)	17_7329130 AceView		Exon 3	Novel mlncRNA
miR-331	12 (10)		A630071D13Rik	Exon 1	Riken mlncRNA

<sup>a</sup>miRNA names listed in parentheses refer to the mouse name if different from the human.<sup>b</sup>The human chromosome is listed first followed by the mouse chromosome in parentheses.<sup>c</sup>Transcript IDs are from ENSEMBL, Vega, FANTOM2, NCBI, or AceView.<sup>d</sup>Exon positions for miRNAs in transcript are listed. If the human/mouse positions are different, the mouse location is presented in parentheses.



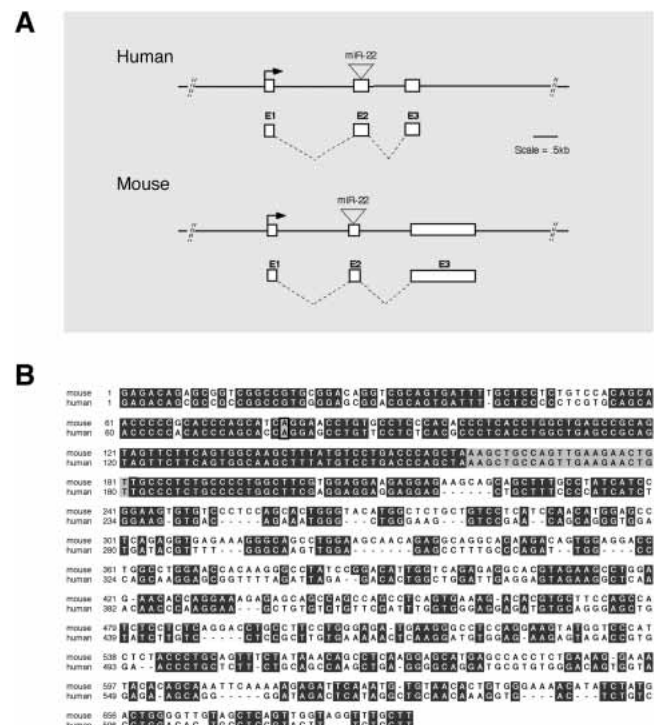
**Figure 1** Genomic organization of the mammalian *miR-100*, *let-7a-2*, *miR-125b-1* locus and sequence alignment of human and mouse *miR-125b-1* host transcripts. (A) Identified transcripts are shown with exons represented as vertical lines and labeled above in numerical order. Dotted lines denote splicing of transcripts. Positions of miRNA precursors are denoted with a triangle. Exons are not to scale. Human/Mouse Transcript 1 corresponds to ENSESTT00000036713 and ENSMUSESTT00000030830, respectively. Transcript 2 corresponds to human ENS-ESTT00000036715 and mouse 2610203C20Rik transcripts. (B) A DNA sequence alignment for Transcript 2 is shown. Identical bases are denoted in black. Consensus polyadenylation sites are boxed in. Note that nucleotides 597–955 of the human Transcript 2 are not shown.

suggests that overlapping transcripts in the genome are likely hosts for miRNAs.

The above result impelled us to derive likely miRNA expression pattern information from all host genes identified here. We curated host gene expression data from a number of publicly available sources including primary literature and from public gene expression pattern databases (see Methods). This analysis allowed the identification of physiologically regulated, signal transduction, and tissue/organ specifically transcribed miRNA hosts (see Supplemental Tables A–D for a full description of expression patterns of miRNA hosts). For example, *miR-33* is in an intron of *SREBP2* (Table 2), an important gene in cholesterol homeostasis that is tightly regulated by sterols at the transcriptional level (Sato et al. 1996). In addition, we identified a cluster of three intronic miRNAs (*miR-25*, *-93*, and *-106b*) that are likely linked to the transcription of *DNA Replication Licensing Factor MCM7* (Table 2), a gene that is quiescent in nondividing cells and is upregulated just prior to S phase of the cell cycle (Fujita et al. 1996). Interestingly, human *miR-7-3* is contained within intron 2 of *Pituitary Gland Specific Factor 1* (Table 2), a gene that is expressed in the pituitary gland (Tanaka et al. 2002). This mature miRNA has two other identical copies in the genome, one of which (*miR-7-1*) is possibly linked to the expression of a ubiquitously expressed gene, *HNRPK* (Supplemental Table E). This illustrates that expression pattern information for identical miRNAs

may be confounded by distinct but overlapping patterns of transcription. In such cases where multiple paralogous miRNAs exist in the genome, it may prove useful to distinguish the expression of individual miR copies by analyzing the expression of their respective host genes (see Supplemental Table E for a list of similar/paralogous microRNAs contained in different host genes).

The current dogma in the miRNA field is that compartmentalized processing of all miRNAs to mature miRs occurs with Drosha in the nucleus and Dicer in the cytoplasm (for review, see Bartel 2004; He and Hannon 2004). In two seminal studies, it was demonstrated that miRNAs are dependent on Drosha for processing in HeLa cells (Lee et al. 2003) and that enrichment of a particular pri-/pre-miRNA occurs in the nucleus while the mature miR is found in the cytoplasm (Lee et al. 2002). However, we note that the miRNAs tested in these studies are intronic class miRNAs (i.e., *miRs-15a*, *-16-1*, *-20*, *-30a*; Supplemental Tables A,B), a ‘mixed’ miRNA (*miR-21*) (Supplemental Table D), and several undefined transcription class miRNAs (data not shown). In lieu of the findings in this study of two types of transcription miRNA classes in mammals, it will be very interesting to revisit these experiments in the context of experimentally verified exonic transcription-class miRNAs. It seems likely that commonalities in processing and maturation by Drosha and Dicer occur for all transcription classes of miRNAs; however, the existence of exonic miRNAs in mammals implies that differences of biogenesis may also take place. In this regard, it is noteworthy that we found



**Figure 2** Genomic organization of the mammalian *miR-22* locus and sequence alignment of human and mouse *miR-22* host transcripts. (A) Exons are shown as white rectangles and labeled above. The human/mouse *miR-22* host transcript corresponds to ENSESTT000000065902 and 2210403K04Rik, respectively. (B) DNA sequence alignments of the *miR-22* human/mouse host transcripts are shown. Identical bases are black. The mature *miR-22* sequence is highlighted in gray. The conserved splice site junction between exons 1 and 2 is also highlighted in gray. Note that nucleotides 688–1658 of the mouse transcript are not shown.

**Table 6.** Comparison of miRNA Expression and Host Transcript Expression

MicroRNA	MicroRNA Type	MicroRNA Expression <sup>a</sup>	Mouse Host TU ID	Host Type	Host Expression <sup>b</sup>
miR-9-3	Intronic	brain	ENSMUSESTG00000007408	Protein Coding	B Li H Lu K S
miR-22	Exonic	brain, liver, heart, lung, kidney, spleen	2210403K04Rik	miRNA	
miR-137	Exonic	brain, liver, spleen	ENSMUSESTG00000012545	miRNA	
miR-153	Intronic	brain	ENSMUSG00000054701	Protein Coding	
miR-219-2	mixed	brain	ENSMUSESTG00000007085 MBuild30	miRNA	

<sup>a</sup>miRNA expression is derived from Sempere et al. (2004).

<sup>b</sup>An ethidium bromide-stained gel picture shows RT-PCR results from mouse adult brain (B), liver (Li), heart (H), lung (Lu), kidney (K), and spleen (S).

*miR-206* within an exon of the synapse-associated *miR-206*, 7H4 (Table 4, Supplemental Table C; Velleca et al. 1994; Numata et al. 2003). 7H4 was originally identified by Velleca et al. (1994) as an RNA selectively found in the motor endplate of the rat skeletal neuromuscular junction. Detection of 7H4 at the neuromuscular junction raises the intriguing possibility that the unprocessed form of *miR-206* may be transported outside of the nucleus as a pri-miR-206/7H4 RNA. Future studies will be required to determine whether some miRNAs can be actively transported to subcellular compartments outside of the nucleus embedded within their host transcripts.

## Conclusion

The findings presented here clearly show that the expression of a large subset of mammalian miRNAs may be transcriptionally linked to the expression of other genes, coding for both proteins and ncRNAs. The identification of two distinct classes of miRNAs, overlapping introns of other transcripts, and encoded in exons, suggest that the maturation of miRNAs may be more complex than previously thought. Mammalian snoRNAs, responsible for modification of ribosomal and spliceosomal RNAs, were shown to be excised from introns of protein-coding (often ribosomal protein) genes (Qu et al. 1995). The suggestion that approximately half of mammalian miRNAs are expressed by a similar mechanism has important and far-reaching implications for our understanding of gene expression, regulation, and the basic definition of what constitutes a gene. Previous studies have demonstrated unexplained levels of conservation in intronic and other apparent noncoding sequence in mammals (Bejerano et al. 2004). There has also been recent discussion about the regulatory possibilities of parallel output arising from exonic and intronic sequences in genes (Mattick 2003). Although we are far from categorical evidence of parallel transcription on a global scale, and the data presented here account for only a small percentage of noncoding conservation, the presence of miRNAs in host genes, together with the reported roles of intronic snoRNAs, provides a tantalizing suggestion of this phenomenon on a larger scale.

## METHODS

### miR Databases

miRNA hairpin sequences were retrieved from the microRNA Registry release 3.1 (<http://www.sanger.ac.uk/Software/Rfam/mirna/>; Griffiths-Jones 2004). Some homologs were identified

using pairwise similarity of the hairpin sequence and by matching of mature miRNAs. The ability of flanking regions to adopt a hairpin structure was determined using the mfold program (<http://www.bioinfo.rpi.edu/applications/mfold/>).

### Genome Analysis and Curation

miRNAs were mapped to the human (NCBI Build 34) and mouse (NCBI Build 32) genome assemblies using WU-BLASTN. To safeguard against concerns about mapping errors in the mouse NCBI Build 32 assembly, we also made use of the mouse NCBI Build 30. ENSEMBL IDs from mouse Build 30 are provided in several cases in which errors in the newer assembly were noted. For ease of manual annotation, the results of these searches were made accessible

via the ENSEMBL genome graphical interface (<http://www.ensembl.org/>). Annotation of mouse and human miRNAs was aided by using a combination of genome annotated TUs, including Vega ([http://vega.sanger.ac.uk/Homo\\_sapiens/](http://vega.sanger.ac.uk/Homo_sapiens/)), ENSEMBL (<http://www.ensembl.org/>), FANTOM2.0 (Okazaki et al. 2002; Numata et al. 2003), NCBI Acembly/AceView (<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/>), NCBI (<http://www.ncbi.nlm.nih.gov/mapview/>), and FLJ transcripts (Ota et al. 2004). Genome coordinate overlaps between miRNAs and annotated TUs were classified according to the exonic or intronic nature of the miRNA. The FatiGo data mining tool (<http://fatigo.bioinformatics.es/>; Al-Shahrour et al. 2004) was used to determine if any GO terms are significantly overrepresented in microRNA host genes.

### mRNA-Like Noncoding RNA Databases

A total of ~10,000 miRNA sequences were collected from the FANTOM Functional Annotation Database (<http://fantom2.gsc.riken.go.jp/>; Numata et al. 2003), the Noncoding RNAs Database (<http://biobases.ibch.poznan.pl/ncRNA/>; Erdmann et al. 2000), and the Full-Length Japan database (<http://www.nedo.go.jp/bio-e/>; Ota et al. 2004), and mapped to the genome using SSAHA. The positions of these ncRNAs were made accessible in the ENSEMBL genome graphical interface via the DAS protocol and analyzed for potential overlap with miRNAs in the human or mouse genome.

### Identification of mRNA-Like Noncoding RNAs

miRNA host transcripts were designated as miRNAs in a process similar to that employed by the FANTOM Genome Exploration Research Group (Okazaki et al. 2002). To identify miRNA candidate miRNA hosts, we eliminated those ENSEMBL or NCBI Acembly AceView Transcripts overlapping with miRNAs with similarity to known protein sequences using BLASTX (E-value < 10<sup>-5</sup>) or those that showed homology to known motifs or domains in the Pfam database (Bateman et al. 2004).

### miRNA Host Gene RT-PCR

Total RNA was prepared from 2–3-wk-old mice using the RNA RT-PCR Miniprep Kit (Stratagene). To safeguard against RT amplification of hnRNA (nuclear precursor RNA), an oligoDT primer was used in the first-strand cDNA synthesis step instead of gene-specific primers. In order to unambiguously distinguish spliced cDNA from genomic DNA contamination, specific primers were designed to amplify across introns of the predicted microRNA

host transcript, and – RT controls reactions were performed. The primers used to detect each of the five microRNA transcription units by PCR were as follows: miR-9-3 host transcript, ENSMUS-ESTG00000007408, 5'-GTGGGCCAGGAGAAAAGAGAAAAG-3' (forward) and 5'-CTCTGTGTTCATCCTGTGGCTTTG-3' (reverse); miR-22 carrier transcript, 2210403K04Rik, 5'-CAGTGA TTTGTCTCTGTGCCAC-3' (forward) and 5'-CAACTGAGC TACAACCCAGTCAT-3' (reverse); miR-137 carrier transcript, ENSMUESTG00000012545, 5'-CAGAGTCCGCATGAAGCAGAG CAA-3' (forward) and 5'-GTCCTTGGCAACCGGGAGCTTTTA-3' (reverse); miR-153 host transcript, ENSMUSG00000054701, 5'-CTGTGCTGACCTATGACCACTC-3' (forward) and 5'-CAATCAG GACGTAGGTTCCACT-3' (reverse); miR-219 host transcript, ENSMUESTG00000007085MBuild30, 5'-AGAGGCGCCCGCTT GGGTTACAG-3' (forward) and 5'-TCATTCCCCAGCTTCTGGG TTT-3' (reverse).

### miRNA Host Gene Expression Pattern Curation

Gene expression pattern information was obtained from the primary literature using NCBI PubMed (<http://www.ncbi.nlm.nih.gov:80/>) as a source for articles. In cases where the host gene/TU had no primary literature citations, expression data were instead obtained from the RIKEN Expression Array Database (<http://read.gsc.riken.go.jp/>; Bono et al. 2003), the Gene Expression Atlas Database (<http://expression.gnf.org/cgi-bin/index.cgi>; Su et al. 2002), the NCBI Acembly/AceView Web site, or from SWISS-PROT (<http://us.expasy.org/sprot/>).

### ACKNOWLEDGMENTS

We thank Tony Cox and James Stalker for technical assistance in generating ENSEMBL DAS sources for genome annotation, Dr. Russell Grocock for useful discussions and technical help, and Drs. Jos Jonkers and Madhuri Warren for critical reading of the manuscript. A.R. is an NIH Postdoctoral Fellow. This work was funded by The Wellcome Trust.

### REFERENCES

- Al-Shahrour, F., Diaz-Uriarte, R., and Dopazo, J. 2004. Fatigo: A web tool for finding significant associations of gene ontology terms with groups of genes. *Bioinformatics* **20**: 578–580.
- Ambros, V. 2000. Control of developmental timing in *Caenorhabditis elegans*. *Curr. Opin. Genet. Dev.* **10**: 428–433.
- Bartel, D.P. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**: 281–297.
- Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., et al. 2004. The Pfam protein families database. *Nucleic Acids Res.* **32**: D138–141.
- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. 2004. Ultraconserved elements in the human genome. *Science* **304**: 1321–1325.
- Bono, H., Yagi, K., Kasukawa, T., Nikaido, I., Tominaga, N., Miki, R., Mizuno, Y., Tomaru, Y., Goto, H., Nitanda, H., et al. 2003. Systematic expression profiling of the mouse transcriptome using RIKEN cDNA microarrays. *Genome Res.* **13**: 1318–1323.
- Carrington, J.C. and Ambros, V. 2003. Role of microRNAs in plant and animal development. *Science* **301**: 336–338.
- Chan, A.S., Thorner, P.S., Squire, J.A., and Zielenska, M. 2002. Identification of a novel gene NCRMS on chromosome 12q21 with differential expression between rhabdomyosarcoma subtypes. *Oncogene* **21**: 3029–3037.
- Chanfreau, G., Legrain, P., and Jacquier, A. 1998. Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J. Mol. Biol.* **284**: 975–988.
- Dostie, J., Mourelatos, Z., Yang, M., Sharma, A., and Dreyfuss, G. 2003. Numerous microRNPs in neuronal cells containing novel microRNAs. *RNA* **9**: 180–186.
- Erdmann, V.A., Szymanski, M., Hochberg, A., Groot, N., and Barciszewski, J. 2000. Non-coding, mRNA-like RNAs database Y2K. *Nucleic Acids Res.* **28**: 197–200.
- French, P.J., Bliss, T.V., and O'Connor, V. 2001. Ntab, a novel non-coding RNA abundantly expressed in rat brain. *Neuroscience* **108**: 207–215.
- Fujita, M., Kiyono, T., Hayashi, Y., and Ishibashi, M. 1996. hCDC47, a human member of the MCM family. Dissociation of the nucleus-bound form during S phase. *J. Biol. Chem.* **271**: 4349–4354.
- Gene Ontology Consortium. 2001. Creating the gene ontology resource: Design and implementation. *Genome Res.* **8**: 1425–1433.
- Griffiths-Jones, S. 2004. The microRNA Registry. *Nucleic Acids Res.* **32**: D109–111.
- He, L. and Hannon, G.J. 2004. MicroRNAs: Small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5**: 522–531.
- Hong, Y.K., Ontiveros, S.D., Chen, C., and Strauss, W.M. 1999. A new structure for the murine Xist gene and its relationship to chromosome choice/counting during X-chromosome inactivation. *Proc. Natl. Acad. Sci.* **96**: 6829–6834.
- Houbaviy, H.B., Murray, M.F., and Sharp, P.A. 2003. Embryonic stem cell-specific microRNAs. *Dev. Cell* **5**: 351–358.
- Kim, J., Krichevsky, A., Grad, Y., Hayes, G.D., Kosik, K.S., Church, G.M., and Ruvkun, G. 2004. Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. *Proc. Natl. Acad. Sci.* **101**: 360–365.
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853–858.
- Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. 2002. Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**: 735–739.
- Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., and Tuschl, T. 2003. New microRNAs from mouse and human. *RNA* **9**: 175–179.
- 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.
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., et al. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**: 415–419.
- Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., and Bartel, D.P. 2003. Vertebrate microRNA genes. *Science* **299**: 1540.
- Mattick, J.S. 2003. Challenging the dogma: The hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* **25**: 930–939.
- Michael, M.Z., O'Connor, S.M., van Holst Pellekaan, N.G., Young, G.P., and James, R.J. 2003. Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol. Cancer Res.* **1**: 882–891.
- Migliazza, A., Bosch, F., Komatsu, H., Cayanis, E., Martinotti, S., Toniato, E., Guccione, E., Qu, X., Chien, M., Murty, V.V., et al. 2001. Nucleotide sequence, transcription map, and mutation analysis of the 13q14 chromosomal region deleted in B-cell chronic lymphocytic leukemia. *Blood* **97**: 2098–2104.
- 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.
- Numata, K., Kanai, A., Saito, R., Kondo, S., Adachi, J., Wilming, L.G., Hume, D.A., Hayashizaki, Y., and Tomita, M. 2003. Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res.* **13**: 1301–1306.
- Okazaki, Y.M., Furuno, T., Kasukawa, J., Adachi, H., Bono, S., Kondo, I., Nikaido, N., Osato, R., Saito, H., Suzuki, I. et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563–573.
- Ota, T.Y., Suzuki, T., Nishikawa, T., Otsuki, T., Sugiyama, R., Irie, A., Wakamatsu, K., Hayashi, H., Sato, K., Nagai, K., et al. 2004. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat. Genet.* **36**: 40–45.
- Qu, L.H., Henry, Y., Nicoloso, M., Michot, B., Azum, M.C., Renalier, M.H., Caizergues-Ferrer, M., and Bachelier, J.P. 1995. U24, a novel intron-encoded small nucleolar RNA with two 12 nt long, phylogenetically conserved complementarities to 28S rRNA. *Nucleic Acids Res.* **23**: 2669–2676.
- Sato, R., Inoue, J., Kawabe, Y., Kodama, T., Takano, T., and Maeda, M. 1996. Sterol-dependent transcriptional regulation of sterol regulatory element-binding protein-2. *J. Biol. Chem.* **271**: 26461–26464.
- Sempere, L.F., Freemantle, S., Pitha-Rowe, I., Moss, E., Dmitrovsky, E., and Ambros, V. 2004. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol.* **5**: R13.
- Sleutels, F., Zwart, R., and Barlow, D.P. 2002. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415**: 810–813.
- Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A., et al. 2002. Large-scale analysis of the human and mouse transcriptomes. *Proc.*

- Natl. Acad. Sci.* **99**: 4465–4470.
- Tam, W. 2001. Identification and characterization of human BIC, a gene on chromosome 21 that encodes a noncoding RNA. *Gene* **274**: 157–167.
- Tanaka, S., Tatsumi, K., Okubo, K., Itoh, K., Kawamoto, S., Matsubara, K., and Amino, N. 2002. Expression profile of active genes in the human pituitary gland. *J. Mol. Endocrinol.* **28**: 33–44.
- Tycowski, K.T., Shu, M.D., and Steitz, J.A. 1993. A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. *Genes & Dev.* **7**: 1176–1190.
- . 1996. A mammalian gene with introns instead of exons generating stable RNA products. *Nature* **379**: 464–466.
- Velleca, M.A., Wallace, M.C., and Merlie, J.P. 1994. A novel synapse-associated noncoding RNA. *Mol. Cell Biol.* **14**: 7095–7104.
- Xu, P., Vernooy, S.Y., Guo, M., and Hay, B.A. 2003. The *Drosophila* microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. *Curr. Biol.* **13**: 790–795.
- Yamamoto, K., Cox, J.P., Friedrich, T., Christie, P.T., Bald, M., Houtman, P.N., Lapsley, M.J., Patzer, L., Tsimaratos, M., Van, T.H.W.G., et al. 2000. Characterization of renal chloride channel (CLCN5) mutations in Dent's disease. *J. Am. Soc. Nephrol.* **11**: 1460–1468.

## WEB SITE REFERENCES

- <http://biobases.ibch.poznan.pl/ncRNA/>; Noncoding RNAs database.
- <http://www.bioinfo.rpi.edu/applications/mfold/>; mfold RNA program.
- <http://www.ensembl.org>; ENSEMBL genome viewer.
- <http://expression.gnf.org/cgi-bin/index.cgi>; Gene expression atlas database.
- <http://fantom2.gsc.riken.go.jp/>; FANTOM functional annotation database.
- <http://fatigo.bioinfo.cnio.es/>; FatiGO Data Mining with Gene Ontology Terms.
- <http://www.ncbi.nlm.nih.gov/mapview/>; NCBI genome viewer.
- <http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/>; NCBI Acembly AceView gene collection.
- <http://www.nedo.go.jp/bio-e/>; Full-Length Japan database.
- <http://read.gsc.riken.go.jp/>; RIKEN expression array database.
- <http://www.sanger.ac.uk/Software/Rfam/mirna/>; miRNA Registry.
- <http://us.expasy.org/sprot/>; SWISS-PROT.
- [http://vega.sanger.ac.uk/Homo\\_sapiens/](http://vega.sanger.ac.uk/Homo_sapiens/); Vega human annotation browser.
- <http://www.geneontology.org/>; Gene Ontology.

Received April 28, 2004; accepted in revised form July 27, 2004.



## Identification of Mammalian microRNA Host Genes and Transcription Units

Antony Rodriguez, Sam Griffiths-Jones, Jennifer L. Ashurst, et al.

*Genome Res.* 2004 14: 1902-1910

Access the most recent version at doi:[10.1101/gr.2722704](https://doi.org/10.1101/gr.2722704)

---

### Supplemental Material

<http://genome.cshlp.org/content/suppl/2004/10/01/gr.2722704.DC1>

### References

This article cites 43 articles, 21 of which can be accessed free at:  
<http://genome.cshlp.org/content/14/10a/1902.full.html#ref-list-1>

### License

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

Affordable, Accurate  
Sequencing.



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---