Correspondence

EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues

Neil R Smalheiser

Address: Department of Psychiatry, University of Illinois at Chicago, 1601 W. Taylor Street MC 912, Chicago, IL 60612, USA.
E-mail: smalheiser@psych.uic.edu

Published: 18 June 2003

Genome ***Biology*** 2003, **4:**403

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2003/4/7/403

© 2003 BioMed Central Ltd

Summary

A significant population of expressed sequence tags (ESTs) encodes chimeric transcripts containing microRNA (miRNA) precursor sequences as well as pieces of adjacent mRNAs in sense orientation. These chimeric transcripts may potentially be involved in miRNA biosynthesis, and/or affect expression of adjacent mammalian mRNAs.

Hundreds of microRNAs (miRNAs) have recently been discovered in species ranging from plants to humans. They are encoded by genes that express transcripts of single or clustered miRNA precursors of around 70 nucleotides in size, which form imperfect hairpin structures and are further processed to 17-23 nucleotide miRNAs by the action of Dicer [1]. The miRNAs appear to have quite diverse roles: some induce translational arrest, whereas others induce RNA interference (RNAi). Although miRNAs are clearly important in the genome, the biology of miRNA precursors and their transcription is still not well understood. I examined 90 previously characterized miRNAs expressed in mouse and human [2,3] for their homology to sequences present in the NCBI Entrez EST database, and found that many expressed sequence tags (ESTs) encoded chimeric miRNA precursor transcripts that also contained pieces of mRNAs.

Less than half the miRNAs examined (41 of 90) had sequences that exactly matched (in either orientation) one or more ESTs in the publicly accessible NIH database; 36 of these miRNAs matched ESTs in human, rat or mouse. This might reflect a lack of coverage of ESTs in this database, but more probably it reflects the fact that EST sequencing strategies favor long, stable, poly(A)$^+$ transcripts which may not be general features of miRNA transcription and processing pathways [4].

About one-third of these miRNA sequences matched exactly to one or more chimeric ESTs encoding both a miRNA precursor and a region of an adjacent mRNA (Table 1). These cannot represent cloning artifacts, because multiple, independent EST isolates were observed for many examples, and in all but one case the miRNA and mRNA sequences could be mapped to genomic clones and mapped to nearby sites on the same chromosome. In half the cases, one or more of the matching ESTs expressed poly(A)$^+$ tails and the EST orientation could be unambiguously assigned (Table 1). These ESTs definitely expressed miRNA precursor hairpins (rather than antisense transcripts). The unassigned ESTs probably encode miRNA precursors as well, as this agrees with the direction expected from database annotation, they all expressed precursor hairpin sequences uninterrupted by splicing, and in all ESTs, the miRNA sequences were in the same orientation as the piece of mRNA.

It is important to emphasize that most of the ESTs were fundamentally different from the reference mRNAs whose sequences they shared; that is, most contained sequences external to the reference mRNA (with no known variants including them), and/or were spliced differently from the reference mRNA. It is known that mammalian miRNA precursors can be located within introns of both protein-coding

Table 1

ESTS encoding chimeric miRNA precursor mRNA transcripts

miRNA	Example EST	Source	Number of ESTs with mRNA	Any with poly(A)$^+$ tail?	mRNA	Length of mRNA contained in nucleotides	Location of mRNA contained
21, 104	BF326048	Human normal amnion	3	No	NM_030938 vesicular membrane protein 1	290	3′ UTR
22	BQ887833	Human pigmented retinal epithelium	13	Yes	AF070569* clone 24659	472	5′ UTR
93, 94	AW990440	Mouse lactating mammary gland	1	No	XM_124678† mini chromosome maintenance deficient 7	159	Coding sequence
123, 126	BI395608	Rat mixed tissues	1	No	NM_139104 estrogen-regulated protein	167	Coding sequence
124a	BF402302	Rat brain	2	Yes	XM_139109 kinesin-like	164	Coding sequence
125b	BG000222	Human normal placenta	1	No	NM_147207 ischemia related factor vof-16	73	5′ UTR
142-s, 142-as	BM994627	Human metastatic chondrosarcoma	6	Yes	XM_173924† hypothetical protein	21	Coding sequence

Each of the miRNAs reported in [2] and [3] were characterized against the NCBI Entrez combined EST database using BLAST (parameters optimized for short sequences: expect = 1,000, word size = 7, no filtering) [14,15]. Each EST that matched a miRNA perfectly in either orientation was characterized by BLAST against the nr database (using default parameters). An EST was deemed to be a likely miRNA precursor if its sequence matched the miRNA exactly and if the mfold secondary-sequence prediction algorithm [16,17] predicted that this sequence lies on the arm of an imperfect hairpin of around 70 nucleotides. Note that several different miRNA sequences are often represented in the same EST. *Although this mRNA is not annotated, it overlaps in sense direction with two other annotated mRNAs BC007813 and NM_032895, allowing it to be assigned unambiguously. †After this paper was initially submitted for publication, these records were removed from GenBank as a result of standard genome annotation processing (though still visible upon query of the database). However, that does not imply that the records are necessarily obsolete or in error. EST 990440 matched not only XM_124678, but numerous other mRNAs that are still in GenBank - for example, NM_008568. Thus, the finding is not restricted to a single rogue mRNA entry. The entire sequence of XM_173924 maps with no discrepancies to two human chromosome 17 genomic clones (for example, AC023992), suggesting that it does not contain sequencing errors.

and noncoding genes [5], so any EST that expresses the mRNA along with retained introns might erroneously appear to be 'chimeric'. Although four of the miRNA sequences described here do reside within introns, in at least three of these the ESTs described do not appear simply to represent mRNA sequences that contain retained introns (the other, miRNA 124a, cannot be assessed because it lacks corresponding genomic clones to identify intronic borders). In three cases (miRNA 21/104, 22 and 125b) the miRNA precursor sequences are located in intergenic regions, beyond the borders of the reference mRNAs. The miRNA precursor hairpin was generally not located at one end of the EST, but had flanking 5′ and 3′ sequences. For each miRNA listed in Table 1, at least one of the corresponding ESTs was derived from normal fetal or adult tissues, though some were also found expressed in cancer tissue. Most involve mRNAs that encode well-characterized protein products, though two correspond to hypothetical proteins.

The ESTs encoding miRNAs and mRNA pieces were apparently transcribed by RNA polymerase II, as many had poly(A)$^+$ tails. RNA polymerase III is unlikely to be responsible for transcribing the chimeric transcripts, as the majority of the ESTs in Table 1 have internal stretches of four or more Ts in sense orientation that are thought to act as termination signals for the polymerase. However, as potential RNA polymerase III termination signals were encountered in only one miRNA precursor hairpin region, this polymerase may still be involved in transcribing primary miRNA precursor transcripts in other situations.

Examination of ESTs is fraught with potential problems, including cloning artifacts, uncertain orientation, and inclusion of unprocessed or aberrantly processed transcripts. However, chimeric miRNA-mRNA transcripts were detected for numerous miRNAs, with multiple EST isolates, and from several different tissues and different species, so they are likely to represent a regular phenomenon. Furthermore, many had poly(A)$^+$ tails and were spliced, indicating that they can be extensively processed.

It is uncertain whether these chimeric transcripts are further processed to functionally active miRNAs. This is a possibility, as Zeng and Cullen reported

that certain miRNA precursors could be processed effectively when they were expressed as RNA polymerase II transcripts containing flanking 5′ and 3′ sequences [6]. Yet this would still not explain why so many of the matching ESTs also transcribed into the neighboring mRNA. Is this a clue to a distinctive function for chimeric transcripts, or does this reflect the nature of controls on their transcription and termination?

Another intriguing question is why, in each case, the miRNA sequence was oriented in the same direction as the mRNA, regardless of whether the miRNA was positioned upstream, within, or downstream of the mRNA sequence itself. Certainly, miRNAs are not in general forbidden from being located on the opposite strand of mRNAs. For example, in the course of the present study, miRNA 127 was found to be encoded by a precursor (EST BE294363) that lies on the opposite strand from the protein-coding region of a transposon-associated polyprotein mRNA that makes Gag protein and reverse transcriptase (XM_090919). A miRNA in this location would be in a position to inhibit transposon function by perfect antisense pairing to the polyprotein mRNA followed by RNAi. One conceivable function of chimeric transcripts that express pieces of mRNA in sense orientation may be to downregulate the endogenous mRNAs via sense co-suppression [7], as previously proposed for chimeric transcripts arising from the antisense promoter of L1 retrotransposons [8,9].

Finally, it is not clear how the RNA polymerase II transcription of these ESTs is regulated. Is it governed by external sequences, perhaps related to the neighboring mRNA, or by internal sequences related to the miRNA precursor hairpins? Of the chimeric ESTs that could be lined up against genomic clones of the same species, I could not identify any obvious polymerase II promoter regions using the ProScan algorithm [10,11] either within or up to 4 kb upstream of the EST sequences (the exception being ESTs matching miRNA 22, which resides upstream of a mRNA 5′ UTR). As an alternative, one can conceive of the possibility that at least some miRNA precursor hairpins (that is, those associated with chimeric transcripts) may bind specific proteins that affect RNA polymerase II transcription. There is already evidence that proteins may recognize the loops of miRNA precursor hairpins to allow their transport into the cytoplasm [12]. Hairpins may also arguably have some activity as internal promoters or enhancers, since Llave *et al*. [13] tested a single plant pre-miRNA construct lacking an exogenous promoter and found that it had detectable, albeit limited, expression when transfected into cells.

Ultimately, the purpose of bioinformatic analyses is to suggest new laboratory experiments. Identifying a population of chimeric ESTs within GenBank is merely the starting point for asking whether one can validate and characterize full-length endogenous chimeric transcripts made within cells. If so, then it will be possible to learn how these relate to other potential biosynthetic routes for miRNA production, how their transcription is regulated, and what functions (if any) they may have.

Acknowledgements
This work was supported by NIH grants LM07292, DA15450 and the Human Brain Project. I thank Vetle Torvik for assistance with batch BLAST searches.

References
1. Ambros V: **microRNAs: tiny regulators with great potential.** *Cell* 2001, **107:**823-826.
2. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T: **Identification of tissue-specific microRNAs from mouse.** *Curr Biol* 2002, **12:**735-739.
3. Mourelatos Z, Dostie J, Paushkin S, Sharma A, Charroux B, Abel L, Rappsilber J, Mann M, Dreyfuss G: **miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs.** *Genes Dev* 2002, **16:**720-728.
4. Lee Y, Jeon K, Lee JT, Kim S, Kim VN: **MicroRNA maturation: stepwise processing and subcellular localization.** *EMBO J* 2002, **21:**4663-4670.
5. Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T**: New microRNAs from mouse and human.** *RNA* 2003, **9:**175-179.
6. Zeng Y, Cullen B**: Sequence requirements for micro RNA processing and function in human cells.** *RNA* 2003, **9:**112-123.
7. Zamore PD: **Ancient pathways programmed by small RNAs.** *Science* 2002, **296:**1265-1269.
8. Nigumann P, Redik K, Matlik K, Speek M: **Many human genes are transcribed from the antisense promoter of L1 retrotransposon.** *Genomics* 2002, **79:**628-634.
9. Speek M: **Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes.** *Mol Cell Biol* 2001, **21:**1973-1985.
10. Prestridge DS**: Predicting Pol II promoter sequences using transcription factor binding sites.** *J Mol Biol* 1995, **249:**923-932.
11. **Web Promoter Scan Service** [http://bimas.cit.nih.gov/molbio/proscan/index.html]
12. Kawasaki H, Taira K: **Short hairpin type of dsRNAs that are controlled by tRNA(Val) promoter significantly induce RNAi-mediated gene silencing in the cytoplasm of human cells.** *Nucleic Acids Res* 2003, **31:**700-707.
13. Llave C, Xie Z, Kasschau KD, Carrington JC: **Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA.** *Science* 2002, **297:**2053-2056.
14. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.
15. **NCBI BLAST Server** [http://www.ncbi.nlm.nih.gov/BLAST]
16. Mathews DH, Sabina J, Zuker M, Turner DH: **Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.** *J Mol Biol* 1999, **288:**911-940.
17. **mfold** [http://www.bioinfo.rpi.edu/applications/mfold/old/rna]

comment reviews reports deposited research refereed research interactions information