Sequence analysis
Evidence for human microRNA-offset RNAs in small RNA sequencing data

David Langenberger1,2, Clara Bermudez-Santana1,2,3, Jana Hertel1,2, Steve Hoffmann1,2, Philipp Khaitovich1 and Peter F. Stadler1,2,5–8,*
1Bioinformatics Group, Department of Computer Science, 2Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtestraße 16-18, D-04107 Leipzig, Germany, 3Department of Biology, Universidad Nacional de Colombia, Carrera 45 No 26-85 -Edificio Uriel Gutierrez, Bogotá, D.C., Colombia, 4CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, 320 Yue Yang Road, 200031 Shanghai, China, 5Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, 6RNomics Group, Fraunhofer Institut for Cell Therapy and Immunology, Perlickstraße 1,D-04103 Leipzig, Germany, 7Institute for Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Vienna, Austria and 8The Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM, USA
Received on May 6, 2009; revised on June 25, 2009; accepted on July 1, 2009
Advance Access publication July 6, 2009
Associate Editor: Ivo Hofacker

ABSTRACT
MicroRNA-offset-RNAs (moRs) were recently detected as highly abundant class of small RNAs in a basal chordate. Using short read sequencing data, we show here that moRs are also produced from human microRNA precursors, albeit at quite low expression levels. The expression levels of moRs are unrelated to those of the associated microRNAs. Surprisingly, microRNA precursors that also show moRs are typically evolutionarily old, comprising more than half of the microRNA families that were present in early Bilateria, while evidence for moRs was found only for a relative small fraction of microRNA families of recent origin.

Contact: studia@bioinf.uni-leipzig.de
Supplementary information: Supplementary data are available at Bioinformatics online and in machine-readable form at http://www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/09-01/1

1 INTRODUCTION
In a recent study, Shi et al. (2009) found that in the tunicate Ciona intestinalis, half of the identified microRNA (miRNA) loci encode up to four distinct, stable small RNAs. These additional RNAs, termed miRNA-offset RNAs (moRs), are generated from sequences immediately adjacent to mature miR and miR* loci. Like mature miRNAs, they are ~20 nt long, developmentally regulated, and appear to be produced by RNase III-like processing from the pre-miRNA hairpin. This observation prompted us to specifically search for analogous pattern in human small RNA sequencing libraries.

2 METHODS
The datasets analyzed here for expression at miRNA loci were produced in the context of other projects and will be published in that context. In brief, total RNA was isolated from the frozen prefrontal cortex tissue using the TRIzol (Invitrogen, USA) protocol with no modifications. Low molecular weight RNA was isolated, ligated to the adapters, amplified and sequenced following the Small RNA Preparation Protocol (Illumina, USA) with no modifications.

Reads were mapped to the human genome (NCBI36.0 Release of July 2008) using a variant of the tool that automatically recognizes blocks of reads. In the first step, a mapped read u with start and end positions u_s and u_e is replaced by a Gaussian density $\rho_u$ with mean $\mu_u = (u_s + u_e)/2$ and variance $\sigma_u^2$. We set $\sigma_u = \sigma_0 (u_e - u_s)/2$, where $\sigma_0$ is a parameter that is used to tune the resolution. For each locus, these Gaussian densities are added up to tune the resolution. For each locus, these Gaussian densities are added up to tune the resolution. The resulting curves $f^+$ and $f^−$ that exhibit pronounced but smooth peaks centered at blocks of reads with nearly identical midpoints (Fig. 1). Since the area under a peak equals the number of reads in the block the height of the peaks provides a meaningful trade-off between the coherence of a block and its expression level. The expression levels of moRs are unrelated to those of the associated microRNAs. Surprisingly, microRNA precursors that also show moRs are typically evolutionarily old, comprising more than half of the microRNA families that were present in early Bilateria, while evidence for moRs was found only for a relative small fraction of microRNA families of recent origin.

© The Author 2009. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
Fig. 1. Decomposition of the cluster of reads at the mir-125b-1 locus (Lee et al., 2005) on chr11 (bottom panel). The blockbuster algorithm replaces each read by Gaussian profile centered at the midpoint of the read. The middle panel shows the superposition f(i) of these profiles for each different width of the Gaussian, here chosen to be fraction of the read lengths L. Clusters (top panel) are identified as sets of reads whose midpoints are located in close to the peaks of f(i). Clusters 2 and 3 correspond to the miR-125b-1 and miR-125b-1*.

Fig. 2. Distribution of short reads at the hsa-mir-425 locus. There are three clearly distinct blocks of reads: the two more abundant ones correspond to miR and miR*, the third one to the 5’-moRNA. Below the conservation pattern is shown. Figure exported from the UCSC Genome Browser.

3 RESULTS

In the brain libraries, we found 78 annotated miRNA loci that exhibit blocks of reads at positions characteristic for moRNAs. For 11 loci, the miRNA Expression Atlas (Landgraf et al., 2007) also contains moRNA reads, see Supplementary Material. In all cases, the reads match perfectly and uniquely to the human genome (hg18), strongly suggesting they are neither technical nor computational artifacts. For 71 of the 78 loci, the moRNAs are conserved together with their miRNA. In contrast to the situation in the urochordate Ciona intestinalis, however, human moRNAs appear to be expressed at very low levels. In particular, at least in the brain libraries examined here, moR levels are systematically below the expression levels of miR and miR* reads.

One of the most prominent regulated moRs in Ciona intestinalis is miR-219 (Shi et al., 2009). Interestingly, three of its human paralogs also produce clear evidence for offset RNAs (miR-219-2; three reads; and miR-124-2, four reads). The offset RNA reads at the 78 human loci share several characteristics with each other and with the moRNAs of urochordates: (i) The moRNA reads are located adjacent to the miRNA reads, and in some cases with only a few nucleotides overlapping the miR or miR*. This conforms with the findings in urochordates and is indicative of processing by a Dicer-like enzyme.
(Shi et al., 2009). (ii) In most cases, the moR sequences are located completely within the predicted hairpin structure. (iii) The moRNA reads are located (almost) entirely in a well-conserved region.

Compared to the 3′-side of the stem, there are more than five times as many moRNA reads on the 5′-side of the stem. This is independent of whether the 3′- or 5′-side is predominantly processed into miRNAs. In fact, exactly half of the 78 loci show a prevalence of the 5′-miR, while the 3′-miR is more abundant in the remaining 39 cases. In contrast, the 5′-moRNA is represented by more reads than the 3′-side in 67 cases (86%). Correspondingly, most of the loci have moRNA reads only on the 5′-side of the precursor hairpin (Fig. 3).

This prevalence for the 5′-side is independent of the expression patterns of the mature miRNAs. Using Fisher’s exact test, we find that there is no significant association of the 5′/3′-bias in the numbers of moRNA- and miR reads, respectively. There is also no significant correlation between the number of moRNA reads and the expression levels of the corresponding mature miRs (Fig. 3).

An investigation of the sequence patterns around the cut-site between miRNA reads and offset reads shows no discernible sequence preferences. Interestingly, there is also no difference in the predicted length of precursor hairpins between miRNAs with and without moRNA reads. For details we refer to the Supplementary Material.

The miRNA families with offset RNAs are significantly over-represented among the oldest animal miRNAs. In fact, more than half of them originated already in the ancestral bilaterian. Among those that have arisen in Mammalia, again the older ones are more likely to exhibit evidence for moRNAs (Fig. 4).

The 78 miRNA loci belong to only 54 distinct families. Of these, four families show moRNAs in three or more paralogs, and seven families have two paralogs with evidence for moRNA expression. As almost all miRNA families with multiple paralogs are evolutionarily old, this observation corroborates the association of moRNAs with an early evolutionary origin.

4 DISCUSSION

Despite the low level of expression compared with miRNAs, our data strongly suggest that many human pre-miRNAs are processed to produce moRNAs in a systematic way. Several lines of evidence suggest that these transcripts are functional, in particular the extreme level of sequence conservation and the example of miR-219 with moRNAs conserved between human and Ciona. The uncorrelated expression of miRs and moRNAs, and the extreme 5′-bias of moRNA reads provides evidence that human moRNAs are not just a random by-product of the miRNA processing pathway. The observation that moRNAs sequences are also found by Sanger sequencing rules out that they are technical artifacts from the next generation sequencing procedures. Taken together, our analysis points to a function independent of that of the miRNAs processed from the same locus. These molecules thus may well form a distinct functional class of miRNA-like agents, e.g. akin to mittrons (Berezikov et al., 2007). This conclusion is supported further by the intriguing observation that the majority of miRNAs with moRNA expression are among the evolutionarily oldest families.

ACKNOWLEDGEMENTS

European Union (EDEN, contract 043251 to P.F.S.); Deutsche Forschungsgemeinschaft under the auspices of SPP-1174 ‘Deep Metazoan Phylogeny’ (to P.F.S.); a PhD scholarship of the DAAD-AleCol program (to C.B.-S.); a formel.1 grant of the Medical Faculty of the University of Leipzig; the Freistaat Sachsen under the auspices of the LIFE project.

Conflict of Interest: none declared.

REFERENCES


