Identification of Novel MicroRNA-Like Molecules Generated from Herpesvirus and Host tRNA Transcripts

Tiffany A. Reese, Jing Xia, L. Steven Johnson, Xiang Zhou, Weixiong Zhang and Herbert W. Virgin

Identification of Novel MicroRNA-Like Molecules Generated from Herpesvirus and Host tRNA Transcripts

Tiffany A. Reese,1 Jing Xia,2 L. Steven Johnson,1 Xiang Zhou,2 Weixiong Zhang,2,3* and Herbert W. Virgin1,4* Department of Pathology and Immunology,1 Department of Computer Science and Engineering,2 Department of Genetics,3 and Midwest Regional Center of Excellence for Biodefense and Emerging Infectious Disease Research,4 Washington University in St. Louis, St. Louis, Missouri

Received 1 April 2010/Accepted 8 July 2010

We applied deep sequencing technology to small RNA fractions from cells lytically infected with murine gammaherpesvirus 68 (γHV68) in order to define in detail small RNAs generated from a cluster of tRNA-related polycistronic structures located at the left end of the viral genome. We detected 10 new candidate microRNAs (miRNAs), six of which were confirmed by Northern blot analysis, leaving four as provisional. In addition, we determined that previously identified and annotated viral miRNA molecules were not necessarily represented as the most abundant sequence originating from a transcript. Based on these new small RNAs and previously reported γHV68 miRNAs, we were able to further describe and annotate the distinctive γHV68 tRNA-miRNA structures. We used this deep sequencing data and computational analysis to identify similar structures in the mouse genome and validated that these host structures also give rise to small RNAs. This reveals a possible convergent usage of tRNA/polymerase III (pol III) transcripts to generate small RNAs from both mammalian and viral genomes.

Data from deep sequencing technologies suggest that the coding complexity of many genomes, from virus and plant to mouse and human genomes, has been underestimated. Researchers have identified a vast array of different types of noncoding RNAs (ncRNAs) emanating from these different genomes. Importantly, some viruses, with their compact genomes and tractable genetics, provide a valuable tool for understanding general principles of eukaryotic genomes and a functional model for defining novel types of ncRNAs as well as their mechanisms of action.

The transcriptional complexity of eukaryotic genomes is exemplified by the increase in the number and types of functions of small ncRNAs. These RNAs can be subdivided to include not only microRNAs (miRNAs) (21 to 25 nucleotides [nt]), small interfering RNAs (siRNAs) (21 to 25 nt), and Piwi-interacting RNAs (piRNAs) (24 to 31 nt) (11, 14) but also the recently identified tRNA-derived RNA fragments (tRFs) (18 to 25 nt) (9). In addition to tRFs, there are a few other examples of small RNAs that are associated with tRNAs and polymerase III (Pol III). These include 19-nt Dicer-dependent RNA fragments derived from tRNAs (4), human small RNAs interspersed among Alu repeats and transcribed by Pol III (2), and tRNA-miRNA polycistrons from the murine gammaherpesvirus 68 (γHV68) (20).

Many DNA viruses, especially herpesviruses, encode miRNAs and other ncRNAs that regulate viral replication and host responses. For example, the adenovirus-associated (VA) RNA is an ~165-nt ncRNA that inhibits the antiviral effects of PKR-mediated translation inhibition (12), the Epstein-Barr virus (EBV)-encoded RNAs (EBERs) disrupt interferon-induced apoptosis of host cells (22), and herpesvirus saimiri encodes seven small nuclear RNAs of the Sm class (HSURs) which are believed to play a role in T-cell activation (5). Additionally, functionally conserved miRNAs from different herpesviruses, including human cytomegalovirus (HCMV), EBV, and Kaposi’s sarcoma herpesvirus (KSHV), help these viruses evade immune recognition by targeting a host transcript, MIBC, a ligand of natural killer (NK) cells which controls herpesvirus replication (16, 23). These different ncRNAs may participate in posttranscriptional gene silencing (by mRNA degradation or translational regulation) or transcriptional gene silencing by acting on chromatin function and accessibility. Together, these ncRNAs constitute new classes of viral genes that may play important roles in viral infection.

γHV68 is a murine gammaherpesvirus genetically related to EBV and KSHV. Similar to these other herpesviruses, γHV68 is known to encode at least 10 miRNAs with as-yet-undefined function (20). These were identified by cDNA cloning, with three of 10 confirmed by Northern blot analysis in infected B-cell lines (20) and nine of 10 detected by reverse transcription (RT)-PCR from infected fibroblasts (6). However, distinct from EBV and KSHV, γHV68 generates its primary miRNA transcripts (pri-miRNA) via Pol III, instead of Pol II, as part of larger transcripts that include tRNA-like elements directly upstream of the predicted miRNAs (3, 6, 20). Moreover, these miRNAs are generated via a noncanonical processing pathway that utilizes tRNase Z instead of Drosha to liberate the pre-miRNA hairpin from the tRNA. The subsequent pre-miRNA is thought to then...
be processed by Dicer, similar to other miRNAs, to generate the mature miRNA (1).

To further expand our understanding of \( \gamma \)HV68-expressed small RNAs, we analyzed small RNA libraries by Illumina sequencing and mapped the small RNAs to the first 7.5 kb of the genome. In the initial identification of the \( \gamma \)HV68 miRNAs, tRNA-miRNA polycistrons from this region were predicted with each of the eight tRNAs followed by two stem-loop hairpins (20). The 10 previously annotated miRNAs mapped to these predicted hairpin structures from tRNA-like transcripts, but many hairpins were without an identified miRNA. From our sequencing data, we demonstrate that most of these hairpins do indeed have candidate miRNAs emanating from them. Also, a tRNA with no previously identified miRNA, tRNA7, has a candidate miRNA following it. Moreover, in some cases the most frequent read from a region of a previously described viral miRNA was shifted by a few nucleotides from the published miRNA sequence. We detected an approximately 22-nt small RNA for six of the 10 candidate miRNAs by Northern blot analysis. Further mining of the sequencing data also revealed mouse transcripts with tRNAs followed by novel small RNAs that could correspond to either miRNAs or putative tRFs recently characterized from prostate cancer cell lines (9), thus suggesting a possible coevolution of virus and host resulting in similar mechanisms of expression of small ncRNAs.

### MATERIALS AND METHODS

**Animals.** C57BL/6 mice were obtained from Jackson Laboratories and bred at Washington University, St. Louis, MO. Stat1KO mice were obtained from Rob-ert Schreiber (13). All mice were housed and bred in a specific-pathogen-free barrier facility in accordance with federal and institutional guidelines. Mice were infected intraperitoneally between 8 and 10 weeks of age with 1 \( \times 10^6 \) PFU of \( \gamma \)HV68 and euthanized 2 days later. Peritoneal cells were collected for RNA isolation.

**Cells.** NIH 3T12 cells (ATCC CRL-164) and Vero cells (ATCC CCL-81) were maintained in DMEM5 (Dulbecco’s modified Eagle’s medium with 5% fetal calf serum [FCS], 2 mM \( L \)-glutamine). Bone marrow-derived macrophages were generated as follows. Bone marrow was flushed from femurs of C57BL/6 mice and cultured in BME (10% FCS, 20% L-cell-conditioned medium derived from a macrophage colony-stimulating factor [M-CSF]-producing fibroblast cell line, 5% horse serum, 2 mM \( L \)-glutamine, and 1 mM sodium pyruvate in DMEM) for the first 7 days of \( \gamma \)HV68 infections. On day 7, adherent cells were scraped and seeded in BM10 medium (10% FCS, 10% L-cell-conditioned medium, 5% horse serum in DMEM). On day 10, cells were infected and maintained in BM10 medium.

**Virus infections.** \( \gamma \)HV68 WUMS (ATCC VR1465) was passaged and titrated on NIH 3T12 cells (19). NIH 3T12 cells and bone marrow-derived macrophages were infected at a multiplicity of infection (MOI) of 10 PFU per cell at 37°C and 5% \( CO_2 \), washed three times with medium, and replated in media. Subsequent RNA extractions on mock and infected cells were done 18 h later.

**RNA preparations.** Total RNA was prepared using the mirVana miRNA isolation kit (Ambion) according to the manufacturer’s instructions for isolation of total RNA. RNA was treated with DNase I (Ambion) and run on a 15% denaturing polyacrylamide gel and stained with ethidium bromide to check for RNA integrity.

**Small RNA preparations for sequencing.** Total RNA libraries for sequencing from mock- or \( \gamma \)HV68-infected cells were prepared using the small RNA sample prep kit (Illumina) by following the Small RNA v.1.5 Sample Preparation Guide. Twenty micrograms of total RNA was used as input, and small RNA was size selected by running total RNA on a 15% denaturing polyacrylamide gel and cutting out the RNA corresponding to 18 to 30 nt.

**Initial processing of sequencing libraries.** The raw sequence reads were first processed to remove reads with no 3’ sequencing adapter, of low quality, or shorter than 18 nt. The sequence adapter trimming was done by an in-house program that recursively searches for the longest subsequence of the adapter from a macrophage colony-stimulating factor [M-CSF]-producing fibroblast cell line, 5% horse serum, 2 mM \( L \)-glutamine, and 1 mM sodium pyruvate in DMEM) for the first 7 days of \( \gamma \)HV68 infection. On day 7, adherent cells were scraped and seeded in BM10 medium (10% FCS, 10% L-cell-conditioned medium, 5% horse serum in DMEM). On day 10, cells were infected and maintained in BM10 medium.

**Identification of candidate miRNAs.** We revised and extended an miRNA identification method we developed previously (24) to find novel miRNAs in \( \gamma \)HV68 and mouse genomes, which may have more mismatches and bulges in miRNA precursor foldback structures than regular miRNAs. We first mapped the adaptor-trimmed sequence reads to the genome using Bowtie (8), merged neighboring loci if they shared overlapping reads, and then examined the folding structures of the (merged) loci. Since the average length of an miRNA precursor is \( \approx 80 \) nt, we took 100 nt as the length of putative pre-miRNAs in our analysis. At each genomic locus to be analyzed, a series of DNA sequence segments covering the sequence reads were extracted for secondary structure analysis. The starting sequence segment extended 80 nt upstream of the sequence reads, and subsequent segments were extracted by a sliding window of 100 nt, with an
increment of 1 nt, until the window reached 80 nt downstream of the sequence reads. Each of these 80-nt segments was folded by the RNA-fold program (7). Segments lacking stems of at least 18 nt and segments lacking sequence reads that mapped to any of their stems were excluded. A representative segment was chosen from those that have the same or a similar folding structure. The top 5 folding structures with a free energy no greater than −18 kCal/mol were further visually inspected. We retained those segments that formed a typical hairpin foldback structure with up to 5 mismatches on one stem region. We examined multiple folding structures in order to identify miRNA candidates, since a computationally predicted optimal RNA folding structure may not be necessarily functionally relevant. We extracted a longer region that included a candidate miRNA and its neighboring tRNA and folded the region using the RNA-fold program.

Northern blot analysis. Northern blotting for miRNAs was done using the enhanced method for detection (18). Twenty micrograms of total RNA was run on a 15% denaturing acrylamide gel (Invitrogen) from mock-infected 3T12 cells, and, in some cases, Vero cells. Decade markers were run on a gel. Following quantitative PCR, products were run on a 3% NuSieve GTG agarose gel to determine the relative size of the amplified products compared with mouse miRNA-16 and the DNA marker V (Roche). Products from mock- and HV68-infected samples were analyzed both by melting curve and by gel electrophoresis. An HCMV RNA (snRNA) RNU6B. Products from mock- and HV68-infected samples were analyzed both by melting curve and by gel electrophoresis. An HCMV RNA (snRNA) RNU6B. Products from mock- and HV68-infected samples were analyzed both by melting curve and by gel electrophoresis. An HCMV RNA (snRNA) RNU6B.

## Table 2. Statistics of 3 small RNA sequencing libraries

<table>
<thead>
<tr>
<th>Category of reads</th>
<th>γHV68-1</th>
<th>γHV68-2</th>
<th>γHV68-3</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>6,411,884</td>
<td>10,722,962</td>
<td>13,792,659</td>
<td>Total raw sequencing reads</td>
</tr>
<tr>
<td>No adaptor</td>
<td>3,946,977</td>
<td>4,346,389</td>
<td>4,568,762</td>
<td>Reads with no adaptor</td>
</tr>
<tr>
<td>Low quality</td>
<td>1,874</td>
<td>1,486,925</td>
<td>756,171</td>
<td>Reads with middle Ns</td>
</tr>
<tr>
<td>Short</td>
<td>1,040,150</td>
<td>1,486,925</td>
<td>756,171</td>
<td>Reads &lt; 18 nt after adapter trimmed</td>
</tr>
<tr>
<td>Total qualified</td>
<td>1,870,881</td>
<td>5,628,518</td>
<td>8,452,638</td>
<td>Total high-quality reads ≥18 nt</td>
</tr>
<tr>
<td>Total mapped to γHV68</td>
<td>6,176</td>
<td>48,289</td>
<td>208,530</td>
<td>Total high-quality reads mapped to γHV68 genome</td>
</tr>
<tr>
<td>Total mapped to mouse</td>
<td>1,229,450</td>
<td>4,295,302</td>
<td>6,329,431</td>
<td>Total high-quality reads mapped to mouse genome</td>
</tr>
<tr>
<td>Unique qualified</td>
<td>188,558</td>
<td>5,628,518</td>
<td>8,452,638</td>
<td>Total high-quality reads mapped to mouse genome</td>
</tr>
<tr>
<td>Unique mapped to γHV68</td>
<td>997</td>
<td>3,388</td>
<td>9,461</td>
<td>Unique high-quality reads mapped to γHV68 genome</td>
</tr>
<tr>
<td>Unique mapped to mouse</td>
<td>28,327</td>
<td>94,931</td>
<td>83,660</td>
<td>Unique high-quality reads mapped to mouse genome</td>
</tr>
</tbody>
</table>

Below that, the numbers of positive and negative strand reads per nucleotide are plotted on a log2 scale. Thus, a peak represents all the reads in the region that overlap a particular nucleotide. Red arrows point to the position at which novel candidate miRNAs map. Data from three independent experiments were combined for the final analysis.

![FIG. 1. Identification of novel candidate viral miRNAs in γHV68-infected fibroblasts. (A) Size-selected small RNA (18 to 30 nt) from 3T12 mouse fibroblasts infected with γHV68 was Solexa sequenced. The resulting reads were mapped to the γHV68 genome, and the first 7.5 kb of the genome is shown. The top line represents the position in the genome, and the annotated positive- and negative-strand ORFs are shown as green arrows. Yellow arrows represent annotated tRNAs on the positive strand, and blue arrows represent annotated miRNAs on the positive strand. Below that, the numbers of positive and negative strand reads per nucleotide are plotted on a log2 scale. Thus, a peak represents all the reads in the region that overlap a particular nucleotide. Red arrows point to the position at which novel candidate miRNAs map. Data from three independent experiments were combined for the final analysis.](http://jvi.asm.org/DownloadedFrom)
FIG. 2. Predicted structure of novel candidate γHV68 miRNAs. Predicted secondary structures of tRNA-miRNA transcripts, with annotated miRNAs and peak reads depicted in red. Novel candidate miRNAs are boxed. Novel and previously annotated miRNAs were analyzed for peak read. #, annotated miRNAs for which the most abundant read is shifted slightly off and does not correspond to annotation. #*, an annotation that differs significantly from the peak read such that the peak read is on the opposite arm from miRNA annotation. Graphs depict read frequency across the nucleotide position of the structures for each polycistron. For tRNA7 and tRNA8, boxed regions depict areas with possible alternate folded structures.
RESULTS

Identification of novel miRNAs from \(\gamma\)HV68-infected cells.

To identify novel \(\gamma\)HV68 miRNAs, we performed Illumina deep sequencing on size-selected (<30-nt) RNA from lyrically infected fibroblasts. Total RNA was harvested after 18 h of infection of 3T12 cells, and small RNA libraries were prepared for sequencing. We obtained more than 30.9 million raw sequence reads from three replicate small RNA libraries, among which 262,995 mapped to the \(\gamma\)HV68 genome. The reads that mapped to the mouse genome were 21 nt or 23 nt long (data not shown). Figure 1 shows the reads that mapped to the \(\gamma\)HV68 genome, with each read colored according to whether the mapped read is the abundant read in its respective miRNA peak (red), or whether it is a peak read that is not the most abundant read in the peak (blue). Novel miRNA peaks indicated by red arrows.

After mapping reads to the \(\gamma\)HV68 genome, we found five novel sequencing peaks directly following the \(\gamma\)HV68 tRNAs, similar but distinct from the 10 known \(\gamma\)HV68 miRNAs (Fig. 1). These novel sequences from the positive strand all mapped to the already-predicted tRNA-miRNA secondary structures (Fig. 2) (20). The few reads from the negative strand are of unknown significance and were not further evaluated. Graphical displays of read frequencies across the tRNA-miRNA transcripts are shown in Fig. 2, and sequences of the novel candidate miRNAs and their read counts are listed in Table 1. Candidate miRNAs shown in Fig. 2 were chosen based on the following criteria: (i) occurrence of sequence reads on the arms of predicted hairpin structures, (ii) frequencies of peak reads on predicted hairpins, (iii) presence of possible miRNA\(^*\) sequence reads, and (iv) presence of possible 2- to 3-nt 3' overhangs on miRNA/miRNA\(^*\) duplexes. miRNA\(^*\) is the sequence from the opposite arm of an miRNA/miRNA\(^*\) duplex formed by the pre-miRNA hairpin. We found two putative folding structures for tRNA7 and tRNA8 with compatible low folding energies, which yielded alternate possible pre-miRNA structures (Fig. 2). When we examined the reads across the region of tRNA7, we found reads that mapped to both possible stem-loops, although the frequencies for any given read for the miR-M1-13b stem-loop were very low (Table 1). tRNA8 was similar; there were reads across the alternate stem-loop encompassing miR-M1-14, but their frequency was low. There was one read that occurred 46 times and included the sequence (AGGGGGAGTGTGTGGTCTGTA) for miR-M1-14-5p; however, that read was 29 nucleotides long (data not shown). We included miR-M1-13b and miR-M1-14 in our subsequent analyses to attempt to confirm them by other methods.

Sequence abundance of previously identified miRNAs. Interestingly, when we examined our sequencing results for the previously identified \(\gamma\)HV68 miRNAs, we found that the most abundant sequence did not always correspond to the annotated miRNA sequence (20). This is indicated in Fig. 2 by the pound symbol (#) next to the name of the miRNA. In all but one case, the peak read for a given stem-loop is shifted off from the annotated miRNA/miRNA\(^*\) duplex (Table 3). For miR-M1-4, -3, -5, -6, and -9, the differences are all at the 3' end and may represent differences in processing or could be an artifact of adapter ligation for sequencing. However, miR-M1-2, -4, and
8 were more substantially different from the previous annotation. For miR-M1-4 and miR-M1-8, both showed differences between the peak read and the previous annotation at the 5' end of the miRNA. This could significantly alter the predicted seed sequences or target mRNAs and change conclusions regarding the function of these two miRNAs. In the case of miR-M1-2, the peak read illustrated in Fig. 2 is actually on the 5' arm of the stem-loop, while the annotated miRNA is supposed to be on the 3' arm of the stem-loop. The miRNA sequence for miR-M1-2 is detected only four times from the sequencing, while the reads from the other arm of the stem-loop are detected at a much higher frequency and are repre-

FIG. 3. Detection of candidate viral miRNAs. Northern blot analysis of novel candidate miRNAs from 3T12 cells mock infected (M) or infected with γHV68 (V). Ethidium bromide staining is shown to indicate RNA loading. Blots are representative of 2 or 3 independent experiments. In many cases, blots were probed for one miRNA and later stripped and reprobed for a different miRNA.

FIG. 4. Detection of candidate viral miRNAs in vitro and in vivo. (A) RT-PCR detection of candidate miRNAs from lytically infected 3T12 cells. (B) RT-PCR detection of candidate miRNAs from bone marrow-derived macrophages infected with γHV68 for 20 h. (C) RT-PCR detection of candidate miRNAs from peritoneal cells isolated from Stat1KO mice infected with γHV68. Data represent results from two or three independent experiments. Error bars are standard errors of the mean. (D) Quantitative PCR products from mock or infected 3T12 cells described for panel A were run on a 3% NuSieve GTG agarose gel. Sizes of markers are indicated on the left. miRNA-amplified products run with the 80- to 89-nt markers.
sent by over 9,700 reads (Table 3). This suggests that miR-M1-2 may be misannotated and the actual miRNA is on the 5' arm, and the 3' sequence, with lower read abundance, is actually miRNA*.

Confirmation of candidate γHV68 miRNAs. Putative novel miRNAs were confirmed using both Northern blot analysis and RT-PCR. Six of the 10 novel candidate γHV68 miRNAs could be detected by Northern blot analysis from 3T12 cells (Fig. 3), and those that could not, probably due to low relative abundance, were also tested by RT-PCR (Fig. 4A). RT-PCR was performed on γHV68-infected 3T12 cells (Fig. 4A), bone marrow-derived macrophages (Fig. 4B), as well as directly ex vivo peritoneal cells of infected Stat1-deficient mice (Fig. 4C), demonstrating that the γHV68 candidate miRNAs were expressed during the course of in vivo infection. PCR products from 3T12 cDNA were run on a gel to confirm specific detection of miRNA products (Fig. 4D). This RT-PCR assay works by polyadenylating all RNA and adding a universal tag during the reverse transcription step. The subsequent PCR is performed using a universal primer and an miRNA-specific primer. Thus, both the mature miRNA sequence and the pre-miRNA can be detected using this method. In multiple cases for our candidate miRNAs, the predominant PCR product was larger than expected for an miRNA (as determined by detection of miR-M1-1 from γHV68 and miR-16 from mouse), suggesting that the PCR may be detecting pre-miRNA products. Even for miR-M1-1, the PCR assay amplified a small band corresponding to the mature miRNA, as well as a large band that could result from amplification of the pre-miRNA (Fig. 4D). It is possible that our predicted miRNA sequences are not exactly correct, and a different primer that is shifted by even 1 to 2 nt could amplify an miRNA-sized product. In the cases of miR-M1-10-5p, miR-M1-12, miR-M1-13a-5p, miR-M1-13a-3p, miR-M1-13b-5p, miR-M1-13b-3p, miR-M1-14-5p, and miR-M1-14-3p, we found the PCR detection method to be inconclusive. However, because of positive detection by Northern blot analysis, we can conclude that an miRNA-sized product exists for miR-M1-10-5p, miR-M1-10-3p, miR-M1-11, miR-M1-12, miR-M1-13a-5p, and miR-M1-13a-3p. miR-M1-13b and miR-M1-14 remain provisional, because we could not confirm their existence by Northern blot analysis or RT-PCR.

Mouse tRNA-small RNA polycistrons. Given the interesting structures of the tRNA-miRNA polycistrons in γHV68 and the assumption that viruses and their hosts utilize similar genetic mechanisms to regulate gene expression, we interrogated the sequencing data for possible mouse tRNA-small RNA polycistrons. Three tRNA-small RNA polycistrons were identified from the sequencing data (Fig. 5). Novel small RNAs mapped to regions just 3' of known tRNAs (Fig. 5) and could be folded into putative tRNA-miRNA polycistrons with structures similar to those in γHV68 (Fig. 6A). Interestingly, small RNA 717/209 mapped to two different chromosomal locations following two different tRNAs, tRNA717 and tRNA209 on chromosomes 1 and 14, respectively (Table 4; Fig. 6A). Small RNA 84 mapped uniquely to the region downstream of tRNA84. Notably, the tRNA717-small RNA is directly preceding a short interspersed element (SINE). Northern blot analysis using the probes shown in Fig. 6A confirmed the existence of a 20- to 22-nt RNA species, as well as a possible precursor of >100 nt for tRNA717 and tRNA209 and a putative tRNA-small RNA band of >130 nt for all three small RNAs, thus indicating the existence of larger transcripts that could include a pre-miRNA hairpin (Fig. 6B). Probes just for the miRNA sequence gave similar results (data not shown). The longer probes for tRNA84 and tRNA209 should bind uniquely to the structures in Fig. 6A, but we were unable to design a unique probe for tRNA717. The small RNA sequence for the putative miRNA is the same as that in tRNA209, and the sequence directly after the small RNA is part of a SINE; therefore, the long probe for the tRNA717 hairpin may have cross-hybridized to tRNA209-small RNA or to a SINE. Finally, while the small RNAs were identified from our sequencing data on infected cells, Northern blot analysis showed that these RNAs are present in both uninfected and infected 3T12 cells but not in the negative-control monkey cell line Vero.

FIG. 5. Identification of novel tRNA-small RNA polycistrons in mouse. Sequencing data from Table 2 were also analyzed for mouse small RNAs associated with tRNAs. Three novel small RNAs (arrow) immediately downstream of tRNA717 on chromosome 1 (A), tRNA84 on chromosome 13 (B), and tRNA209 on chromosome 14 (C) were identified.

DISCUSSION

Using Illumina deep sequencing of γHV68-infected 3T12 cells, we identified 10 novel candidate miRNAs, six of which were confirmed by Northern blot analysis. These novel miRNAs emanate from the same Pol III transcripts previously identified to generate the 10 known γHV68 miRNAs. These
data increase the number of γHV68 miRNAs substantially and further define the unusual γHV68 tRNA-miRNA structures that generate pre-miRNA stem-loops from larger Pol III transcripts. While the previous work to identify γHV68 miRNAs was done using a cDNA cloning method (20), our data using deep sequencing technology provide a more extensive picture of miRNA expression from the γHV68 genome. Moreover, this technique allowed us to examine small RNA expression from the mouse genome in parallel and to identify similar tRNA-small RNA polycistrons in mouse.

**TABLE 4. Novel mouse tRNA-associated small RNAs**

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Sequence</th>
<th>No. of reads</th>
<th>Associated tRNA(s)</th>
<th>Confirmation</th>
</tr>
</thead>
<tbody>
<tr>
<td>tRNA717/209 small RNA</td>
<td>UUGGACUCUACUCUCUUUA</td>
<td>278</td>
<td>tRNA717 and -209</td>
<td>Northern blot analysis</td>
</tr>
<tr>
<td>tRNA84 small RNA</td>
<td>UCUGACACAGUGCGUUUUU</td>
<td>717</td>
<td>tRNA84</td>
<td>Northern blot analysis</td>
</tr>
</tbody>
</table>
One interesting conundrum revealed by this sequencing data is the incongruity between previous miRNA annotation using different methods and miRNA identification using deep sequencing. As discussed above, one of the main criteria for prediction of an miRNA, after position on a predicted stem-loop, was relative sequence abundance. We and others (21, 24) have used the peak read within the arm of a stem-loop to identify an miRNA. However, when we applied that criterion to already-annotated miRNAs from the γHV68 genome, we found that only two (miR-M1-7-5p and miR-M1-7-3p) actually met that standard. All of the rest were either slightly shifted off the most abundant sequence or, in the case of miR-M1-2, on the complete opposite arm of the stem-loop. Similar results were observed for Illumina sequencing of embryonic stem cells (15), and these sequence variations from the “reference” sequence were termed isomiR sequences. It is possible that isomiR sequences can vary across tissue or cell type (the original clones were from a B-cell line, and our results are from fibroblasts) and that 3’ modifications and RNA editing represent differences in cell type or time point of infection. Along the same lines, miR-M1-2 could be biologically active in latently infected B cells but otherwise represents the inactive miR-M1-2* in latently infected cells (17). isomiR sequences may also derive from variability in a Dicer cleavage site. This raises the question of whether the current γHV68 annotations are actually representative of the most prevalent isomiR sequence. In the cases of miR-M1-10 and miR-M1-13a, we were able to detect both the 5’ and 3’ arms by Northern blot analysis, which were similar to results from extensive sequencing of KSHV that revealed abundant miRNA* sequences that were detectable by Northern blot analysis, along with their “non-star” miRNAs (10). In the case of KSHV, some miRNA* sequences were found to be biologically active. Further work will be required to determine if the same is true for all the derivatives of γHV68 miRNAs.

Also important to note is the relationship between read count and relative abundance. Our data demonstrate that read abundance does not always correlate with actual abundance of a transcript or with ease of detection by other methods (e.g., Northern blotting). For example, miR-M1-11 and mir-M1-13a-3p were present at a relatively low read frequency but were readily detected by Northern blot analysis. On the other hand, other candidate miRNAs, such as mir-M1-12 and mir-M1-13a-5p, have greater numbers of reads associated with them but are barely detectable by Northern blot analysis (Table 2 and Fig. 3). While there could be a difference in probe affinity and labeling, this was reproducible with different preparations of probes. This indicates that read number alone is not necessarily a quantitative measure of miRNA abundance.

In addition to highlighting the abundance of miRNAs emanating from the γHV68 genome, these sequencing data also reveal the possible relationship between the γHV68 tRNA-miRNA structures and similar structures in mouse. While we have not shown whether the mouse small RNAs identified in Fig. 5 to 7 are miRNAs or some other small RNA, such as a tRF (9), the structures predicted for these tRNA-small RNAs are strikingly similar to those for the γHV68 miRNAs. Moreover, some tRFs are thought to be processed by tRNAse Z (9) from a Pol III transcript consisting of a tRNA-tRF, similar to the γHV68 tRNA-miRNAs (1). The mouse tRNA-small RNAs we found all have Pol III termination signals (stretches of multiple U’s) at the end of the small RNAs. However, they also have subsequent termination signals, suggesting that a larger transcript could exist. Our Northern blot analysis confirms that both small and large transcripts are produced (Fig. 6B). Further work will be required to determine the function of these small RNAs and their relationship to the viral transcripts. Moreover, additional investigation into the possible presence of these types of transcripts in other viruses and hosts is under way.

ACKNOWLEDGMENTS
This work was supported by the ARRA administrative supplement MRCE 2 U54 AI057160 (H.W.V. and W.Z.), NIH grant AR058681 (W.Z.), and NSF grant DBI-0743797 (W.Z.). T.A.R. is supported by the Damon Runyon Cancer Research Foundation.

REFERENCES


