A combined computational and microarray-based approach identifies novel microRNAs encoded by human gamma-herpesviruses

ADAM GRUNDHOFF,[1,3] CHRISTOPHER S. SULLIVAN,[2,3] and DON GANEM[2]

[1]Heinrich-Pette Institut für experimentelle Virologie und Immunologie an der Universität Hamburg, D-20251 Hamburg, Germany
[2]Howard Hughes Medical Institute, Departments of Microbiology and Medicine, and G.W. Hooper Foundation, University of California, San Francisco, California 94143-0552, USA

ABSTRACT

We have developed an approach to identify microRNAs (miRNAs) that is based on bioinformatics and array-based technologies, without the use of cDNA cloning. The approach, designed for use on genomes of small size (<2 Mb), was tested on cells infected by either of two lymphotropic herpesviruses, KSHV and EBV. The viral genomes were scanned computationally for pre-miRNAs using an algorithm (VMir) we have developed. Candidate hairpins suggested by this analysis were then synthesized as oligonucleotides on microarrays, and the arrays were hybridized with small RNAs from infected cells. Candidate miRNAs that scored positive on the arrays were then subjected to confirmatory Northern blot analysis. Using this approach, 10 of the known KSHV pre-miRNAs were identified, as well as a novel pre-miRNA that had earlier escaped detection. This method also led to the identification of seven new EBV-encoded pre-miRNAs; by using additional computational approaches, we identified a total of 18 new EBV pre-miRNAs that produce 22 mature miRNA molecules, thereby more than quadrupling the total number of hitherto known EBV miRNAs. The advantages and limitations of the approach are discussed.

Keywords: microRNA; miRNA; herpesvirus; KSHV; EBV

INTRODUCTION

MicroRNAs (miRNAs) are small, noncoding, ~22-nt RNAs that bind directly to mRNA targets and inhibit their ability to express protein products (for review, see Du and Zamore 2005; Hammond 2005; Kim 2005). In animals, miRNAs are predominantly thought to silence gene expression by binding with imperfect base-pairing to the 3′-UTR of target RNAs, preventing their translation, and reducing their steady-state levels by a mechanism that is still not completely understood (Bagga et al. 2005; Lim et al. 2005). Additionally, miRNAs can bind to a target mRNA with perfect complementarity, thereby directing cleavage of the target in an analogous fashion to siRNA-mediated cleavage (Hutvagner and Zamore 2002; Zeng et al. 2003; Mansfield et al. 2004; Yekta et al. 2004).

miRNAs are initially transcribed as a primary precursor molecule (pri-miRNA) that can be several hundred to thousands of nucleotides long. All pri-miRNAs contain at least one characteristic hairpin structure; these hairpins are recognized by a protein complex containing the nuclear RNase enzyme Drosha and processed into ~70-nt hairpin intermediates (pre-miRNAs) (Lee et al. 2002, 2003; Denli et al. 2004; Gregory et al. 2004; Han et al. 2004; Landthaler et al. 2004; Zeng et al. 2005). The pre-miRNA is then exported to the cytoplasm (Yi et al. 2003; Bohnsack et al. 2004; Lund et al. 2004; Zeng and Cullen 2004), where the hairpin structure is recognized by a complex containing the RNase Dicer, which processes the pre-miRNA into the mature effector ~22-nt miRNA molecule (Bernstein et al. 2001; Grishok et al. 2001; Hutvagner et al. 2001; Ketting et al. 2001; Chendrimada et al. 2005; Forstemann et al. 2005; Gregory et al. 2005; Jiang et al. 2005; Saito et al. 2005).

For the most part, miRNAs have been identified by a modified version of the Rapid Amplification of cDNA Ends (RACE) protocol (Elbashir et al. 2001; Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001). Total RNA is size-fractionated by denaturing polyacrylamide gel electrophoresis, excised from the gel, and eluted, and then

[3]These authors contributed equally to this work.

Reprint requests to: Adam Grundhoff, Heinrich-Pette Institut für experimentelle Virologie und Immunologie an der Universität Hamburg, Martinistrasse 52, D-20251 Hamburg, Germany; e-mail: adam.grundhoff@hpi.uni-hamburg.de; fax: +49-40-48051-296.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.2326106.

oligonucleotides are ligated onto the 5′- and 3′-ends of the RNA. Ligated products are then reverse-transcribed, amplified by PCR, digested with a restriction enzyme, concatemerized, and sequenced. This approach has been enormously successful; in fact, the majority of known miRNAs were identified using this technique (miRBase [http://microrna.sanger.ac.uk/sequences/index.shtml]; see Griffiths-Jones 2004). However, there are some limitations to this approach. Cloning is laborious, and the number of amplicons that must be sequenced to saturate the screen is high (i.e., >1000). Therefore, miRNAs of low abundance may be missed. Additionally, it has been suggested there may be biases in the sequence composition of cloned miRNAs (Elbashir et al. 2001). Finally, cloning requires hundreds of micrograms of total RNA starting material, which is not practical in cases where the amount of sample tissue is limiting. For these reasons, we set out to develop a methodology to identify miRNAs that does not rely on cloning, with the particular goal of discovering new virally encoded miRNAs.

Several computational algorithms have been developed that take advantage of the characteristic hairpin structure found in pri-miRNA and pre-miRNA precursor structures to predict the existence of miRNAs from primary sequence (for review, see Bentwich 2005; Brown and Sanseau 2005). However, the majority of these approaches use evolutionary conservation as a preliminary filter to limit the number of candidate miRNAs called by the program (Grad et al. 2003; Lai et al. 2003; Lim et al. 2003a,b; Berezikov et al. 2005). For many viruses, only very distant evolutionary orthologs are known; thus, for our purposes we required a methodology that does not rely on evolutionary sequence conservation. Recently, Pfeffer et al. (2005) developed a program that predicts miRNAs without the need for an evolutionary conservation filter. This program has been successful in identifying miRNAs in viral genomes when candidates are confirmed by cloning.

We have developed a strategy that combines computational prediction and microarray analysis to identify candidate miRNAs encoded by viruses, with Northern blotting used as a confirmatory test. (A similar approach has recently been used by Bentwich et al. [2005] for identification of miRNAs encoded by the human genome, although their confirmatory test involved cDNA cloning.) The computational algorithm we use (VMir) is an updated version of our earlier reported miRNA prediction algorithm (virMir; Sullivan et al. 2005). Because we have the added filter of considering those candidate miRNAs that score positive on a microarray, our method has the advantage of being able to look at all possible predicted hairpins, not just the ones that score highly in the VMir algorithm. Additionally, because we use Northern blotting rather than cDNA cloning to confirm candidate miRNAs, our approach eliminates any chance of cloning biases.

We have used our approach to analyze miRNA production by two human gamma-herpesviruses: the Kaposi's sarcoma-associated herpesvirus (KSHV) and Epstein-Barr virus (EBV). Both viruses have been previously found by the cDNA cloning method to produce miRNAs. Three groups have recently reported extensive cloning analyses of KSHV miRNAs (Cai et al. 2005; Pfeffer et al. 2005; Samols et al. 2005); using our method, we were able to identify most of the KSHV miRNAs previously identified by cloning. Additionally, even though this virus was already extensively mined for miRNAs, we identify a new, previously undetected KSHV pre-miRNA, bringing the total number of known KSHV pre-miRNAs to 12. Moreover, we describe 22 new miRNAs encoded by 18 novel EBV pre-miRNAs, thereby increasing the number of known EBV-encoded miRNAs by a factor of more than five. The success of identifying new miRNAs in these select gamma-herpesviruses suggests that our approach could be applicable to other organisms with small genome sizes (up to ~2 Mb).

RESULTS

Refinement of VMir, a program to predict viral miRNAs

Recently, we developed a computer algorithm (virMir) to predict virally encoded miRNAs and successfully applied the program to identify miRNAs encoded by Simian Virus 40 (Sullivan et al. 2005). In order to allow the screening of larger viral genomes, we have further refined the program. The refined version, called VMir, features an updated scoring algorithm (see Materials and Methods for details) as well as the incorporation of several quality filters designed to reduce the complexity of the prediction. The program slides a window of predefined size over the viral genome, advancing each window by a given step size (the default values are 500 nt for window size and 10 nt for step size). The secondary structures of RNAs corresponding to each window are predicted using the RNAfold algorithm (Hofacker et al. 1994), and individual hairpin structures are detected. Hairpins with a size above a given cutoff value (default value 45 nt) are scored according to structural features and resemblance to known miRNA precursors (see Materials and Methods). The position, structure, and score of the hairpins as well as the individual sequence windows in which they can be detected are recorded.

After completion of the prediction, the recorded hairpins are compared to one another and categorized in order to reduce the number of potential candidates. Hairpins that form at the same genomic location and are of identical structure within a core region of maximally 85 nt (calculated from the terminal loop) are considered local variations of one another. The largest hairpin within such a collection is designated the Main Hairpin (MHP), while all others are considered partial variations thereof and are designated Subsidiary Hairpins (SHPs). Likewise, hairpins of identical sequence and structure but different genomic

location are considered Repeated Hairpins (RHPs) of a given MHP. The results are written to a file, and the hairpins can subsequently be visualized using a graphical user interface (GUI) and filtered according to various properties such as category, size, score, and minimal number of sliding windows (hereafter referred to as "window count") in which a given hairpin is detected.

VMir analysis of the genomes of KSHV and EBV

We first applied the VMir program to the genome of Kaposi's sarcoma-associated herpesvirus (KSHV). Without further filtering, the program initially detected a total of 3046 MHPs (for simplicity, we refer to these MHPs as "hairpins" in the following). Filter values for minimal scores and window counts were set to 115 and 35, respectively. (These values were selected so that 98% of a reference set of 175 known human miRNA precursors passed the filters [data not shown].) Of the initially selected hairpins, 267 passed the window filter, and 146 of these achieved a score of 115 or above. Figure 1A shows the locations and VMir scores for these hairpins. While the

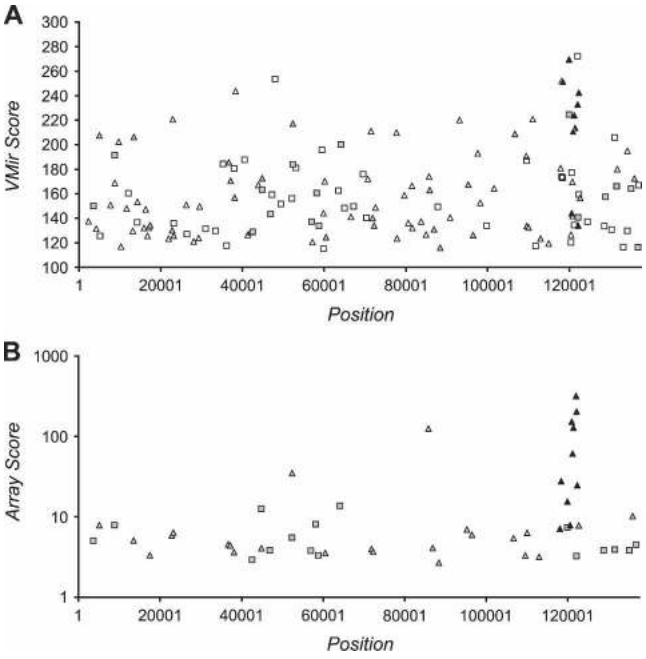

FIGURE 1. (*A*) VMir analysis of the KSHV genome; shown are all hairpins that fold in 35 or more windows and achieved a VMir score of 115 or above. Hairpins are plotted according to genomic location and VMir score. (*B*) Array analysis of potential KSHV-encoded pre-miRNAs. Only hairpins scoring under stringent conditions are shown, plotted according to position and the score achieved in the array analysis (see text for details). Positions indicated on the *X*-axis refer to the analyzed sequence that represents the complete long unique region of KSHV and was assembled from GenBank entries KSU86667, U93872, and KSU85269 (see Materials and Methods for details). (Squares) Hairpins in direct orientation; (triangles) hairpins in reverse orientation; (filled symbols) hairpins that scored in the array analysis (gray) and hairpins representing confirmed pre-miRNAs (black).

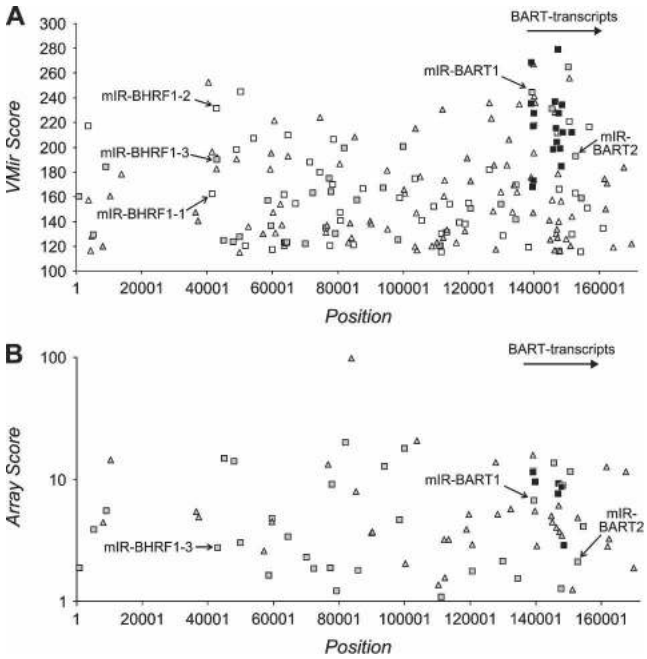

FIGURE 2. (*A*) VMir analysis of the EBV genome; shown are all hairpins that fold in 35 or more windows and achieved a VMir score of 115 or above. Hairpins are plotted according to genomic location and VMir score. (*B*) Array analysis of potential EBV-encoded pre-miRNAs. Only oligonucleotides with normalized ratios of 1 or above were considered in the array analysis, and only hairpins that scored under these conditions are shown, plotted according to position and the score achieved in the array analysis (see text for details). The five pre-microRNAs previously identified by Pfeffer et al. (2004) are marked (we did not attempt to confirm these miRNAs by Northern blotting). Positions given on the *X*-axis refer to the analyzed sequence representing the complete wild-type EBV genome (GenBank entry AJ507799). (Squares) Hairpins in direct orientation; (triangles) hairpins in reverse orientation; (filled symbols) hairpins that scored in the array analysis (gray) and hairpins representing confirmed pre-miRNAs (black).

candidates are overall widely dispersed across the viral genome, a cluster of eight high scoring hairpins (with scores between 211 and 269) in reverse orientation, located between nucleotides 118,000 and 122,400, was immediately apparent.

Next, we analyzed the prototypical wild-type genome of EBV (GenBank entry AJ507799). This sequence represents the complete genome of the B95-8 strain; however, the B95-8 strain has a deletion removing part of the BART (Bam-H1-A rightward transcript) region; this was repaired using the corresponding sequence from the Raji strain. Analysis of the EBV genome was carried out using the same filter settings as for KSHV. Out of a total number of 3479 individual hairpins detected by VMir, 322 passed the filter for minimal window counts and 205 of these in addition had a VMir score of <115 (Fig. 2A). Previously, Pfeffer and colleagues (Pfeffer et al. 2004) had identified five EBV-encoded miRNAs via cloning of small RNAs from B95-8-infected cells. Three of these miRNAs (miR-BHRF1-1

through -3) are located within the 5′- and 3′-untranslated regions (UTRs) of the BHRF1 (Bam H1 rightward open reading frame 1) transcript, while the remaining two (miR-BART1 and -2) map to introns of the BART transcripts. In our VMir analysis, the pre-miRNA hairpins encoding these miRNAs achieved scores ranging between 162 and 245, corresponding to ranks 8 and 97 in the VMir prediction for the lowest (MD254, corresponding to miR-BHRF1-1) and highest (MD1449, corresponding to miR-BART1) scoring hairpins, respectively (see Fig. 2A). As for KSHV, the majority of the predicted hairpins in EBV appeared dispersed across the genome (Fig. 2A). (Note that the region seemingly devoid of hairpins [nucleotide positions 12,001–35,355] represents the major internal repeat region 1 [IR1] of the EBV genome. Only the first representative hairpin is shown, while additional repeats of this hairpin were classified as RHPs and filtered out.) However, closer examination revealed a cluster of 23 hairpins with VMir Scores of 180 or above located approximately between nucleotide positions 139,000 and 157,000. While these hairpins are observed in a similar genomic location as the high scoring KSHV cluster, they are in opposite, i.e., direct, orientation. The cluster comprises 62% of all hairpins in direct orientation with a score of 180 or higher, concentrated in a segment representing 10% of the viral genome. Interestingly, the cluster maps to the region of the BART transcripts (indicated in Fig. 2A) and includes pre-miR-BART1 and BART2, with 17 of the hairpins located within the segment deleted in the B95-8 strain of EBV. A detailed representation of the hairpins within the cluster is shown in Figure 7, below.

Confirmation of KSHV pre-miRNAs by microarray analysis and Northern blotting

In order to confirm candidate miRNA precursors predicted by VMir, we performed a microarray analysis using custom arrays that contain 12,000 oligonucleotides directly synthesized on the chips. Since we did not want to limit our analysis to the top candidates from our VMir prediction, we also included lower scoring hairpins as well as hairpins folding in <35 windows in the array design. Accordingly, oligonucleotides complementary to the full set of 3046 hairpin structures initially detected by the folding algorithm were selected. The oligonucleotides had a maximum size of 50 nt and were designed such that each hairpin was represented by at least three overlapping oligonucleotides: two oligonucleotides covering the left and the right arms of the hairpin (extending from the apex position of the terminal loop in the 5′- and 3′-directions; for longer hairpins, additional oligonucleotides with a minimum overlap of 25 nt were tiled along the arms) and one oligonucleotide centered across the terminal loop. We reasoned that this design would allow us to distinguish between miRNAs originating from the 5′-arm or 3′-arm of a given hairpin. For control purposes, we also selected oligonucleotides in the identical manner for the 175 known human miRNA precursors that were represented in the microRNA database (miRBase [http://microrna.sanger.ac.uk/sequences/index.shtml]; see Griffiths-Jones 2004) at the time. As additional controls for random degradation, the array also contains oligonucleotides tiled across coding regions of several housekeeping genes (GAPDH, β-actin, α-tubulin) as well as 18S ribosomal RNA.

In addition to the array described above (referred to as the "hairpin array" hereafter), we also designed a second array independent of our VMir prediction. This array (termed the "tile array") contains 50-mers tiled across the complete coding region of the KSHV genome in both forward and reverse orientations, with an overlap of 25 nt between adjacent oligonucleotides. Additionally, the tile array also carries the same control oligonucleotides as the hairpin array. In pilot experiments, we hybridized both arrays with size-selected RNA (average size 20 nt) from the EBV/KSHV negative Burkitt's lymphoma cell line, BJAB. While these experiments showed that the arrays generally allowed the identification of endogenous miRNAs, we also observed a direct correlation between the GC content of the oligonucleotides and the intensity of hybridization signals (data not shown). Given the small size of the hybridized material, this relationship was to be expected; in subsequent experiments, we therefore analyzed only ratios between our sample material and appropriate controls (see the following paragraph). In addition, oligonucleotides that showed a GC content of 85% or higher within any window of 21 nt shifted along the oligonucleotide were excluded from our analysis.

To detect miRNAs encoded by KSHV, we analyzed small RNAs isolated from the latently infected PEL cell line BCBL1. As reference/negative control material for both arrays, we used two different samples, for a total of four different experimental setups. Since B-cells are refractory to infection with KSHV in vitro, we were not able to derive samples from an isogenic control (i.e., from a B-cell line before and after infection with KSHV). In one set of experiments, we therefore hybridized the (KSHV-positive) BCBL1 samples together with RNAs of the same size isolated from (KSHV-negative) BJAB cells. In another set of experiments, two different size-selected samples were derived from the same BCBL1 material and hybridized to the arrays: one with an average size of 20 nt and the other with an average size of 30 nt. We reasoned that the main source of background hybridization in our experiments derives from random RNA degradation products, which are abundant in the typical size range of miRNAs. However, in contrast to the miRNAs enriched in the ~20-nt range, such cross-hybridizing degradation products should be equally abundant in fractions of larger size as well and therefore should result in approximately equal hybridization signals

for both samples. In the following, we refer to the two different hybridization schemes described above as BCBL1 versus BJAB and 20-mer versus 30-mer, respectively.

Primary analysis of the hybridized arrays was carried out using a standard software package (Genepix). To allow comparison between different array experiments, the obtained ratios were normalized to a mean value of 0 and unit standard deviation (thus, normalized values of 0 and 1 represent ratios that were exactly the mean of all oligonucleotide ratios across the array or onefold of the standard deviation above the mean ratio, respectively; see Materials and Methods for details). Based on the array results, we then calculated scores for all of the hairpins that had passed our VMir filter criteria. For this purpose, oligonucleotides were first matched to the various hairpins. In order to be considered for a given hairpin, an oligonucleotide had to show an overlap of at least 15 nt. (For the hairpin arrays, this condition is necessarily met by all oligonucleotides specifically designed for that hairpin; however, oligonucleotides designed for a neighboring hairpin could also potentially be considered since many adjacent hairpins folding in different windows showed significant overlap.) Because miRNAs are generally located directly adjacent to the terminal loops of the precursor structures, oligonucleotides that were >10 nt away from the terminal loop were not further considered for the analysis of that hairpin. For the remaining oligonucleotides, the normalized ratios from the hybridization data were retrieved and oligonucleotides with normalized ratios above a certain threshold (1.75; corresponding to a ratio 1.75-fold above the standard deviation across the array) were selected. To exemplify the general concept of array design and analysis, Figure 3 shows the text-based output for one of the confirmed miRNA precursors, based on the analysis of hairpin and tile arrays hybridized with a 20-mer versus 30-mer sample. As can be seen, oligonucleotides mapping to the right arm of the hairpin showed strong hybridization signals (between 12.7-fold and 34.2-fold above the standard deviation) in both arrays. In addition, signals, albeit much weaker, were also detected for two oligonucleotides mapping to the left arm of the hairpin: MR2892_L on the hairpin array and 04928_F on the tile array, with normalized ratios of 1.1 and 2.1, respectively. (Due to the ratio cutoff of 1.75, however, oligonucleotide MR2892_L was not included in the analysis.)

To rank our candidates according to their probability of encoding miRNAs, we assigned each hairpin a score based on a combination of the array results and VMir scores. For the calculation of this score, we first normalized the VMir scores to a median value of 1. For each hairpin, we then calculated an array score consisting of the sum of ratios from all oligonucleotides with ratios above the cutoff of 1.75 (70.53 in the case of the example shown in Fig. 3). This value was then multiplied with the normalized VMir score to calculate the final score. Since the highest scoring hairpin of our prediction achieved a normalized VMir score of 2.9, the array scores could be maximally tripled by an outstanding VMir score. No multiplication was performed if the normalized VMir score for a given hairpin was <1; in other words, the value calculated from the array hybridizations could only be improved for hairpins with a VMir score above, but not decreased by a VMir score below the median value.

In Figure 1B, we present the results of our analysis from the four different array hybridizations performed with RNA from KSHV-positive BCBL1 cells. The scores were calculated as described above and plotted on a logarithmic scale against the genomic location of the hairpins. While a total of 50 hairpins achieved positive scores, 10 of the top 20 candidates mapped to the region of the cluster recognized during our initial analysis of the VMir prediction.

To elucidate whether the observed array signals were indeed due to the hybridization of viral miRNAs, we performed Northern blotting experiments with RNA from KSHV-positive BCBL1 and KSHV-negative BJAB cells for all of the 50 scoring hairpins. Probes were designed starting at the boundaries of the terminal loop and extending 35 nt into the left/5p or the right/3p arm of the hairpins. The 5p, the 3p, or both arms were chosen based on the location of oligonucleotides scoring in the arrays, as shown in the textual output presented in Figure 3. Bands within the size range expected for mature miRNAs were observed for eight of the 10 clustered hairpins that were in the top 20 of our array analysis (see Fig. 4, blots labeled MR2894, -2892, -2887, -2883, -2868, -2847, -2885, and -2844). For two of these hairpins (MR2892 and -2868), we detected miRNAs from the 5p as well as the 3p arm. With the exception of one hairpin, MR2079, all other Northern blots were negative. Hairpin MR2079, located within an intron of the ORF29a/b transcript, achieved ranks 14 and 20, respectively, in our VMir and array analysis. While the size of the detected band is in perfect accord with a miRNA (see Fig. 4, MR2079), we also observed a faint signal of same or similar size in KSHV-negative BJAB cells (too weak to be seen in Fig. 4). We therefore cannot exclude the possibility that the signal results from cross-hybridization of a cellular miRNA with MR2079 sequences.

We next considered the possibility that additional pre-miRNAs were missed due to stringent filtering. We therefore reanalyzed our array data using less stringent criteria. In particular, we allowed hairpins folding in a minimum of 10 instead of 35 windows into the analysis and lowered the ratio cutoff value for the arrays from 1.75 to 1.5. Scores were calculated as described above, and the 40 top-ranking candidates from the resulting list were chosen for further consideration (data not shown). By performing Northern blots for the 19 candidates that had not already been analyzed during the previous experiment, we were able to identify an additional two pre-miRNAs (Fig. 4, blots labeled MR2896 and -2889). These two hairpins were

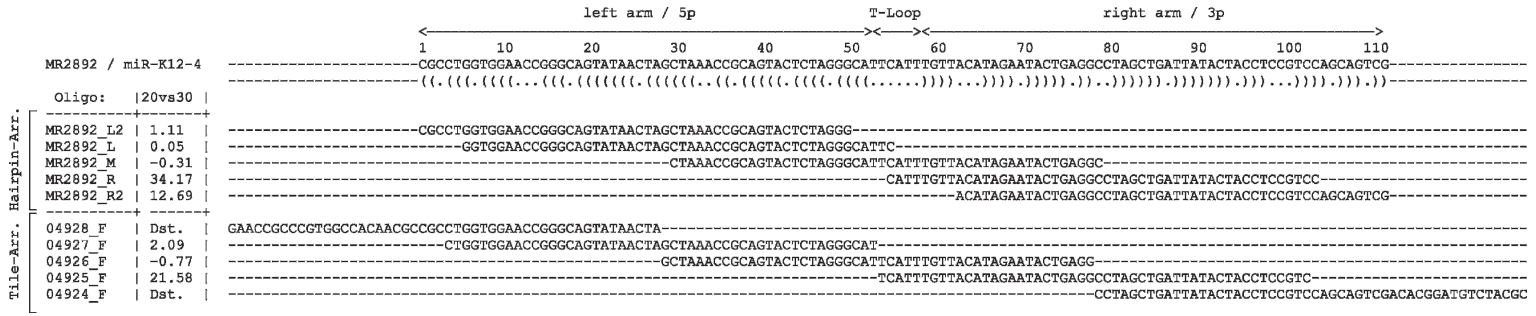

FIGURE 3. Example of the array analysis for one of the confirmed KSHV pre-miRNAs (MR2892/miR-K12-4), using data of the 20-mer vs. 30-mer sample hybridized to the hairpin and tile arrays. The hairpin structure is indicated *below* its sequence by brackets (paired nucleotides) and dots (unpaired nucleotides), and regions constituting *left* and *right* arms as well as the terminal loop are indicated *above*. The names of oligonucleotides from the hairpin and tile arrays are shown to the *left*. The normalized ratios observed for arrays hybridized to the 20-mer vs. 30-mer sample are shown to the *right* of the oligonucleotide names. Dst. is shown instead of ratios for oligonucleotides that are >10 nt away from the terminal loop; these oligonucleotides were not considered for the analysis of MR2892.

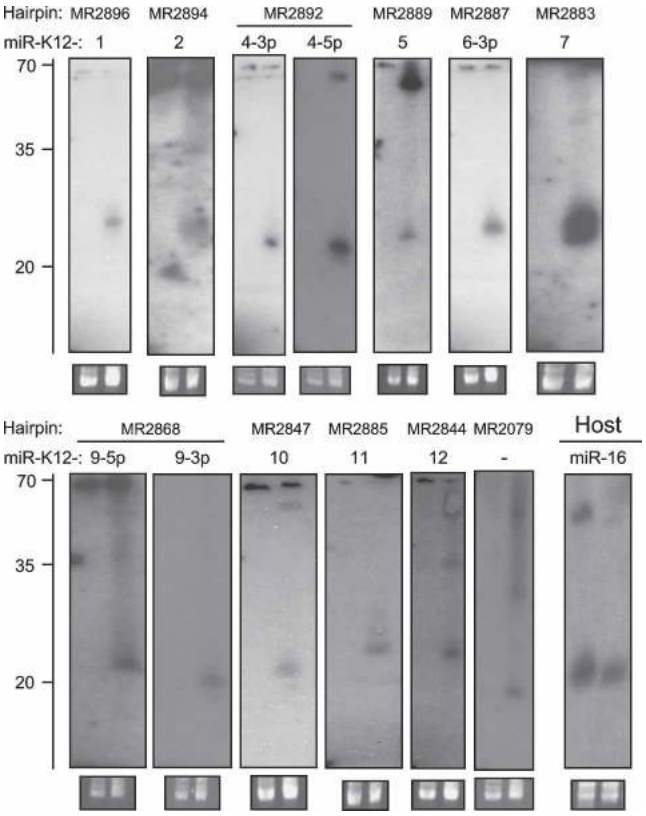

FIGURE 4. Northern Blot confirmation of KSHV miRNAs. Total RNA from KSHV-positive BCBL1 cells (*right* lanes in each blot) and KSHV-negative BJAB cells (*left* lanes in each blot) was probed with radioactively labeled oligonucleotides. Probes were selected based on the arrays analysis and correspond to the bold sequences in Figure 5. The names of the hairpins as detected by VMir are shown at the *top*, while names of the miRNAs are given *below*. No miRNA name was assigned to hairpin MR2079 since a weak signal (too faint to be seen in the figure) was observed in the BJAB lane (see text for details). The *lower right* blot served as a positive control and was probed for human miR-16. As a load control, an ethidium bromide stain of the low-molecular-weight RNA is shown *under* each blot. A ladder showing the approximate migration of oligonucleotides is shown to the *left*.

located in the same cluster as the eight pre-miRNAs identified during the first round of Northern blotting. Re-examination of the data revealed they had been initially missed because they fold only in 10 (MR2889) and 13 (MR2896) windows, respectively. Thus, our analysis detected a total of 12 miRNAs produced from 10 different pre-miRNAs. The data from our analysis are summarized in Table 1. In columns 2 and 3, the table shows the names of the hairpins predicted by the VMir program, along with their scores and ranks. The next six columns to the right show the ranks achieved by these hairpins during our array analysis. The first column gives the ranks from the initial analysis across all four arrays, using stringent filtering criteria. In the next column, the ranks for the second analysis using the relaxed filter settings are shown. To estimate the contribution of each of the four different arrays to the overall result, we also analyzed each of the arrays individually using the relaxed filter criteria; these data are shown in the next four columns.

As noted in the Introduction, three groups have reported the identification of KSHV-encoded miRNAs via the cloning of small RNAs (Cai et al. 2005; Pfeffer et al. 2005; Samols et al. 2005). Together, these groups identified a total of 11 different pre-miRNAs (each group individually identified 10). While nine of these pre-miRNAs were also identified by us, the remaining two represent the two candidates from the top 20 of our array analysis (MR2878 and -2893; ranks 17 and 2, respectively), which we were unable to confirm by Northern blotting. An additional pre-miRNA identified during our analysis, MR2844, was identified only in our study. The rightmost columns of Table 1 show the pre-miRNAs identified by the three groups together with the names assigned by each group. In the following, we refer to the pre-miRNAs in adherence to the naming scheme used by Pfeffer et al. (2005) as shown in the leftmost column of Table 1. Since MR2894 was not identified by Pfeffer and colleagues, we refer to this pre-miRNA as KSHV-miR-K12-2. Accordingly, the pre-miRNA exclusively identified during our study was named KSHV-miR-K12-12. In Figure 5, we show the predicted structures for the confirmed KSHV pre-miRNAs as well as the hairpin (MR2079) showing cross-hybridization (perhaps with a cellular miRNA) in KSHV-negative cells. The regions to which Northern probes were designed based on the array results are shown in bold. In addition, the positions of the miRNAs cloned by the three other groups are indicated by solid brackets. Since our method does not rely on cloning of miRNA sequences, we cannot precisely map the position of a mature miRNA within its precursor hairpin. For the novel KSHV-miR-K12-12 that was exclusively identified in our study but not cloned by the other groups, the position of the miRNA therefore can only be estimated; we have indicated this estimated position by the dotted bracket in Figure 5.

Figure 6 shows a map of the KSHV miRNA cluster within the KSHV genome. All pre-miRNAs map to a segment of ~4.5 kb that is located directly upstream of the major KSHV latency cluster consisting of ORFs 71, 72, and 73. The segment includes one open reading frame (ORF K12) and three repeat regions (DR3, DR4, and LIR′). Sequences between LIR1′ and ORF71, furthermore, can serve as an origin of replication during the lytic lifecycle of KSHV (Lin et al. 2003; AuCoin et al. 2004; Wang et al. 2004). Recently, a transcript that is likely to represent a pri-miRNA for the various pre-miRNAs has been identified (Li et al. 2002; Pearce et al. 2005; see Fig. 6, 2.5-kb LTd-transcript). The transcript is coterminal with a previously described mRNA transcript for Kaposin (Sadler et al. 1999); however, it starts several kilobases further upstream and encompasses all of the pre-miRNAs (as well as ORFs 71 and 72). While KSHV-miR-K12-1 through -9 as well as -11 are

TABLE 1. KSHV microRNAs

				Array hits (ranks)									
				Filter 1[a]	Filter 2[b]								
	VMir					Hairpin array		Tiled array		Pfeffer et al.[c]			
Pre-miR Name	Hairpin	Score	Rank[d]	All arrays	All arrays	20 vs. 30	BCBL1 vs. BJAB	20 vs. 30	BCBL1 vs. BJAB	Predicted	Cloned	Cai et al.	Samols et al.
miR-K12-1	MR2896	120	371*	na	30	—	—	14	51	np	miR-K12-1	miR-K1	KSHV-miRNA-1
miR-K12-2	MR2894	242	7	13	12	13	23	7	—	np	nc	miR-K2	nc
miR-K12-3	MR2893	134	102	2	2	2	8	3	6	np	miR-K12-3	miR-K3	KSHV-miRNA-2
miR-K12-4	MR2892	233	8	1	1	1	1	1	1	8	miR-K12-4	miR-K4	KSHV-miRNA-3
miR-K12-5	MR2889	181	81*	na	31	—	—	26	11	np	miR-K12-5	miR-K5	KSHV-miRNA-4
miR-K12-6	MR2887	214	15	4	4	4	3	6	8	4	miR-K12-6	miR-K6	KSHV-miRNA-5
miR-K12-7	MR2883	211	17	3	3	5	2	2	—	7	miR-K12-7	miR-K7	KSHV-miRNA-7
miR-K12-8	MR2878	144	84	17	40	21	—	35	—	np	miR-K12-8	miR-K8	KSHV-miRNA-8
miR-K12-9	MR2868	269	2	9	10	—	4	—	—	3	miR-K12-9	miR-K9	KSHV-miRNA-9
miR-K12-10	MR2847	251	5	8	8	8	—	5	—	np	miR-K12-10	miR-K10	KSHV-miRNA-10
miR-K12-11	MR2885	224	10	6	6	7	11	4	—	6	miR-K12-11	nc	KSHV-miRNA-6
miR-K12-12	MR2844	252	4	20	47	—	—	15	—	5	nc	nc	nc
	MR2079	217	14	7	7	6	14	—	—	np	nc	nc	nc

Rows 1–12, pre-microRNAs identified during our study; row 13, the candidate hairpin MR2079, which could not be unambiguously confirmed.
(na) Not analyzed, (np) not predicted, (nc) not cloned.
[a]Filter 1: Hairpins folding in <35 windows and oligonucleotides with normalized ratios <1.75 were not considered.
[b]Filter 2: Hairpins folding in <10 windows and oligonucleotides with normalized ratios <1.5 were not considered.
[c]Shown are the ranks of predicted (Predicted) and the names of confirmed (Cloned) pre-miRNAs from the studies of Pfeffer et al. (2004, 2005).
[d]Ranks are given according to a VMir prediction using Filter 1; hairpins marked with an asterisk (*) did not pass Filter 1, and the indicated ranks refer to a prediction using Filter 2.

FIGURE 5. Predicted hairpin structures of confirmed KSHV pre-miRNAs and hairpin MR2079. Hairpins longer than 100 nt were truncated; these hairpins are indicated by a double slash preceding the stem. (In the context of 500-nt windows, MR2896/KSHV-miR-K12-1 was detected only as a short [51-nt] hairpin; sequences shown in lowercase represent an extension of the hairpin obtained after folding of a smaller [80-nt] window centered on the hairpin.) The positions of the sequences within the KSHV genome are shown in brackets; all positions are given according to GenBank entry U93872. The regions complementary to which the oligonucleotide Northern probes were designed are shown in bold. The positions of mature miRNAs that were cloned by other groups (Cai et al. 2005; Pfeffer et al. 2005; Samols et al. 2005) are marked by solid brackets; dotted brackets indicate suspected miRNA positions (see text for details).

contained within the large intron of the transcript, KSHV-miR-K12-10 resides within the coding region of ORF K12. The novel KSHV-miR-K12-12 is located within the 3′-UTR, 265 nt away from the end of the transcript.

Identification of novel EBV-encoded pre-miRNAs

The experiments described above have allowed us to discover a novel miRNA that was not identified by several groups carrying out extensive cloning experiments of small RNAs from KSHV-infected cells. Only one study so far has attempted to clone viral miRNAs from EBV-infected cells (Pfeffer et al. 2004) and therefore, especially considering the cluster of high scoring hairpins within the BART region observed during our VMir analysis (see Fig. 7), we reasoned that EBV might encode additional miRNAs besides the five previously identified by Pfeffer and colleagues (2004). Accordingly, we designed a DNA microarray for all of the EBV hairpins predicted by VMir in the identical way as described above for KSHV. (Because the coding region of EBV is considerably larger than that of KSHV, we were not able to design a tiled array for EBV as well.) The hairpin array was hybridized with size-selected RNAs (~20 nt) from the EBV-positive Burkitt's lymphoma cell line Jijoye (Kohn et al. 1967). As for KSHV, we used two different reference samples: (1) size-selected RNAs of ~20 nt from EBV-negative BJAB cells (Jijoye vs. BJAB) and (2) RNAs with an average size of 30 nt from Jijoye cells (20-mer vs. 30-mer). The array analysis was carried in the same way and generally using the same filter parameters as described for our initial analysis of the KSHV arrays; however, since the dynamic range of the EBV arrays was overall lower (presumably due to a higher abundance of random degradation products) (data not shown), we decreased the cutoff value for normalized oligonucleotide ratios from 1.75 to 1. The results of the array analysis are shown in Figure 2B.

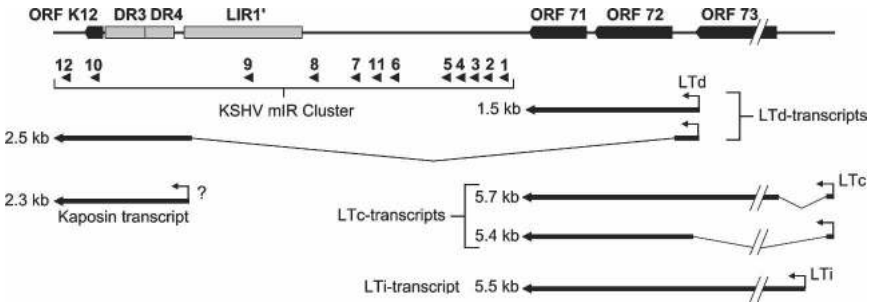

FIGURE 6. Map of the KSHV genome between ORFs K12 and 73. Open reading frames and repeat regions are shown as black and as gray boxes, respectively. The positions of the KSHV-pre-miRNAs are indicated by black arrowheads. Several latently expressed mRNAs are shown *below* (Dittmer et al. 1998; Sadler et al. 1999; Li et al. 2002; Matsumura et al. 2005; Pearce et al. 2005). The question mark indicates a 2.3-kb "Kaposin" mRNA (Sadler et al. (1999); it is unclear to what extent this transcript contributes to the overall abundance of Kaposin-encompassing transcripts in latently infected cells.

Compared to the KSHV analysis, the array results for EBV appeared generally noisier, and a high scoring hairpin cluster was not as easy to identify. Nonetheless, closer examination revealed that 10 out of the top 25 candidates mapped to the BART region. While one of these hairpins represented miR-BART1 (rank 25), the remaining nine had not previously been identified as pre-miRNAs (miR-BART2 and -BHRF1-3 achieved ranks 65 and 61, respectively; we could not detect signals for miR-BHRF1-1 and -2 during our initial analysis). To investigate whether our array analysis had, indeed, identified novel pre-miRNAs, we performed Northern blotting experiments for the top 24 candidates with probes designed as described above for KSHV (we did not perform a Northern blot for miR-BART1, which ranked 25th). Indeed, we could confirm expression of miRNAs for seven of the nine candidates mapping to the BART region (see top panel in Table 2 and corresponding Northern blots in Fig. 8). The Northern blots for the remaining two hairpins from the BART region as well as the other 15 hairpins were negative.

Because the signal:noise ratio in our EBV arrays was lower than in the KSHV arrays, we suspected that we might have missed additional pre-miRNAs with weak hybridization signatures. In the case of KSHV, an additional two pre-miRNAs were identified after lowering the stringency of our analysis filters. However, due to the higher noise levels, we deemed a similar approach less suitable for the EBV arrays. Instead, we decided to rely more on our VMir analysis and performed Northern blots for all hairpins with VMir scores of at least 150 that map to the region of the BART transcripts. Accordingly, we designed oligonucleotide probes for the 16 hairpins that had not already been analyzed during the first round of Northern blots. As before, probes were selected so that 35 nt from the terminal loop structures were covered. However, since we did not have information about the location of potential miRNAs, Northern blots for both arms of the hairpins were performed. Indeed, this second round of Northern blots confirmed an additional 11 hairpins as pre-miRNAs (middle panel in Table 2 and corresponding Northern blots in Fig. 8). One of the corresponding hairpins, MD1540, was also detected during our array analysis but had not been considered due to its low rank (rank 58) (see Table 2).

Thus, a total of 18 novel pre-miRNAs encoded by EBV were identified during our study. Since four of the pre-miRNAs produced mature miRNAs from both arms, the total of novel mature miRNAs is 22. In accordance with the scheme proposed by Pfeffer et al. (2004), we have named the pre-miRNAs miR-BART3 through -20. Figure 9 shows the predicted hairpin structures as well as suspected locations of mature miRNAs.

A summary of our analysis of the EBV genome can be found in Table 2. The table shows the novel pre-miRNAs identified during the first and second rounds of Northern blotting (top and middle panels, respectively) as well as the five pre-miRNAs previously identified by Pfeffer et al. (2004) (bottom panel), along with their VMir scores, VMir ranks, and the ranks achieved during the array analysis. (For comparison, we have also included additional lists based on individual analyses of the Jijoye vs. BJAB and 20-mer vs. 30-mer arrays.) Figure 10 shows a map of the BART region of the EBV genome. Seven of the newly discovered pre-miRNAs are found within close proximity of the previously described miR-BART1, while the remaining 11 are located ~6.5 kb further downstream. A total of 14 of the novel pre-miRNAs reside within the region deleted in the B95-8

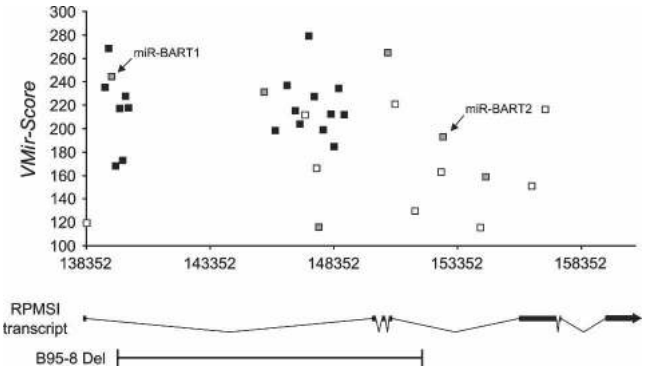

FIGURE 7. Detailed depiction of VMir-predicted hairpins within the BART region of EBV. Only hairpins predicted in the same (i.e., forward) orientation as the BART transcripts are shown. (Black boxes) Pre-miRNA hairpins that were confirmed by Northern blotting in this study (we did not attempt Northern blots for the previously identified pre-miRNAs miR-BART1 and -2). (Gray boxes) Hairpins that achieved positive scores during the array analysis but were not confirmed as pre-miRNAs by Northern blotting. The prototypical RPMS1 transcript as described by Smith et al. (2000) is shown *below*. The region deleted in the B95-8 strain of EBV is indicated.

TABLE 2. EBV microRNAs

	VMir			Array hits (ranks)			Pfeffer et al.[a]	
miR Name	Hairpin	Score	Rank	Both arrays	Jijoye vs. BJAB	20 vs. 30	Predicted	Cloned
miR-BART3	MD1446	235	13	17	12	18	5	nc
miR-BART4	MD1448	269	2	16	11	17	np	nc
miR-BART7	MD1456	173	80	18	26	13	n/a	
miR-BART12	MD1518	215	31	24	—	7	n/a	
miR-BART13	MD1521	204	41	19	34	9	n/a	
miR-BART16	MD1533	199	44	22	36	12	n/a	
miR-BART17	MD1535	212	33	21	23	16	n/a	
miR-BART5	MD1452	168	85	—	—	—	np	nc
miR-BART6	MD1453	217	29	—	—	—	n/a	
miR-BART8	MD1457	228	21	—	—	—	n/a	
miR-BART9	MD1458	218	27	—	—	—	n/a	
miR-BART10	MD1510	198	45	—	—	—	n/a	
miR-BART11	MD1515	237	10	—	—	—	n/a	
miR-BART14	MD1526	279	1	—	—	—	n/a	
miR-BART15	MD1528	228	22	—	—	—	n/a	
miR-BART18	MD1537	185	63	—	—	—	n/a	
miR-BART19	MD1540	234	15	58	—	40	n/a	
miR-BART20	MD1542	212	34	—	—	—	n/a	
miR-BART1	MD1449	245	8	25	20	36	np	c
miR-BART2	MD1591	193	51	65	—	50	np	c
miR-BHRF1-1	MD254	162	97	—	—	—	np*	c
miR-BHRF1-2	MD273	232	16	—	—	—	6	c
miR-BHRF1-3	MD275	190	56	61	—	43	4	c

First two panels, novel pre-miRNAs identified in our study; the first and second panels show the novel pre-miRNAs confirmed during our first and second rounds of Northern blotting, respectively. The five pre-miRNAs first identified by Pfeffer et al. (2004) are shown in the bottom panel.
[a]Shown are ranks of the pre-miRNAs in the computational prediction of Pfeffer et al. (2005) and whether the pre-miRNAs were cloned.
(np) Not predicted, (n/a) not applicable (the pre-miRNAs reside in a region deleted in the B95-8 genome analyzed by Pfeffer and colleagues), (*) a prediction was made for the complementary strand of miR-BHRF1-1, (nc) not cloned, (c) cloned.

strain of EBV, and an additional pre-miRNA (pre-miR-BART6) is partially affected by the deletion, explaining why these 15 pre-miRNAs had been missed in the study of Pfeffer et al. (2004). However, pre-miR-BART3, -BART4, and -BART5 precede the deleted region and thus are likely to be expressed in B95-8-infected cells. The BART locus can give rise to a multitude of differentially spliced transcripts originating from several promoters (Hitt et al. 1989; Gilligan et al. 1990, 1991; Chen et al. 1992, 2005; Brooks et al. 1993; Smith et al. 1993, 2000; Sadler and Raab-Traub 1995; Sugiura et al. 1996). The coding potential of many of these transcripts is uncertain; likewise, their potential to serve as pri-miRNAs remains unknown. In Figure 10, we have indicated the position and splicing pattern of the prototypical RPMS1 and RPMS1A transcripts according to the study of Smith et al. (2000). While miR-BART2 is positioned within the fourth, all of the remaining pre-miRNAs map to the first intron, and thus it appears possible that all miRNAs are expressed simultaneously from these transcripts.

DISCUSSION

Here we report the identification of several novel pre-miRNAs encoded by KSHV and EBV, using an approach that combines computational prediction and subsequent microarray analysis. Our prediction program (VMir) was refined from a previous version developed to identify pre-miRNAs in the genomes of SV40 and Polyomavirus (Sullivan et al. 2005; C. Sullivan, A. Grundhoff, R. Treisman, C. Sung, T. Benjamin, and D. Ganem, unpubl.); the refinements were designed to optimize the scoring algorithm itself as well as to make the program more suitable for the screening of larger viral genomes. Several algorithms for the identification of miRNAs have been published in the past (Grad et al. 2003; Lai et al. 2003; Lim et al. 2003a, 2005; Bentwich et al. 2005; Berezikov et al. 2005; Pfeffer et al. 2005). However, most methods rely on comparison of local sequence conservation, especially within noncoding regions, between closely related species. In such an analysis, pre-miRNAs are identified due to conservation of the hairpin structures, with a high degree of sequence homology typically observed in the region encoding the mature miRNA. In contrast, noncoding sequences surrounding the hairpin structures are often less conserved (Berezikov et al. 2005). While these approaches have been highly successful in the identification of microRNAs in the genomes of, for example, *Caenorhabditis elegans* or *Drosophila melanogaster*,

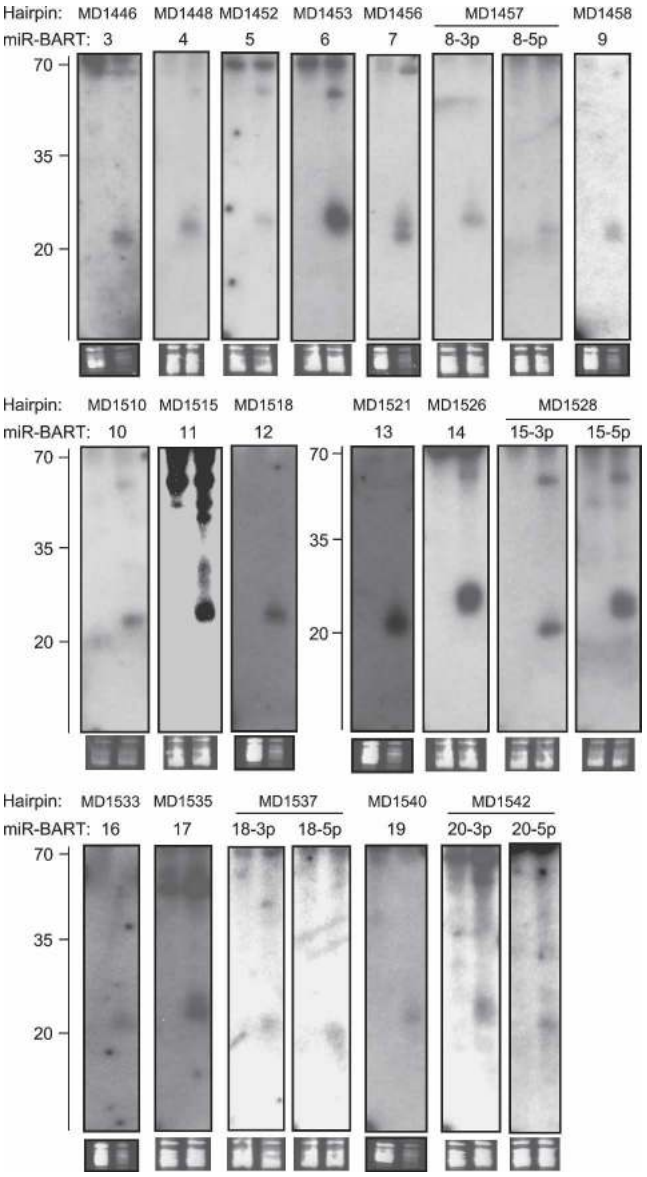

FIGURE 8. Northern Blot confirmation of EBV miRNAs. Total RNA from EBV-positive Jijoye cells (*right* lanes in each blot) and EBV-negative BJAB cells (*left* lanes in each blot) was probed with radioactively labeled oligonucleotides. Probes were selected based on the arrays analysis and correspond to the bold sequences in Figure 9. The names of the hairpins as detected by VMir are shown at the *top*, while names of the miRNAs are given *below*. As a load control, an ethidium bromide stain of the low-molecular-weight RNA is shown *under* each blot. A ladder showing the approximate migration of oligonucleotides is shown to the *left*.

they are less suited for the analysis of viral genomes. Since viruses are highly adapted to their specific host cell environment and also frequently capture host genes, even closely related viruses frequently show large sequence blocks with little or no homology. For example, among the human herpesviruses, KSHV represents the closest known relative of EBV; both viruses infect similar cell types and can establish a long-term latent infection within the memory B-cell reservoir. Nevertheless, none of the genes expressed during latent infection is conserved between KSHV and EBV, and each virus has captured its own set of host genes.

To our knowledge, besides VMir, only two programs not relying on comparison of cross-species conservation to identify pre-miRNAs have been described (Bentwich et al. 2005; Pfeffer et al. 2005). In general, all three programs have the common principle of comparing structural features of candidate hairpins (e.g., bulge size and symmetry, percentages of paired vs. unpaired nucleotides, etc.) against a reference set of known pre-miRNAs; however, they differ substantially in the way in which such factors are weighted. As neither of the two other algorithms is publicly available, it is difficult to perform a direct comparison with VMir. However, counting this work, KSHV has been subjected to four independent studies aiming at the identification of virally encoded miRNAs. Given this coverage, one can reasonably assume that the majority of KSHV miRNAs have been discovered and estimate the efficiency of the VMir prediction. Since VMir was primarily developed for the analysis of viruses (rather than organisms with large genomes) and we furthermore planned to perform a validation screen using microarray analysis, our algorithm was deliberately designed to over- rather than underpredict possible pre-miRNAs. Nevertheless, eight out of the 12 KSHV pre-microRNAs were among the top 20 candidates of the VMir prediction when using stringent filter criteria (see Table 1). For all practical purposes, however, the remaining four would have been missed had we based our study on the VMir prediction alone. (It is more difficult to judge VMir's efficacy in identifying EBV miRNAs since only one cloning study from EBV-infected cells has been reported to date; nonetheless, 11 of the currently known pre-miRNAs ranked within the top 30).

The algorithm developed by Pfeffer and coworkers (2005) was used to predict pre-miRNAs in the genomes of several viruses, including those of EBV and KSHV. The traditional cloning method was used to identify microRNAs expressed by these viruses and thus validate the prediction method; however, Pfeffer et al. (2005) did not attempt to confirm predicted pre-miRNAs for which no mature microRNAs were cloned. A total of eight potential miRNAs were predicted for KSHV, and five of these were confirmed via cloning experiments. Two of the five were also predicted in the opposite (i.e., direct) orientation (the hairpins antisense to pre-miR-K12-4 and -9; due to the partial inverted nature of the hairpin arms, high scoring hairpins are often predicted in both orientations; the respective hairpins were also detected in our prediction and achieved ranks 1 and 9, respectively). Interestingly, the remaining predicted (though unconfirmed) hairpin represents the novel KSHV-miR-K12-12 identified during our study (see Table 1). For EBV, a total of seven potential pre-miRNAs were predicted: Two of these are the previously reported miR-BHRF1-2 and –3,

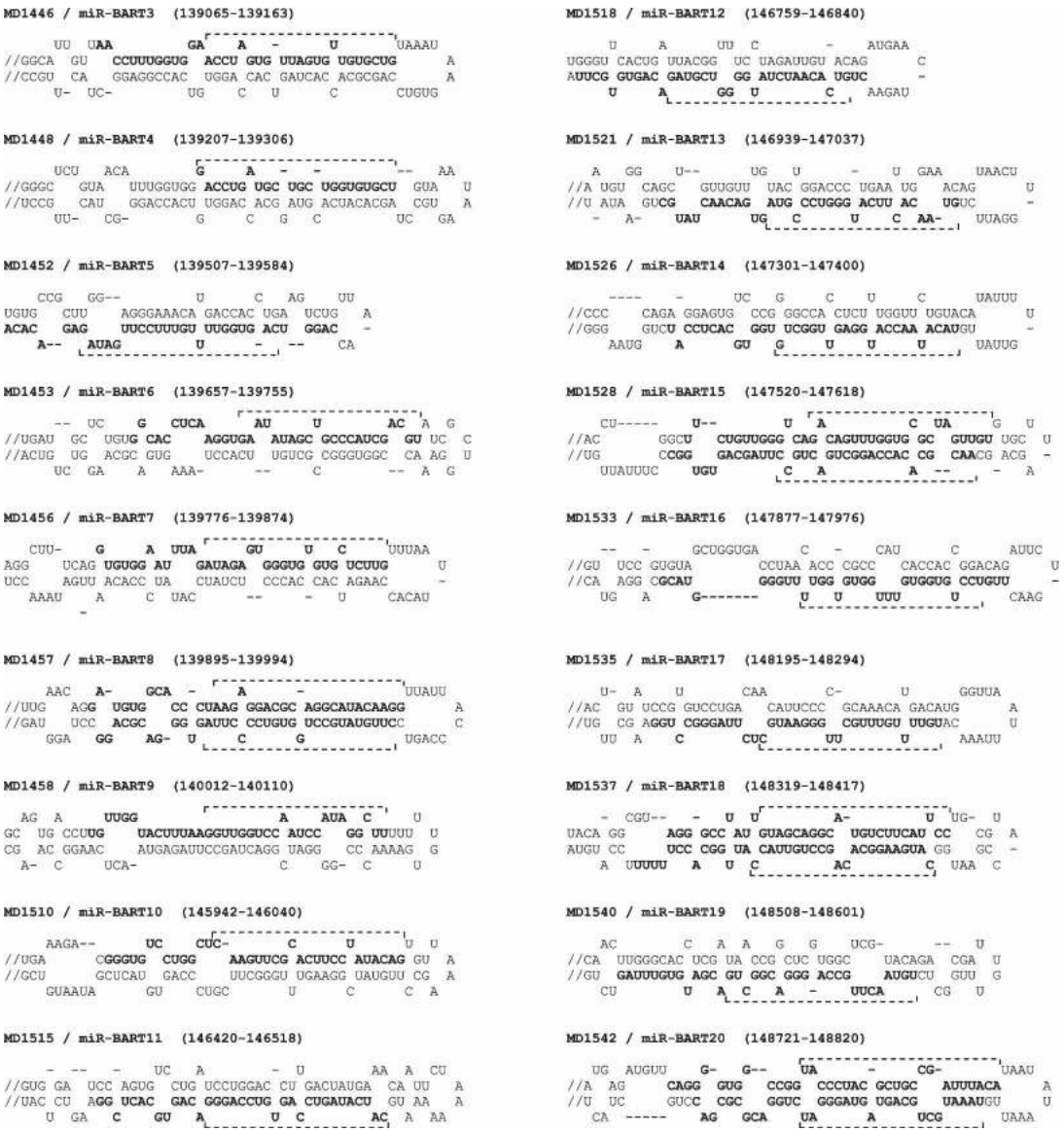

FIGURE 9. Predicted hairpin structures of novel EBV pre-miRNAs. Hairpins longer than 100 nt were truncated; these hairpins are indicated by a double slash preceding the stem. The positions of the depicted sequences within the EBV genome are shown in brackets; all positions are given according to GenBank entry AJ507799. Regions complementary to which the oligonucleotide Northern probes were designed are shown in bold. The suspected positions of the mature miRNAs are indicated by dotted brackets.

while a third (not confirmed via cloning) indeed conforms to the novel pre-miR-BART3 (see Table 2).

Our approach to detect viral miRNAs is very similar to the one developed independently by Bentwich et al. (2005) for the identification of cellular miRNAs. Bentwich and colleagues performed bioinformatics searches to predict pre-miRNA candidates; these candidates were then analyzed using hybridization of size-selected RNA with DNA microarrays carrying suspected miRNA sequences. However, there are notable differences between the two approaches. First, due to the much larger scope of the work of Bentwich and colleagues, their computational algorithm necessarily had to be much more stringent, as only a limited number of hairpins can be tested even with high throughput array analysis. In contrast, the relatively small size of the viral genomes analyzed here also allowed us to test hairpins that were not considered good candidates by computational means only; this enabled us to identify, for example, KSHV-pre-miRNA-1. Regarding the hybridization protocol, Bentwich et al. (2005) used size-selected RNA that was ligated to adaptors, amplified, and transcribed using a T7 promoter to obtain the material used for the hybridization; as a control, non-size-selected RNA processed in the same way was used. In contrast, we directly labeled RNA of ~20 nt derived from infected cells and hybridized it together with one of two different control

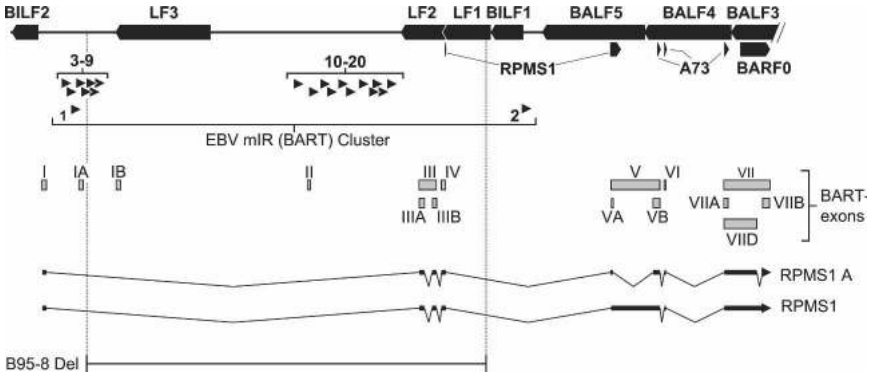

FIGURE 10. Genomic map of the BART locus of EBV. Leftward open reading frames are shown as black boxes at the *top*. Black boxes *below* indicate the rightward open reading frame BARF0 as well as sequences coding for the RPMS1 and A73 proteins. The positions of EBV-pre-miRNAs are indicated by black arrowheads pointing *right*. (The previously identified pre-miR-BART1 and -2 [Pfeffer et al. 2004] are shown *below* the 18 novel pre-miRNAs identified in this study.) The various exons of the differentially spliced BART transcripts are shown as gray boxes and are labeled according to the study of Smith et al. (2000). Two alternatively spliced RPMS1-encoding transcripts identified in the same study are also shown.

samples: either RNA of the same size from uninfected BJAB cells or slightly larger RNA fractions from the infected cells. For KSHV, we also used a tiled array besides of the array carrying the hairpins predicted by VMir. While our selection of pre-miRNA candidates was based on a combined analysis across all arrays, we have also analyzed each of the arrays individually in order to allow identification of the best method (see Tables 1, 2). Clearly, the 20-mer versus 30-mer hybridizations performed better than those using BJAB RNA as reference material.

Our method of directly labeling the RNA has the advantage of being rapid. In addition, it does not depend on potentially biased ligation procedures, which might miss certain microRNAs (discussed further below); likewise, the problem of sufficient coverage (which can be difficult to achieve during cloning experiments) does not apply. However, as we confirm potential pre-miRNAs by Northern blotting and do not generate a miRNA-enriched library, we are not able to obtain information about the precise miRNA sequence. We consider this the main disadvantage of our method.

In some cases, this can make it difficult to decide whether a given candidate represents a false positive, as exemplified by hairpin MR2079 (see Table 1; Fig. 4). MR2079 achieved high ranks in both VMir and array analysis (4 and 20, respectively), and a band in the size range expected for a mature microRNA was readily detectable using Northern probes complementary to the scoring arm. Furthermore, the hairpin is located within an intron (a common source of pre-miRNAs). However, we also observed a faint signal of similar or same size in KSHV-negative BJAB cells. Therefore, although we have not observed significant similarity between the MR2079 probe and known cellular miRNAs, it is possible that the probe cross-reacted with a cellular miRNA that is expressed at much higher levels in BCBL1 as compared to BJAB cells. An alternative explanation is that MR2079 represents a bona fide pre-miRNA that mimics a hitherto unidentified cellular counterpart; however, without further analysis, we cannot distinguish between the two possibilities. Due to the absence of specific sequence information, we are also unable to pinpoint the precise location of a miRNA within its precursor. Although the arrays generally allow the identification of the 5′-arm, the 3′-arm, or both arms as the source of miRNAs, our method therefore identifies pre-miRNAs rather than mature miRNAs. Within certain limits, the location of the mature miRNA can be predicted based on the hairpin structure. Assuming an average size of 22 nt and a terminal loop structure of at least 10 nt (Zeng et al. 2005), we have indicated putative miRNA positions for the pre-miRNAs that were exclusively identified in our study (dotted brackets in Figs. 5, 9). However, given miRNA size heterogeneity as well as imperfect prediction of hairpin structures, these can be considered only as rough estimates. (We have therefore also refrained from performing a target prediction for the novel miRNAs identified in our study.)

Lastly, although certainly not unique to our approach, we encountered the problem of false negatives, i.e., candidates that were negative in either Northern blotting or array analysis. With regard to the first kind of false negative, this is in large part due to the relatively insensitivity of the Northern blot procedure. Of note, our array analyses suggested expression of miR-K12-3 and expression of 5p- as well as 3p-miRNAs from KSHV-pre-miR-K12-6 and -8. While all of these miRNAs were cloned and thus independently confirmed in other studies (Cai et al. 2005; Pfeffer et al. 2005; Samols et al. 2005), by Northern blotting we were unable to confirm expression of miR-K12-3 and -8 and could detect only the 3p miRNA of KSHV-miR-K12-6. The second kind of false negatives were encountered during our analysis of EBV, where only eight out of the total of 18 novel pre-miRNAs achieved positive signals on our arrays. In large part, this is probably due simply to a lower quality of the samples or higher abundance of RNA degradation products. However, we note that, as judged by Northern blotting, the miRNAs that were negative on our array analysis are not necessarily those of low abundance (see, e.g., miR-BART-6 in Fig. 8). Thus, it is possible that additional factors influence the performance of a given oligonucleotide on the array, such as the relative location of the mature miRNA within the oligonucleotide or abundance of cytosine residues, as suggested by previous studies (Barad et al. 2004; Bentwich et al. 2005).

Despite these limitations, the feasibility of our combined approach is documented by the fact that, in addition to the pre-miRNAs previously reported by others, we discovered one novel KSHV-pre-miRNA and 18 novel EBV-pre-miRNAs. In both cases, we suspected the location of the pre-miRNAs because of the observation of clusters of high scoring hairpins in the VMir prediction. In the case of KSHV, this observation was reassuring but not necessarily required, as the array analysis alone was sufficient enough to identify all pre-miRNAs; 10 of the 12 miRNAs were among the top 20 candidates of our initial analysis, while the identification of the remaining two required a second analysis using less stringent filter settings. As discussed above, our EBV arrays performed less well, and we observed relatively high levels of background noise. Of the known EBV microRNAs, we could detect signals only for miR-BART1, miR-BART2, and miR-BHRF1-3, which achieved ranks 25, 65, and 61, respectively. (A Northern blot for miR-BHRF1-2 showed that this miRNA is expressed in Jijoye cells, albeit at much lower levels than in B95-8 cells [data not shown].) In order to test the reliability of our array, we performed Northern blots initially only for the top 25 array candidates, using a low cutoff for normalized oligonucleotide ratios. This analysis showed that indeed seven novel pre-microRNAs were among the tested candidates, all of which mapped to the BART region. As mentioned previously, the BART region contains 62% of all hairpins in direct orientation that achieved VMir scores of 180 or higher, located within a segment that comprises 1/10 of the viral genome. Reasoning that this observation merited special attention, we therefore focused our efforts on the BART cluster instead of relying exclusively on the array analyses as we had done for KSHV. In our second round of Northern blots, we probed for each hairpin that mapped to the region defined by the BART transcripts, as long as it was in the same (i.e., direct) orientation as the transcripts and had achieved a VMir score of at least 150. By doing so, we could identify an additional 11 novel pre-microRNAs. Thus, while our KSHV work was primarily based on the array analysis, for EBV, the initial array data served as a catalyst for a subsequent analysis based on the VMir prediction and consideration of EBV biology (i.e., the location of the BART transcripts).

While our work has identified several novel pre-microRNAs in EBV and KSHV, both viruses have been subjected to cloning studies aiming at the discovery of miRNAs before. Why have these pre-microRNAs been missed in the previous studies? In the case of KSHV, close examination of the KSHV-miR-K12-12 precursor hairpin revealed a recognition site for the restriction enzyme BanI within the (suspected) region of the mature miRNA. This enzyme was used in all three studies to concatemerize cDNA-products, thus explaining why KSHV-miR-K12-12 was not cloned. In contrast to KSHV, only one study so far attempted to clone small RNAs from EBV-infected cells (Pfeffer et al. 2004). While this study used cells infected with the B95-8 strain of EBV, most of the novel pre-miRNAs are located within a region that is deleted in B95-8; thus, the miRNAs could not have been cloned. (Neither could the pre-miRNAs have been predicted in the subsequent computational analysis of the B95-8 strain by Pfeffer et al. [2005]. Since latency is stably maintained in B95-8, lytic replication can be induced in these cells and the strain is also fully transforming in vitro, the missing miRNAs must be nonessential for these processes or have essential functions that are duplicated by other miRNAs from this region.) It is less clear, however, why miR-BART3, -4, and -5 were missed, as they map outside of the deletion and thus are very likely to be expressed in B95-8 cells. It has been proposed that there might be a bias toward the cloning of miRNAs with a certain sequence composition (Elbashir et al. 2001). If so, and especially if such a bias is responsible for the fact that the three miRNAs were missed during the previous study, this would underscore the importance of alternative approaches such as ours that do not depend on the cloning of miRNAs.

In summary, we have identified novel gamma-herpesvirus-encoded microRNAs by an approach that does not rely on cloning. Each of the hitherto available methods for the identification of miRNAs has its own limitations: While the cloning of small RNAs might be biased against certain miRNAs, array-based methods might be unable to identify select miRNAs as well. Likewise, computational approaches are always limited by inaccuracies of the underlying structural prediction algorithms. We therefore emphasize that we do not consider the method described here as being superior to cloning but rather as representing a complementary approach. In fact, for a primary analysis, we would advise researchers to use the traditional cloning method, as it directly provides miRNA sequence information and is technically less complex. However, our method has allowed us to identify novel pre-miRNA in viruses that had been subjected to cloning efforts before, and it is here that we can most readily see the utility of our approach: in the detailed analysis of genomes of comparatively small size, where every hairpin counts.

MATERIALS AND METHODS

Cell lines

The KSHV-positive primary effusion cell line BCBL1 (Renne et al. 1996) was cultured in RPMI-1640 (GIBCO/Invitrogen Corp.) supplemented with 10% FCS, 5 mM 2-mercaptoethanol, 1 mM sodium pyruvate, and 2 mM L-glutamine. The EBV-negative Burkitt's lymphoma cell line BJAB (Klein et al. 1974) was grown in RPMI-1640 medium with 10% FCS. EBV-positive Jijoye cells (Kohn et al. 1967) were maintained in RPMI 1640 medium with 10% FCS and 2 mM L-glutamine adjusted to contain 1.5 g/L sodium bicarbonate, 4.5 g/L glucose, 10 mM HEPES, and 1.0 mM sodium pyruvate.

Computational prediction of pre-microRNA candidates

We used VMir, a refined version of a program described previously (Sullivan et al. 2005) to predict microRNAs encoded by KSHV and EBV. For EBV, we analyzed the prototypical wild-type sequence deposited in GenBank (GenBank entry AJ507799). The sequence representing the long coding region of the KSHV genome (KSHV LCR; 137,559 bp) was assembled from three overlapping GenBank entries. The nucleotide positions flanking the three segments in the assembled and source sequences, respectively, are as follows: nucleotides 1–97 of KSHV LCR; nucleotides 295–391 of GenBank entry KSU86667; nucleotides 88–133,748 of KSHV LCR; nucleotides 1–133,661 of GenBank entry U93872; nucleotides 133,481–137,559 of KSHV LCR; and nucleotides 1–4079 of GenBank entry KSU85269. Note that, with the exception of Figure 1, all nucleotide positions given in figures and text refer to GenBank entry U93872, which harbors all of the KSHV miRNAs.

The VMir program (available from the authors upon request) uses several predefined but adjustable parameters; in the following, the default values used in this study are given in brackets. The program slides a window of adjustable size (500 nt) over the sequence of interest, advancing each window by a given step size (10 nt). The secondary structure of RNAs corresponding to each window is predicted using the RNAFold algorithm (Hofacker et al. 1994). Hairpins with a size above a certain threshold (45 nt) are then identified and scored (see following paragraph). The hairpins detected in each of the windows are recorded and, after completion of the structural analysis, are classified as Main, Subsidiary, or Repeated Hairpins (MHPs, SHPs, or RHPs, respectively). First, all recorded hairpins are compared to one another and grouped into local and repeat families. For comparison purposes, a core region of maximally 85 nt (from the terminal loop structure) is considered. Local groups consist of hairpins that form at the same location and are structurally identical within the core region, but can differ in flanking regions due to their folding in different sequence windows. The longest hairpin within such a group is designated the MHP, while all others are classified as the SHPs of the local family. Repeat groups contain hairpins that are of identical sequence and structure (within the core region) but form at different genomic locations. The 5′-proximal hairpin within a repeat group is designated the MHP of the group; all others are classified as RHPs. Individual hairpins, i.e., hairpins that do not belong to a group, are always classified as MHPs. The results of the analysis are written to a file, and the detected hairpins can later be filtered according to various parameters such as the size, the score, and the number of windows in which a hairpin can be detected (window count). For the calculation of the latter, the hairpin of interest as well as all of its subsidiary hairpins are considered (e.g., if a MHP folds in 10 windows and has SHPs folding in five and three windows, the total window count for that hairpin is 18).

For scoring purposes, hairpins are first trimmed to a predefined maximum size (120 nt). A basic score is then calculated as follows: Each paired nucleotide within the hairpin is awarded two points. From this score, for hairpins with a terminal loop of >17 nt, one point is subtracted for each additional nucleotide. For each symmetric bulge, a value corresponding to the size of the bulge multiplied by factors of 1 or 1.5 is subtracted for bulges with a size ≤4 or >4, respectively, while for each asymmetric bulge, a value corresponding to bulge size multiplied by a factor of 2 is subtracted. To calculate the final score, the basic score is multiplied with a factor Ip, which indicates the probability that a given hairpin represents a pre-miRNA. The algorithm for the calculation of Ip was developed based on the statistical comparison of two data sets: a reference set of 175 known human pre-miRNAs and a training set of 5500 unrelated hairpins from randomly chosen sequences (a detailed description of the algorithm can be obtained from the authors). These data sets were analyzed for (1) the percentage of paired versus unpaired nucleotides within the stem, (2) the size of the longest helix within the stem, (3) the relative position of the terminal loop within the hairpin, (4) the size of the terminal loop, and (5) the occurrence of bulges within the stem–loop region that is known (reference set of known pre-miRNAs) or would be expected (training set) to encode the mature miRNA. The resulting Ip values range between 0 and 2 for the lowest and highest probability, respectively, that a given hairpin represents a pre-miRNA (a neutral value of 1 means that the hairpin's features do not discriminate it from either the reference or the training set).

Northern blot analysis and microarray hybridization

Total RNA was harvested using RNA-Bee (Iso-Tex Diagnostic, Inc.) according to the manufacturer's instructions. Northern blots were conducted as previously described (Sullivan et al. 2005). Briefly, total RNA was electrophoresed through a 15% acrylamide urea denaturing gel and electroblot-transferred to Zeta-Probe GT membrane (BioRad). Blots were hybridized to radiolabeled antisense oligonucleotide probes in ExpressHyb (BD Biosciences Clontech) hybridization buffer and then exposed to autoradiography film to visualize bands.

Oligonucleotide microarrays were custom synthesized using the CombiMatrix 12k semiconductor-based custom array platform (CombiMatrix Corp.). RNA was size-fractionated through a 15% acrylamide urea denaturing gel, eluted overnight, precipitated, and concentrated. RNA was labeled with the MirVana miRNA labeling kit (Ambion Inc.) and Cy3 or Cy5 d(UTP) (Cyscribe, Amersham Biosciences). Arrays were hybridized in 1× miRNA MirVana hybridization buffer (Ambion) and washed according to the manufacturer's directions.

Microarray analysis

Arrays were scanned using a GenePix 4000B Scanner (Molecular Devices Corp.) and the GenePix Pro 6 software package.

Ratios were calculated based on median Cy3 and Cy5 fluorescence intensities, and the resulting ratio data sets were normalized by subtracting from each individual ratio the mean of all ratios across the array and subsequent division of the resulting value by the standard deviation (normalization to 0 mean and unit standard deviation). Oligonucleotides that had a GC content of 85% or higher in any window of 21 nt, showed stretches of more than seven consecutive G or C residues, stretches of more than 10 consecutive A or T residues, or segments of 8 or more nt or more consisting exclusively of G and C residues were flagged and not further considered in the analysis. Scores were awarded to the hairpins predicted by VMir as follows: First, the normalized ratios of all oligonucleotides that (1) showed an overlap of at least 15 nt with the hairpins of interest and (2) were located no further than 10 nt away from the terminal loop of the hairpin were retrieved.

A primary array score was then calculated by summation of all normalized ratios above a given threshold (see text for threshold values used in the individual analyses). In cases in which the analysis included more than one array, the sum was calculated from the ratios of scoring oligonucleotides across all arrays. The primary score was subsequently multiplied by a factor that was based on the VMir score of the hairpin of interest. For the calculation of this factor, the VMir scores of all predicted hairpins were first normalized to a mean value of 1. For all hairpins with a normalized score of >1 (i.e., hairpins with a score above the mean value of the prediction), the multiplication factor consisted of the normalized score, whereas the factor was set to 1 (and therefore essentially no multiplication was performed) for hairpins with a normalized score <1.

ACKNOWLEDGMENTS

We thank Jill Bechtel for her contribution of some of the cells and RNA used in this study. C.S.S. is supported by NIH training grant T02.

Received December 13, 2005; accepted January 19, 2006.

REFERENCES

AuCoin, D.P., Colletti, K.S., Cei, S.A., Papouskova, I., Tarrant, M., and Pari, G.S. 2004. Amplification of the Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8 lytic origin of DNA replication is dependent upon a *cis*-acting AT-rich region and an ORF50 response element and the *trans*-acting factors ORF50 (K-Rta) and K8 (K-bZIP). *Virology* **318:** 542–555.

Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., and Pasquinelli, A.E. 2005. Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* **122:** 553–563.

Barad, O., Meiri, E., Avniel, A., Aharonov, R., Barzilai, A., Bentwich, I., Einav, U., Gilad, S., Hurban, P., Karov, Y., et al. 2004. MicroRNA expression detected by oligonucleotide microarrays: System establishment and expression profiling in human tissues. *Genome Res.* **14:** 2486–2494.

Bentwich, I. 2005. Prediction and validation of microRNAs and their targets. *FEBS Lett.* **579:** 5904–5910.

Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., et al. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.* **37:** 766–770.

Berezikov, E., Guryev, V., van de Belt, J., Wienholds, E., Plasterk, R.H., and Cuppen, E. 2005. Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* **120:** 21–24.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409:** 363–366.

Bohnsack, M.T., Czaplinski, K., and Gorlich, D. 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10:** 185–191.

Brooks, L.A., Lear, A.L., Young, L.S., and Rickinson, A.B. 1993. Transcripts from the Epstein-Barr virus BamHI A fragment are detectable in all three forms of virus latency. *J. Virol.* **67:** 3182–3190.

Brown, J.R. and Sanseau, P. 2005. A computational view of microRNAs and their targets. *Drug Discov. Today* **10:** 595–601.

Cai, X., Lu, S., Zhang, Z., Gonzalez, C.M., Damania, B., and Cullen, B.R. 2005. Kaposi's sarcoma-associated herpesvirus expresses an array of viral microRNAs in latently infected cells. *Proc. Natl. Acad. Sci.* **102:** 5570–5575.

Chen, H.L., Lung, M.M., Sham, J.S., Choy, D.T., Griffin, B.E., and Ng, M.H. 1992. Transcription of BamHI-A region of the EBV genome in NPC tissues and B cells. *Virology* **191:** 193–201.

Chen, H., Huang, J., Wu, F.Y., Liao, G., Hutt-Fletcher, L., and Hayward, S.D. 2005. Regulation of expression of the Epstein-Barr virus BamHI-A rightward transcripts. *J. Virol.* **79:** 1724–1733.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. 2005. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436:** 740–744.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432:** 231–235.

Dittmer, D., Lagunoff, M., Renne, R., Staskus, K., Haase, A., and Ganem, D. 1998. A cluster of latently expressed genes in Kaposi's sarcoma-associated herpesvirus. *J. Virol.* **72:** 8309–8315.

Du, T. and Zamore, P.D. 2005. microPrimer. The biogenesis and function of microRNA. *Development* **132:** 4645–4652.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. 2001. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes & Dev.* **15:** 188–200.

Forstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D. 2005. Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* **3:** e236.

Gilligan, K., Sato, H., Rajadurai, P., Busson, P., Young, L., Rickinson, A., Tursz, T., and Raab-Traub, N. 1990. Novel transcription from the Epstein-Barr virus terminal EcoRI fragment, DIJhet, in a nasopharyngeal carcinoma. *J. Virol.* **64:** 4948–4956.

Gilligan, K.J., Rajadurai, P., Lin, J.C., Busson, P., Abdel-Hamid, M., Prasad, U., Tursz, T., and Raab-Traub, N. 1991. Expression of the Epstein-Barr virus BamHI A fragment in nasopharyngeal carcinoma: Evidence for a viral protein expressed in vivo. *J. Virol.* **65:** 6252–6259.

Grad, Y., Aach, J., Hayes, G.D., Reinhart, B.J., Church, G.M., Ruvkun, G., and Kim, J. 2003. Computational and experimental identification of *C. elegans* microRNAs. Mol. *Cell* **11:** 1253–1263.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432:** 235–240.

Gregory, R.I., Chendrimada, T.P., Cooch, N., and Shiekhattar, R. 2005. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* **123:** 631–640.

Griffiths-Jones, S. 2004. The microRNA Registry. *Nucleic Acids Res.* **32:** D109–D111.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106:** 23–34.

Hammond, S.M. 2005. Dicing and slicing: The core machinery of the RNA interference pathway. *FEBS Lett.* **579:** 5822–5829.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. 2004. The Drosha–DGCR8 complex in primary microRNA processing. *Genes & Dev.* **18:** 3016–3027.

Hitt, M.M., Allday, M.J., Hara, T., Karran, L., Jones, M.D., Busson, P., Tursz, T., Ernberg, I., and Griffin, B.E. 1989. EBV gene expression in an NPC-related tumour. *EMBO J.* **8:** 2639–2651.

Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, S., Tacker, P., and Schuster, P. 1994. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.* **125:** 167–188.

Hutvagner, G. and Zamore, P.D. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297:** 2056–2060.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293:** 834–838.

Jiang, F., Ye, X., Liu, X., Fincher, L., McKearin, D., and Liu, Q. 2005. Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes & Dev.* **19:** 1674–1679.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes & Dev.* **15:** 2654–2659.

Kim, V.N. 2005. MicroRNA biogenesis: Coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* **6:** 376–385.

Klein, G., Lindahl, T., Jondal, M., Leibold, W., Menezes, J., Nilsson, K., and Sundstrom, C. 1974. Continuous lymphoid cell lines with characteristics of B cells (bone-marrow-derived), lacking the Epstein-Barr virus genome and derived from three human lymphomas. *Proc. Natl. Acad. Sci.* **71:** 3283–3286.

Kohn, G., Mellman, W.J., Moorhead, P.S., Loftus, J., and Henle, G. 1967. Involvement of C group chromosomes in five Burkitt lymphoma cell lines. *J. Natl. Cancer Inst.* **38:** 209–222.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294:** 853–858.

Lai, E.C., Tomancak, P., Williams, R.W., and Rubin, G.M. 2003. Computational identification of *Drosophila* microRNA genes. *Genome Biol.* **4:** R42.

Landthaler, M., Yalcin, A., and Tuschl, T. 2004. The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Curr. Biol.* **14:** 2162–2167.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294:** 858–862.

Lee, R.C. and Ambros, V. 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294:** 862–864.

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.

Li, H., Komatsu, T., Dezube, B.J., and Kaye, K.M. 2002. The Kaposi's sarcoma-associated herpesvirus K12 transcript from a primary effusion lymphoma contains complex repeat elements, is spliced, and initiates from a novel promoter. *J. Virol.* **76:** 11880–11888.

Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., and Bartel, D.P. 2003a. Vertebrate microRNA genes. *Science* **299:** 1540.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. 2003b. The microRNAs of *Caenorhabditis elegans*. *Genes & Dev.* **17:** 991–1008.

Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433:** 769–773.

Lin, C.L., Li, H., Wang, Y., Zhu, F.X., Kudchodkar, S., and Yuan, Y. 2003. Kaposi's sarcoma-associated herpesvirus lytic origin (ori-Lyt)-dependent DNA replication: Identification of the ori-Lyt and association of K8 bZip protein with the origin. *J. Virol.* **77:** 5578–5588.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. 2004. Nuclear export of microRNA precursors. *Science* **303:** 95–98.

Mansfield, J.H., Harfe, B.D., Nissen, R., Obenauer, J., Srineel, J., Chaudhuri, A., Farzan-Kashani, R., Zuker, M., Pasquinelli, A.E., Ruvkun, G., et al. 2004. MicroRNA-responsive sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nat. Genet.* **36:** 1079–1083.

Matsumura, S., Fujita, Y., Gomez, E., Tanese, N., and Wilson, A.C. 2005. Activation of the Kaposi's sarcoma-associated herpesvirus major latency locus by the lytic switch protein RTA (ORF50). *J. Virol.* **79:** 8493–8505.

Pearce, M., Matsumura, S., and Wilson, A.C. 2005. Transcripts encoding K12, v-FLIP, v-cyclin, and the microRNA cluster of Kaposi's sarcoma-associated herpesvirus originate from a common promoter. *J. Virol.* **79:** 14457–14464.

Pfeffer, S., Zavolan, M., Grasser, F.A., Chien, M., Russo, J.J., Ju, J., John, B., Enright, A.J., Marks, D., Sander, C., et al. 2004. Identification of virus-encoded microRNAs. *Science* **304:** 734–736.

Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F.A., van Dyk, L.F., Ho, C.K., Shuman, S., Chien, M., et al. 2005. Identification of microRNAs of the herpesvirus family. *Nat. Methods* **2:** 269–276.

Renne, R., Zhong, W., Herndier, B., McGrath, M., Abbey, N., Kedes, D., and Ganem, D. 1996. Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture. *Nat. Med.* **2:** 342–346.

Sadler, R.H. and Raab-Traub, N. 1995. Structural analyses of the Epstein-Barr virus BamHI A transcripts. *J. Virol.* **69:** 1132–1141.

Sadler, R., Wu, L., Forghani, B., Renne, R., Zhong, W., Herndier, B., and Ganem, D. 1999. A complex translational program generates multiple novel proteins from the latently expressed kaposin (K12) locus of Kaposi's sarcoma-associated herpesvirus. *J. Virol.* **73:** 5722–5730.

Saito, K., Ishizuka, A., Siomi, H., and Siomi, M.C. 2005. Processing of pre-microRNAs by the Dicer-1–Loquacious complex in *Drosophila* cells. *PLoS Biol.* **3:** e235.

Samols, M.A., Hu, J., Skalsky, R.L., and Renne, R. 2005. Cloning and identification of a microRNA cluster within the latency-associated region of Kaposi's sarcoma-associated herpesvirus. *J. Virol.* **79:** 9301–9305.

Smith, P.R., Gao, Y., Karran, L., Jones, M.D., Snudden, D., and Griffin, B.E. 1993. Complex nature of the major viral polyadenylated transcripts in Epstein-Barr virus-associated tumors. *J. Virol.* **67:** 3217–3225.

Smith, P.R., de Jesus, O., Turner, D., Hollyoake, M., Karstegl, C.E., Griffin, B.E., Karran, L., Wang, Y., Hayward, S.D., and Farrell, P.J. 2000. Structure and coding content of CST (BART) family RNAs of Epstein-Barr virus. *J. Virol.* **74:** 3082–3092.

Sugiura, M., Imai, S., Tokunaga, M., Koizumi, S., Uchizawa, M., Okamoto, K., and Osato, T. 1996. Transcriptional analysis of Epstein-Barr virus gene expression in EBV-positive gastric carcinoma: Unique viral latency in the tumour cells. *Br. J. Cancer* **74:** 625–631.

Sullivan, C.S., Grundhoff, A.T., Tevethia, S., Pipas, J.M., and Ganem, D. 2005. SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature* **435:** 682–686.

Wang, Y., Li, H., Chan, M.Y., Zhu, F.X., Lukac, D.M., and Yuan, Y. 2004. Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: Cis-acting requirements for replication and ori-Lyt-associated RNA transcription. *J. Virol.* **78:** 8615–8629.

Yekta, S., Shih, I.H., and Bartel, D.P. 2004. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304:** 594–596.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Dev.* **17:** 3011–3016.

Zeng, Y. and Cullen, B.R. 2004. Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res.* **32:** 4776–4785.

Zeng, Y., Yi, R., and Cullen, B.R. 2003. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc. Natl. Acad. Sci.* **100:** 9779–9784.

———. 2005. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J.* **24:** 138–148.

A combined computational and microarray-based approach identifies novel microRNAs encoded by human gamma-herpesviruses

Adam Grundhoff, Christopher S. Sullivan and Don Ganem

RNA 2006 12: 733-750