

Identification of microRNAs of the herpesvirus family

Sébastien Pfeffer^{1,12}, Alain Sewer^{2,12}, Mariana Lagos-Quintana¹, Robert Sheridan³, Chris Sander³, Friedrich A Grässer⁴, Linda F van Dyk⁵, C Kiong Ho^{6,11}, Stewart Shuman⁶, Minchen Chien⁷, James J Russo⁷, Jingyue Ju^{7,8}, Glenn Randall⁹, Brett D Lindenbach⁹, Charles M Rice⁹, Viviana Simon¹⁰, David D Ho¹⁰, Mihaela Zavolan² & Thomas Tuschl¹

Epstein-Barr virus (EBV or HHV4), a member of the human herpesvirus (HHV) family, has recently been shown to encode microRNAs (miRNAs). In contrast to most eukaryotic miRNAs, these viral miRNAs do not have close homologs in other viral genomes or in the genome of the human host. To identify other miRNA genes in pathogenic viruses, we combined a new miRNA gene prediction method with small-RNA cloning from several virus-infected cell types. We cloned ten miRNAs in the Kaposi sarcoma-associated virus (KSHV or HHV8), nine miRNAs in the mouse gammaherpesvirus 68 (MHV68) and nine miRNAs in the human cytomegalovirus (HCMV or HHV5). These miRNA genes are expressed individually or in clusters from either polymerase (pol) II or pol III promoters, and share no substantial sequence homology with one another or with the known human miRNAs. Generally, we predicted miRNAs in several large DNA viruses, and we could neither predict nor experimentally identify miRNAs in the genomes of small RNA viruses or retroviruses.

miRNAs are conserved, endogenous 21–23-nucleotide (nt) noncoding RNA molecules that regulate protein-coding gene expression in plants and animals via the RNA silencing machinery (for reviews see refs. 1 and 2). Depending on the extent of complementarity, miRNAs can either direct cleavage or regulate translation of their target mRNAs. Candidate targets of animal miRNAs have been computationally predicted by various groups (reviewed in ref. 1), but only a very small subset of these predictions have been validated.

Typically, miRNAs are derived from long pol II primary transcripts³ that are processed in the nucleus by the ribonuclease III enzyme Droscha together with the double-stranded RNA (dsRNA)-binding domain containing the protein DGCR8 (reviewed in ref. 4). The excised ~70-nt fold-back dsRNA precursor (pre-miRNA) is then exported to the cytoplasm by the export factor Exportin 5 (ref. 5). The pre-miRNA is later cleaved by the ribonuclease III enzyme Dicer to

excise the miRNA in the form of a small interfering RNA (siRNA)-like duplex⁶, and this is followed by asymmetric assembly of the miRNA strand into effector complexes containing the Argonaute protein.

The majority of miRNAs were identified by cDNA cloning, whereas others have been identified by genetic analysis or by computer prediction. Plant and invertebrate genomes encode just over a hundred miRNA genes; vertebrate genomes encode several hundred miRNA genes. Most recently, several miRNA genes have been discovered in EBV, a member of the herpesvirus family⁷. A searchable database containing all known miRNA sequences is available⁸ (<http://www.sanger.ac.uk/Software/Rfam/mirna/>).

The discovery of miRNAs in a herpesvirus indicates that viruses have evolved to exploit RNA silencing for regulation of both host and viral genes⁷. The EBV miRNAs have no notable sequence similarity with known host cell miRNAs. Additionally, sequence homology searches against the genomes of other human herpesviruses have not revealed any orthologous miRNAs. This observation indicated that, unlike host miRNAs, viral miRNAs are difficult to identify using existing miRNA gene prediction software, which heavily relies on cross-species sequence conservation^{9,10}. Here we present a new computational method that predicts likely locations of miRNA precursors in individual genomes using only information about the local sequence composition and about the predicted secondary structure of RNA transcripts. Applying this method to a subset of pathogenic viruses, we predicted that DNA viruses, and in particular the family of herpesviruses, would be the most likely to encode miRNAs. By small-RNA cloning, we confirmed that KSHV and MHV68, two members of the gammaherpesvirus subfamily, and HCMV, a member of the betaherpesvirus subfamily, encode miRNAs. At the same time, we neither predicted nor experimentally detected miRNAs in human cells infected by small-genome RNA viruses, including yellow fever virus (YFV), hepatitis C virus (HCV) and human immunodeficiency virus 1 (HIV1).

¹Laboratory of RNA Molecular Biology, The Rockefeller University, 1230 York Avenue, Box 186, New York, New York 10021, USA. ²Biozentrum der Universität Basel, Klingelbergstr 50-70, CH-4056 Basel, Switzerland. ³Computational Biology Center, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA. ⁴Institut für Mikrobiologie und Hygiene, Abteilung Virologie, Haus 47, Universitätskliniken, D-66421 Homburg/Saar, Germany. ⁵Microbiology and Immunology Departments, University of Colorado Health Science Center, Aurora, Colorado 80045, USA. ⁶Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA. ⁷Columbia Genome Center, Russ Berrie Pavilion, 1150 St. Nicholas Avenue, New York, New York 10032, USA. ⁸Department of Chemical Engineering, Columbia University, 500 West 120th Street, New York, New York 10027, USA. ⁹Center for the Study of Hepatitis C, Laboratory of Virology and Infectious Disease, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA. ¹⁰Aaron Diamond AIDS Research Center, The Rockefeller University, 455 First Avenue, New York, New York 10016, USA. ¹¹Present address: Department of Biological Sciences, 648 Cooke Hall, North Campus, State University of New York at Buffalo, Buffalo, New York 14260, USA. ¹²These authors contributed equally to this work. Correspondence should be addressed to M.Z. (mihaela.zavolan@unibas.ch) or T.T. (ttuschl@rockefeller.edu).

RESULTS

Computational method of pre-miRNA prediction

Our computational method assumes that miRNA precursors have specific composition and secondary structure features that are recognized by the miRNA-processing enzymes during miRNA biogenesis. We therefore described each candidate stem in terms of a set of properties (detailed in Methods) and then trained a support vector machine (SVM) to separate known miRNA precursors from a set of stem-loops unlikely to contain miRNAs. To report specific miRNA precursor predictions, we set a score threshold, which we chose as

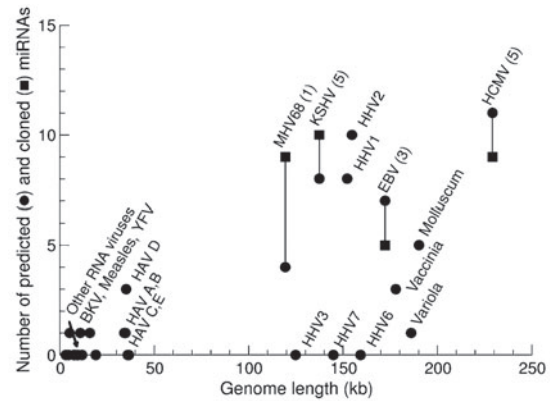
to achieve a good tradeoff between sensitivity and specificity in the training data, namely to retrieve 71% of the positive miRNAs in the training set while misclassifying only 3% of the negative examples. Candidate stems to which the SVM assigns a score above this threshold are considered miRNA precursor predictions. To prioritize our cloning studies, we also calculated the probability that at least one of the candidate stems in a given viral genome is a miRNA precursor. This indicator takes into account the full distribution of SVM scores assigned to all the candidate stems in the genome and is therefore not dependent on a score threshold.

Table 1 | Summary of pre-miRNA predictions for selected human and mouse DNA and RNA viruses

Family (subfamily)	Genus	Species	Accession ^a	Prob. (%) ^b	Pred. ^c	Valid. ^d	Cloned ^e
dsDNA viruses							
Herpesviridae							
(Alphaherpesvirinae)	Simplexvirus	HHV1	NC_001806	99.9	8 (7)	NT	NT
		HHV2	NC_001798	100	10 (6)	NT	NT
	Varicellovirus	HHV3	NC_001348	54.7	0	NT	NT
(Betaherpesvirinae)	Cytomegalovirus	HHV5/HCMV	AC146907	99.9	11 (9)	5 (5)	9
		Roseolovirus	HHV6	NC_001664	84.2	0	NT
			HHV7	NC_001716	45.8	0	NT
(Gammaherpesvirinae)	Lymphocryptovirus	HHV4/EBV	NC_001345	99.0	7 (6)	2 (3)	5
		Rhadinovirus	HHV8/KSHV	NC_003409	100	8 (6)	5 (5)
			MHV68	NC_001826	97.6	4 (2)	1 (1)
Poxviridae							
(Chordopoxvirinae)	Orthopoxvirus	Vaccinia virus	U94848	88.9	3 (3)	NT	NT
		Variola virus	L22579	92.1	1 (1)	NT	NT
		Molluscipoxvirus	Molluscum contagiosum virus	NC_001731	99.1	5 (5)	NT
Adenoviridae	Mastadenovirus	HAVA	NC_000942	40.6	1 (1)	NT	NT
		HAVB	NC_004001	46.5	1 (1)	NT	NT
		HAVC	NC_001405	43.7	0	NT	NT
		HAVD	NC_002067	67.7	3 (3)	NT	NT
		HAVE	NC_001454	36.8	0	NT	NT
Papillomaviridae	Papillomavirus	HPV18	f	0.90	0	NT	NT
Polyomaviridae	Polyomavirus	BK virus	NC_001538	31.9	1 (1)	NT	NT
ssDNA viruses							
Parvoviridae							
(Parvovirinae)	Dependovirus	AAV 2	NC_001401	5.60	0	NT	NT
DNA and RNA reverse transcribing viruses							
Hepadnaviridae	Orthohepadnavirus	Hepatitis B virus	NC_003977	1.40	0	NT	NT
Retroviridae	Lentivirus	HIV1	NC_001802	2.50	0	0	0
(-) ssRNA viruses							
Filoviridae	'Marburg-like' virus	Marburg virus	NC_001608	20.8	0	NT	NT
		'Ebola-like virus'	NC_002549	18.2	0	NT	NT
Paramyxoviridae							
(Paramyxovirinae)	Morbillivirus	Measles virus	NC_001498	26.9	1 (1)	NT	NT
(+) ssRNA viruses							
Picornaviridae	Enterovirus	Poliovirus	NC_002058	2.00	0	NT	NT
Flaviviridae	Flavivirus	YFV	NC_002031	24.1	1 (1)	0	0
		Hepacivirus	HCV	NC_004102	43.1	0	0
Togaviridae	Alphavirus	Sindbis virus	NC_001547	7.30	0	NT	NT

HAV, human adenovirus; AAV, adeno-associated virus; NT, not tested. ^aGenBank accession numbers. ^bEstimate of the probability that at least one miRNA precursor is present in the considered genome (see Methods). ^cPredicted precursors. Numbers in parenthesis indicate the number of unique genomic loci irrespective of transcript orientation. ^dExperimentally confirmed precursors. Numbers in parenthesis indicate the number of validated genomic loci irrespective of transcript orientation. ^eTotal identified pre-miRNAs based on cloning. ^fThe genome sequence for HPV was extracted from the Los Alamos papillomavirus database (<http://hpv-web.lanl.gov/stdgen/virus/hpv/>).

Figure 1 | Display of the miRNA predictions and cloning results as a function of the virus genome size. The horizontal axis represents the sizes of the viral genomes studied. Circles correspond to the number of predicted pre-miRNAs for a particular virus and squares to the total number of cloned miRNAs. The number of predicted pre-miRNAs that were verified by cloning is indicated in parentheses.



We applied our computational method to at least one representative member of each family of human pathogenic viruses and predicted that many members of the herpesvirus family encode miRNAs (Table 1). In contrast, for RNA viruses we generally predict low probabilities of encoding miRNAs. This is not a mere consequence of their smaller genomes, as can be inferred from the fact that the calculated probability of encoding at least one miRNA varies among herpesviruses from 45.8% for HHV7 to 100% for HHV2 and HHV8, in spite of their large and comparable genome sizes.

We also used our method to predict the specific location of miRNA precursors in the viral genomes, using a threshold score as described above. The number of stems with scores higher than the threshold is also not a direct function of the genome size (Fig. 1). For EBV, seven pre-miRNAs were predicted to originate as transcripts from five dis-

tinct genomic loci (Supplementary Table 1 online). Two of these loci had pre-miRNAs predicted from the sense and antisense transcripts. Of the five cloned miRNAs from latent-stage infection⁷, two were predicted correctly, and a third corresponded to a predicted complementary strand. It is conceivable either that the other predicted miRNAs were expressed at other stages of the infection cycle and therefore escaped cloning or that they were false-positive predictions.

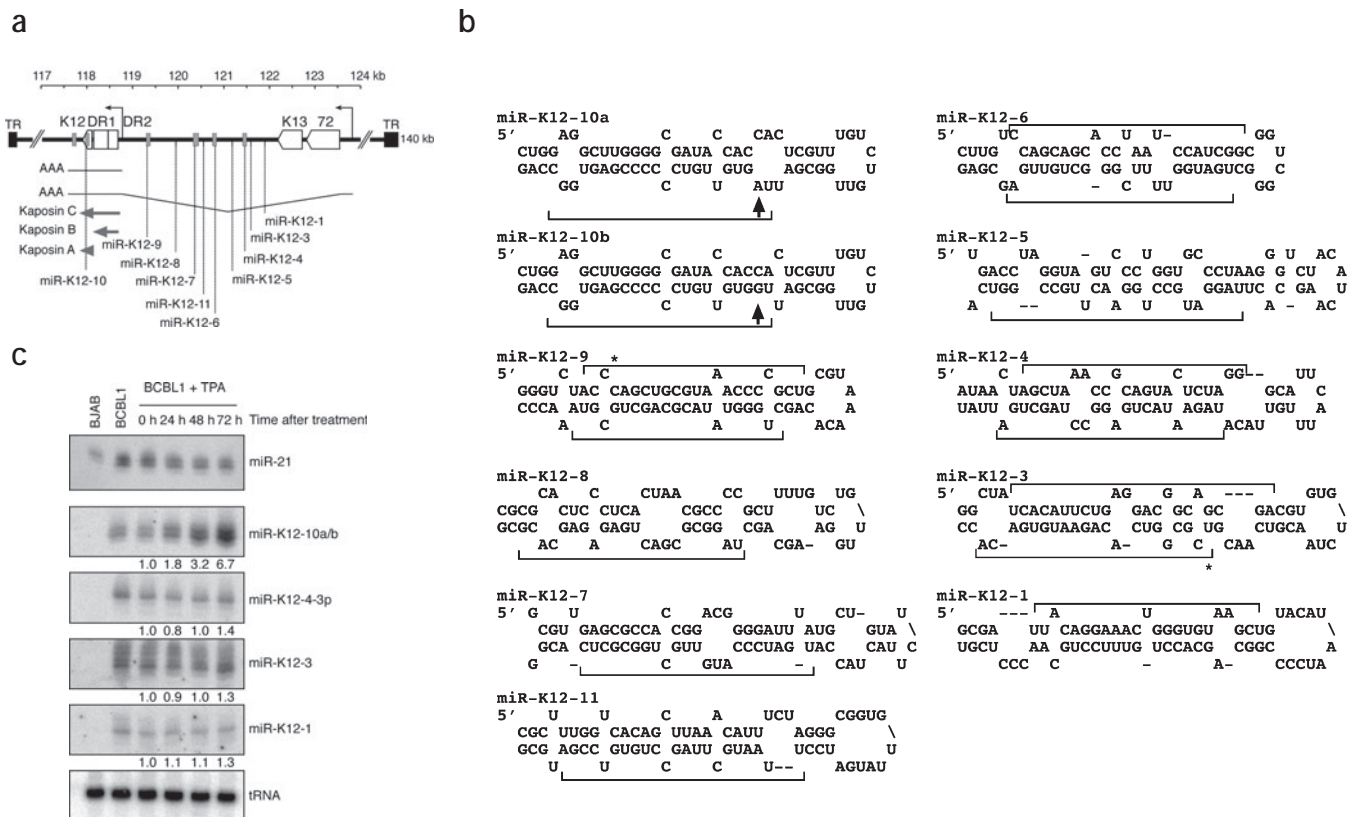


Figure 2 | Genomic positions and predicted secondary structure of KSHV pre-miRNAs. (a) Diagram of the miRNA-containing fragment of the KSHV genome. Terminal repeats (TR) are shown as black boxes. DR1 and DR2 indicate (G+C)-rich direct repeats 1 and 2. Cloned miRNAs are shown as dotted lines, predicted miRNAs as gray rectangles, the two alternate promoters for kaposin transcripts as black arrows, the kaposin transcripts as black lines, and the intronic region in the larger transcript as a break in the line. The large open arrows indicate known ORFs. The thick gray arrows below show ORFs for kaposin isoforms A, B and C. (b) Predicted KSHV pre-miRNAs. The predominantly cloned miRNAs are indicated with a black bar, asterisks highlight low-abundance small RNAs, and the black arrow indicates the A-to-I edited residue. (c) Northern blots for KSHV-miR-K12-10a, miR-K12-10b, miR-K12-4, miR-K12-3 and miR-K12-1 from total RNA isolated from a KSHV-negative cell line (BJAB) and from BCBL1 cells at the indicated times after TPA treatment. The expression of human miR-21 was also examined for reference. The numbers below the miR signal indicate relative signal intensity with respect to the 0 h signal after normalizing for gel loading by using the tRNA signal.

Experimental isolation of new viral miRNAs

Guided by our computational predictions we cloned and sequenced small RNAs from cells infected with HCMV, KSHV and the KSHV-related MHV68¹¹. Additionally, we studied the population of small RNAs isolated from cells infected with small-genome RNA viruses, namely YFV, HCV and HIV1. We used our previously established small-RNA cloning protocol¹² for the preparation of YFV, HCV and HIV1 small-RNA libraries and a modified version for the preparation of KSHV, CMV and MHV68 libraries. The modified cloning protocol required the use of the N-terminal domain of T4 RNA ligase 2, Rnl2(1–249) ligase¹³, for joining a pre-adenylated 3' adapter oligonucleotide to the gel-purified small RNAs. Rnl2(1–249) ligase, like T4 RNA ligase 1, can join a preadenylated 3' adapter oligonucleotide to the 3' hydroxyl-containing small RNA molecules, but the undesired circularization of small RNAs caused by adenylate transfer from the 3' adapter to the 5' phosphate-containing small RNAs is notably less than that seen with T4 RNA ligase 1. We then sequenced at least 1,000 independently cloned small RNAs from each library. Annotation of the cloned RNAs revealed a viral content ranging from 0 to 21% depending on the virus studied (Supplementary Table 2 online).

We identified viral miRNAs as the sequences that were cloned multiple times and that mapped perfectly to the viral genome at a location adjacent to an approximately 30-nt highly complementary sequence, which gives rise to the pre-miRNA (Figs. 2, 3 and 4). In some instances, the existence of such a dsRNA precursor was directly confirmed by cloning small RNAs from both arms of the pre-miRNA (Figs. 2b and 4b). We identified ten KSHV miRNAs by cloning from the latently KSHV-infected body cavity-based B lymphoma BCBL1 cell line¹⁴ (Fig. 2b and Supplementary Table 3 online). Our computational method predicted a total of eight pre-miRNAs, five of which overlap with the experimental set (Supplementary Table 1 online).

Furthermore, we identified nine MHV68 miRNAs by cloning small RNAs from latently MHV68-infected mouse B lymphoma S11 cells¹⁵ (Fig. 3b and Supplementary Table 4 online). From four predicted pre-miRNAs that were confined to the (+) and (–) strands of only two genomic locations, we experimentally confirmed one miRNA. But we observed that another four of the cloned miRNAs had a score immediately below the prediction threshold owing to an unusual feature caused by their pol III tRNA promoters. Finally, we identified nine HCMV miRNAs cloned from small RNAs isolated from lytically infected primary human fibroblasts (Fig. 4b and Supplementary Table 5 online). We cloned five miRNAs from 11 predicted miRNA genes (Supplementary Table 1 online).

As expected from our computational prediction, we did not experimentally identify miRNAs derived from HCV, and identified only two and four different small RNAs in (+)-strand orientation from HIV1 and YFV, respectively (Supplementary Table 6 online). The corresponding genomic regions surrounding the cloned RNAs were not predicted to form stable base-paired structures, suggesting that these small RNAs are random degradation products of the turnover of longer viral transcripts. These indirect observations are consistent with the idea that HCV, YFV and HIV1 do not elicit an RNA interference (RNAi) response. Alternatively, these viruses could have evolved inhibitors of RNAi. In the case of HCV, we examined if infected cells were still able to carry out RNA silencing using synthetic siRNAs or hairpin RNA against the lamin A/C gene. Infected and uninfected cells supported lamin A/C gene silencing with equal efficiency, arguing against the expression of a viral inhibitor of RNAi (Supplementary Fig. 1a online). Consistent with the unperturbed RNA silencing, northern blot analysis of several host cell miRNAs revealed no differences between infected and noninfected cells (Supplementary Fig. 1b online).

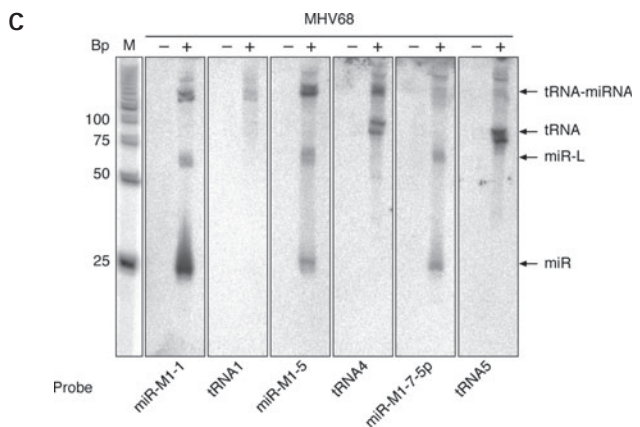
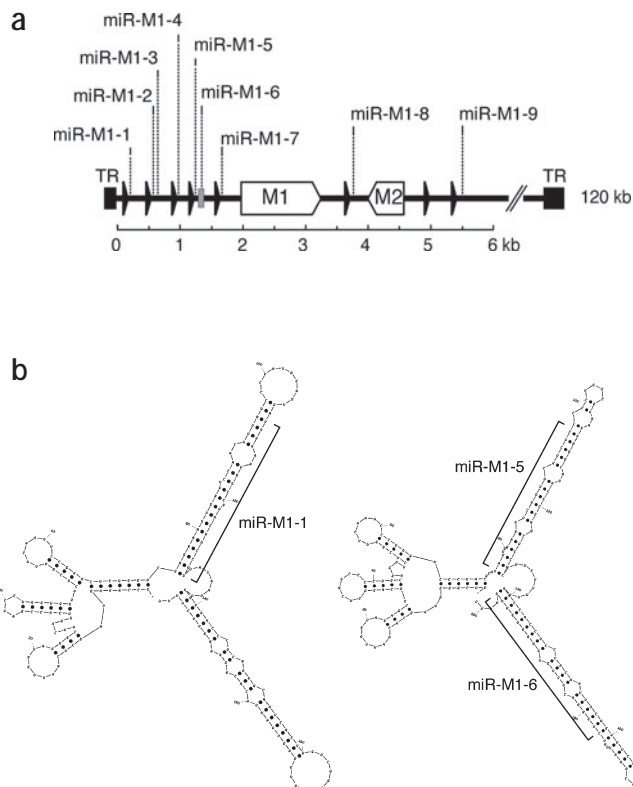


Figure 3 | Genomic positions and predicted secondary structures of MHV68 pre-miRNAs. (a) Diagram of the miRNA-containing locus of the MHV68 genome. Terminal repeats (TR) are shown as black boxes. Cloned miRNAs are shown as dotted lines, predicted miRNAs as gray rectangles, and previously characterized tRNA-like molecules¹⁹ as black arrows. The large open arrows indicate known ORFs. (b) Examples of predicted fold-back of MHV68 pre-miRNA primary transcripts. The tRNA is on the left, and a black bar indicates the predominantly cloned miRNAs. (c) MHV68 miRNAs are expressed from the tRNA primary transcripts. Northern blot for MHV68-miR-M1-1, tRNA1, miR-M1-5, tRNA4, miR-M1-7-5p and tRNA5 from total RNA isolated from the MHV68-negative A20 cell line (–) and from the MHV68-positive S11 cell line (+). M, 5' radiolabeled 25-base-pair DNA ladder.

Features of the KSHV miRNAs

All the cloned and predicted KSHV miRNAs are confined to a 5-kb region of the 140-kb viral genome, which also encodes the transforming protein-coding kaposin gene¹⁶. Two alternate transcripts for kaposin have been described^{17,18}, which each give rise to three possible isoforms, A, B and C, of the kaposin protein depending on the translation codon used¹⁸ (**Fig. 2a** and **Supplementary Table 3** online). KSHV-miR-K12-10 is located within the open reading frame (ORF) K12 so that its excision coincides with the destruction of a fraction of the kaposin mRNA. The other KSHV miRNAs are located within the intronic region of the larger kaposin primary transcript (**Fig. 2a**). But, as it is unclear if the large kaposin transcript is expressed in BCBL1 cells¹⁷, it is also conceivable that these miRNAs are expressed from a promoter located immediately upstream of the miRNA cluster.

We also noticed two approximately equally frequently cloned nucleotide sequence variants of KSHV-miR-K12-10, which differ in having either an A or a G at position 2 of the cloned sequence (**Fig. 2b** and **Supplementary Table 3** online). We refer to the A-containing sequence as KSHV-miR-K12-10a and the G-containing sequence as KSHV-miR-K12-10b. Database analysis identified perfect matches of KSHV-miR-K12-10a and KSHV-miR-K12-10b to different viral isolates. When we PCR-amplified and sequenced ten clones from the miRNA-containing segment of the BCBL1-borne KSHV genomic DNA, we were able to recover perfect matches only to KSHV-miR-K12-10a, suggesting that KSHV-miR-K12-10a pre-miRNA is partially edited by a dsRNA-specific deaminase that converts the adenosine at position 2 into inosine. This editing event would also change a glycine to a serine in the kaposin A and C proteins. To assess the extent of A-to-I conversion in the kaposin mRNA/pre-KSHV-miR-K12-10 transcripts, we prepared a poly(A)-specific cDNA library from our BCBL1 cells. Sequencing of 14 clones of PCR-amplified kaposin mRNA/pre-KSHV-miR-K12-10 segment revealed 12 A-to-I substitutions.

BCBL1 cells harbor replication-competent KSHV, and upon treatment with tetradecanoyl phorbol acetate (TPA) these cells undergo the complete program of KSHV gene expression, resulting ultimately in viral replication and the release of mature virions¹⁴. We asked whether KSHV miRNAs might be differentially expressed upon induction of the virus lytic cycle. We isolated total RNA from BCBL1 cells at various times after TPA treatment and examined KSHV miRNA expression by northern blot analysis. We found that KSHV-miR-K12-10 expression was increased up to 6.7-fold in TPA-induced cells relative to untreated cells, whereas the expression of the miRNAs confined to the intronic region was unaffected (**Fig. 2c**).

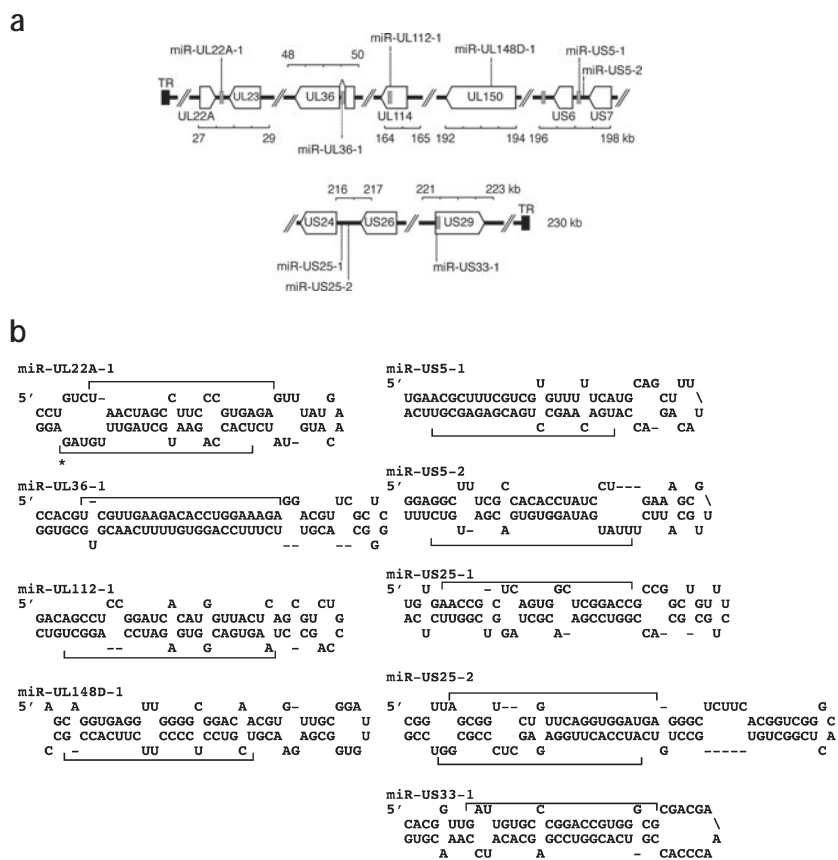


Figure 4 | Genomic positions and predicted secondary structures of HCMV pre-miRNAs. (a) Diagram of the miRNA-containing fragments of the HCMV genome. Cloned miRNAs are shown as dotted lines and predicted miRNAs as gray rectangles. The miRNAs encoded from the (+) strand of the genome are shown above the genome and those deriving from the (-) strand of the genome are below. The large open arrows indicate the orientation of the viral ORFs. (b) Predicted HCMV pre-miRNAs. The predominantly cloned miRNAs are indicated with a black line bracket. An asterisk is used to highlight low-abundance small RNAs.

Features of the MHV68 miRNAs

To address evolutionary relationships in the gammaherpesvirus family, we studied MHV68, which is the closest relative to human EBV and KSHV¹¹. All of the nine cloned miRNAs cluster within a 6-kb region (**Fig. 3a**). This region also contains eight previously noted tRNA genes, which, surprisingly, were not aminoacylated¹⁹. Because all of the cloned MHV68 miRNAs are located immediately downstream of these tRNA sequences (**Fig. 3a**), it suggests that the pre-miRNAs are transcribed by pol III. The pol III tRNA promoter of a conventional tRNA gene consists of an A box (positions +8 to +19) and a B box (positions +52 to +62), followed by a run of oligo(T) as a terminator signal²⁰. Secondary structure prediction of the putative tRNA-pre-miRNA primary transcripts (**Fig. 3b**) indicates that one or two 20–25-base-pair pre-miRNA hairpins follow the tRNA structure. In each case a run of oligo(T) is identified at the end of the first or the second pre-miRNA. Sequence analysis indicated that the eight tRNA promoters may express a total of 14 distinct pre-miRNAs (**Supplementary Table 4** online), 9 of which we identified by cloning.

To confirm that MHV68 miRNAs are embedded in the pol III tRNA primary transcripts, we performed a northern blot analysis on the three most abundant MHV68 miRNAs using probes

matching to either the mature miRNA or the first 20 nt of the upstream tRNA (Fig. 3c). The probes complementary to miR-M1-1, miR-M1-5 and miR-M1-7-5p detected bands of ~20 nt, ~60 nt and ~200 nt (Fig. 3c, panels 1, 3 and 5). The ~200-nt band was also detected with probes specific for the upstream tRNAs (Fig. 3c, panels 2, 4 and 6) in addition to the ~75-nt band of the mature tRNA (Fig. 3c, panels 4 and 6).

Features of the HCMV miRNAs

HCMV is a member of the beta-herpesvirus subfamily²¹. In contrast to the viruses described above that were studied during their latency, we cloned the HCMV small RNAs from RNA isolated from lytically infected primary cells. The proportion of cloned rRNA degradation was notably higher than with the libraries we prepared from latently infected cells, which may be due to an increased RNA turnover in apoptotic cells (Supplementary Table 2 online). We cloned 424 small RNAs that mapped to the virus genome; 60% of these sequences were cloned multiple times and corresponded to nine miRNAs. In contrast to KSHV and MHV68, the HCMV miRNAs are spread across the viral genome (Fig. 4a). Three of the nine miRNAs are transcribed from the complementary strand of known ORFs, five miRNAs are located in intergenic regions and one is located within an intron (Fig. 4a). The remaining sequences were cloned mostly once, sometimes twice (Supplementary Table 6 online), and did not represent miRNAs but were rather random degradation products of the abundant viral transcripts.

DISCUSSION

We developed a new prediction algorithm that provides a reliable starting point for guiding experimental identification of viral miRNAs. In contrast to other miRNA prediction software, our program does not rely on evolutionary conservation to identify miRNA genes. Consistent with low prediction scores, we were unable to identify miRNAs by cloning in several small-genome RNA viruses, but we confirmed many predicted miRNA genes by cloning the miRNAs from large-genome DNA viruses.

We determined the sensitivity (fraction of cloned miRNAs for which we predicted precursors) and specificity (fraction of predictions for which mature miRNAs were cloned) of our prediction method for different values of the prediction threshold for all the viruses for which we obtained cloning data (Supplementary Fig. 2 online). For three of the four viruses from which we cloned miRNAs, the specificity increased markedly when the score threshold reached a value of 0, still keeping the sensitivity above 40%. Thus, the threshold score that we set for prediction based on the training sets provides a good tradeoff between sensitivity and specificity for the virus data. Using this score threshold we predicted the location of 13 (39%) of the 33 precursors of cloned miRNAs, and we did not predict miRNA precursors in the large genomes of HHV3, HHV6 and HHV7. This recovery rate for viral miRNAs is lower than for the training set, in part because of some unusual features of viral miRNAs, such as the relatively short stems of the pol III-transcribed MHV68 miRNAs. But MHV68 was still predicted to contain miRNAs with a probability of 97.6%. In contrast, the probability of encoding miRNAs was considerably lower for the similar-sized genomes of HHV3, HHV6 or HHV7 because all of the stem-loops predicted from these viruses had relatively low scores. If we disregard the direction of transcription, we made 23 unique predictions, of which 14 (61%) were experimentally confirmed. It is

conceivable that some predicted miRNAs were undetectable because they may be under the control of promoters that are only active at one of the various latent or lytic stages of infection or depend on the specific infected cell type. For example, EBV, KHSV and MHV68 can infect both B cells and epithelial cells^{11,22,23}, which have distinct gene expression programs.

In many organisms, RNA silencing is part of an innate immune response against RNA viruses (reviewed in refs. 24 and 25) and transposable elements^{26–28}. In response, counter-defense strategies to thwart the host response were found in plant viruses, the insect Flock house virus and were proposed for vaccinia and influenza viruses (reviewed in ref. 25). But we could not detect any viral siRNA in HCV-, YFV- or HIV1-infected cells, and showed that HCV does not inhibit RNAi of a cellular gene. Thus, it is likely that human RNA viruses effectively hide dsRNA replication intermediates and do not interfere with the RNA silencing machinery.

The human gammaherpesviruses EBV and KSHV pose an important health problem and are closely associated with various types of cancer^{22,23}. The new discovery of miRNAs may shed some light on the mechanism of herpesvirus pathogenesis. All of the KSHV miRNAs were found in a locus previously reported to encode transforming properties^{16,29}. The most interesting segment of this transforming region is the kaposin gene, which is expressed in three different isoforms, kaposin A, B and C, depending on the translation initiation codon used¹⁸ (Fig. 2a). Notably, KSHV-miR-K12-10 is located within the kaposin mRNA transcript, in a position overlapping the K12 ORF. Excision of the K12-10 pre-miRNA presumably abolishes protein expression from the remaining fragments, which are missing the cap or poly(A) tail. Similar to the EBV *BHRF1* gene and its miRNAs, whose expression increases along with lytic replication, kaposin and its embedded KSHV-miR-K12-10 are upregulated by induction of the lytic stage. This indicates that viruses can express both noncoding and coding sequences simultaneously, and presumably, depending on the expression levels, a certain fraction of transcripts may escape miRNA processing to then serve as template for protein synthesis.

The KSHV-miR-K12-10 is edited at position 2 of the mature sequence. It is unclear whether this editing has regulatory function, considering that the A-to-I change is within the region of an miRNA shown to be the most critical region involved in target recognition³⁰. The sequence and structural context of the edited A is optimal for recognition by adenosine deaminases acting on RNA, which prefer a 5' uracil and a mismatched cytosine on the opposite strand³¹. Editing of a cellular miRNAs has been reported previously, but at a substantially lower frequency than observed for this viral miRNA³².

Analysis of many primary transcripts of miRNAs indicated that they generally originate as pol II transcripts³, which permits tissue-specific and developmental regulation of miRNA transcripts by the same mechanisms as govern the expression of protein-coding genes. It was therefore surprising to identify miRNA transcription of MHV68 miRNAs initiated from a tRNA-specific pol III promoter, which is largely an unregulated promoter active in any cell type. The discovery that tRNA promoters drive MHV68 miRNA expression explains the puzzling detection of uncharged viral tRNAs in germinal centers of MHV68 latently infected mice¹⁹, and suggests that these tRNA-like molecules are nonfunctional remnants from the expressed miRNA promoter region. Furthermore, the MHV68 pre-miRNAs have an unusual short-hairpin structure, suggesting that their processing and nuclear export mechanism may also be distinct from host cell miRNA maturation.

Almost all of the identified viral miRNA genes are located within regions of the genome that are specific for each virus analyzed, and the cloned miRNAs share no sequence similarity among themselves or with host cell miRNAs, suggesting a short and independent evolutionary history. Only HCMV miRNAs seem to be partially conserved with respect to the chimpanzee CMV, but not the more distant mouse CMV. The detection of unusual pol III–driven miRNAs further supports our hypothesis that viral miRNAs evolved independently of each other.

The cellular targets for most of these viral miRNAs remain to be identified. Those miRNAs that are generated from a strand opposite a protein-coding segment may act like siRNAs in regulating the level of viral protein. CMV-miR-UL112-1 might target UL114, a homolog of the mammalian uracyl-DNA glycosylase. UL114 is required for efficient viral DNA replication³³, suggesting that the miRNA expression may control viral replication. A similar function has previously been attributed to a miRNA expressed in EBV⁷. Because there is no substantial conservation between viral miRNAs or between viral miRNAs and the host cell miRNAs, the computational identification of regulated host or viral genes will be challenging.

METHODS

miRNA predictions. One essential feature of pri- and pre-miRNAs appears to be a stem-loop secondary structure that is recognized at various steps of biogenesis. Therefore, in the first step of our algorithm we identified genomic regions (50–100 nt long) that, when transcribed, are predicted to assume the same stem-loop secondary structures, independent of the precise start and end of the transcript. The detailed calculation of conserved stem-loops is described in the **Supplementary Methods** online.

We then used the SVM methodology to predict new miRNA precursors among a set of conserved stem-loops. The SVM requires that each data point (in our case a stem-loop) be described in terms of a vector of features, and we included the free energy of folding, the length, the length of the longest symmetrical stem, the count of A, C, G and U nucleotides in the symmetrical stem, and the number of A-U, G-C and G-U pairs in the predicted minimal energy structure, among other criteria. We trained the classifier using conserved stem-loops from human miRNA loci as positives and sequences that are not likely to be processed by the RNAi machinery—mRNAs, rRNAs, tRNAs, random regions of the human genome and viral genomes—as negatives. Although our data sets of positives and negatives are not linearly separable, using a quadratic kernel and penalizing misclassification of positives eight times more than the misclassification of negatives allowed to obtain a model that assigns a positive score to 71% of the true positives and to only 3% of false positives.

Given the distributions of scores for the set of positive and negative examples, and the scores for the stable stems identified in a given genome, we estimated the minimal number of miRNA precursors expected in each genome as follows. We assume that the probability p that a conserved stem contains a miRNA is only a function of the score assigned to the stem by the SVM. We expect that the function $p(s)$ has a step-like monotonic behavior, being close to zero for negative values of s and becoming asymptotically close to 1 for large positive scores. To concretely specify $p(s)$, we first choose a suitable parametric functional expression (for instance, $p(s) = 0.5 \times (1 + \tanh(as + b))$, in which a and b are the parameters) and then fix its parameters by maximizing the likelihood of the training set, defined as $\prod_{i \in \text{positives}} p(s_i) \times \prod_{j \in \text{negatives}} (1 - p(s_j))$. The expression for the probability of having

exactly m miRNAs is $P(m) = \sum_{\text{all } m\text{-tuples of candidates}} p(s_1) \dots p(s_m) \times \prod_{i \notin m\text{-tuple}} (1 - p(s_i))$. The fifth column in **Table 1** contains the quantity $\sum_{m \geq 1} P(m)$, which corresponds to the probability of finding at least one miRNA precursor in the set of candidates.

RNA preparation and small-RNA cloning. Total RNA extraction and cloning was performed as described previously¹² with the following modifications for the KSHV, MHV68 and HCMV libraries. We spiked 200 μg total RNA with 0.2 nM radiolabeled 19- and 24-nt oligoribonucleotides containing the *PmeI* restriction site (5'-CGUACGCGGGUUUAAACGA-3' and 5'-CGUACGCGGAAUAGUUUAAACUGU-3'), size-fractionated and ligated to a 5'-adenylated 3' adapter oligonucleotide (5'-AppTTTAACCGCAATTCAG-L-3'; Ap, Adenylate; p, phosphate; L, C6-amino linker (Dharmacon)) using the Rnl2 (1–249) ligase¹³ in the absence of ATP. The reaction was performed in a final volume of 20 μl containing 0.1 μg of the Rnl2 (1–249) ligase, 5 μM of the 3' adapter oligonucleotide, 0.01 M MgCl_2 , 0.01 M 2-mercaptoethanol, 0.05 M Tris-HCl (pH 7.6) and 0.1 mg acetylated bovine serum albumin. After gel purification, the ligation product was directly joined to the 5' adapter chimeric DNA/RNA oligonucleotide (5'-ACGGAATTCCTCACTrArArA-3'; r, ribonucleotide) using T4 RNA ligase 1 (NEB) in the presence of ATP and in the same reaction solution as above. The final ligation product was reverse transcribed using the primer 5'-GACTAGCTGGAATTCGCGGTAAA-3' and PCR amplified using the reverse transcription primer and the primer 5'-CAGCCAACGGAATTCCTCACTAAA-3'. After purification, the PCR product was digested with *PmeI* to eliminate size marker sequences and an additional PCR was performed to introduce the nonpalindromic *BanI* restriction sites with primers 5'-CAGCCAA CAGCACCGAATTCCTCACTAAA-3' and 5'-GACTAGCTTGGT GCCGAATTCGCGGTAAA-3'. The *BanI*-digested PCR products were then concatemerized using T4 DNA ligase and later ligated into the commercial T/A cloning vector pCR2.1 (Invitrogen). Individual colonies were screened for the presence of inserts by PCR and the insert-containing PCR products were sequenced as described elsewhere⁷.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank members of the Tuschl laboratory for helpful comments on the manuscript, and D. Marks, B. John and E. Clambey for discussions. M.Z. thanks E. van Nimwegen for useful suggestions. S.P. is supported by a grant from the Breast Cancer Alliance and by the Leukemia and Lymphoma Society (5554-05). A.S. is supported by the Swiss National Science Foundation grant 205321-105945 to M.Z.; B.D.L. and C.M.R. are supported by a grant from the US Public Health Service (CA85883). G.R. is supported by a grant from the American Cancer Society (PF-02-016-01-MBC). L.F.v.d. is supported by the University of Colorado Health Science Center School of Medicine and Howard Hughes Medical Institutes Investigator Award and the US National Institutes of Health (CA103632). The research was also funded by a grant from the National Institutes of Health (P01 GM073047-01).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 27 December 2004; accepted 10 February 2005
Published online at <http://www.nature.com/naturemethods/>

1. Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
2. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
3. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060 (2004).

4. Tomari, Y. & Zamore, P.D. MicroRNA biogenesis: Drosha can't cut it without a partner. *Curr. Biol.* **15**, R61–R64 (2005).
5. Yi, R., Qin, Y., Macara, I.G. & Cullen, B.R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**, 3011–3016 (2003).
6. Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834–838 (2001).
7. Pfeffer, S. *et al.* Identification of virus-encoded microRNAs. *Science* **304**, 734–736 (2004).
8. Griffiths-Jones, S. The microRNA registry. *Nucleic Acids Res.* **32**, D109–D111 (2004).
9. Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B. & Bartel, D.P. Vertebrate microRNA genes. *Science* **299**, 1540 (2003).
10. Lai, E.C., Tomancak, P., Williams, R.W. & Rubin, G.M. Computational identification of *Drosophila* microRNA genes. *Genome Biol.* **4**, R42 (2003).
11. Speck, S.H. & Virgin, H.W. Host and viral genetics of chronic infection: a mouse model of gamma-herpesvirus pathogenesis. *Curr. Opin. Microbiol.* **2**, 403–409 (1999).
12. Pfeffer, S., Lagos-Quintana, M. & Tuschl, T. in *Current Protocols in Molecular Biology* (eds. Ausubel, F.M. *et al.*) 26.24.21–26.24.18 (Wiley Interscience, New York, 2003).
13. Ho, C.K., Wang, L.K., Lima, C.D. & Shuman, S. Structure and mechanism of RNA ligase. *Structure (Camb.)* **12**, 327–339 (2004).
14. Renne, R. *et al.* Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture. *Nat. Med.* **2**, 342–346 (1996).
15. Usherwood, E.J., Stewart, J.P. & Nash, A.A. Characterization of tumor cell lines derived from murine gammaherpesvirus-68-infected mice. *J. Virol.* **70**, 6516–6518 (1996).
16. Muralidhar, S. *et al.* Identification of kaposin (open reading frame K12) as a human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus) transforming gene. *J. Virol.* **72**, 4980–4988 (1998).
17. Li, H., Komatsu, T., Dezube, B.J. & Kaye, K.M. 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 (2002).
18. Sadler, R. *et al.* 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 (1999).
19. Bowden, R.J., Simas, J.P., Davis, A.J. & Efstathiou, S. Murine gammaherpesvirus 68 encodes tRNA-like sequences which are expressed during latency. *J. Gen. Virol.* **78**, 1675–1687 (1997).
20. Schramm, L. & Hernandez, N. Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* **16**, 2593–2620 (2002).
21. Roizman, B. & Pellet, P.E. The family herpesviridae: a brief introduction in *Fields Virology* (eds. Fields, B.N. *et al.*) 2381–2396 (Lippincott, Williams & Wilkins, Philadelphia, USA, 2001).
22. Kieff, E. & Rickinson, A.B. Epstein-Barr virus and its replication in *Fields Virology* (eds. Fields, B.N. *et al.*) 2511–2551 (Lippincott, Williams & Wilkins, Philadelphia, USA, 2001).
23. Moore, P.S. & Chang, Y. Kaposi's sarcoma-associated herpesvirus in *Fields Virology* (eds. Fields, B.N. *et al.*) 2803–2833 (Lippincott, Williams & Wilkins, Philadelphia, USA, 2001).
24. Baulcombe, D. RNA silencing in plants. *Nature* **431**, 356–363 (2004).
25. Ding, S.W., Li, H., Lu, R., Li, F. & Li, W.X. RNA silencing: a conserved antiviral immunity of plants and animals. *Virus Res.* **102**, 109–115 (2004).
26. Aravin, A.A. *et al.* The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* **5**, 337–350 (2003).
27. Sijen, T. & Plasterk, R.H. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* **426**, 310–314 (2003).
28. Wu-Scharf, D., Jeong, B.R., Zhang, C. & Cerutti, H. Transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-box RNA helicase. *Science* **290**, 1159–1162 (2000).
29. Kliche, S. *et al.* Signaling by human herpesvirus 8 kaposin A through direct membrane recruitment of cytohesin-1. *Mol. Cell* **7**, 833–843 (2001).
30. Doench, J.G. & Sharp, P.A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
31. Bass, B.L. RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **71**, 817–846 (2002).
32. Luciano, D.J., Mirsky, H., Vendetti, N.J. & Maas, S. RNA editing of a miRNA precursor. *RNA* **10**, 1174–1177 (2004).
33. Courcelle, C.T., Courcelle, J., Prichard, M.N. & Mocarski, E.S. Requirement for uracil-DNA glycosylase during the transition to late-phase cytomegalovirus DNA replication. *J. Virol.* **75**, 7592–7601 (2001).