Published online 3 May 2007

Nucleic Acids Research, 2007, Vol. 35, No. 10 ***e73***
doi:10.1093/nar/gkm244

Subtractive hybridization identifies novel differentially expressed ncRNA species in EBV-infected human B cells

Jan Mrázek[1], Simone B. Kreutmayer[1], Friedrich A. Grässer[2], Norbert Polacek[1,*] and Alexander Hüttenhofer[1,*]

[1]Innsbruck Biocenter, Division of Genomics and RNomics—Innsbruck Medical University, Fritz-Pregl-Strasse 3, 6020 Innsbruck, Austria and and [2]Institut für Mikrobiologie und Hygiene, Abteilung Virologie, Haus 47, Universitätskliniken, D-66421 Homburg/Saar, Germany

Received March 7, 2007; Accepted April 3, 2007

ABSTRACT

Non-protein-coding RNAs (ncRNAs) fulfill a wide range of cellular functions from protein synthesis to regulation of gene expression. Identification of novel regulatory ncRNAs by experimental approaches commonly includes the generation of specialized cDNA libraries encoding small ncRNA species. However, such identification is severely hampered by the presence of constitutively expressed and highly abundant 'house-keeping' ncRNAs, such as ribosomal RNAs, small nuclear RNAs or transfer RNAs. We have developed a novel experimental strategy, designated as subtractive hybridization of ncRNA transcripts (SHORT) to specifically select and amplify novel regulatory ncRNAs, which are only expressed at certain stages or under specific growth conditions of cells. The method is based on the selective subtractive hybridization technique, formerly applied to the detection of differentially expressed mRNAs. As a model system, we applied SHORT to Epstein–Barr virus (EBV) infected human B cells. Thereby, we identified 21 novel as well as previously reported ncRNA species to be up-regulated during virus infection. Our method will serve as a powerful tool to identify novel functional ncRNAs acting as genetic switches in the regulation of fundamental cellular processes such as development, tissue differentiation or disease.

INTRODUCTION

In the recent past, the importance of the surprisingly diverse class of non-coding RNA molecules (ncRNAs) has been widely recognized (1–4). The key feature of all ncRNAs is that they are not translated into proteins but rather exert their functions at the RNA level. NcRNAs have been identified in unexpectedly large numbers, with present estimates—based on bioinformatical approaches—in the range of many thousands per eukaryal genome (5,6). They play key roles in a variety of fundamental processes in all three domains of life, that is Eukarya, Bacteria and Archaea. Their functions include DNA replication and chromosome maintenance, regulation of transcription, RNA processing, translation and stability of mRNAs, and even regulation of stability and translocation of proteins (7–10). Many ncRNAs have been discovered fortuitously, suggesting they merely represent the tip of the iceberg.

One of the most prominent and abundant ncRNA classes consists of 21–23-nt long RNA species designated as microRNAs (miRNA) (7). MiRNAs play a crucial role in the posttranscriptional regulation of protein-coding genes during development, differentiation and cellular growth (7). In addition, viral or cellular encoded miRNAs were found to be involved in infections by DNA or RNA viruses (11–14). In the past, only very few ncRNAs, in addition to miRNAs, have been identified, which might be involved in the regulation of viral infections. We thus investigated the small ncRNA transcriptome of human B cells to identify ncRNAs expressed from the viral or cellular genome upon Epstein–Barr virus (EBV) infection.

Epstein–Barr virus, a DNA virus of the herpesviridae family with a linear double-stranded genome of ~172 kb, is associated with a heterogeneous group of human tumors including Burkitt's lymphoma (15). Until now, a total of 25 EBV-encoded ncRNAs, including 23 EBV-encoded miRNAs (16,17) as well as two ~170-nt long ncRNA species, designated as EBER 1 and EBER 2 (18,19), have been identified.

*To whom correspondence should be addressed. Tel: +43 (0)512 9003 70250; Fax: +43 (0)512 9003 73100; Email: alexander.huettenhofer@i-med.ac.at
Correspondence may also be addressed to Norbert Polacek. Email: norbert.polacek@i-med.ac.at

In general, identification of novel regulatory ncRNA species can be achieved by bioinformatical or experimental approaches or a combination of both (1,2,20). For experimental approaches, the generation of specialized cDNA libraries has been shown to represent a powerful tool (20). However, identification of novel ncRNAs is often masked by a background of highly abundant cellular 'house-keeping' ncRNAs, such as ribosomal RNAs, small nuclear RNAs or tRNAs. Thus, for identification of novel differentially expressed ncRNA candidates, e.g. those expressed upon viral infections only, novel experimental methods are required.

MATERIALS AND METHODS

Cells and EBV strains

The non-adherent EBV-transformed B cell lines B95.8-CBL (21) and BL41 (22) were cultured in RPMI 1640 supplemented with 10% FCS and antibiotics (100 U penicillin ml^{-1} and 100 µg streptomycin ml^{-1}). BL41 is an EBV-negative Burkitt's lymphoma line, B95.8-CBL was established by transformation of primary human cord blood lymphocytes (CBLs) using the B95.8 strain of EBV.

Subtractive hybridization of ncRNA transcripts (SHORT)

Total RNA was isolated from the Burkitt's lymphoma cell line BL41 (*driver* RNA) and from CBLs immortalized with the Epstein–Barr virus strain B95-8 (*target* RNA) by the TRIzol method. From 150 to 200 µg total RNA were separated by denaturating 8% polyacrylamide gels and size-fractionated by the excision of RNAs in the size range between 20–70 nt (small fraction) and 70–500 nt (large fraction). RNAs were passively eluted into 12 ml 0.3 M NaOAc (pH 5.2) over night at 4°C, the eluate filtered trough a 0.2 µm syringe filter and concentrated to ~200 µl with the vivaspin 15R concentrator (Sartorius, Goettingen, Germany) at 4°C. Following ethanol precipitation and lyophilization of the RNA pellet, the 3′-ends of 5 µg of the size-selected RNAs were modified using 5 units poly(A) polymerase (Invitrogen, Carlsbad, CA, USA) at 37°C for 1 h in 50 µl buffer containing 50 mM Tris/Cl (pH 8.0), 200 mM NaCl, 12 mM $MgCl_2$, 2 mM $MnCl_2$, 1 mM DTT, 0.4 mM EDTA and 40 units ribonuclease inhibitor (Fermentas, Hanover, MD, USA).

The *driver* RNA was A-tailed in the presence of 4 mM ATP, while the *target* RNA was C-tailed by using 4 mM CTP instead. Subsequently, the 3′-tailed RNAs were phenol/chloroform extracted, ethanol precipitated and resuspended in 7 µl H_2O. Then the RNAs were treated with 10 units tobacco acid pyrophosphatase (Epicentre) for 1 h at 37°C in 10 µl buffer containing 50 mM NaOAc (pH 6.0), 1 mM EDTA, 0.1% β-mercaptoethanol, 0.01% Triton X-100 and 40 units ribonuclease inhibitor. Subsequent to phenol/chloroform extraction and ethanol precipitation, the RNA pellets were dissolved in 10 µl H_2O and an adaptor was ligated to the 5′-ends in a 20 µl reaction containing 50 mM HEPES/NaOH (pH 8.0), 10 mM $MgCl_2$, 10 mM DTT, 0.1 mg/ml BSA, 20 µM adaptor, 1 mM ATP, 40 units ribonuclease inhibitor and 10 units T4 RNA ligase (Fermentas) at 4°C over night. The sequences of the 5′-DNA/RNA-hybrid-adapter were ACGGAATTCCTCACTGAG and GTCAGCAATCCC TAACGAG for the *driver* and *target* RNA, respectively, whereas RNA nucleosides are underlined. After a second phenol/chloroform extraction and ethanol precipitation, the *driver* RNA was reverse transcribed employing 5 µM of the anchored primer TTTTTTTTTTTTTTTTTTTN (primer a), while for the *target* RNA the anchored primer AGGAGCCATCGTATGTCGGGGGGGGGN (primer b) was used in a 20 µl reaction with 200 units reverse transcriptase (Invitrogen) at 42°C for 1 h. 1/40 volume of the *driver* cDNA was PCR amplified in 20 cycles by using the two 5′-biotinylated primers ACGGAATTCCTCAC TGAG (primer c) and primer a. Subsequent to PCR product purification (QIAGEN, Valencia, CA, USA), the DNA was eluted into 50 µl H_2O. 25 µl of this primary PCR was then used in an 'unequal PCR' reaction with 10 µM of the 5′-biotinylated primer c as the sole primer, which linearly amplified the (+)-strand that corresponds to the sequence of the original *driver* RNA. After 20 cycles the 5′-biotinylated DNA was purified (QIAGEN), eluted into 50 µl H_2O and lyophilized. 15 µl of the reverse transcribed *target* RNA was diluted with H_2O to 45 µl and the *target* RNA template was hydrolyzed by adding 5 µl 1 M NaOH and incubation at 70°C for 20 min. Subsequently to the neutralization by adding 45 µl H_2O and 5 µl 1 M HCl, the single stranded *target* cDNA was purified (QIAGEN), eluted into 50 µl H_2O, lyophilized and dissolved in 15 µl H_2O.

For the subtractive hybridization, 5 µl of the single stranded *target* cDNA was used to redissolve the pellet of the 5′-biotinylated *driver* DNA. The molar ratio between *driver* and *target* was calculated to be ~50000:1. After the addition of 5 µl of H_2O, the *driver* and *target* were allowed to hybridize by heating to 98°C for 3 min, followed by incubations at 65°C for 2 min and 42°C for 15 min. Subsequently the volume was increased to 20 µl and the buffer adjusted to 5 mM Tris/Cl (pH 7.5), 1.0 M NaCl and 0.5 mM EDTA before mixing the sample with 10 µl strepavidin coated magnetic beads (Dynal Biotech, Oslo, Norway). The sample was vortexed for 15 min at 25°C before the tube was placed in a magnet for 3 min to allow bead separation. The supernatant, containing single stranded *target* cDNA that was not captured via the 5′-biotinylated *driver*, was removed and used for a consecutive round of subtractive hybridization using fresh 5′-biotinylated *driver* DNA. After the fifth round, the *target* cDNA was PCR amplified using the forward and reverse primer pair GTCAGCAATCCCTAACGAG (primer d) and AGGAGCCATCGTATGTCG (primer e), respectively, and cloned into the pGEM-T vector (Promega, Mannheim, Germany). This experimental design, applying five rounds of SHORT, can be performed in one working day. Theoretically more rounds can be envisioned, however it is of note that with every round of SHORT, the problem of target cDNA dilution increases. In addition we found that the bands in the control PCRs done after every round of SHORT did not change significantly after the fourth round (data not shown).

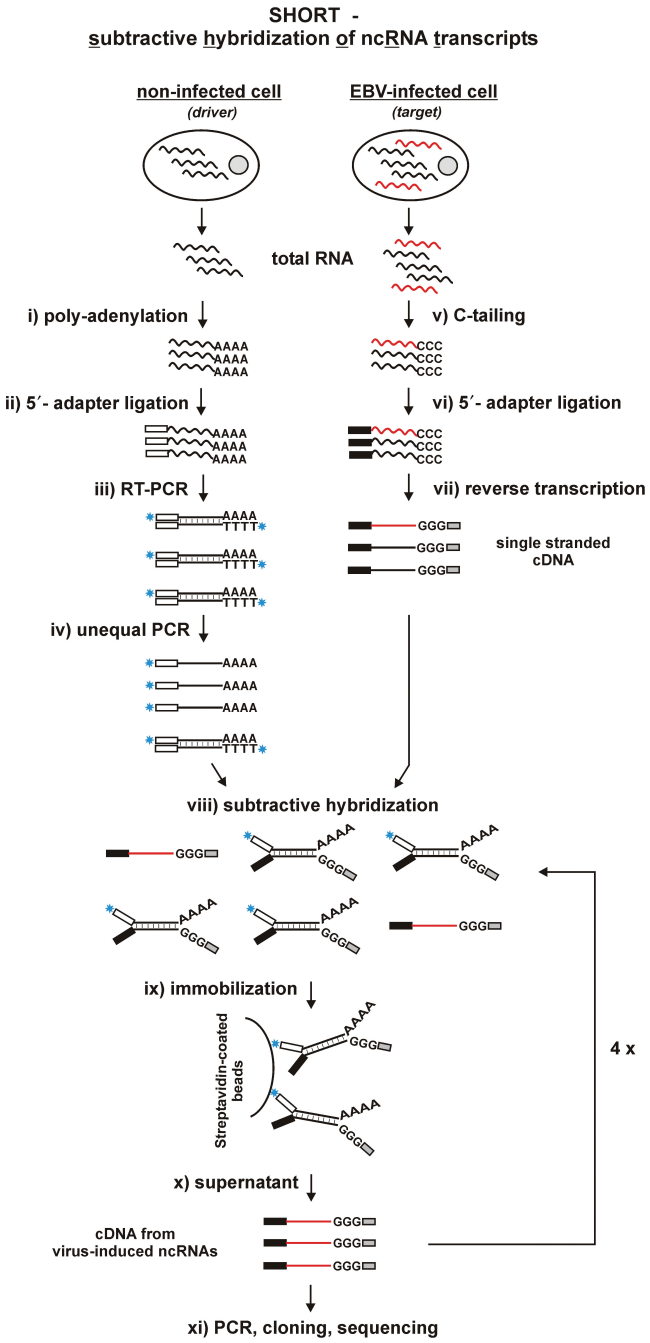


Figure 1. The principle of SHORT. Total *driver* RNA from non-infected cells was size-selected by denaturing PAGE to obtain RNAs in the size range from 20 to 500 nt. Size-selected RNAs were 3′ poly-adenylated using poly(A) polymerase (step i). Subsequently, a DNA/RNA adaptor oligonucleotide was ligated to the 5′-ends of the ncRNAs which served as a 5′-primer site in an RT-PCR reaction (step ii). RT-PCR was performed using an anchored oligo (dT)-primer complementary to the 3′ poly(A)-tail and a 5′-biotinylated adaptor primer as forward primer (step iii). The biotin moiety is shown as blue asterisk. Resulting cDNAs were used as templates in an unequal PCR reaction using excess amounts of 5′-biotinylated adaptor primer to linearly amplify single-stranded and 5′-biotinylated sense cDNA copies (step iv). *Target* RNA, derived from EBV-infected cells, was C-tailed at the 3′-ends (step v). The 5′-ends of the *target* ncRNAs were ligated to an RNA/DNA oligonucleotide linker that differs in sequence from the one used for amplification of the driver RNA (step vi). C-tailed *target* RNAs were subsequently reverse transcribed into single stranded cDNAs by an anchored oligo d(G) primer (step vii) and allowed to hybridize with excess of the single stranded and biotinylated *driver* DNA (step viii). Transcripts that are equally expressed in both *driver* and *target* cells form duplexes and thus were captured via the 5′-biotin moiety of the driver-DNA on streptavidin-coated magnetic beads (step ix). *Target* cell specific transcripts (red) remained single stranded and were therefore enriched in the supernatant (step x). In order to increase the selective efficiency of SHORT, subtractive hybridization (steps viii to x) was repeated four times.

DNA sequencing and sequence analysis

cDNA clones were sequenced and the sequences were analyzed with the LASERGENE sequence analysis program package (DNASTAR, Madison, WI, USA) as described (23).

Northern blot analysis

Total RNA (20–30 μg) from the non-infected B cell line BL41 or the EBV-infected cell line B95.8-CBL were separated on 8% denaturing polyacrylamide gels (7 M urea, 1× TBE buffer), transferred onto nylon membranes and probed with 5′-[^{32}P] end-labeled antisense DNA probes as described (23). Northern blot signals were quantified using a Molecular Dynamics Storm PhosphorImager (Image quant software version 5.0). All quantified northern blot signals were normalized to the band intensities of the 5.8 S rRNA loading control. Comparison of the quantified northern blot bands form EBV-infected to the uninfected control revealed the level of up-, or down-regulation.

RESULTS

In order to identify differentially expressed ncRNAs (from the EBV genome or from the cellular genome), we have adapted an experimental approach, designated as 'subtractive hybridization', which previously has been applied to the identification of differentially expressed mRNAs (24). By this new method, which we named 'SHORT', differentially expressed ncRNAs, in the size range from 20 to 500 nt, are significantly enriched in a size selected cDNA library encoding small ncRNAs (for a detailed description see Figure 1 and Materials and Methods section). NcRNAs were subdivided into two fractions, in the size range between 20–70 nt (small fraction) and 70–500 nt (large fraction), since smaller cDNA species are known to be more efficiently cloned into plasmid vectors than longer cDNAs.

The general principle of the subtractive hybridization technique is the use of excess amounts of RNA transcripts from control cells (in our case from uninfected B cells, also designated as *driver* cells) for hybridization with limiting amounts of single-stranded cDNAs generated from RNA transcripts of the *target* cells (i.e. EBV-infected human B cells). The conceptual novelty of SHORT can be highlighted as follows: since ncRNAs are usually not poly-adenylated, for reverse transcription of RNAs poly(A) or poly(C) tails, respectively, were added at their 3′-ends (20) (Figure 1). In addition, adaptor sequences were added to the 5′-ends of both *driver* and *target* transcripts to generate PCR priming sites. As many

ncRNAs are highly structured, which might impede hybridization and since, in addition, RNA is known to be thermally more unstable than DNA, we converted both the *driver* and *target* RNAs into single-stranded sense- or anti-sense cDNAs, respectively. The conversion of the *driver* RNA into cDNA has the additional advantage that driver transcripts can be amplified by RT-PCR prior to conversion into single-stranded sense cDNAs. Therefore, a large access of *driver* DNA over target DNA can be employed in the selective hybridization step. This might be especially important in cases where the amount of cellular RNA is limited. Subsequently, *driver* and *target* cDNAs were allowed to hybridize (Figure 1). cDNA species derived from ncRNAs that are equally expressed in both cells form duplexes. Since *driver* cDNAs were generated to contain a biotin moiety at their 5′-ends, cDNA duplexes could be captured on streptavidin-coated magnetic beads (Figure 1). *Target* cell specific transcripts remained single stranded and therefore became enriched in the supernatant.

To verify selective amplification of virus-induced ncRNAs, cDNAs in the supernatant were PCR-amplified and size-separated by PAGE and compared to the unsubtracted target cDNA pool from step vii) of Figure 1. In both analyzed size fractions (i.e. from 20 to 70 nt: small fraction and from 70 to 500 nt: large fraction), PCR-derived DNA patterns differed considerably in size and abundance compared to the unsubtracted target control sample (Figure 2a). This is consistent with the SHORT method reducing or eliminating abundant cDNA species, present in both *driver* and *target* pools.

After five subtraction rounds of SHORT, we cloned PCR-amplified cDNAs and sequenced a small number of ~250 clones from each cDNA library (20). In total, we analyzed 506 sequences, which were grouped into 208 unique sequences (designated as contigs). Sequence analysis confirmed that SHORT changed the cDNA profile of the subtracted cDNA library compared to the unsubtracted control cDNA library: while in the unsubtracted library of the large ncRNA fraction 85% of all sequences originated from known 'house-keeping' ncRNAs (such as tRNAs, 28S rRNA, 5.8S rRNA and 5S rRNA), this group was significantly reduced to only 23% after SHORT (Figure 2b). Conversely, the fraction of potential novel ncRNA candidates and EBV-encoded ncRNAs increased from 10 to 40% of analyzed sequences. The fact that 7% of sequences in the subtracted cDNA library originated from the EBV-encoded transcripts EBER 1 and 2 served as a valuable internal control for specific ncRNA enrichment, since in the unsubtracted cDNA library control (in which cDNAs were also derived from EBV-infected cells), not a single EBV-encoded cDNA sequence could be identified (Figure 2b). Sequence analysis of the small fraction also revealed that house-keeping ncRNAs present in both *driver* and *target* pools could be reduced (data not shown). Notably, from 217 analyzable sequences of the small fraction, 81 sequences (or 37%) belong to the class of miRNAs. Twenty-three cDNA sequences (11%) could not be assigned to any known class of ncRNAs and thus represent potentially candidates for novel ncRNA species.

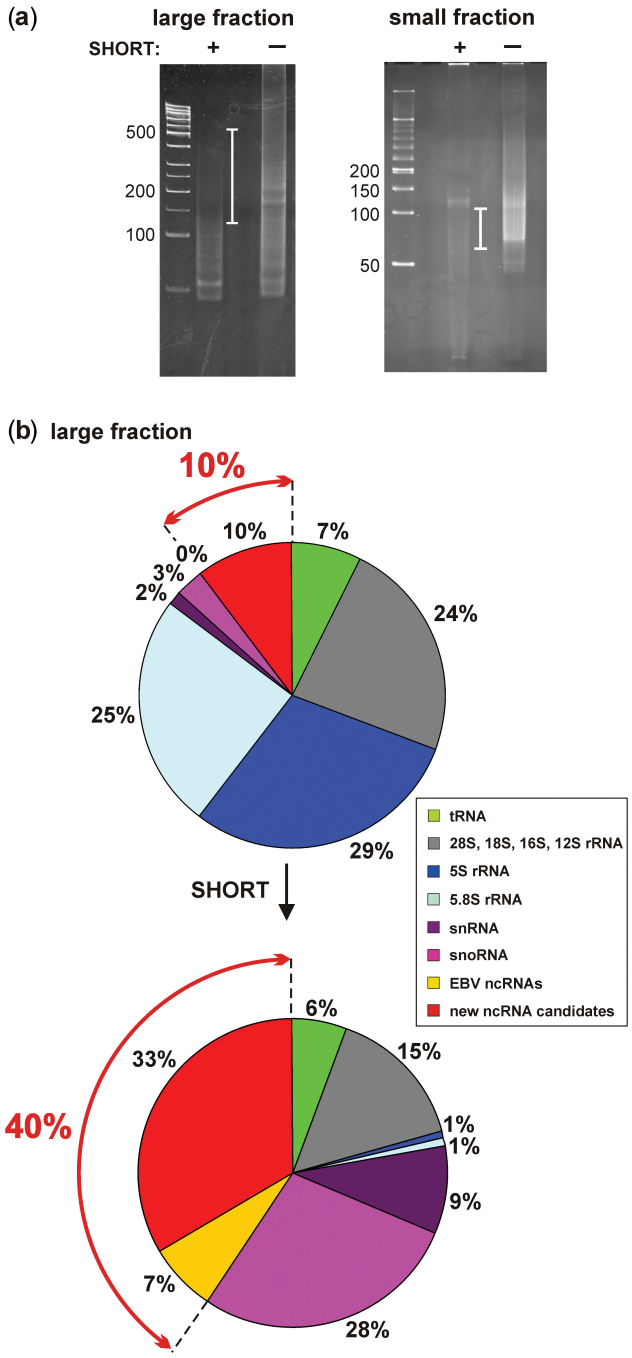


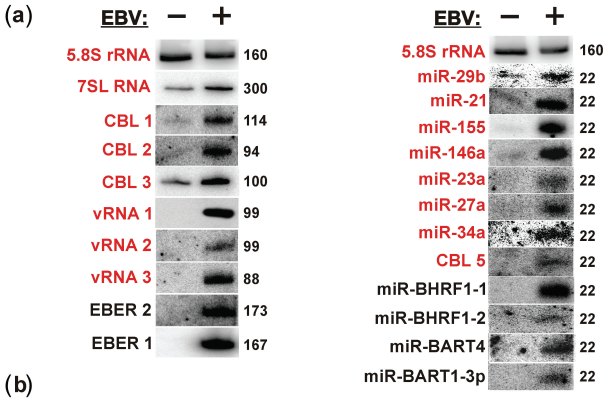
Figure 2. SHORT changes the small ncRNA transcriptome. (**a**) cDNAs originating from ncRNAs of EBV-infected B cells were PCR amplified before (−) or after (+) five rounds of subtractive hybridization (SHORT) and applied on a denaturing polyacrylamide gel. The cDNAs in the size range between ~120–500 nt (large fraction) and ~70–120 nt (small fraction) that were cut out, eluted and cloned into the pGEM-T vector are indicated by white bars. (**b**) Sequencing results of 250 clones of the large fraction library (originating from ncRNAs in the size range between 70 and 500 nt) of EBV-infected cells after five rounds of SHORT is compared to ~70 clones of the unsubtracted EBV-infected library. SHORT changed the ncRNA transcriptome significantly since the fraction of potentially interesting ncRNAs (including new ncRNA candidates and EBV-derived sequences) increased from 10% in the unsubtracted, to 40% in the subtracted library. Concomitantly, the fraction of known 'house-keeping' ncRNAs (especially tRNAs and rRNAs) decreased dramatically (from 85 to 23%).

To investigate whether ncRNAs, identified by SHORT, are indeed differentially expressed upon EBV infection, we chose 14 candidates from the large and 23 candidates from the small fraction from the 'non-house-keeping' ncRNA classes and investigated their expression levels by northern blot analysis. Thereby, expression of 29% of ncRNAs from the large and 26% of the sequences from the small fraction libraries were shown to be significantly up-regulated in the EBV-infected cell line compared to uninfected cells (Figure 3).

In particular, host cell-encoded vault associated ncRNAs 1–3 (98, 88 and 88 nt in size, respectively), were up-regulated by at least 1200-, 20- and 40-fold, respectively, in EBV-infected cells (Figure 3). Interestingly, in our screen we also have identified a novel ncRNA candidate (termed CBL-3), which were 4-fold up-regulated in EBV-infected cells compared to uninfected cells (Figure 3 and Table 1). CBL-3 shares compelling sequence and structural similarities with the known vault RNAs (vRNAs) (Figure 4). Furthermore, vRNA 1 and this putative novel vRNA (CBL-3) were found to be significantly up-regulated also in another B cell line, namely when comparing non-infected to EBV-infected BL2 cells (data not shown). Expression of vRNAs does not appear to be a general cellular response mechanism to viral infections, since in HIV infected T-cells we could not detect a comparable up-regulation of vRNAs (data not shown).

In addition to vault ncRNAs, we also identified two other novel ncRNA candidates from the large cDNA library up-regulated in EBV-infected B cells. The first RNA is a 144-nt long ncRNA transcript encoded on chromosome 19 (CBL-1) whose expression was stimulated 4-fold upon EBV infection (Figure 3a and Table 1). Interestingly, 14 different gene copies of CBL-1 could be identified on chromosome 19 mapping within a region of 2.2 Mb which are encoded on the plus or minus strand, respectively (Figure 5). Sequence identities of the various paralogs range from 98.3 to 100%. The second novel

(a)



(b)

Name	ncRNA class	↑ in EBV$^+$ cells
large fraction		
7SL RNA	SRP RNA	**3x**
CBL 1	unknown	**4x**
CBL 2	rpt repeat	**4x**
CBL 3	unknown	**4x**
vRNA 1	vault RNA	**> 1200x**
vRNA 2	vault RNA	**> 20x**
vRNA 3	vault RNA	**> 40x**
EBER 1	EBV ncRNA	**+**
EBER 2	EBV ncRNA	**+**

Name	ncRNA class	↑ in EBV$^+$ cells
small fraction		
miR-29b	micro RNA	**2x**
miR-21	micro RNA	**7x**
miR-155	micro RNA	**29x**
miR-146	micro RNA	**10x**
miR-23a	micro RNA	**3x**
miR-27a	micro RNA	**3x**
miR-34a	micro RNA	**3x**
miR-BHRF1-1	micro RNA	**+**
miR-BHRF1-2	micro RNA	**+**
miR-BART4	micro RNA	**+**
miR-BART1-3p	micro RNA	**+**
CBL 5	unknown	**3x**

Figure 3. Northern blot analysis of differentially expressed ncRNA candidates. (**a**) Northern blot scans of ncRNAs in uninfected BL41 (EBV−) and EBV-infected CBL cells (EBV+) are shown for the large (RNAs between 70 and 500 nt) and the small (20–70 nt) fraction. The size of the ncRNAs, as estimated by comparison with an internal RNA marker, is indicated on the right. Names of ncRNA candidates are in red and black for host- and EBV-encoded species, respectively. Novel ncRNA candidates identified in this study are referred to as CBL-1–5 (for cord blood lymphocyte derived ncRNAs). The 5.8S rRNA serves as internal RNA loading control. (**b**) Quantification of ncRNA expression of differentially expressed transcripts. The name and class of all differentially expressed ncRNAs identified in this study are indicated as well as the extend of up-(↑) regulation in the EBV-infected cell line (EBV$^+$) compared to the uninfected control. Novel ncRNA candidates are shown with grey background and EBV-encoded transcripts are underlined.

Table 1. Novel ncRNA candidates

	Sequence	cDNA (nt)	N. blot (nt)	Remarks
CBL-1	GGAGCCATTGTGGCTCAGGCCGGTTGCGCCTGCCCTCGG GCCCTCACGGAGGCGGGGGTTCCAGGGCACGAGTTCGA GGCCAGCCTGGTCCACATGGGTCGGAAAAAAGGATTT	114	114	Repeated locus on chromosome 19; intergenic
CBL-2	ATAGACAGCGTGGTCTAGTGGTTAAGAGCACAAGGCTT GGGAGCCAGACTTCCTGGGTTTAAACTCGAGTACCACC ATTAACTAGCCATGCATT	94	94	Encoded in an intergenic locus on chromosome 8
CBL-3	CCCGGGTCGGAGTTAGCTCAAGCGGTTACCTCCTCATG CCGGACTTTCTATCTGTCCATCTCTGTGCTGGGGTTCG AGACCCGCGGGTGCTTACTGACCCTTTT	104	100	Encoded in an intergenic locus on chromosome 5; similarities to vault RNAs
CBL-4	AACAGCAGCCAATAGCTGGTTGGCATTCTGGCCCTGG TTCATGCCAACTCTTGTGTTGACTACCCCAGGATG CCAGCATAGTT	83	133	Encoded in the second intron of the MATR3 gene on chromosome 5
CBL-5	GCTGTCATTGCTACACTGGAGTCA	24	22	Intergenic locus on chromosome 3

The sequence lengths of the five novel ncRNA candidates (CBL-1–5) in the cDNA library as well as the estimated lengths of the bands on the northern blot scans (N. blot) are indicated.

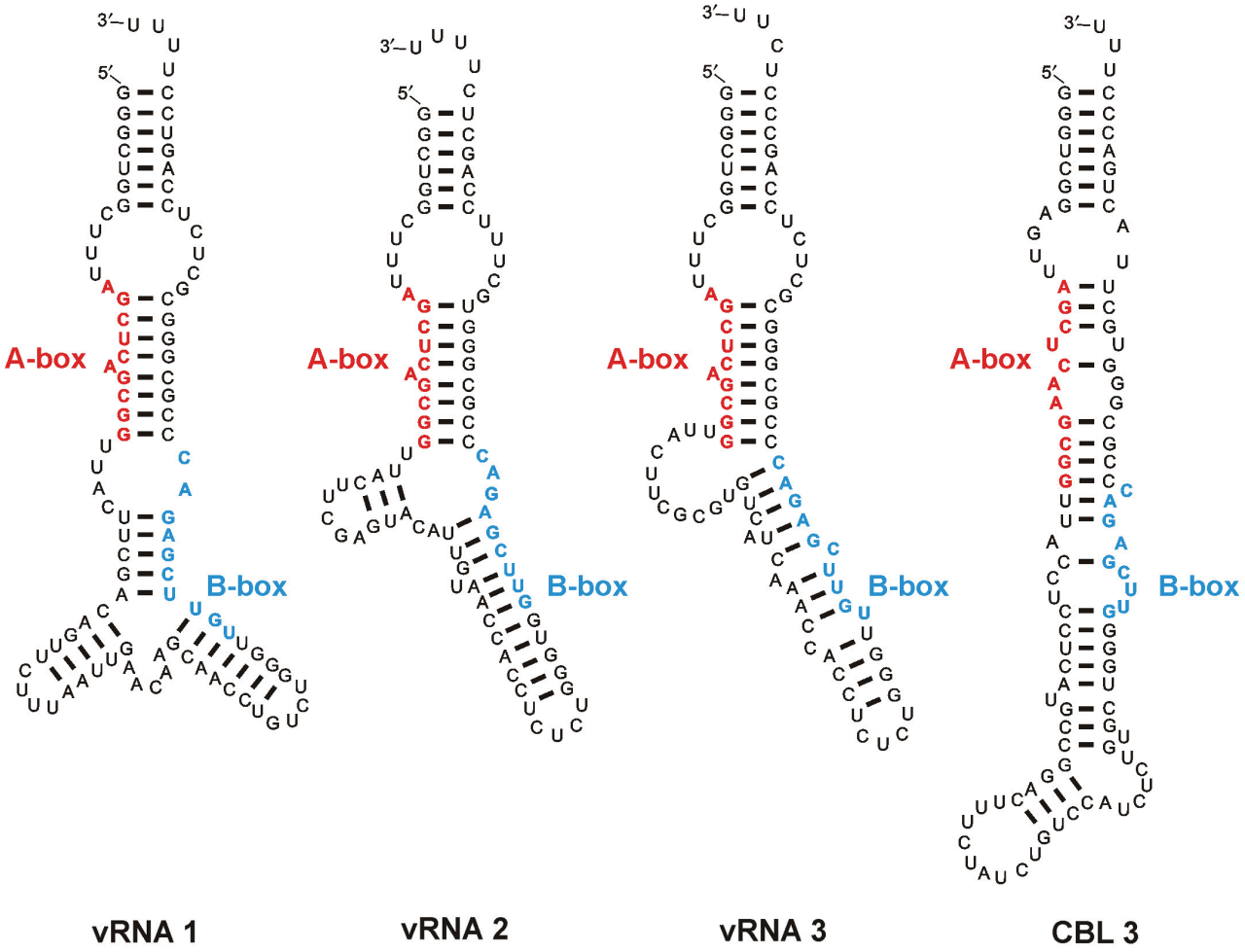


Figure 4. Secondary structure predictions of vault RNA 1–3 and CBL-3. The three known human vault associated RNAs (vRNA 1-3) and the novel ncRNA candidate CBL-3 were folded using the RNAfold algorithm developed by Hofacker and *et al.* (25). The characteristic internal polymerase III promoter elements, A-box and B-box, are colored red and blue, respectively.

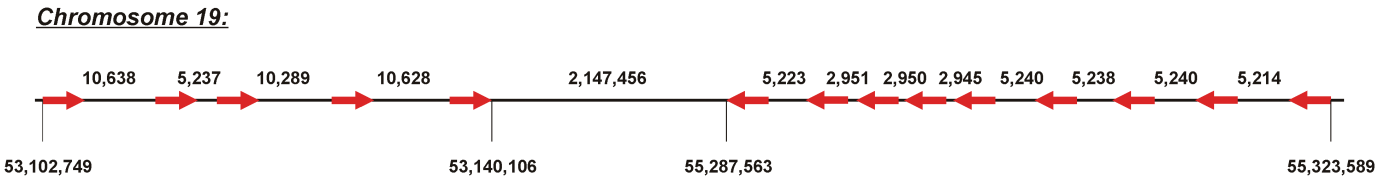


Figure 5. Genomic location of the new ncRNA candidate CBL-1. CBL-1 is encoded in 14 copies (red arrows) on human chromosome 19 spanning a region of 2.2 Mb. The location on the chromosome is given according to the UCSC Genome Browser coordinates. The distances between the individual CBL-1 repeats are given above the arrows.

ncRNA (CBL-2) exhibits a size of 94 nt and was up-regulated 4-fold in EBV-infected B cells (Figure 3). CBL-2 is predicted to map to a locus encoding a MIR subclass of small interspersed nuclear elements (SINE), however expression of the repeat has not been experimentally confirmed, up till now.

From the small-size cDNA library (derived from ncRNAs sized from 20 to 70 nt), a highly up-regulated ncRNA (by 29-fold) in the EBV-infected cell line was identified as the host-encoded miRNA-155 (Figure 3). This miRNA species is processed form a larger RNA precursor transcript, previously designated as *BIC* RNA, and has recently been shown to be up-regulated up to 30-fold in human B cell lymphomas (26). The abundance of miRNA-155 in our SHORT cDNA library (represented by 25% of all cDNA clones from the small fraction library) clearly points to the potential of this method to selectively amplify, and thus identify, known over-expressed ncRNA transcripts of *target* B cells.

Moreover, miRNA-21 was also found to be up-regulated (by 7-fold) in EBV-infected cells (Figure 3). miRNA-21 is encoded on chromosome 17 and its up-regulation has been suggested to play a general role in cell differentiation (27,28). Recently it has been reported that miRNA-21 functions as an oncogene and modulates tumorgenesis in breast tumor tissues (29). To our knowledge, our findings are the first evidence of miRNA-21 being over-expressed upon viral infection in human lymphocytes.

Other up-regulated miRNAs were miRNA-146a (10-fold) and miRNA-29b (2-fold) (Figure 3). In addition to those known miRNA we also identified one potential novel miRNA candidate. Northern blotting revealed that the 22-nt long ncRNA CBL-5 was up-regulated upon EBV infection by 3-fold (Figure 3). The ncRNA candidate is encoded on chromosome 3, but lacks the canonical fold of a *bona fide* miRNA precursor structure (data not shown).

Besides host-encoded miRNAs, we also identified four of the previously described EBV-encoded miRNAs (13,16) (BART4, BART1-3p, BHRF1-1 and BHRF1-2) in our subtracted library (Figure 3). Lack of identification of the remaining 19 known EBV-encoded miRNAs is largely due to the EBV strain used (B95-8) lacking a genomic segment encoding 15 known miRNAs (17).

In addition to ncRNA species, up-regulated upon viral infection, we also observed novel ncRNA candidates in the SHORT subjected library whose expression did not change (CBL-4) (Table 1) or whose expression was even found to be down-regulated. MiRNAs miR-20b, miR-15a and miR-15b were found to be 3- to 4-fold less efficiently expressed in the EBV-infected cell line. This might be explained by the elimination of abundant 'house-keeping' ncRNAs by SHORT, which in turn generally increases the probability of identifying novel ncRNA species in cDNA libraries.

DISCUSSION

By applying the SHORT method to an EBV-infected B-cell line and subsequent sequence analysis of a rather small number of 506 cDNA clones, we have identified 208 unique ncRNA sequences, including five novel ncRNA candidates (Table 1). From these, we selected 37 ncRNAs and analyzed their expression levels by northern blot analysis. 21 ncRNAs (57%) were shown to be up-regulated upon EBV infection (Figure 3). It is strongly anticipated that SHORT in combination with high-throughput deep-sequencing analysis of the subtracted cDNA library will result in the identification of an even higher number of differentially expressed ncRNA candidates.

Especially intriguing was the identification of the B cell-encoded three vault ncRNAs shown to be extensively up-regulated during EBV infection (by up to 1200-fold). vRNAs have been reported previously to serve as integral parts of a so-called vault complex, a large hollow barrel-shaped ribonucleo-protein complex (RNP) with a size of 13 MDa (30). The vault RNP particle is found in numerous higher eukaryal species as diverse as mammals, avians, amphibians, fish, echinoderms, mollusks as well as in lower eukarya such as the slime mold *Dictyostelium discoideum* and in protozoa (30). The vault complex consists of multiple copies of three different highly conserved proteins and at least six copies of vRNAs (31). At present the expression of three vRNA genes (*HVG1-3*) has been demonstrated experimentally while a fourth putative vRNA gene (*HVG4*) has been identified by homology search whose expression could not be observed (32). The putative novel vRNA (CBL-3) that we have identified this study (see above and Figure 4), however, is not related to the *HVG4* locus.

vRNAs have been reported to locate at the two cap structures closing the barrel-shaped mid-section of the vault complex. RNase-treatment of vault complexes has been shown to have no apparent effect on morphology of the vault RNP, implying that the vRNAs might possess functional rather than structural roles in the vault complex (31). The vault RNP has previously been proposed to be involved in nucleo-cytoplasmic transport and implicated in intracellular detoxification processes and hence in multidrug resistance in cancer cells (33,34). For localization of vRNAs, we performed fluorescence *in situ* hybridization experiments employing antisense probes directed against vRNA 1. These experiments demonstrated that vRNA 1 localizes close to the nucleus in EBV-infected B cells (data not shown), consistent with earlier findings indicating that vault RNP associates with nuclear pores in various mammalian cell lines (30,35–38). It is tempting to speculate that vRNAs might be involved in anti-viral defense and/or transport mechanisms.

We also observed up-regulation (by up to 29-fold) from seven previously reported miRNAs, as well as a novel miRNA candidate not previously annotated. Two of these miRNA species, miR-21 and miR-155, have previously been implicated in human cancers (26,29). Our data point to the possibility that they might also be involved in viral infections. MiRNAs have been shown to regulate the gene expression in mammals, including humans, by binding to the 3′-untranslated region (UTR) of mRNAs thus repressing their translation (7). Thereby, single miRNA species were suggested to target up to 100–200 different UTRs within mRNAs, however for the majority of predicted targets experimental evidence is lacking (39,40). For example, by employing the miRNA-target search tool from the Sanger Centre (http://microrna.sanger.ac.uk/targets/v4/), genes involved in tumor suppression, such as the tropomyosin 1 gene (*TPM1*), were identified as potential targets for miR-21. In addition, RGS1, a gene encoding B cell activation protein BL34, was identified as a potential target for miR-23a. It will be interesting to assess experimentally whether predicted mRNAs, targeted by miRNAs identified in our screen, are directly involved in viral infection and/or defense mechanisms.

In summary, the considerable advantage of SHORT, compared to standard experimental RNomics approaches (20), is that it focuses on differentially expressed ncRNA genes and therefore avoids the time-consuming and cost-intensive multiple sequencing of already known and constitutively expressed ncRNAs. Furthermore, removal

of the majority of these 'house-keeping' transcripts by SHORT increases the chance to identify low abundant and so far unidentified ncRNA candidates, an advantage that cannot be used in ncRNA micro array screens.

We demonstrate that SHORT can be applied to enrich ncRNAs of human B cells in the size range from 22 to 300 nt, a range that covers many of the known functional small ncRNA classes from miRNAs (21 nts), snoRNAs (~80–200 nt) up to 7SL RNA (~300 nts). Since SHORT is performed exclusively *in vitro*, requires only minimal amounts of starting ncRNAs and does not depend on the source of the ncRNA transcripts, it is reasonable to assume that this technique can be applied to other cells, tissues or organisms as well. Thus, this procedure will likely result in the identification of novel regulatory ncRNA species involved in diseases like cancer as well as those which regulate fundamental cellular processes such as development or tissue differentiation.

ACKNOWLEDGEMENTS

We thank Heribert Stoiber, Monika Madej, Werner Kapferer, Yuuichi Soeno, Beatriz Campo-Fernández and Dimitriy Shcherbakov for valuable suggestions and help. This work was supported from the Austrian Science Foundation FWF (P171370 to A.H. and Y315 to N.P.) and the bm:bwk (GEN-AU Project D1042-011-011 to A.H and N.P.). Funding to pay the Open Access publication charges for this article was provided by the Austrian FWF (Fonds zur Förderung der wissenschaftlichen Forschung) grant Y315.

Conflict of interest statement. None declared.

REFERENCES

1. Mattick,J.S. (2001) Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.*, **2**, 986–991.
2. Eddy,S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.*, **2**, 919–929.
3. Huttenhofer,A., Schattner,P. and Polacek,N. (2005) Non-coding RNAs: hope or hype? *Trends Genet.*, **21**, 289–297.
4. Storz,G. (2002) An expanding universe of noncoding RNAs. *Science*, **296**, 1260–1263.
5. Washietl,S., Hofacker,I.L. and Stadler,P.F. (2005) Fast and reliable prediction of noncoding RNAs. *Proc. Natl Acad. Sci. USA*, **102**, 2454–2459.
6. Washietl,S., Hofacker,I.L., Lukasser,M., Huttenhofer,A. and Stadler,P.F. (2005) Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat. Biotechnol.*, **23**, 1383–1390.
7. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
8. Tuschl,T. (2003) Functional genomics: RNA sets the standard. *Nature*, **421**, 220–221.
9. Ambros,V. (2001) MicroRNAs: tiny regulators with great potential. *Cell*, **107**, 823–826.
10. Volpe,T.A., Kidner,C., Hall,I.M., Teng,G., Grewal,S.I. and Martienssen,R.A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, **297**, 1833–1837.
11. 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.
12. Sarnow,P., Jopling,C.L., Norman,K.L., Schutz,S. and Wehner,K.A. (2006) MicroRNAs: expression, avoidance and subversion by vertebrate viruses. *Nat. Rev. Microbiol.*, **4**, 651–659.
13. Pfeffer,S., Zavolan,M., Grasser,F.A., Chien,M., Russo,J.J., Ju,J., John,B., Enright,A.J., Marks,D. *et al.* (2004) Identification of virus-encoded microRNAs. *Science*, **304**, 734–736.
14. Pfeffer,S., Sewer,A., Lagos-Quintana,M., Sheridan,R., Sander,C., Grasser,F.A., van Dyk,L.F., Ho,C.K., Shuman,S. *et al.* (2005) Identification of microRNAs of the herpesvirus family. *Nat. Methods*, **2**, 269–276.
15. Straathof,K.C., Bollard,C.M., Rooney,C.M. and Heslop,H.E. (2003) Immunotherapy for Epstein–Barr virus-associated cancers in children. *Oncologist*, **8**, 83–98.
16. Cai,X., Schafer,A., Lu,S., Bilello,J.P., Desrosiers,R.C., Edwards,R., Raab-Traub,N. and Cullen,B.R. (2006) Epstein–Barr virus microRNAs are evolutionarily conserved and differentially expressed. *PLoS Pathog.*, **2**, e23.
17. Grundhoff,A., Sullivan,C.S. and Ganem,D. (2006) A combined computational and microarray-based approach identifies novel microRNAs encoded by human gamma-herpesviruses. *RNA*, **12**, 733–750.
18. Fok,V., Friend,K. and Steitz,J.A. (2006) Epstein–Barr virus noncoding RNAs are confined to the nucleus, whereas their partner, the human La protein, undergoes nucleocytoplasmic shuttling. *J. Cell Biol.*, **173**, 319–325.
19. Fok,V., Mitton-Fry,R.M., Grech,A. and Steitz,J.A. (2006) Multiple domains of EBER 1, an Epstein–Barr virus noncoding RNA, recruit human ribosomal protein L22. *RNA*, **12**, 872–882.
20. Huttenhofer,A. and Vogel,J. (2006) Experimental approaches to identify non-coding RNAs. *Nucleic Acids Res.*, **34**, 635–646.
21. Frech,B., Zimber-Strobl,U., Suentzenich,K.O., Pavlish,O., Lenoir,G.M., Bornkamm,G.W. and Mueller-Lantzsch,N. (1990) Identification of Epstein–Barr virus terminal protein 1 (TP1) in extracts of four lymphoid cell lines, expression in insect cells, and detection of antibodies in human sera. *J. Virol.*, **64**, 2759–2767.
22. Lenoir,G.M., Vuillaume,M. and Bonnardel,C. (1985) The use of lymphomatous and lymphoblastoid cell lines in the study of Burkitt's lymphoma. *IARC Sci. Publ.*, 309–318.
23. Tang,T.H., Polacek,N., Zywicki,M., Huber,H., Brugger,K., Garrett,R., Bachellerie,J.P. and Huttenhofer,A. (2005) Identification of novel non-coding RNAs as potential antisense regulators in the archaeon Sulfolobus solfataricus. *Mol. Microbiol.*, **55**, 469–481.
24. Sagerstrom,C.G., Sun,B.I. and Sive,H.L. (1997) Subtractive cloning: past, present, and future. *Annu. Rev. Biochem.*, **66**, 751–783.
25. Hofacker,I.L., Fontana,W., Stadler,P.F., Bonhoeffer,S., Tacker,P. and Schuster,P. (1994) Fast folding and comparison of RNA secondary structure. *Monatsh. Chem.*, **125**, 167–188.
26. Eis,P.S., Tam,W., Sun,L., Chadburn,A., Li,Z., Gomez,M.F., Lund,E. and Dahlberg,J.E. (2005) Accumulation of miR-155 and BIC RNA in human B cell lymphomas. *Proc. Natl Acad. Sci. USA*, **102**, 3627–3632.
27. Houbaviy,H.B., Murray,M.F. and Sharp,P.A. (2003) Embryonic stem cell-specific MicroRNAs. *Dev. Cell*, **5**, 351–358.
28. Kasashima,K., Nakamura,Y. and Kozu,T. (2004) Altered expression profiles of microRNAs during TPA-induced differentiation of HL-60 cells. *Biochem. Biophys. Res. Commun.*, **322**, 403–410.
29. Si,M.L., Zhu,S., Wu,H., Lu,Z., Wu,F. and Mo,Y.Y. (2006) miR-21-mediated tumor growth. *Oncogene*, 1–5.
30. van Zon,A., Mossink,M.H., Scheper,R.J., Sonneveld,P. and Wiemer,E.A. (2003) The vault complex. *Cell Mol. Life Sci.*, **60**, 1828–1837.
31. Kong,L.B., Siva,A.C., Kickhoefer,V.A., Rome,L.H. and Stewart,P.L. (2000) RNA location and modeling of a WD40 repeat domain within the vault. *RNA*, **6**, 890–900.
32. van Zon,A., Mossink,M.H., Schoester,M., Scheffer,G.L., Scheper,R.J., Sonneveld,P. and Wiemer,E.A. (2001) Multiple human vault RNAs. Expression and association with the vault complex. *J. Biol. Chem.*, **276**, 37715–37721.
33. Kitazono,M., Okumura,H., Ikeda,R., Sumizawa,T., Furukawa,T., Nagayama,S., Seto,K., Aikou,T. and Akiyama,S. (2001) Reversal of LRP-associated drug resistance in colon carcinoma SW-620 cells. *Int. J. Cancer*, **91**, 126–131.

34. Gopinath,S.C., Matsugami,A., Katahira,M. and Kumar,P.K. (2005) Human vault-associated non-coding RNAs bind to mitoxantrone, a chemotherapeutic compound. *Nucleic Acids Res.*, **33**, 4874–4881.
35. Chugani,D.C., Rome,L.H. and Kedersha,N.L. (1993) Evidence that vault ribonucleoprotein particles localize to the nuclear pore complex. *J. Cell Sci.*, **106**, 23–29.
36. Abbondanza,C., Rossi,V., Roscigno,A., Gallo,L., Belsito,A., Piluso,G., Medici,N., Nigro,V., Molinari,A.M. *et al.* (1998) Interaction of vault particles with estrogen receptor in the MCF-7 breast cancer cell. *J. Cell Biol.*, **141**, 1301–1310.
37. Berger,W., Elbling,L. and Micksche,M. (2000) Expression of the major vault protein LRP in human non-small-cell lung cancer cells: activation by short-term exposure to antineoplastic drugs. *Int. J. Cancer*, **88**, 293–300.
38. van Zon,A., Mossink,M.H., Houtsmuller,A.B., Schoester,M., Scheffer,G.L., Scheper,R.J., Sonneveld,P. and Wiemer,E.A. (2006) Vault mobility depends in part on microtubules and vaults can be recruited to the nuclear envelope. *Exp. Cell Res.*, **312**, 245–255.
39. Lewis,B.P., Burge,C.B. and Bartel,D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.
40. Krek,A., Grun,D., Poy,M.N., Wolf,R., Rosenberg,L., Epstein,E.J., MacMenamin,P., da Piedade,I., Gunsalus,K.C. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.