Supplementary Material

Widespread context-dependency of microRNA-mediated regulation

Florian Erhard [1], Jürgen Haas [2,3], Diana Lieber [2,4], Georg Malterer [2], Lukasz Jaskiewicz [5], Mihaela Zavolan [5], Lars Dölken [6,#] and Ralf Zimmer [1,#]

[1]Institut für Informatik, Ludwig-Maximilians-Universität München, Amalienstraße 17, 80333 München, Germany, [2]Max-von-Pettenkofer Institut, Virologie, Ludwig-Maximilians-Universität München, Pettenkoferstrasse 9a, 80336 München, Germany [3]Division of Pathway Medicine, University of Edinburgh, 49 Little France Crescent, Edinburgh EH17 8TR, UK [4]Institut für Virologie, Universitätsklinikum Ulm, Albert-Einstein-Allee 11, 89081 Ulm, Germany [5]Biozentrum, University of Basel and Swiss Institute of Bioinformatics, Klingelbergstr. 50/70, CH-4056, Basel, Switzerland [6]Department of Medicine, University of Cambridge, Box 157, Level 5, Addenbrookes Hospital, Hills Road, CB20QQ Cambridge, UK

1 Supplementary Methods

1.1 PAR-CLIP and sequencing

PAR-CLIP on DG75 and BCBL1 were performed by the Zavolan laboratory as described (Kishore et al., 2011; Jaskiewicz et al., 2012). Briefly, a total of $3 * 10^8$ cells per replicate were grown and treated with 4-thiouridine (Sigma) for 14 hours (final concentration 100 uM). Cells were pelleted and washed in cold PBS. Aliquots of $5 * 10^7$ cells were resuspended in 5 ml of cold PBS, placed in a 15 cm petri dish and irradiated at 365 nm with 100 mJ twice on ice, with 30 s break in between. Cross-linked cells were collected, pelleted and snap-frozen. PAR-CLIP was performed using 11A9 anti-Ago2 monoclonal antibody (Rüdel et al., 2008). The PAR-CLIP sequencing data for BC1 and BC3 from (Gottwein et al., 2011) have been downloaded from GEO (accession number: GSE32113). We applied PARma (Erhard et al., 2013) to the whole collection of all PAR-CLIP datasets as described.

1.2 PARma

PARma is specifically designed to accurately determine target sites and to determine which microRNA is responsible for each target site and is described in a separate paper (Erhard et al., 2013). Briefly, it estimates seed activity probabilities and the parameters of a generative model for the PAR-CLIP data simultaneously in an iterative manner. The model and probabilities are then used to accurately determine the seed position within each PAR-CLIP cluster, and to compute a cluster confidence score (C-score) and a microRNA assignment confidence score (MA-score). The C-score can be used to exclude false positive clusters (i.e. clusters that do not correspond to a target site of any microRNA), whereas the MA-score can be used to judge whether the assigned microRNA is indeed targeting a given site.

The PAR-CLIP expression value for each cluster is computed for each experiment by counting the reads overlapping the main crosslinking site (Erhard et al., 2013). For proper comparison across experiments, the expression values for all clusters are normalized using the same strategy as in Anders and Huber (2010), i.e. by dividing each count value by the geometric mean across experiment and taking the median of all these values from a specific experiment as the normalization factor for this experiment.

To analyze the positional distribution, we subdivided each transcript into 60 bins and counted the number of target sites of cellular, EBV and KSHV microRNA, respectively, belonging to each bin. In order to compare cellular and viral frequencies, the number of target sites within each bin was divided by the total number of cellular, EBV or KSHV target sites, respectively.

Context-dependent target sites were determined by applying stringent cutoffs: More than 10 normalized reads in all replicates in the active context and less than 5 in all other experiments.

1.3 Correcting for sampling noise

In order to estimate the contribution of mRNA levels to the cellular context for microRNA-mediated regulation, we first inspected the correlation between the mRNA fold changes and PAR-CLIP read fold changes (see Figure S5A). Replicate counts were summed and a pseudocount of 1 was used to circumvent division by zero. These fold changes were correlated to some extent, but there were also many exclusive (i.e. context-dependent) target sites present, that did not show any or only a very modest mRNA fold change. In order to properly estimate the fraction of non-correlated target sites and to handle the sampling noise of the low-count data and pseudocounts, we took the following approach:

First, we estimated the variance of PAR-CLIP fold changes based on replicate experiments. Because the number of replicates was extremely low ($n = 2$), no reliable estimates can be compute in a target site-wise manner, and, thus, we took a population based approach similar to methods that estimate significance of differential expression in RNA-seq data (Anders and Huber, 2010; Robinson et al., 2010). Since the variance is not equal for strong target sites and weak target site (as measured by the number of PAR-CLIP reads) due to sampling noise, variance was estimated conditional on the target site strength. Then, for each target site, we checked, whether the mRNA fold change was within a critical region as defined by significance levels of 1% and 5%.

To estimate read count fold change variances, we considered the absolute difference of read counts from replicate measurements (see Figure S5). For a given target site strength (as measured by the geometric mean of read counts across replicates), the distribution of these absolute differences resembled a gamma distribution by visual inspection. Thus, by using a running window approach, we estimated the distribution of absolute read count differences by fitting a gamma distribution to each window of 1000 target sites along the target site strength (i.e. the red line in Figure S5) using the *fitdistr* function from the R package MASS. We plotted the rate and shape parameters of the gamma distribution as fitted for different windows along the target site strength (see Figure S5D) and noticed that the shape parameter was relatively constant and the rate parameter increased linearly in log space with the target site strength. For robustness of the fits, we therefore computed the median shape parameter S across all windows and computed a robust linear fit for the rate parameters $R(s)$ against the logarithmized target sites strengths s. Thus, our model describes the absolute read count difference of a target site with strength s (i.e. the geometric mean of read counts across all experiments) by a gamma distribution with rate and shape parameters $R(s)$ and S.

This conditional gamma distribution allows us to compute the distribution of absolute read count differences for a given target site strength. As illustrated in Figure S5E, this conditional gamma distribution nicely reflects the variances for replicate measurements.

2 Supplementary Discussion

2.1 Technical bias

When analyzing high-throughput data obtained from experiments performed in different laboratories a certain extent of differences in between given data sets can be expected. In our case, distinct clusters of target sites may also be consequences of erroneously assigned microRNAs, bias introduced by differing sample preparation methods or insufficient sequencing depth/sequencing library complexity.

The first question we addressed was whether the KSHV microRNA target sites identified only in one or two but not all three cell lines actually represented viral microRNA binding sites. We could properly address this since one of the PAR-CLIP data sets analyzed in this study is derived from a KSHV negative B-cell line which does not express the KSHV microRNAs. Therefore, the PAR-CLIP clusters attributed to the KSHV microRNAs by PARma should not be present in this cell line. Overall, only very few reads mapped to the identified target sites of KSHV microRNAs in KSHV negative cells. As we observed random reads spread across a multitude of transcripts at low frequency, these reads presumably result from infrequent unspecific immunoprecipitates or insufficient removal of background total RNA rather than microRNA-specific signatures (see also Figure 2 in main text). This is supported by a significantly lower frequency of C to T conversions and lower consistency across replicates for these reads. In conclusion, erroneously assigned microRNAs are unlikely to present a main factor for our observations of poor correlation and exclusive sites.

Differing sample preparation methods are the most likely cause of bias and differences in microRNA targets obtained by PAR-CLIP data. As such, the RNase used to trim the Argonaute-2 protected microRNA target sequences has recently been shown to be a major source of bias (Kishore et al., 2011). Other sources may include differing immunoprecipitation efficiencies, RNAse treatment times, sequencing adapters or any other slight variation in the PAR-CLIP protocol which may all result in a target site to be identified in one experiment but not in another. Such bias could be controlled for, if PAR-CLIP data for one or more cell lines were available that haven been measured in multiple labs. And only considering the high correlation of replicate measurements does exclude such technical bias. Nevertheless, we can, to some extent, use the inverse argument: BC1 and BC3 were analyzed in the same lab using the same protocol. Thus, if technical bias was responsible for poor correlation and exclusive sites and not context-specific microRNA targeting, the correlation between BC1 and BC3 should be as high as for replicate experiments of either cell line. As illustrated in Figure S2C, this is not the case. However, the correlations between BCBL1 and BC3 are even lower than between BC1 and BC3, for instance. There may be two reasons for that: Either BC1 and BC3 are more similar to each other than BCBL1 and BC3 with respect to their cellular context for microRNA-mediated regulation or between-lab comparison of PAR-CLIP target sites is indeed influenced by technical bias to some extent. But nevertheless, technical bias cannot explain the relatively low correlations between BC1 and BC3 which provides first evidence that the observed differences may indeed not only be due to technical bias.

Insufficient saturation of PAR-CLIP libraries or sequencing depth may result in seemingly cell-type specific and thus context dependent microRNA/target interactions may also be a reason for our observations: If only a small fraction of target sites was detected (either due to poor immunoprecipitation efficiency or insufficient sequencing depth), sampling effects would play a severe role: Only due to limited sampling, a cluster may get very few or no reads in one sample and many reads in another, even if the target site is strongly associated with a microRNA in both experiments. Considering only a single experiment, this cannot be excluded by simply counting reads and computing statistical significance under a naive probabilistic model: There are many sequences that are identified multiple times (see Figure S1), which may have two reasons: Either there was only a single RNA molecule of this sequence in the library and the amplification before sequencing gave rise to these multiply sequenced reads (indicating insufficient saturation), or there were multiple copies of such an RNA already in the library. Particularly for PAR-CLIP data, the latter is highly probable, since RNase T1, which is used in the PAR-CLIP protocol, cleaves in a sequence-specific way downstream of guanosines (Pace et al., 1991). Since the target mRNA seems highly accessible for cleavage outside of the microRNA target site, the number of possible distinct sequencing reads for a cluster is highly constrained. However, when we consider replicate measurements of targets sites for a specific microRNA, e.g. for BC3 (see Figure S2B), we observe that they are highly correlated (median $\rho > 0.77$ across all microRNAs for all replicate pairs, see Figure S2C). Therefore, all sequencing data utilized in this meta-analysis were found to be of sufficient saturation not to inflict major bias to our analysis.

2.2 Other contributors to the cellular context

RISC is a highly modular protein complex (Frohn et al., 2012). Therefore, proteins that interact with RISC may influence the effects of microRNA/target interactions: For instance, the NHL family protein LIN41 has been found to suppress let-7 and miR-124 activity by ubiquitilation of AGO2 (Rybak et al., 2009) and several other NHL proteins have been implicated in the regulation of RISC activity (Pasquinelli, 2012). In addition to ubiquitilation, AGO2 is susceptible for several other types of modification including hydroxylation (Qi et al., 2008), phosphorylation (Rüdel et al., 2011) and poly(ADP)-ribosylation (Leung et al., 2011).

Another layer of complexity in microRNA-mediated regulation is induced by mutual microRNA-target regulation. Very strong microRNA binding sites in an mRNA (Franco-Zorrilla et al., 2007), a pseudogene (Cazalla et al., 2010), a non-coding RNA (Cesana et al., 2011) or viral RNAs (Marcinowski et al., 2012) may sequester RISCs containing a specific microRNA acting as a microRNA sponge. It has also been hypothesized that the microRNA target sites in a cell as a whole allow for crosstalk between expressed transcripts giving rise to an intricate post-transcriptional regulatory network based on mutually competing microRNA/target interactions (Salmena et al., 2011). The complexity of such a regulatory network is further underlined by the nonlinearity of the regulatory outcome of a microRNA/target interaction (Mukherji et al., 2011): Depending on the exact copy numbers of microRNA and mRNA and the affinity of the microRNA for the target site, protein expression may be either completely abolished or only fine-tuned.

It has been suggested that sequestering of RISCs by the expression of transcripts containing one or multiple strong target sites for a specific microRNA may derepress its targets (Franco-Zorrilla et al., 2007; Cazalla et al., 2010; Cesana et al., 2011; Marcinowski et al., 2012). Thus, such microRNA sponges may also be important contributors to the cellular context for microRNA-mediated regulation. In such a setting, weak target sites should disappear first. By comparing binding energies for our set of context-dependent target sites, we tested whether such effects play a role in our datasets. However, we could not identify any microRNA where exclusive binding sites had significantly different binding energies than constitutive target sites (data not shown). This may be due to deficiencies of the current RNA energy model to describe microRNA/target duplexes, or because microRNA sponges do not play an important role for our cell lines. And indeed, all of the selected cellular microRNAs exhibit target sites that are exclusive in DG75 and other target sites exclusively present in BCBL1, which would not be expected if a microRNA sponge is active in one of these cell lines (see Supplementary Figures S7 and S8).

2.3 Functional considerations of context-dependent regulation

Based on results from concurrent research on transcriptional regulation (The ENCODE Project Consortium, 2012; Wang et al., 2012b,a; Yanez-Cuna et al., 2012), context-dependency in post-transcriptional regulation should not come as a surprise: It is known that transcription factors bind to their target sites in a context-dependent manner. Therefore, context-dependency of regulatory mechanisms presumably is beneficial in an evolutionary sense, and this is a widespread phenomenon for transcriptional regulation. Here, we argue that evolution also has invented this additional layer of complexity for microRNA-mediated regulation as well.

One evolutionary benefit of the additional layer of complexity by context-dependent microRNA/target interaction may be the greater flexibility in regulation: Modulating the expression level of a microRNA would alter the regulation of hundreds of targets and therefore potentially influence a multitude of cellular processes. In contrast, using context-dependent regulation, for instance by activating or inactivating an RNA binding protein (RBP), smaller groups of targets could be activated or inactived in a much more focused manner. The combinatorics that unfolds when multiple RBPs, multiple target sites or other factors contribute to the overall regulation provides opportunities for evolutionary forces to achieve the desired expression levels for individual genes.

Our analysis of target sites of constitutively expressed cellular microRNAs revealed that a large fraction of context-dependent targets may be due to induced mRNA levels. For instance, a gene may get transcribed at high rates in BCBL1 as compared to DG75, leading to elevated mRNA levels. At the same time, microRNA target sites are more active in BCBL1 than in DG75, leading to an induced degradation as compared to DG75. Importantly, this dependency between microRNA and mRNA is not necessarily linear, as pointed out above. Thus, these constitutive microRNAs seem to limit the expression levels of their target mRNAs: If targets have high enough expression levels, they become subject to microRNA-mediated regulation thus providing an upper bound for the target mRNA levels.

2.4 Consequences of context-dependency

Widespread dependency of microRNA targeting on cellular context has significant consequences for both computational and experimental approaches for microRNA target discovery.

MicroRNA target prediction algorithms may not as bad as their reputation (Thomas et al., 2010; Sethupathy et al., 2006; Ritchie et al., 2009): False positive as well as false negative predictions simply may be due to a wrong context used when evaluating the predictions. Thus, none of the apparently inconsistent evaluations of microRNA target prediction algorithms may be wrong: Each of these was evaluated on a different cellular context and, consequently, differing prediction methods seemed more accurate than others. However, the problem of microRNA target prediction may be defined in an incorrect way and as long as prediction methods do not incorporate the cellular context, predicted targets are of limited use. Thus, we expect that future development of microRNA target prediction methods will mainly depend on integrating features of cellular context into the prediction algorithms. Such an approach is obviously heavily dependent on progress in unraveling contributing factors to the cellular context.

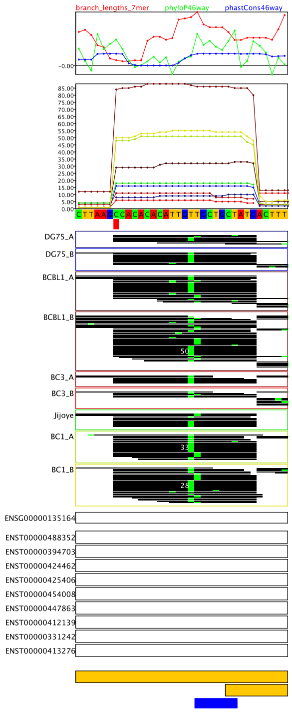
Another consequence of a general context-dependency of microRNA targeting is that experimental assays for microRNA target discovery and validation must be interpreted with care. In various studies, either a single or a pool of microRNAs has been transfected into a cell line and gene expression has been measured genome-wide either on the mRNA level using microarrays or RNA-seq (Lim et al., 2005; Linsley et al., 2007; Grimson et al., 2007; He et al., 2007; Xu et al., 2010) or on the protein level using mass spectrometry (Selbach et al., 2008; Baek et al., 2008) differentially for transfected and control cells. In addition to the well-known problem of secondary regulation (Tu et al., 2009; Naeem et al., 2011), there are several reasons why downregulated genes should not generally be taken as the set of targets for the transfected microRNA: First, they may be targets exclusively active in the cell line investigated in the study. Second, the transfected microRNA may have copy numbers at levels never occurring in physiological conditions. And third, the transfection itself may lead to an altered cellular context, e.g. by the induction of cellular stress pathways. Also, further conclusions drawn from such experiments, e.g. that microRNAs generally lead to widespread but only modest regulation should be revisited: This may only be true in the context of the experiment, which is not a cellular context that evolved naturally but has been forced onto a cell line artificially. Furthermore, later studies revealed that the particular outcome or strength of regulation is dependent on the exact mRNA and microRNA copy numbers (Mukherji et al., 2011).

Also, the most widely used validation assay for microRNA targets is based on fluorescence reporter genes that are fused to target 3'-UTRs and co-transfected with the microRNA into a particular cell line usually lacking this microRNA (Kiriakidou et al., 2004; Gottwein et al., 2011). Obviously, not only is the microRNA-target pair introduced into a non-natural context and probably expressed at non-physiological conditions, but also the microRNA target site itself is expressed in a non-natural context, i.e. the 3'-UTR of a fusion gene. Thus, both outcomes of such an experiment may not hold for the microRNA target pair under different conditions, i.e. if the reporter assay does not confirm regulation, the target site may still be highly functional in the right context and if the reporter assay validates regulation, this may not be true for the context under consideration. In conclusion, while luciferase assays provide good evidence that a certain gene can be regulated by a given microRNA it does not allow any claims about whether this interaction is of relevance in the biological context of interest.

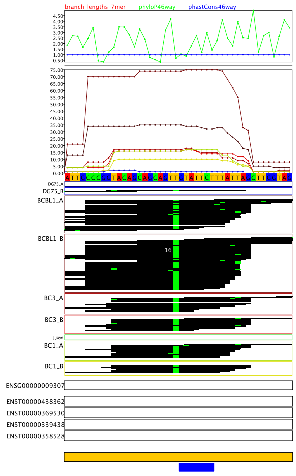
Consortium-driven endeavors to unravel context-dependent transcriptional regulation have been started years ago, as soon as the human genome project has been finished (The ENCODE Project Consortium, 2012). Context-dependent transcription factor binding sites have been determined in an overwhelming variety of conditions and it is one of the most intriguing results of the ENCODE project that context-dependency is one of the key features of transcriptional regulation.

Here, we show that context-dependency is also an important factor in post-transcriptional regulation. We propose that a similar approach as for transcriptional regulation must also be taken for microRNA-mediated regulation and that a lot of additional experiments are necessary to further investigate both, microRNA targets specific to certain contexts and key contributors that determine cellular context.

3 Supplementary Figures

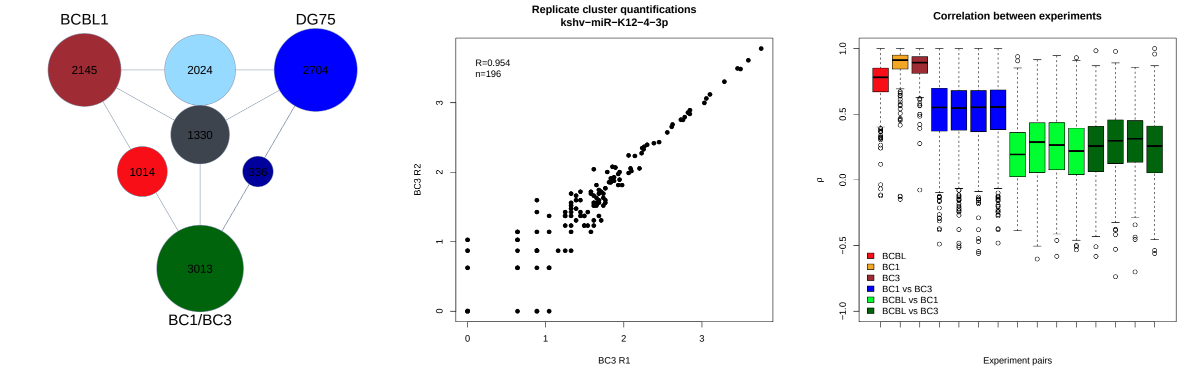


(A)



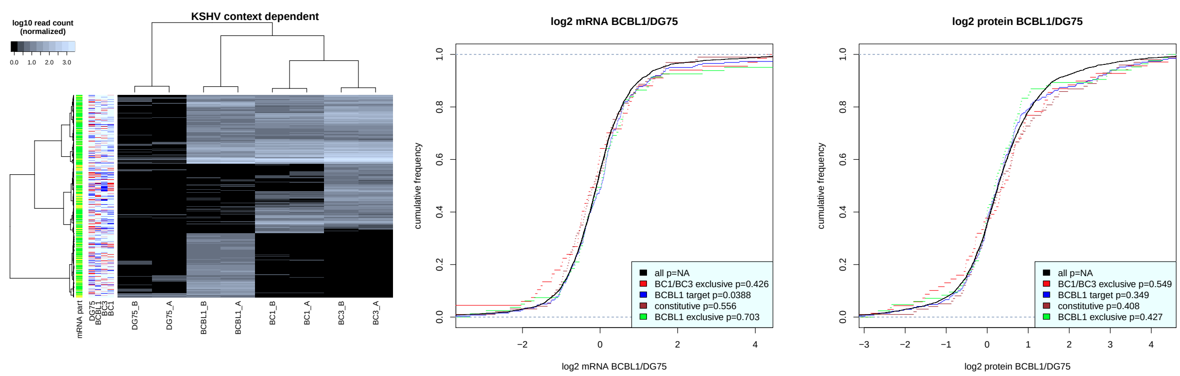
(B)

Figure S1: PAR-CLIP cluster visualization; related to Figure 1. From top to bottom in both panels, illustrated are conservation scores (branch lengths of 7-mers as described by Friedman et al. (2009) and the widely used phyloP (Pollard et al., 2010) and phastCons (Siepel et al., 2005) scores, all computed for the 46-way vertebrate multiple alignment obtained from the UCSC genome browser (Dreszer et al., 2012), the read coverage in each experiment and the genomic sequence of cluster. Below the sequence, SNP positions according to the 1000 genomes project are indicated in red (here only in Figure S1A) and the actual sequencing reads are shown as black bars for each of the experiments. Mismatches are color-coded as in the genomic sequence on the top (i.e. in both clusters, there are T to C conversions only). The height of the bar directly corresponds to the read count in the PAR-CLIP experiment up to a count of 15 reads and more than 15 reads are indicated in white. Ensembl genes and transcripts are shown below the reads (here only in Figure S1A), in addition to PAR-CLIP clusters in yellow and seed site assignments in blue. In Figure S1A an experimentally validated targets site of hsa-miR-15 in the 3'UTR of DMTF1 is shown. Figure S1B depicts a cluster which a KSHV microRNA has been assigned to. Both clusters exhibit a clear pattern of PAR-CLIP sequencing reads (RNase T1 cleavage positions 3' of a guanosine, a very strong T to C conversion upstream of the seed site), that is also shared by many microRNA seed containing clusters. These characteristics are used by PARma to compute confidence scores. The target site of hsa-miR-15 illustrates that in each experiments, several PAR-CLIP reads fall into the cluster. For the KSHV microRNA, only in KSHV positive cell lines (BCBL1, BC1 and BC3) reads are observed with the exception of a single read in the second replicate of DG75. This single read presumably is an example of random background reads that are scatter across various transcripts in all experiments.



(A) (B) (C)

Figure S2: Correlations of PAR-CLIP cluster quantifications; related to Figure 2. Figure S2A illustrates the number of human microRNA target sites observed only in individual cell lines (or BC1/BC3 combined; outermost labeled circles), in two cell lines (circles on the edges between cell lines) and in all four cell lines (center circle). As in Figure 2A, relatively few target sites appear to be active in multiple cell lines. Figure S2B shows the normalized quantifications of all target sites of kshv-miR-K12-4-3p as a scatter plot (see also Figure 2). Quantifications are highly correlated. In Figure S2C, the distributions for all microRNAs of Spearman's ρ (non-parametric correlation coefficient) are shown for each possible comparison of KSHV positive cell lines. Replicate correlations are drastically higher than between experiment correlations, indicating context specific microRNA targeting.



(A) (B) (C)

Figure S3: Context dependent target sites of KSHV microRNAs; related to Figure 3. Figure S3A visualizes the differential analysis of all sets of considered KSHV microRNAs. None of the target sites has a significant amount of reads in the uninfected control cell line DG75. The top third corresponds to constitutive target sites that are active in all three KHSV positive cell line ($n = 162$), the middle third are target sites exclusively active in BC1 or BC3 and not in BCBL1 ($n = 151$) and the bottom third shows BCBL1 exclusive active target sites ($n = 151$). Figures S3B and S3C show the fold change distributions of mRNAs and proteins between BCBL1 and DG75, respectively. When the log fold changes of mRNAs and proteins are considered individually, no significant shift for any set of context-specific microRNA targets is observed. Thus, in spite of the fact that mRNA half-lives are significantly decreased by KSHV microRNAs, there is no observable effect on steady-state levels of neither mRNA nor proteins. This highlights the need to look at RNA stability rather than total RNA levels when studying the effect of microRNAs on their targets when stable cell lines are compared.

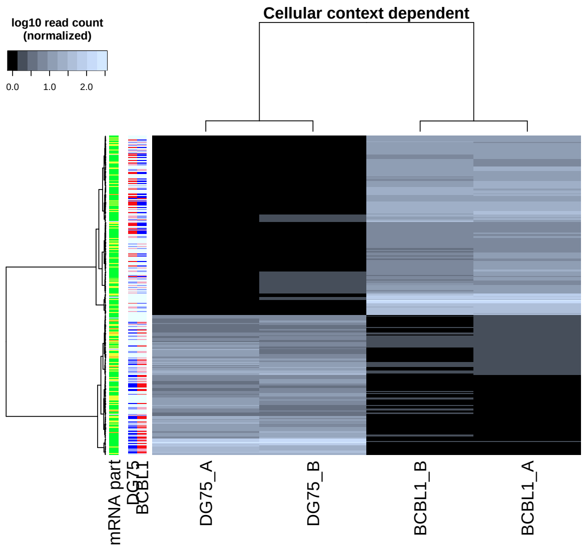
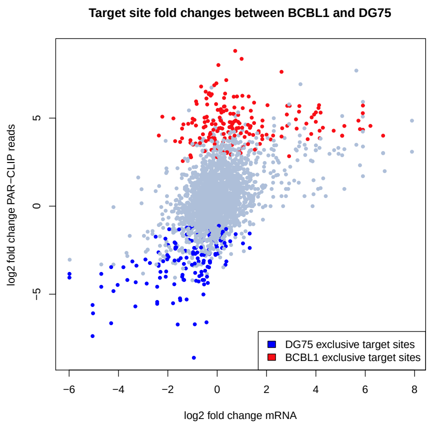
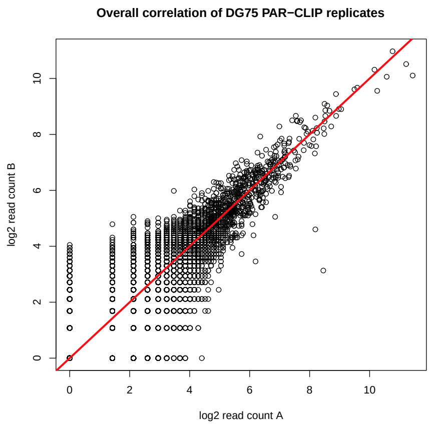


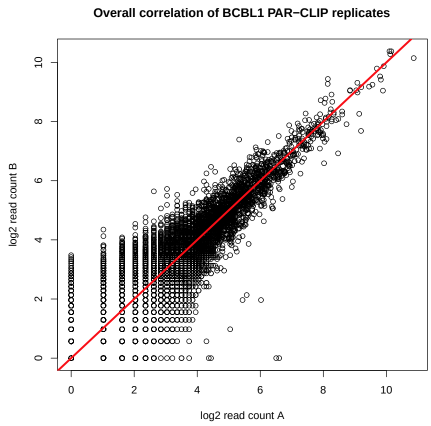
Figure S4: Context-dependent target interactions of human microRNAs; related to Figure 4. The differential PAR-CLIP analysis for all target sits of cellular microRNAs is visualized. The top part corresponds to BCBL1 specific target sites of constitutively expressed cellular microRNAs ($n = 184$) whereas the bottom half represents target sites exclusively active in DG75 and not in BCBL1 ($n = 137$). Importantly, all these patterns of context-dependency are highly reproducible across replicates.



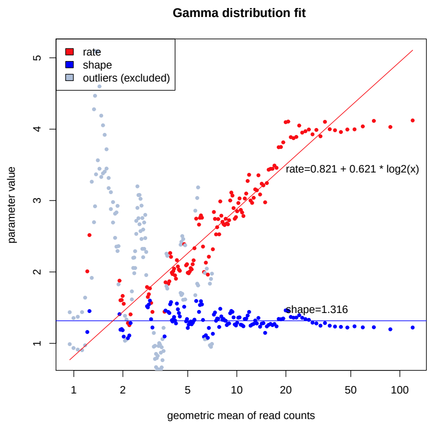
(A)



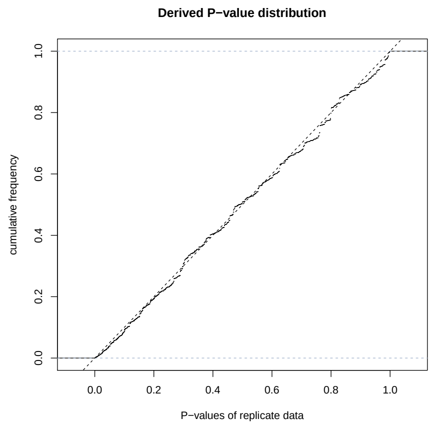
(B)



(C)

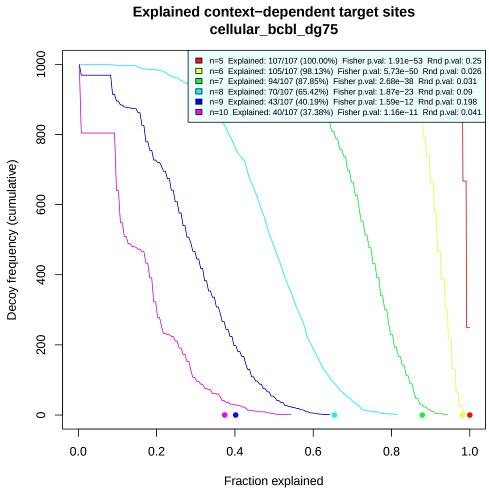


(D)

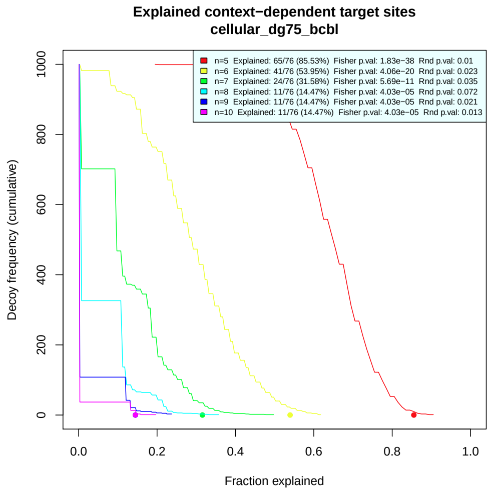


(E)

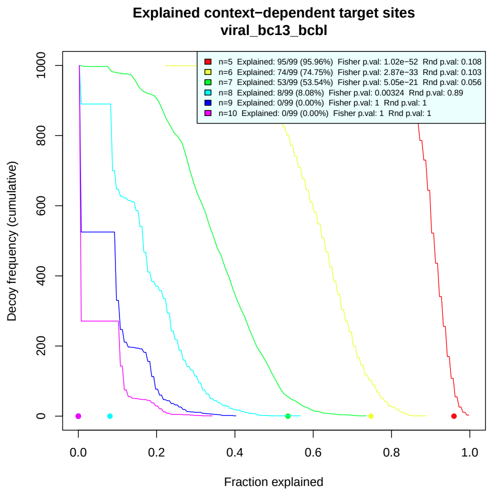
Figure S5: Read count correlation and conditional gamma distribution fit; related to Figure 5. Figure S5A shows the correlation of PAR-CLIP reads and mRNA fold change. Replicate counts were summed and a pseudocount of 1 was used to circumvent division by zero. In red and blue, context-dependent target sites of cellular microRNAs are shown. In Figures S5B and S5C, the correlation of replicate PAR-CLIP read counts are shown in log scale for the DG75 and BCBL1 experiments, respectively (using a pseudocount of 1). The red line indicates the main diagonal. Deviations from the diagonal are obviously larger for weak targets sites (bottom left), indicative for sampling noise inherent to low count data such as PAR-CLIP. Figure S5D shows the fitted gamma distribution parameters against the target site strength. Outliers from the conditional model are indicated in gray (see Supplementary methods). In Figure S5E the p-value distribution of the gamma model applied to replicate measurements is shown. It closely resembles a uniform distribution, which indicates that our model accurately resembles the observed deviations in replicate measurements.



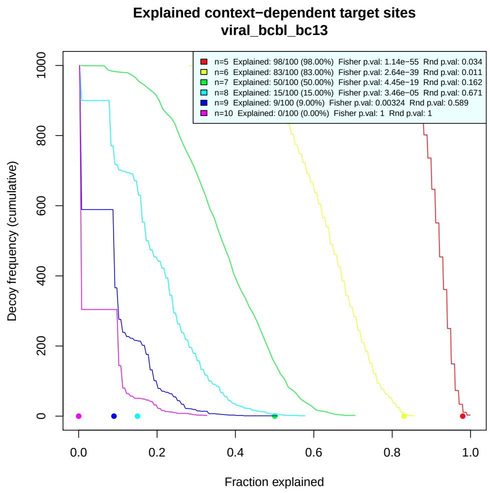
(A)



(B)



(C)



(D)

Figure S6: Motif randomization results; related to Figure 6. For each comparison, positive and negative labels were randomly permuted 1000 times. MERCI was run on each randomized instance and the number of sequences containing an n-discriminative k-mer was counted for n between 5 and 10. A k-mer is n-discriminative, if it occurs in at least n sequences in the positive set and does not occur in the negative set. We plotted the distributions of the fractions of explained sequences and compared them to the actual fractions in the true positive and negative sets (points in the plots).

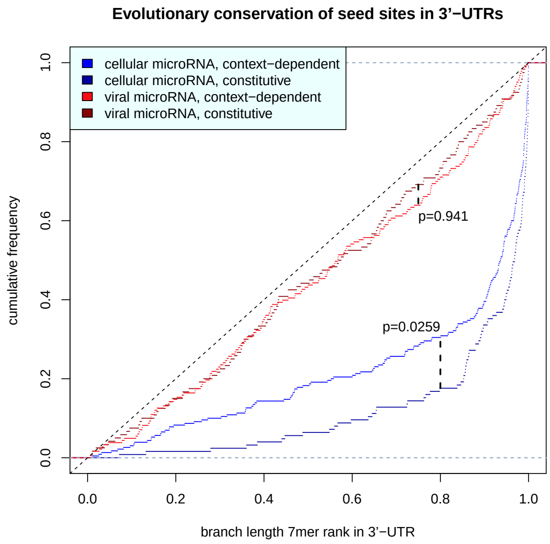
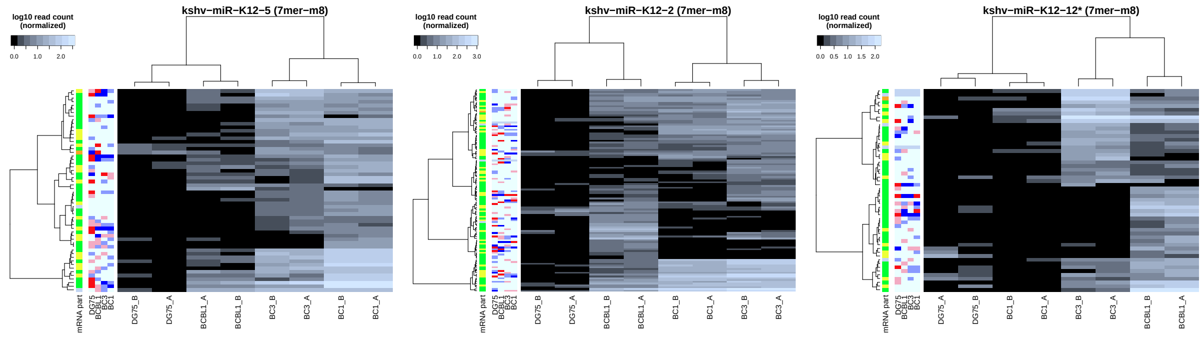
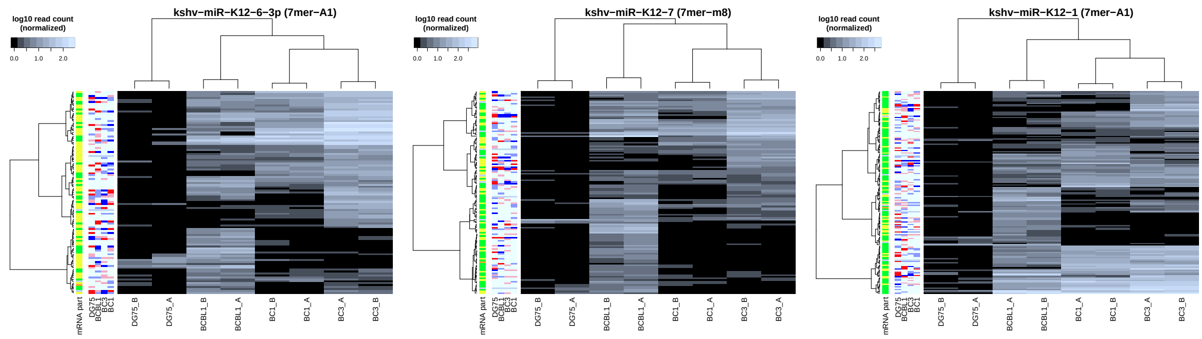


Figure S7: Conservation of target sites; related to Figure 7. Distributions of normalized branch lengths of target sites are illustrated (see main text for a definition of branch lengths). To correct for overall mRNA conservation, branch lengths were computed for all 7-mers in the 3'-UTR and the rank of the seed (normalized between 0 and 1) among all these branch lengths was used as conservation score. All cellular microRNAs considered here are conserved in vertebrates. Constitutive target sites of these microRNAs are significantly more conserved than context-dependent sites, and this stronger conservation is not due to a stronger overall conservation level of the 3'-UTR.



(A) (B) (C)



(D) (E) (F)

Figure S8: Heatmaps of KSHV microRNAs. A PAR-CLIP target site read count heatmap for 6 of 12 KSHV microRNA used in the analyses is shown. The remaining 6 microRNAs can be found in Figure S9. Each row corresponds to a single target site where PARma assigned the corresponding microRNA seed match; each column corresponds to one of the individual PAR-CLIP experiments. The additional annotations on the left side indicate the part of the transcript, where a cluster is located (orange: 5'-UTR; yellow: coding; green: 3'-UTR; gray: not located on known mRNA) and the expression of the transcript in all experiments (red, at least 2-fold lower expression than the mean expression value for this transcript across all experiments; light red, at least 1.4-fold lower expression than the mean; light blue, at least 1.4-fold higher expression; blue, at least 2-fold higher expression).

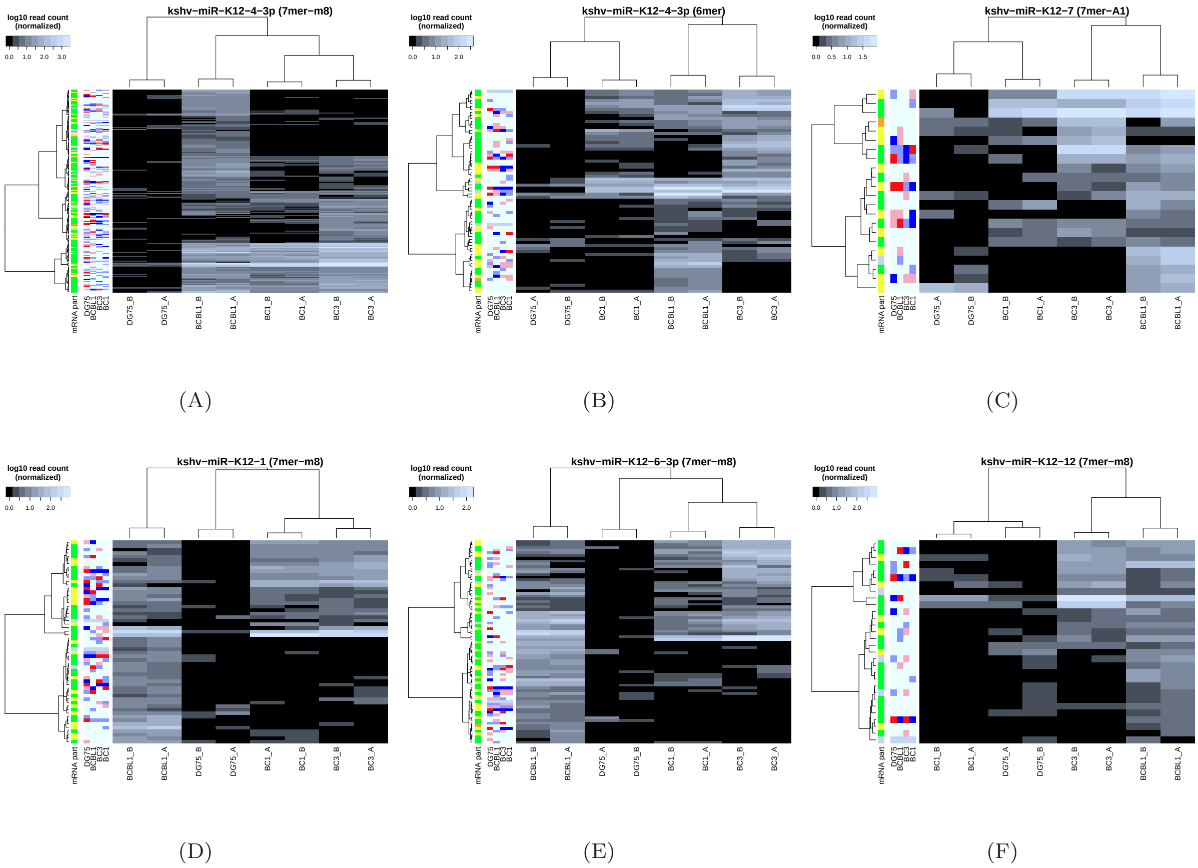


Figure S9: Heatmaps of KSHV microRNAs, continued. See Figure S8 for description.

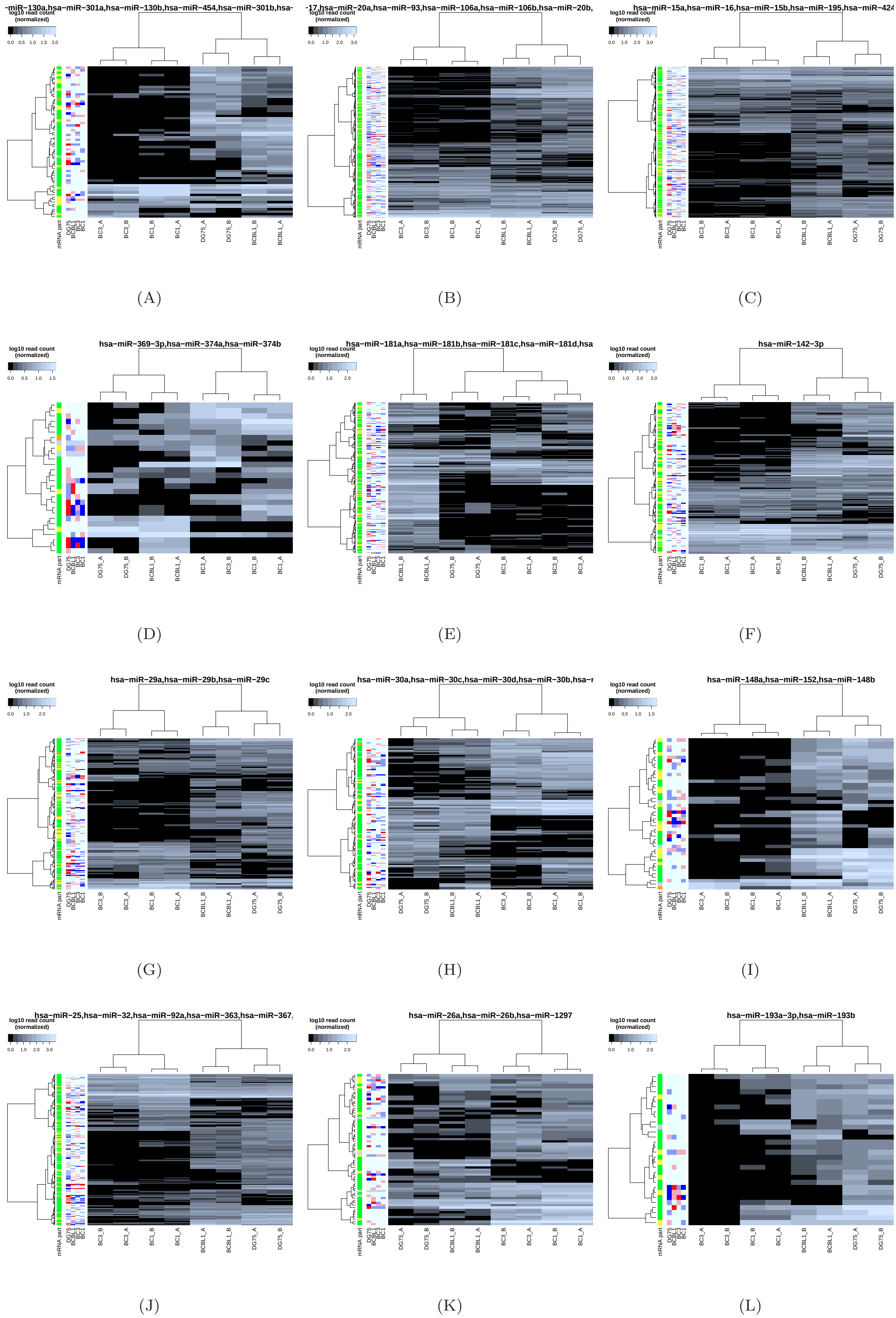
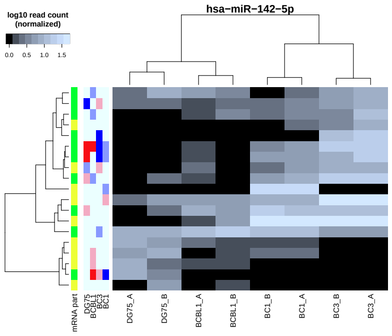
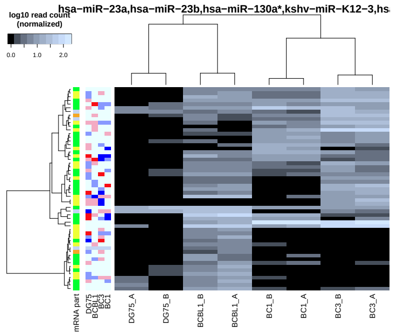


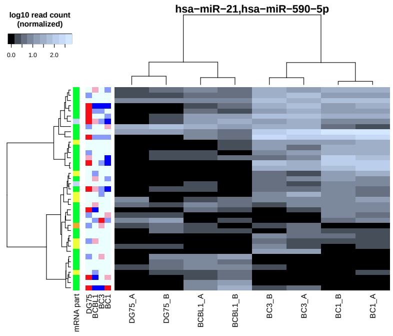
Figure S10: Heatmaps of human microRNAs. A PAR-CLIP target site read count heatmap for 12 of the 20 human microRNA used in the analyses is shown. The remaining 8 are depicted in Figure S10; see Figure S8 for more information about these heatmaps.



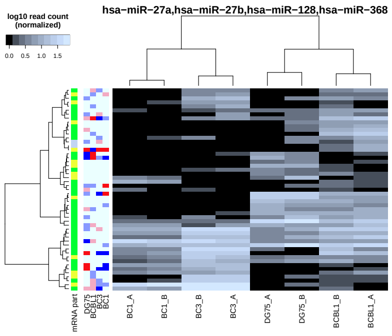
(A)



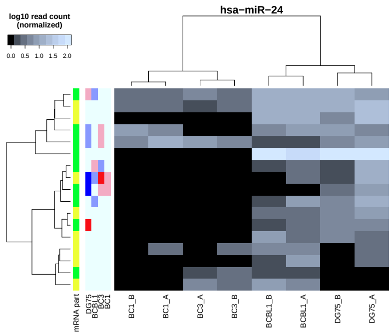
(B)



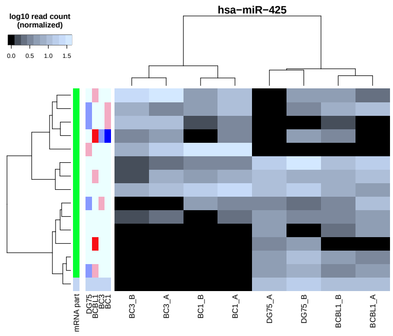
(C)



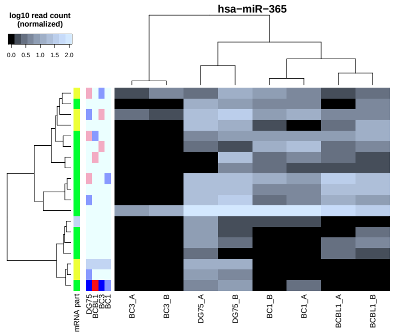
(D)



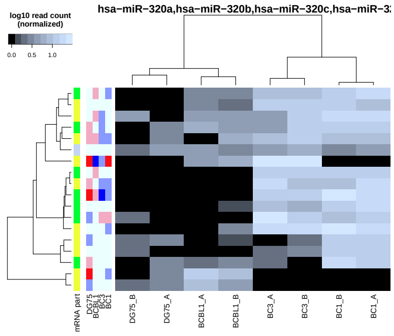
(E)



(F)



(G)



(H)

Figure S11: Heatmaps of human microRNAs, continued. A PAR-CLIP target site read count heatmap for 8 of the 20 human microRNA used in the analyses is shown. The remaining 12 are depicted in Figure S9; see Figure S8 for more information about these heatmaps.

References

Anders, S. and Huber, W., 2010. Differential expression analysis for sequence count data. *Genome Biology*, **11**(10):R106.

Baek, D., Villen, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P., 2008. The impact of microRNAs on protein output. *Nature*, **455**(7209):64–71.

Cazalla, D., Yario, T., and Steitz, J. A., 2010. Down-regulation of a host MicroRNA by a herpesvirus saimiri n oncoding RNA. *Science*, **328**(5985):1563–1566.

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I., 2011. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*, **147**(2):358–369.

Dreszer, T. R., Karolchik, D., Zweig, A. S., Hinrichs, A. S., Raney, B. J., Kuhn, R. M., Meyer, L. R., Wong, M., Sloan, C. A., Rosenbloom, K. R., *et al.*, 2012. The UCSC genome browser database: extensions and updates 2011. *Nucleic acids research*, **40**(Database issue):D918–923.

Erhard, F., Dölken, L., Jaskiewicz, L., and Zimmer, R., 2013. PARma: identification of microRNA target sites in Argonaute PAR-CLIP data. *Genome Biology - in press*, .

Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., Leyva, A., Weigel, D., Garcia, J. A., and Paz-Ares, J., *et al.*, 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature Genetics*, **39**(8):1033–1037.

Friedman, R. C., Farh, K. K.-H., Burge, C. B., and Bartel, D. P., 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, **19**(1):92 –105.

Frohn, A., Eberl, H. C., Stöhr, J., Glasmacher, E., Rüdel, S., Heissmeyer, V., Mann, M., and Meister, G., 2012. Dicer-dependent and -independent argonaute2 protein interaction networks in mammalian cells. *Molecular & Cellular Proteomics*, **11**(11):1442–1456.

Gottwein, E., Corcoran, D. L., Mukherjee, N., Skalsky, R. L., Hafner, M., Nusbaum, J. D., Shamulailatpam, P., Love, C. L., Dave, S. S., Tuschl, T., *et al.*, 2011. Viral MicroRNA targetome of KSHV-Infected primary effusion lymphoma cell lines. *Cell Host & Microbe*, **10**(5):515–526.

Grimson, A., Farh, K. K.-H., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P., 2007. MicroRNA targeting specificity in mammals: Determinants beyond seed pairing. *Molecular Cell*, **27**(1):91–105.

He, L., He, X., Lim, L. P., de Stanchina, E., Xuan, Z., Liang, Y., Xue, W., Zender, L., Magnus, J., Ridzon, D., *et al.*, 2007. A microRNA component of the p53 tumour suppressor network. *Nature*, **447**(7148):1130–1134.

Jaskiewicz, L., Bilen, B., Hausser, J., and Zavolan, M., 2012. Argonaute CLIP - a method to identify in vivo targets of miRNAs. *Methods (San Diego, Calif.)*, **58**(2):106–112.

Kiriakidou, M., Nelson, P. T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., and Hatzigeorgiou, A., 2004. A combined computational-experimental approach predicts human microRNA targets. *Genes & development*, **18**(10):1165–1178.

Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., and Zavolan, M., 2011. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nature Methods*, **8**(7):559–564.

Leung, A. K. L., Vyas, S., Rood, J. E., Bhutkar, A., Sharp, P. A., and Chang, P., 2011. Poly(ADP-ribose) regulates stress responses and microRNA activity in the cytoplasm. *Molecular cell*, **42**(4):489–499.

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**(7027):769–773.

Linsley, P. S., Schelter, J., Burchard, J., Kibukawa, M., Martin, M. M., Bartz, S. R., Johnson, J. M., Cummins, J. M., Raymond, C. K., Dai, H., *et al.*, 2007. Transcripts targeted by the MicroRNA-16 family cooperatively regulate cell cycle progression. *Molecular and Cellular Biology*, **27**(6):2240–2252.

Marcinowski, L., Tanguy, M., Krmpotic, A., Rädle, B., Lisni, V. J., Tuddenham, L., Chane-Woon-Ming, B., Ruzsics, Z., Erhard, F., Benkartek, C., *et al.*, 2012. Degradation of cellular miR-27 by a novel, highly abundant viral transcript is important for efficient virus replication in vivo. *PLoS Pathog*, **8**(2):e1002510.

Mukherji, S., Ebert, M. S., Zheng, G. X. Y., Tsang, J. S., Sharp, P. A., and Oudenaarden, A. v., 2011. MicroRNAs can generate thresholds in target gene expression. *Nature Genetics*, **43**(9):854–859.

Naeem, H., Küffner, R., and Zimmer, R., 2011. MIRTFnet: analysis of miRNA regulated transcription factors. *PLoS ONE*, **6**(8):e22519.

Pace, C. N., Heinemann, U., Hahn, U., and Saenger, W., 1991. Ribonuclease t1: Structure, function, and stability. *Angewandte Chemie International Edition in English*, **30**(4):343360.

Pasquinelli, A. E., 2012. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nature Reviews Genetics*, **13**(4):271–282.

Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., and Siepel, A., 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome research*, **20**(1):110–121.

Qi, H. H., Ongusaha, P. P., Myllyharju, J., Cheng, D., Pakkanen, O., Shi, Y., Lee, S. W., Peng, J., and Shi, Y., 2008. Prolyl 4-hydroxylation regulates argonaute 2 stability. *Nature*, **455**(7211):421–424.

Ritchie, W., Flamant, S., and Rasko, J. E. J., 2009. Predicting microRNA targets and functions: traps for the unwary. *Nature Methods*, **6**(6):397–398.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K., 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**(1):139–140.

Rüdel, S., Flatley, A., Weinmann, L., Kremmer, E., and Meister, G., 2008. A multifunctional human argonaute2-specific monoclonal antibody. *RNA (New York, N.Y.)*, **14**(6):1244–1253.

Rüdel, S., Wang, Y., Lenobel, R., Körner, R., Hsiao, H.-H., Urlaub, H., Patel, D., and Meister, G., 2011. Phosphorylation of human argonaute proteins affects small RNA binding. *Nucleic acids research*, **39**(6):2330–2343.

Rybak, A., Fuchs, H., Hadian, K., Smirnova, L., Wulczyn, E. A., Michel, G., Nitsch, R., Krappmann, D., and Wulczyn, F. G., 2009. The let-7 target gene mouse lin-41 is a stem cell specific e3 ubiquitin ligase for the miRNA pathway protein ago2. *Nature Cell Biology*, **11**(12):1411–1420.

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P., 2011. A ceRNA hypothesis: The rosetta stone of a hidden RNA language? *Cell*, **146**(3):353–358.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N., 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**(7209):58–63.

Sethupathy, P., Megraw, M., and Hatzigeorgiou, A. G., 2006. A guide through present computational approaches for the identification of mammalian microRNA targets. *Nature Methods*, **3**(11):881–886.

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., *et al.*, 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome research*, **15**(8):1034–1050.

The ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**(7414):57–74.

Thomas, M., Lieberman, J., and Lal, A., 2010. Desperately seeking microRNA targets. *Nat Struct Mol Biol*, **17**(10):1169–1174.

Tu, K., Yu, H., Hua, Y.-J., Li, Y.-Y., Liu, L., Xie, L., and Li, Y.-X., 2009. Combinatorial network of primary and secondary microRNA-driven regulatory mechanisms. *Nucleic Acids Research*, **37**(18):5969–5980.

Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., *et al.*, 2012a. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Research*, **22**(9):1680–1688.

Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T. W., Greven, M. C., Pierce, B. G., Dong, X., Kundaje, A., Cheng, Y., *et al.*, 2012b. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Research*, **22**(9):1798–1812.

Xu, G., Fewell, C., Taylor, C., Deng, N., Hedges, D., Wang, X., Zhang, K., Lacey, M., Zhang, H., Yin, Q., *et al.*, 2010. Transcriptome and targetome analysis in MIR155 expressing cells using RNA-seq. *RNA*, **16**(8):1610–1622.

Yanez-Cuna, J. O., Dinh, H. Q., Kvon, E. Z., Shlyueva, D., and Stark, A., 2012. Uncovering cis-regulatory sequence requirements for context specific transcription factor binding. *Genome Research*, .