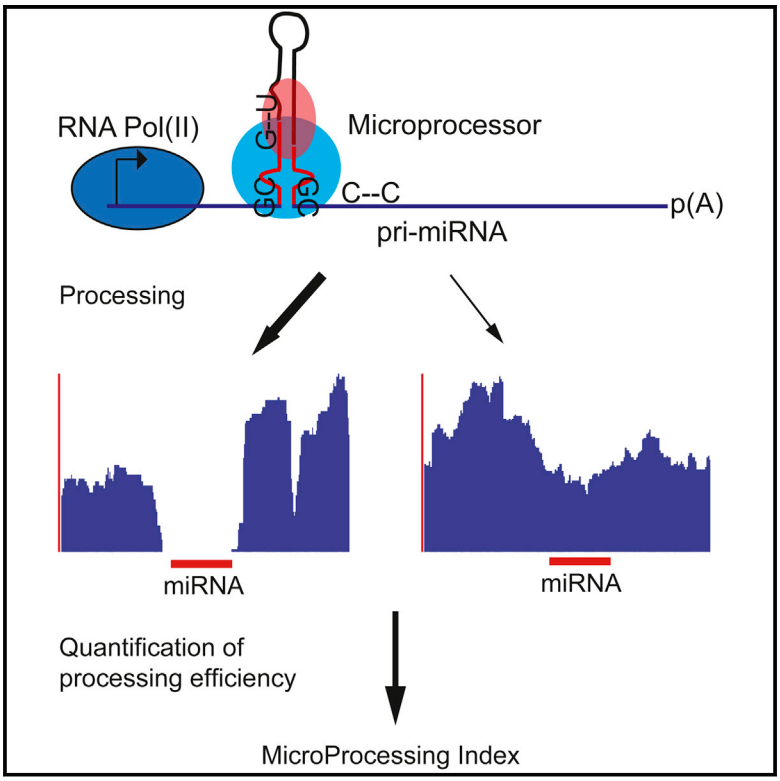
Article

Cell Reports

Microprocessor Activity Controls Differential miRNA Biogenesis In Vivo

Graphical Abstract



Authors

Thomas Conrad, Annalisa Marsico, Maja Gehre, Ulf Andersson Ørom

Correspondence

oerom@molgen.mpg.de

In Brief

Conrad et al. describe an approach for the transcriptome-wide assessment of pri-miRNA processing into pre-miRNA. The authors use this to quantify the processing efficiencies of individual miRNAs and show processing to be more important than primary transcript expression itself to determine the level of mature miRNAs within a cell.

Highlights

In vivo insight into miRNA biogenesis using next-generation RNA sequencing

Transcriptome-wide estimation of in vivo pri-miRNA processing efficiencies

Differential Microprocessor cleavage as a key regulatory step in miRNA biogenesis

Uncoupling of miRNA biogenesis from host gene expression by Microprocessor

Accession Numbers

GSE56862

Conrad et al., 2014, Cell Reports 9, 542–554
October 23, 2014 ©2014 The Authors
http://dx.doi.org/10.1016/j.celrep.2014.09.007

OPEN ACCESS
CellPress

Cell Reports
Article

Microprocessor Activity Controls Differential miRNA Biogenesis In Vivo

Thomas Conrad,[1,5] Annalisa Marsico,[2,3,5] Maja Gehre,[1,4] and Ulf Andersson Ørom[1,*]
[1]Otto Warburg Laboratories, Noncoding RNA Research Group, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[2]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[3]Department of Mathematics and Informatics, Free University of Berlin, 14195 Berlin, Germany
[4]Department of Biochemistry, Free University of Berlin, 14195 Berlin, Germany
[5]Co-first author
*Correspondence: oerom@molgen.mpg.de
http://dx.doi.org/10.1016/j.celrep.2014.09.007

SUMMARY

In miRNA biogenesis, pri-miRNA transcripts are converted into pre-miRNA hairpins. The in vivo properties of this process remain enigmatic. Here, we determine in vivo transcriptome-wide pri-miRNA processing using next-generation sequencing of chromatin-associated pri-miRNAs. We identify a distinctive Microprocessor signature in the transcriptome profile from which efficiency of the endogenous processing event can be accurately quantified. This analysis reveals differential susceptibility to Microprocessor cleavage as a key regulatory step in miRNA biogenesis. Processing is highly variable among pri-miRNAs and a better predictor of miRNA abundance than primary transcription itself. Processing is also largely stable across three cell lines, suggesting a major contribution of sequence determinants. On the basis of differential processing efficiencies, we define functionality for short sequence features adjacent to the pre-miRNA hairpin. In conclusion, we identify Microprocessor as the main hub for diversified miRNA output and suggest a role for uncoupling miRNA biogenesis from host gene expression.

INTRODUCTION

MicroRNAs (miRNA) are small RNAs that posttranscriptionally regulate gene expression (Kim et al., 2009). miRNAs are expressed as long primary miRNA (pri-miRNA) transcripts that are processed in the nucleus to precursor miRNA hairpins (pre-miRNA) by the Microprocessor complex. The pre-miRNAs are exported to the cytoplasm by Exportin 5 (Yi et al., 2003) and are further processed into mature miRNAs by Dicer and incorporated into the RNA-induced silencing complex (RISC) where they exert their predominant functions in regulating translation and the stability of target mRNAs (Kim et al., 2009). miRNAs recognize their targets by base-pair complementarity. The nucleotides 2–7 of the mature miRNA, termed the seed, have been identified as the most essential region for target recognition (Lewis et al., 2003, 2005). Each miRNA is believed to regulate several hundred targets (Lewis et al., 2005; Miranda et al., 2006; Sood et al., 2006), making up extensive gene expression regulatory networks.

The Microprocessor complex was characterized 10 years ago as the minimal complex responsible for pri-miRNA processing and shown to consist of the RNase III enzymes Drosha and DGCR8 (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004). Sequence analysis of pri-miRNA substrates coupled with in vitro studies has revealed a characteristic stem loop structure of the pre-miRNA region to be required for Microprocessor cleavage, which mediates a characteristic 2 nucleotide offset cut on adjacent strands of the helix one helical turn distal to the base of the precursor hairpin (Han et al., 2006). How the Microprocessor is able to specifically recognize and process pri-miRNAs while avoiding the widespread similar RNA stem loop structures that can form across the transcriptome is still largely enigmatic. A recent study used a screening approach in vitro to identify short sequence motifs in the region flanking the pre-miRNA that affect human miRNA biogenesis and can also function in vitro as enhancers for processing of *C. elegans* pri-miRNA in human cells (Auyeung et al., 2013). Although the global impact of these flanking motifs on endogenous pri-miRNA processing has not been defined, their general occurrence in human pri-miRNA sequences suggests that they play an important role for processing.

Besides their role in defining Microprocessor targets in general, variations in pri-miRNA sequence composition may help to fine-tune miRNA expression levels. Several miRNAs belong to families, the most prominent being the let-7 family, consisting of nine different members (Obad et al., 2011). miRNA families are, due to their identical seed regions, assumed to target the same mRNAs for translation regulation. Sequence variations in primary transcripts of individual miRNA family members might affect pri-miRNA cleavage, leading to differential expression within miRNA families. Furthermore, it has been demonstrated that miRNAs arranged in clusters and sharing the same primary transcript can undergo differential processing (Chaulk et al., 2011). However, despite their prominent role in the first step of miRNA biogenesis, a global view on the in vivo sequence determinants of Microprocessor activity has not yet been obtained.

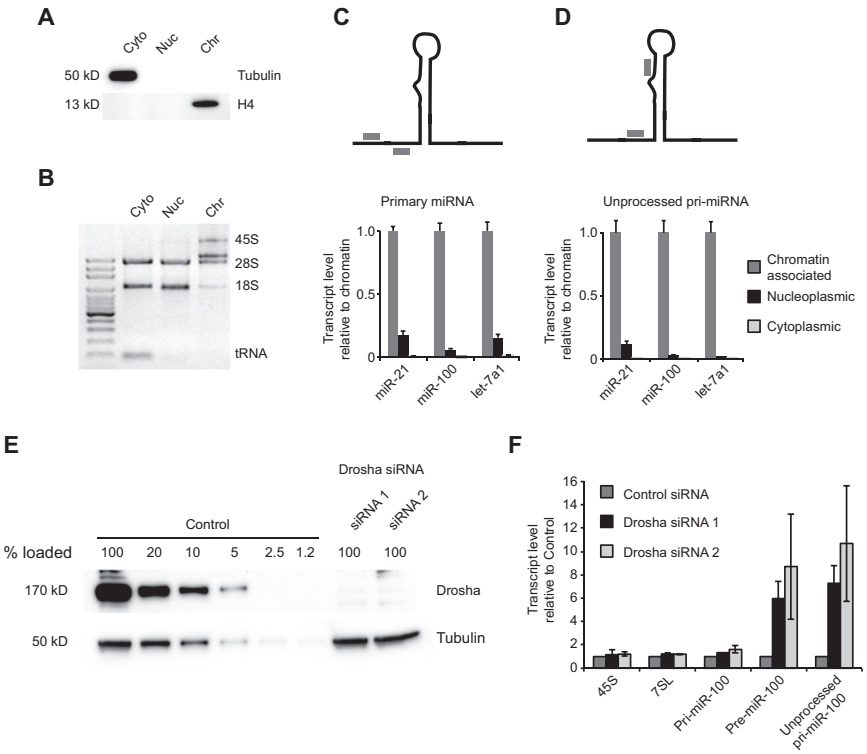


Figure 1. Chromatin Purification of Pri-miRNA Transcripts

(A) Markers for chromatin (histone H4) and cytoplasm (Tubulin) are shown by western blot.

(B) Subcellular distribution of distinct RNA species shown by agarose gel electrophoresis and staining with SYBR safe. Cyto, cytoplasm; Nuc, nucleoplasm; Chr, chromatin.

(C and D) Relative abundances of the indicated pri-miRNA species in the three cellular compartments, determined by real-time qPCR. The top shows the primer locations for detection of the unprocessed and processed pri-miRNA transcripts, respectively. Transcript levels were normalized to the chromatin fraction for each primer pair.

(E) Knockdown efficiency of Drosha using two different siRNAs determined by western blot.

(F) Real-time qPCR of chromatin-associated RNA shows accumulation of pre-mir-100 and unprocessed pri-mir-100 upon Drosha RNAi, whereas 45S rRNA, 7SL noncoding RNA, and basal pri-mir-100 levels remain unaffected. Same amounts of chromatin-RNA were used for RT in each experiment, and ct values were directly compared to the control siRNA. Error bars represent SDs of triplicate experiments.

See also Figure S1.

Although the core Microprocessor components Drosha and DGCR8 are sufficient to cleave pri-miRNAs in vitro (Gregory et al., 2004), a growing number of cofactors have recently been suggested to impact on Microprocessor function in vivo. Some cofactors promote processing of individual pri-miRNAs in specific cell types or developmental stages (Siomi and Siomi, 2010), including the DEAD box helicases DDX5 and DDX17 (Fukuda et al., 2007; Mori et al., 2014), KSRP (Trabucchi et al., 2009), hnRNP A1 (Guil and Cáceres, 2007), BRCA1 (Kawai and Amano, 2012), and Fus (Morlando et al., 2012). Accordingly, a number of factors have been shown to suppress cleavage of individual pri-miRNAs by the Microprocessor, including NF90-NF45 (Sakamoto et al., 2009), Lin28B (Piskounova et al., 2011), MSI2 and HuR (Choudhury et al., 2013), QKI (Wang et al., 2013), and MeCP2 (Cheng et al., 2014). Processing of pri-miRNAs is beginning to emerge as an actively regulated process that is responsive to TGFß signaling via SMADs (Davis et al., 2008); DNA damage via Microprocessor interaction with p53 (Suzuki et al., 2009); or cell density via the Hippo pathway (Mori et al., 2014). Nevertheless, the above studies have focused on individual miRNAs and in vitro assays. To date the overall impact and dynamic range of endogenous Microprocessor activity remains unclear.

In this work, we show that the endogenous Microprocessor activity toward individual pri-miRNAs can be determined using RNA sequencing. We identify the Microprocessor cleavage signature and define the MicroProcessing Index (MPI) as a measure for processing efficiency. We provide experimental evidence that processing efficiency is highly variable among canonical pri-miRNAs and a major determinant for the expression levels of individual mature miRNAs. We show that the processing of individual pri-miRNAs is similar between the three cell lines included in this study, suggesting that the observed diversity in processing is largely dictated by the diverse substrate sequences. We finally use the endogenous pri-miRNA cleavage activities to derive specific sequence motifs that are associated with increased processing.

RESULTS

Pri-miRNAs Can Be Captured by Chromatin Purification

The transcription of pri-miRNA transcripts and their processing to pre-miRNAs are both nuclear events of miRNA biogenesis. Because processing of pri-miRNA transcripts has been proposed to occur cotranscriptionally (Morlando et al., 2008; Pawlicki and Steitz, 2008), it likely happens associated with chromatin. To enrich for pri-miRNA transcripts, we established a cellular fractionation protocol to obtain highly purified chromatin fractions from cells grown in tissue culture (Figures 1A and 1B and S1A–S1D). RNA from these fractions was sequenced as described below and in Experimental Procedures. To determine relative pri-miRNA abundances in the cellular fractions, we used two different sets of primers (Figures 1C and 1D). One set covering an unprocessed region to detect global primary transcript levels reveals 10%–20% localization to the nucleoplasm compared to chromatin-associated RNA (Figure 1C). Using primers that span the pre-miRNA processing sites, we find only 2%–5% of the unprocessed transcript in the nucleoplasm

(Figure 1D). These data suggest that most pri-miRNA transcripts found in the nucleoplasm have been processed by Microprocessor. To assess the processing of individual pri-miRNAs, we used small interfering RNAs (siRNAs) to knock down Drosha (Figure 1E), which did not affect the levels of ribosomal RNA (Figure 1F). At the same time, the expression levels of pri-miRNAs in general are only slightly induced upon Drosha knockdown, whereas a severalfold increase in uncleaved processing sites is observed within the primary transcript (Figure 1F). This specific increase upon Drosha knockdown recapitulates reduced processing into mature miRNAs in the absence of a functional Microprocessor, resulting in more of the unprocessed pri-miRNA transcript. The difference in transcript response to Drosha knockdown between the full-length pri-miRNA and the unprocessed cleavage sites indicates that under normal conditions the cleaved pri-miRNA is more stable than anticipated. Based on these data, we propose that a comprehensive coverage of pri-miRNA transcripts and processing should be obtained from sequencing of RNA associated with the chromatin fraction.

Primary Transcript Sequencing Reveals the Microprocessor Signature

A poly(A) selection step is often used in preparation of RNA for next-generation RNA sequencing to obtain samples with a minimal content of ribosomal RNAs. We used next-generation high-throughput sequencing to generate deep coverage RNA profiles of HeLa cells (~180 M paired-end reads for each sample). Both poly(A)-enriched and non-poly(A)-enriched (rRNA-depleted) protocols were applied to chromatin-associated RNA from HeLa cells transfected with either a control nontargeting siRNA or an siRNA that specifically targets Drosha (Drosha siRNA 2) (for overview of the sequencing approach, see Figure 2A). Assessing read coverage at miRNA loci in sequencing data from poly(A)-selected samples reveals a loss of the transcript upstream of the 3′ Microprocessor cleavage site as a consequence of the poly(A) selection step, both in total cellular RNA and in chromatin-associated RNA (Figure 2B, top). Following the knockdown of Drosha, this termination is abrogated, and the full-length pri-miRNA transcripts can be sequenced and identified (Figure 2B, bottom). Using non-poly(A)-selected RNA sequencing, pri-miRNA sequences 5′ of the processing sites can be detected (Figure 2C), suggesting that the processed pri-miRNA transcript remains associated with chromatin after processing despite being cleaved. In summary, the combination of chromatin purification to enrich for pri-miRNAs and sequencing without poly(A) selection to obtain expression information on discontinuous transcripts provides an unprecedented detailed view into the nature of pri-miRNAs. As a comparison, sequencing reads covering the GAPDH transcript are shown, where distinct exons are evident due to the continuity of the processed transcript (Figures 2B and 2C). Also, for the special class of miRNAs called miRtrons, distinct processing properties can be observed, reflecting the Drosha independent cleavage mechanism (Figures 2B and 2C). Shown in Figure 3A are detailed views on the let-7a-1, let-7d, and let-7f pre-miRNA loci, which reside in the same polycistronic pri-miRNA. Despite being processed from the same pri-miRNA, the Microprocessor cleavage signatures appear differently pronounced at the three loci (Figure 3A, top). Quantitative PCR (qPCR) measurements across the respective cleavage sites reveals a differential susceptibility to Drosha depletion that correlates with the extent of the Microprocessor cleavage signatures, suggesting that the extent of this Microprocessor signature reflects differential in vivo pri-miRNA processing efficiencies (Figure 3B).

The MPI Is a Measure for Genome-wide pri-miRNA Processing

The quantitative properties of the Microprocessor signature as determined in Figures 3A and 3B implicate the possibility for a genome-wide assessment of Microprocessor activity, similar to studies using read coverage at exon-intron junctions to estimate splicing efficiency (Tilgner et al., 2012). To obtain a quantitative measure for processing efficiency, we defined the MicroProcessing Index (MPI). The MPI is taking into account the expression level of the pri-miRNA, as determined by sequencing reads adjacent to the pre-miRNA, and the read density in the precursor region (Figure 3C). The MPI represents the relative extent of the Microprocessor signature. The MPI is computed for each individual miRNA, excluding loci that overlap exon junctions or transcription start sites to avoid confounding signals (see the Supplemental Experimental Procedures). We applied this measure to the sequenced chromatin-associated RNA from HeLa, A549, and HEK293 cells and calculated the MPI value for all pri-miRNA transcripts that show enough read coverage in the regions surrounding the precursor (expression of the primary transcript at least 1.0 RPKM) (Tables S1 and S2). Because the processing is a negative event regarding the transcript level of the pri-miRNA, a significant depletion of read density in the pre-miRNA region (MPI < -1.0; $p < 0.01$), i.e., negative MPI, corresponds to efficiently processed pri-miRNAs, whereas MPI values close to 0 or positive values correspond to inefficiently processed pri-miRNAs. To avoid inclusion of nonauthentic miRNA hairpins, we performed small RNA sequencing and restricted the further analyses to miRNAs that meet the following criteria from Chiang et al. (2010): (1) at least 20 reads per miRNA in two replicates; (2) the precursor region folding into a hairpin; (3) absence of other annotated noncoding RNA species; (4) presence of reads corresponding to a miRNA* species; and (5) homogeneous reads in small RNA sequencing data for both the 5′ and 3′ mature miRNAs (Table S3). Broadly conserved miRNAs had to fulfill four criteria and weakly conserved and nonconserved miRNAs had to fulfill all of these five criteria to be included in the further analysis. Because the subsequent analyses address both efficiently and inefficiently processed pri-miRNAs, we applied a low expression cutoff of 1 RPM (approximately ten reads per replicate) to include miRNAs at all expression levels. This results in a filtered set of 229 annotated miRNA precursors that are used for further analysis.

To test the dependency of processing efficiency on Microprocessor levels, we performed deep sequencing of chromatin-associated transcripts isolated from Drosha-depleted HeLa cells and quantified the effect on individual pri-miRNAs by means of a deltaMPI value (see the Supplemental Experimental Procedures; Figure 3D). As expected for Drosha-dependent processing, pri-miRNAs with high processing efficiency in the control cells show

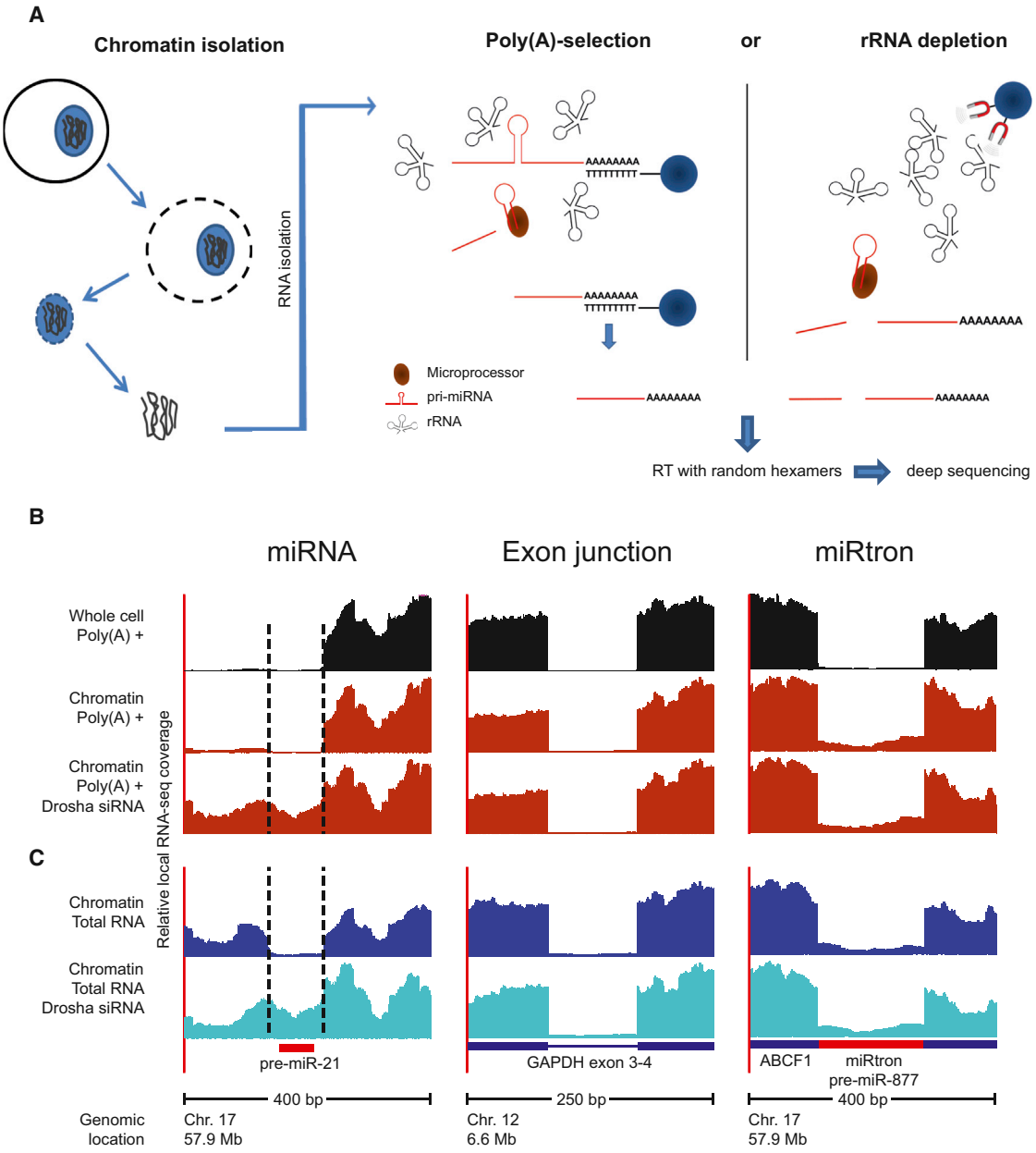


Figure 2. In Vivo miRNA Processing Can Be Visualized by Deep Sequencing.

(A) Overview of the sequencing approach. Chromatin-associated RNA is isolated by cellular fractionation and Trizol extraction to enrich miRNA primary transcripts. Pri-miRNAs cleaved by Microprocessor will be only partially recovered during subsequent poly(A) selection. Alternatively, selective rRNA depletion also retains pri-miRNA fragments upstream of the cleavage site. Random hexamers are used as RT primers to convert all isolated pri-miRNA fragments to cDNA for sequencing.

(B) RNA sequencing reads after poly(A) selection shown for the miR-21 locus, a GAPDH exon-intron junction, and a miRtron, miR-877. Shown are tracks from whole cells, purified chromatin, and chromatin upon Drosha RNAi (top to bottom). Dotted lines indicate Microprocessor cleavage sites.

(C) Sequencing reads for the same loci as in (B) for chromatin-associated RNA upon rRNA depletion (Chromatin total RNA) with and without Drosha RNAi.

a greater response to Drosha depletion compared to noneffi-ciently cleaved pri-miRNAs.

To provide further support for the specificity of our approach, we compared the MPI distribution of pri-miRNAs with a set of 2,010 unrelated pseudohairpins that have been selected based on their similarity to miRNA hairpins (the Supplemental Experimental Procedures). We observe a distinct shift toward negative MPI values that is specific for pri-miRNAs ($p < 2.2 \times 10^{-16}$, Figure 3E). This negative MPI distribution is dependent on the presence of the Microprocessor complex and abrogated upon Drosha knockdown. In contrast, pseudohairpins show no sign of endogenous processing, with MPI values centered around 0 in both conditions.

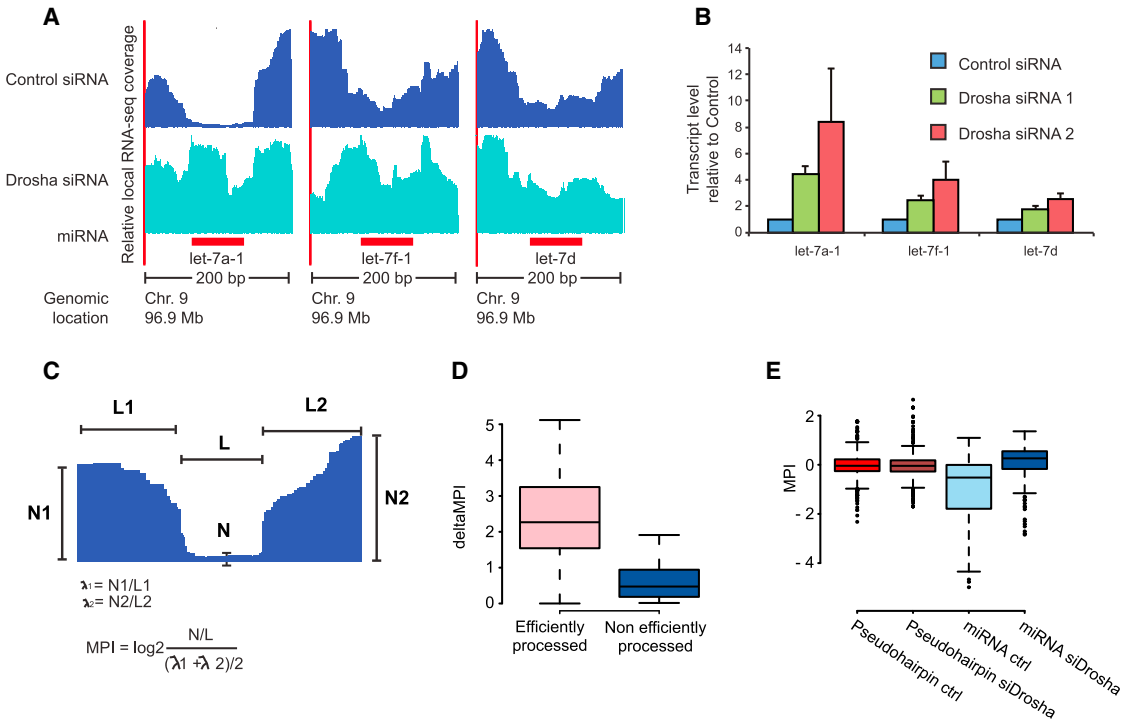


Figure 3. The MicroProcessing Index

(A) Chromatin total RNA read counts at the let-7a-1, let-7f-1, and let-7d pre-miRNA loci with and without Drosha RNAi. All three miRNAs originate from the same polycistronic primary transcript.

(B) Differential increase in uncleaved processing sites of the same pre-miRNAs upon Drosha RNAi determined by real-time qPCR. Primers were designed as depicted in Figure 1D. Error bars represent SD of triplicate experiments.

(C) Definition of the MPI. Shown is a schematic representation of a Microprocessor signature. The MPI is the log2 ratio of the read density in the precursor region and the flanking 100 bp.

(D) Fold change of the MPI upon Drosha RNAi for efficiently and nonefficiently processed pri-miRNAs.

(E) The MPI distribution for pseudohairpins and pri-miRNAs with and without knockdown of Drosha. Solid lines indicate the median.

See also Tables S1, S2, and S3.

Replicate analysis of the MPIs from HeLa chromatin-associated RNA shows a high correlation (Pearson correlation 0.87, p value $<2.2 \times 10^{-16}$), indicating that the values for processing efficiencies derived from sequencing data are highly reproducible (Figure 4A). Comparative analyses show that the MPI of individual miRNAs vary little between cell lines (Figures 4B, S2A, and S2B), with a Pearson correlation of 0.88, 0.82, and 0.82, respectively (p value $\leq 2.2 \times 10^{-16}$ in all cases). Correlations between cell lines are still very stable when considering only broadly conserved miRNAs in the analysis (Pearson correlation of 0.85, 0.87, 0.83, and 0.81, p value $<2.2 \times 10^{-16}$ in all cases, Figures S3A–S3D). Such correlations also hold when comparing only miRNAs with MPI values less than −1.0 or higher than 0.5 in order to discard the possibility that high correlation values result from the clustering of points with MPI close to 0 (R = 0.75, R = 0.76, R = 0.58, R = 0.66 between HeLa replicates HeLa and A549, HeLa and HEK293, and A549 and HEK293, respectively; all p values $<1.0 \times 10^{-5}$; Figures S3E–S3H). The similarity in processing between cell lines suggests that invariant features, such as primary sequence and secondary structure of the pri-miRNA transcripts, are the major determinants of endogenous pri-miRNA processing.

We observe a wide range of Microprocessor activities on individual pri-miRNAs, ranging from very low processing (e.g. mir-573; Figure 4C) to almost complete processing for mir-181a1 (Figure 4D). Ninety seven pri-miRNAs display an MPI below −1.0 in at least one cell line. The biological significance of this threshold is apparent upon comparison with mature miRNA levels derived from small RNA sequencing with a median mature RPM of 38 for 114 inefficiently, and 937 for 72 efficiently processed miRNAs in HeLa (Figure 4E; Table S1). Processing efficiencies are distributed over a wide range, suggesting that the regulation of Microprocessor activity exceeds the simple discrimination of pri-miRNAs from non-pri-miRNA hairpins. Differences in MPI between cell lines could reveal active regulation at the Microprocessor level and thus provide valuable cues for further study of individual pri-miRNAs. We therefore determined differentially processed pri-miRNAs that are expressed in two cell lines and show a difference in MPI of at least 1 (log2 scale). Only a fraction of pri-miRNAs shows differential processing between the tested cell lines according to these criteria (6.3% in HeLa versus A549; 16.7% in HeLa versus HEK293; 22.8% in A549 versus HEK293), with HEK293 cells appearing more different from HeLa and A549 (Tables S1 and S2). We observe

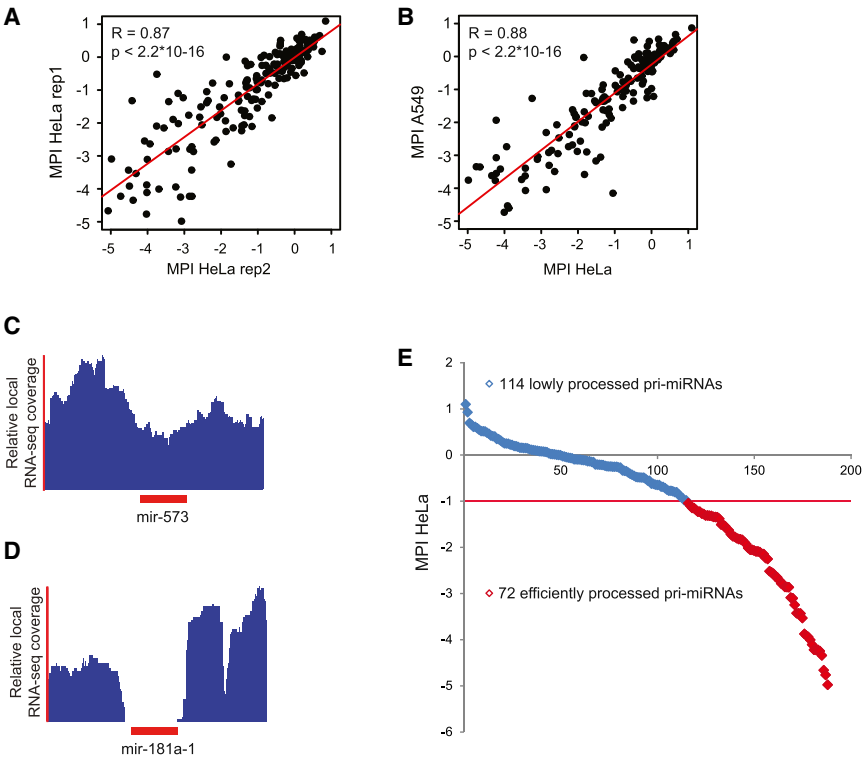


Figure 4. The MPI Reflects Processing Efficiency

(A) Pearson correlation of MPIs from two independent HeLa RNA-seq experiments.

(B) Correlation between MPIs measured in HeLa and A549 cells.

(C) Read distribution across the inefficiently processed pri-mir-573 locus.

(D) Read distribution across the efficiently processed pri-mir-181-a1 locus.

(E) Distribution of in vivo processing efficiencies. Plotted are the ranked MPI values measured in HeLa for expressed pri-miRNAs.

See also Figures S2 and S3.

the highest deltaMPI for pri-mir-21 (3.46 in HEK293 versus HeLa), which is known to be regulated at the Microprocessor level (Davis et al., 2008). Overall differences in processing show a median deltaMPI for differentially processed pri-miRNAs of 1.22.

Pri-miRNA Processing Is a Major Determinant of miRNA Expression

Many miRNAs are expressed from polycistronic transcripts or reside in introns of protein coding host genes. Accordingly, we observe only a limited correlation between pri-miRNA and mature miRNA expression, reflecting substantial uncoupling of host gene expression and miRNA production (R = 0.33, p value 9.3×10^{-6}; Figure 5A). It is still unclear at what stage of the multistep miRNA biogenesis pathway steady-state miRNA levels are primarily controlled. We thus used miRNA sequencing data from HeLa, A549, and HEK293 cells to investigate the relationship between Microprocessor activity and mature miRNA abundance. The correlation between mature miRNA expression and primary transcript levels is increased when only efficiently processed transcripts (MPI < −1) are included in the analysis (Pearson correlation 0.44, p value 1.1×10^{-7}; Figure 5A). Furthermore, we observe a significant correlation between the MPI and mature miRNA expression levels in all three cell lines (Pearson correlations −0.46, −0.44, and −0.52 in HeLa, A549, and HEK293, respectively; all p values < 1.0×10^{-10}; Figures 5B, S4A, and S4B). This suggests that mature miRNA steady-state levels can be modeled using a combination of pri-miRNA expression and the MPI. Importantly, there is no significant correlation between pri-miRNA transcript levels and the MPI (Pearson correlation 0.038, p value 0.61; Figure 5C), underlining that processing efficiency is not simply defined by transcript abundance and that the MPI is an unbiased estimator of pri-miRNA processing. We derived a quantitative linear model to explain mature miRNA expression from both variables in HeLa cells. We evaluated the performance of the model by determining the Pearson correlation coefficient between modeled and measured mature expression (Figures 5D and 5E). Considering that additional steps such as Dicer cleavage and miRNA stability affect miRNA expression after initial processing, the analysis shows a surprisingly robust agreement between modeled and measured expression values (R = 0.55, p = 8.8×10^{-16}). Interestingly, this model returns a higher predictive value for the processing efficiency than for the primary transcript level itself (coefficient for primary transcription = 0.31; MPI = −0.45). In addition, we verify that this relationship is general, as we can predict mature miRNA expression in the other cell types (A549 and HEK293) using a model trained on values from HeLa cells (R = 0.48 and R = 0.50, p = 2.9×10^{-11} and p = 1.9×10^{-13}; Figures 5F and 5G). When considering only conserved miRNAs, these correlations are R = 0.67, R = 0.43, R = 0.40, and R = 0.38, respectively, and all p values are < 5.3×10^{-4} (Figures S5A–S5D). Correlations between predicted and measured miRNA expression are still significant when we apply this analysis to only miRNAs with MPI values less than −1.0 or higher than 0.5 indicating the robustness of our results also with a reduced data set size (R = 0.45, R = 0.48, R = 0.48, R = 0.44, all p values are <1.0×10^{-5}; Figures S5E–S5H). These data suggest that differential Microprocessor activity is a key regulatory mechanism to uncouple miRNA from host gene expression and achieve diversity in mature miRNA expression.

Nucleotide Motifs at the Microprocessor Cleavage Site

The possibility to group miRNAs by MPI allows us to define sequence features and motifs in the pre-miRNA flanking regions associated with enhanced processing in an endogenous context. miRNAs were divided into efficiently and nonefficiently processed transcripts for further analysis (see Experimental

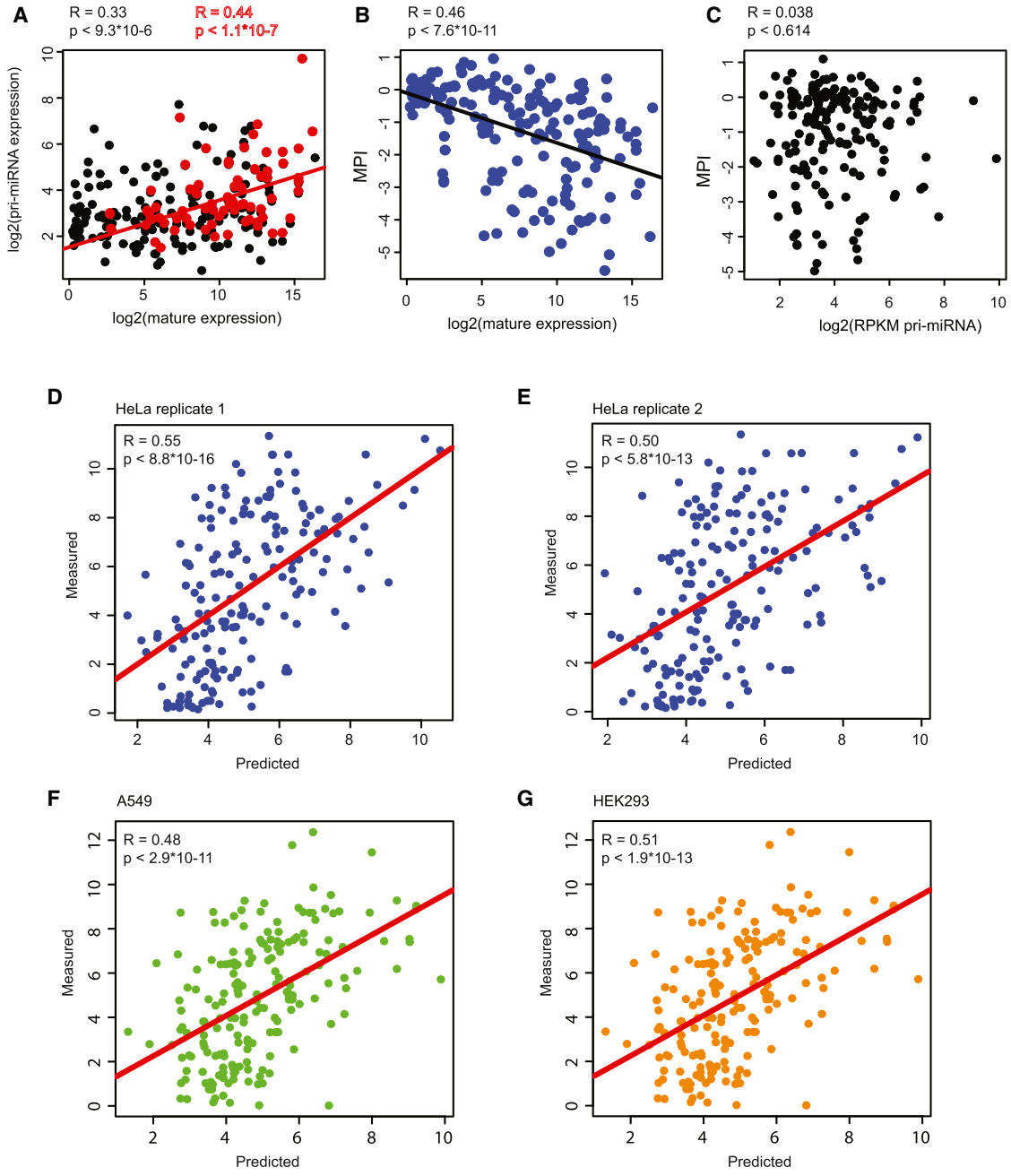


Figure 5. The MPI Is a Predictor of miRNA Expression

(A) Correlation between pri-miRNA expression and mature miRNA abundance for all pri-miRNAs (R = 0.33) or efficiently processed pri-miRNAs (MPI < −1; red circles; R = 0.44).

(B) Correlation between MPI and mature miRNA abundance in HeLa.

(C) MPI values do not correlate with pri-miRNA expression levels in HeLa.

(D–G) Prediction of mature miRNA expression levels from combined pri-miRNA expression and MPI using a linear model in HeLa (D and E), A549 (F), and HEK293 (G) cells. The red lines indicate the linear fit between predicted and measured expression ($y = 0.31x_1 - 0.45x_2 + 1.2 \times 10^{-16}$), where x_1 is the standardized pri-miRNA expression and x_2 is the standardized MPI value. Predicted and observed miRNA expressions are significantly correlated in all three cell lines ($R > 0.48$ $p \leq 10^{-10}$).

See also Figures S4 and S5.

Procedures). We find comparable frequencies of each nucleotide between the two groups; i.e., no general sequence bias in the regions flanking pre-miRNA sequences (data not shown). We used a regularized log-linear model (Supplemental Experimental Procedures) to identify dinucleotide and spaced dinucleotide motifs that are predictive of high miRNA processing efficiency, regardless of the position in the upstream and downstream pre-miRNA flanking regions (Figures 6A, 6B, and S6A). We then examined if the derived sequence motifs exhibit any strong positional enrichment in the upstream or downstream flanking regions, rather than a broad distribution along pri-miRNA sequences. To this end, we plotted the frequencies of significant motifs as function of the sequence position for both upstream and downstream miRNA flanking regions (Figures 6C–6E, S6B, and S6C). Although most of the initial motifs do not show any positional preference, the upstream GNNU motif is highly enriched at a position ending with the first base of the 5′ miRNA (Figure 6C). Considering that the frequently occurring U as the first base of miRNAs is involved in downstream functions (Fellmann et al., 2011), enrichment of the GNNU motif suggests that the G at position −3 can impact on processing efficiency dependent on the sequence context of the flanking stem loop region. Downstream, we identify the CNNC motif between positions 17 and 21 as the most enriched spaced dinucleotide in the efficiently processed pri-miRNAs. These data are in agreement with recent studies that identified the CNNC motif as an enhancer of pri-miRNA processing in an in vitro screen and as target of p72 binding (Auyeung et al., 2013; Mori et al., 2014) and add further experimental support to the biological significance of sequence motifs for in vivo pri-miRNA processing. Also downstream, we identify a three-letter motif, UNC, that is underrepresented in efficiently processed pre-miRNA (Figure S6). Underrepresentation of a motif could indicate an inhibitory effect for the binding of factors that promote processing, or an affinity for factors that inhibit processing, as has been reported (Cheng et al., 2014; Choudhury et al., 2013; Piskounova et al., 2011; Sakamoto et al., 2009).

A corresponding analysis of dinucleotide motifs reveals positive and negative correlations with processing (Figure 6B). Dinucleotides GA and CU are enriched in the upstream flanking and stem region, respectively, of efficiently processed miRNAs, whereas the dinucleotide AA shows a considerable depletion in the stem and downstream flanking regions (Figure 6B). We find that the dinucleotide GC is highly enriched in the 3′ stem region, and base pairs with a GC-motif with increased occurrence at position −13 in the 5′ stem (Figure 6E), coinciding with the base of the pre-miRNA stem loop. Interestingly, the GC motif shows the reversed pattern in nonefficiently processed pri-miRNAs, with depletion at position −13 but increased occurrence within the upstream stem region. Together, this points to an activating effect on processing of the GC dinucleotide motif at and around the base of the stem loop. The previously mentioned in vitro screen proposed the occurrence of a UG motif involved in enhancing processing of pri-miRNAs (Auyeung et al., 2013), which we find in our analysis as well (Figure 6E).

We repeated the motif analysis using a reduced set of pseudohairpins for comparison (see the Supplemental Experimental Procedures). GC dinucleotides and CNNC and GNNU motifs are also enriched in efficiently processed pri-miRNAs over this background model (Figure S7), highlighting the specificity of these motifs for Microprocessor-mediated miRNA processing.

With these data, we demonstrate that endogenous RNA-sequence-derived processing efficiencies faithfully represent endogenous Microprocessor kinetics and can be used to identify functional motifs despite the short length and limited number of input sequences.

An independent qualitative validation of the MPI was done using a reporter assay to test the in vitro processing efficiency of selected pri-miRNA transcripts (Allegra and Mertens, 2011). Here, pri-miRNA sequences are inserted in the 3′ UTR of a Luciferase reporter gene. A reduction in Luciferase activity then serves as a readout for Microprocessor cleavage of the luciferase mRNA. As a proof of concept, we inserted the ∼300 nt sequence surrounding miR-296, a miRNA not expressed in any of the cell lines analyzed but with important roles in stem cells (Tay et al., 2008). Introducing pri-miR-296 into the reporter vector decreased Luciferase activity (increased processing; Figure 7A) and led to an increase in mature miR-296 expression (Figure 7B). This effect was recapitulated with pri-miR-21 and pri-miR-34a and abrogated upon Drosha knockdown (Figure 7C), confirming that the assay recapitulates some aspects of pri-miRNA processing. Testing ten pri-miRNA sequences inserted into the Luciferase vector shows a Pearson correlation of 0.77 between the reporter assay and the in-vivo-determined MPI (Figure 7D), underlining a significant relationship between in vitro processing assays and in vivo processing. Although in vitro reporter assays do not necessarily reflect all the regulatory events taking place in vivo, the high correlation between in vitro measurements and the processing efficiencies determined in vivo provides substantial support for the reported approach and an endogenous function of the identified motifs. However, using these reporter assays, mutations of the GNNU motif, while maintaining base-pairing in the stem, do not change the processing efficiency significantly (data not shown). This possibly reflects differences between in vitro and in vivo processing. To test the effect of the identified basal GC motif on processing efficiency in vitro, we used miR-100. miR-100 is an efficiently processed miRNA that contains a −13 GC sequence. The effect of single mutations at −13G and −12C in the miR-100 pri-miRNA sequence is striking, almost completely disrupting the processing (Figures 7E and 7F). Given that the GC motif is located within the base of the double-stranded stem region, it is important to differentiate between structural and real sequence determinants of Microprocessor activity. To address this, we introduced compensatory mutations at the complementary positions in the 3′ end of the stem loop (Figures 7E and 7F). When altering the GC motif while maintaining base-pairing at the base of the stem, we still observe a significant reduction in processing efficiency, arguing for a sequence specific effect. Nevertheless, mutations at either side of the stem loop that disrupt the paired base of the hairpin are more detrimental to processing, suggesting that the structural properties of pri-miRNA hairpins are more important in specifying processing than the sequence motifs.

Taken together, reporter assays mimicking the pri-miRNA of miR-100 could recapitulate the hypotheses derived from in vivo

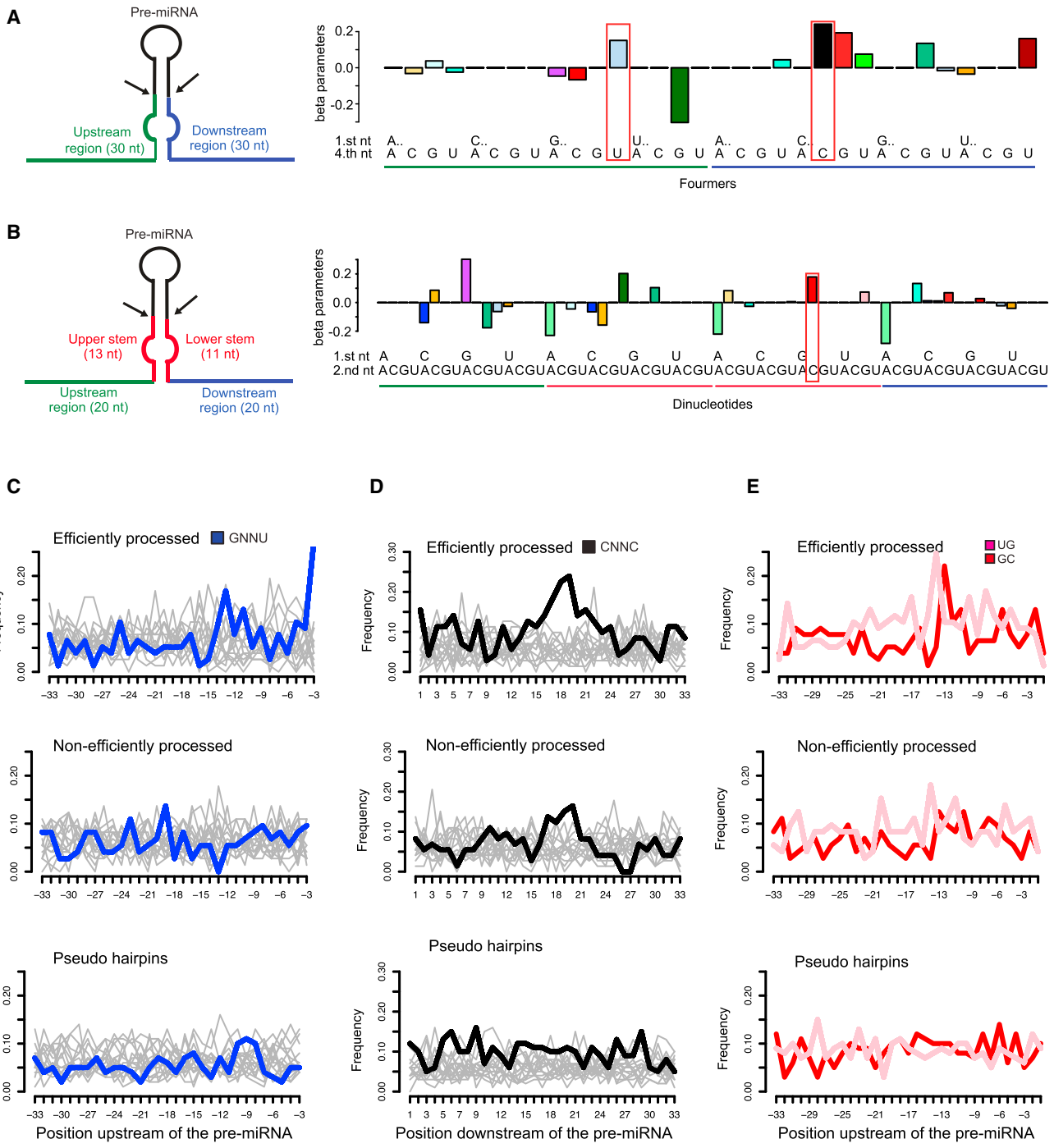


Figure 6. Regulatory Sequence Motifs in the miRNA Flanking Regions

(A and B) Regression coefficients for enriched (positive values) or depleted (negative values) spaced dinucleotides (A) and dinucleotides (B) in the stem and flanking regions of efficiently processed compared to the nonefficiently processed pri-miRNAs. Marked in red are motifs that also show positional enrichment.

(C and D) Frequency profiles (gray lines) of all spaced dinucleotide motifs for efficiently processed miRNAs (top), nonefficiently processed miRNAs (middle), and pseudo hairpins (bottom). The GNNU spaced dinucleotide in the upstream region (C) and the CNNC spaced dinucleotide in the downstream region (D) are colored in light blue and black, respectively.

(E) Distribution of dinucleotide motifs UG and GC in the upstream regions of efficiently processed miRNAs (top), nonefficiently processed miRNAs (middle), and pseudo hairpins (bottom).

See also Figures S6 and S7.

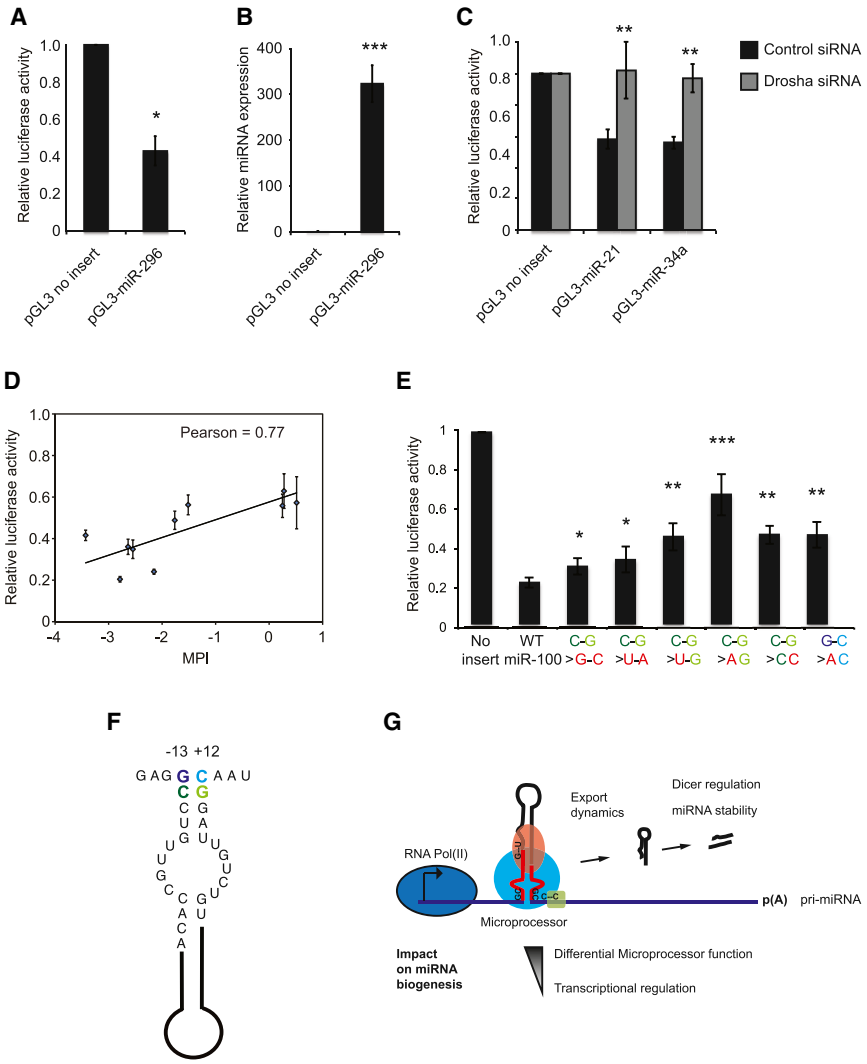


Figure 7. Reporter Validation and Model

Reporter assay validation of pri-miRNA processing.

(A) Proof of concept for Luciferase processing assay using miR-296.

(B) Mature miR-296 expression (normalized to U6) for control and miR-296 vector.

(C) Processing efficiency for miR-21 and miR-34a with and without Drosha knockdown.

(D) Correlation between reporter assay and MPI for let-7a-1, mir-31, mir-125a, mir-100, mir-21, mir-224, let-7i, mir-4427, mir-675, and mir-34a.

(E and F) Analysis of processing efficiency of a reporter vector containing a pri-miR-100 insert with the mutations indicated in (F), showing a schematic of the miR-100 hairpin. y axes in (E) indicate relative processing efficiency.

(G) Model of miRNA biogenesis, with the proposed hierarchy of Microprocessor function and transcriptional regulation. *p < 0.05, **p < 0.01, ***p < 0.001.

Microprocessor signatures and sequence analyses. Although maintaining secondary structure is known to be critical for pri-miRNA processing, the −13 GC motif at the base of the pre-miRNA hairpin appears to support endogenous processing of pri-miRNAs. Together, sequence motifs and secondary structure requirements highlight the importance of the region around position −13 and the base-pairing to the 3′ GC motif. The downstream CNNC motif has been extensively validated and is present in the majority of conserved pri-miRNA transcripts (Auyeung et al., 2013).

DISCUSSION

miRNA Biogenesis In Vivo Is Primarily Regulated by Microprocessor

Here, we report an approach that allows a global and quantitative view on in vivo pri-miRNA processing. In contrast to previous studies that relied on reporter systems and in vitro processing assays, visualization of the Microprocessor cleavage events within chromatin associated pri-miRNA transcripts by next-generation sequencing allows us to simultaneously determine the processing kinetics for hundreds of pri-miRNAs in their endogenous context. We define the MPI as a measure that faithfully reflects endogenous processing efficiencies, promising numerous applications of the described approach to study in vivo how cofactors and stress responses globally affect pri-miRNA processing. The sequencing data that we have generated reveal an unanticipated complexity of Microprocessor function, promising further insight into miRNA biogenesis when data for additional cell lines and tissues have been generated to derive a general picture of pri-miRNA processing.

Although it has been suggested that tissue-specific miRNA expression is controlled at the transcriptional level (Gao et al., 2011), the present study reveals differential Microprocessor activity as a major determinant of steady-state miRNA levels (Figures 5D–5G). The lack of correlation between in vivo processing efficiencies and basal transcription (Figure 5C) demonstrates that both processes are independently controlled steps in the miRNA biogenesis pathway. Several miRNAs have been shown to be extensively regulated at a post-Microprocessor step (e.g., as members of the let-7 family by LIN28A via uridylation and inhibition of Dicer processing) (Heo et al., 2008), and steady-state miRNA levels are further affected by nuclear export kinetics and the stability of the mature miRNA in the cytoplasm. This inherent complexity in the multistep miRNA biogenesis pathway prevents the precise prediction of mature miRNA abundance from a single variant. At the same time, these considerations further highlight the surprisingly good correlation we observe between MPI and mature miRNA levels. This finding fits well with the widespread occurrence of miRNAs encoded within protein coding genes and in polycistronic transcripts.

Here, independent control of the cleavage by Microprocessor allows a partial uncoupling from host gene expression and diversification of miRNA expression from shared primary transcripts. The latter has been demonstrated for mir-18a, which is transcribed from the polycistronic mir-17-92 locus but whose processing is specifically stimulated by hnRNP A1 (Guil and Cáceres, 2007). We find that the uncoupling from basal transcription levels by differential Microprocessor cleavage is a global phenomenon that is a major contributor to the observed diversity in miRNA abundance. The wide range of observed processing efficiencies in vivo puts further emphasis on evidence suggesting active regulation of the initial miRNA processing step in response to various signaling pathways and developmental processes (Davis et al., 2008; Mori et al., 2014; Suzuki et al., 2009), given that it opens ample possibilities to fine-tune individual miRNA expression in response to additional signals and stresses. There is evidence that the Microprocessor step in the miRNA biogenesis pathway is largely affected in physiological processes such as cancer and inflammation (Gregory and Shiekhattar, 2005; Mori et al., 2014). Understanding the dynamics and the factors that determine Microprocessor-mediated processing can therefore help shedding light on the misregulation of miRNAs in disease.

Slow Processing Can Be a Feature of Bona Fide Microprocessor Targets

Much effort has been put into annotating genome-wide miRNAs by predictions and expression analysis using deep sequencing of a large collection of tissues (Friedländer et al., 2014; Landgraf et al., 2007). Although the catalog of annotated miRNAs keeps expanding, a recent study also points toward many nonreal miRNAs being wrongly annotated (Chiang et al., 2010), indicating that the extensive computational effort for genome-wide annotation of miRNAs could inflate the number of real miRNAs. This raises the question of whether the large number of pri-miRNAs we observe with MPIs around 0 could include a substantial number of falsely annotated hairpins. Based on several considerations, we believe that most of the inefficiently processed pri-miRNAs reported in our study resemble bona fide Microprocessor targets that are processed at a very low rate. First, we performed small RNA sequencing (RNA-seq) and restricted our analysis to pri-miRNAs with robust evidence for authentic expressed miRNA hairpins according to criteria in Chiang et al. (2010) as specified in the Results and Experimental Procedures. Second, our motif analysis revealed a distinct set of sequence features that are shared by efficiently and inefficiently processed miRNAs but are not found in pseudohairpins, like the enrichment of GC dinucleotides in the 5′ stem of pre-miRNA (Figures 6 and S7). Furthermore, we find well-characterized miRNAs with slow processing kinetics in our analysis, one example being pri-mir-7-1, which is exclusively processed in neuronal lineages (Choudhury et al., 2013). Interestingly, we observe the lowest MPI value for this pri-miRNA in HEK293 cells, which share some characteristics of neuronal cells.

In most cases, low processing efficiency translates into low abundance of the associated mature miRNA. Nevertheless, we observe some pri-miRNAs with no signs of processing despite high mature miRNA levels in the same cell. This could be due to several reasons. The mature miRNAs could be particularly stable, so that substantial levels are maintained while the turnover is low. Alternatively, processing could in few cases occur after release of the primary transcript from chromatin and escape detection by our assay. Finally, although miRNA processing generally precedes splicing (Kataoka et al., 2009; Kim and Kim, 2007; Morlando et al., 2008), some pre-miRNAs in introns of protein coding genes may be processed after the intron is spliced out and released into the nucleoplasm. Some hint toward one of the latter possibilities may come from the observation that most abundant miRNAs with low signs of processing reside in protein coding host genes, whereas independent pri-miRNAs are depleted within this group. To unravel the processing pathway of these miRNAs will be an interesting focus of future research.

The Hierarchy of Processing Determinants

One of the most significantly enriched sequence motifs that we find in flanking regions of efficiently cleaved pre-miRNAs is the downstream CNNC motif that has been derived by Auyeung et al. (2013) from a large-scale in vitro screen using variants of four model pri-miRNA transcripts and was shown to bind SRP20, underlining the potential importance of this motif for pri-miRNA processing both in vitro and in vivo. Also, the DEAD box helicase p72 has been found to bind a related VCAUCH motif and affect the processing of pri-miRNAs (Mori et al., 2014). We now provide evidence that the CNNC motif generally contributes to processing efficiency in humans. In addition, we identify sequence elements that are associated with efficient processing in vivo, such as the regulatory GC motif. Functional studies on the GC motif emphasize that the structural properties of the pre-miRNA stem loop are essential prerequisites for recognition by Microprocessor, whose activity is further modulated by sequence features.

Overall, the effects of motifs and secondary structure determinants point toward the base of the extended hairpin as a regulatory hub for miRNA processing. The combination of structural features, sequence elements, and Microprocessor cofactors finally results in highly differential processing of pri-miRNA transcripts, where the Microprocessor regulatory step is the predominant determinant of mature miRNA levels (Figure 7G). Recent progress in the analysis of in vivo RNA folding has revealed an extensive plasticity of endogenous RNA secondary structures (Ding et al., 2014; Rouskin et al., 2014; Wan et al., 2014), underlining the important contributions from novel methodologies to the study of in vivo RNA biology. The combination of several in vivo approaches will expand our understanding about how RNA regulates cellular processes and help to reveal the molecular interplay between sequence, structure, and processing.

EXPERIMENTAL PROCEDURES

Tissue Culture and Transfection Conditions

HeLa cells were maintained under standard conditions at 37°C and 5% CO_2 in Dulbecco's modified Eagle's medium containing 10% FCS and antibiotics. For RNAi, 5×10^{-5} cells were seeded per 10 cm dish, transfected after 24 hr with 60 nM siRNA, harvested by trypsinization 72 hr after transfection, and fractionated as described below.

Cellular Fractionation and Chromatin RNA Isolation

Cellular fractionation was done as described in Bhatt et al. (2012) with minor modifications as described in the Supplemental Information.

Quantitative Real-Time PCR

RNA (500 ng) was reverse transcribed using the High-Capacity RNA-to-cDNA Kit (Invitrogen 4387406). cDNA was quantified on a 7900HT Fast Real-Time PCR system (Applied Biosystems) using the SYBR Green PCR Master Mix (Invitrogen 4364344).

RNA Sequencing

Library preparation was performed using the TruSeq Stranded Total RNA Kit (Illumina) for chromatin associated RNA and the True Seq Small RNA Kit (Illumina) for mature miRNAs. Sequencing was performed on an Illumina HiSeq 2500 instrument.

Analysis of Sequencing Data

Alignments to the hg19 assembly of the human genome were performed with TopHat 2 (Trapnell et al., 2009) with default options. We allowed the reads to map up to five multiple positions in the genome to account for miRNA isoforms and allowed up to one mismatch per sequence in the alignments.

The expression of mature miRNAs was determined from the small RNA-seq data by means of the mirDeep2 software (Friedländer et al., 2012). The reads were mapped to the hg19 assembly of the human genome by using the short-read alignment software Bowtie (Hatem et al., 2013), and reads that mapped perfectly to nucleotide position 1 to 18 of each annotated miRNA were used to quantify the miRNA expression. The sum of raw read counts from the 5p and 3p arms for each miRNA was converted to RPM values and the log2 of the average RPM from the two replicates was used as miRNA expression level. For conserved miRNAs, we required the hairpin structure to be conserved in the orthologous members of the gene family (as defined in mirBase) including mouse or other mammals and the seed to be conserved in more than 50% of the orthologous genes according to ClustalW alignments. The final list of the miRNAs used for the analysis of this paper includes 229 miRNAs, 138 classified as broadly conserved, 52 classified as weakly conserved, and 39 as nonconserved.

Collection of Pseudo Hairpins

A publicly available data set of pseudo hairpins (Xue et al., 2005) was downloaded at http://bioinfo.au.tsinghua.edu.cn/software/mirnasvm/Triplet-svm-predictor.htm and filtered as described in the Supplemental Experimental Procedures.

Determining the Drosha Processing Sites and MPI

To determine the exact site of the 5′ and 3′ Microprocessor cleavage, we used the annotation of the 5p and 3p miRNA strands from miRBase and mapped them onto the sequence of the pre-miRNA. In order to assess the significance of the Microprocessor signature, we modeled the read-count distribution along the genome with a Poisson distribution. Instead of using a uniform λ parameter estimated from the background read density across the whole genome, we used a dynamic parameter λlocal for each precursor, estimated as $\lambda\text{local} = \text{mean}(\lambda 1, \lambda 2)$. $\lambda 1$ and $\lambda 2$ were estimated from the 100 bp genomic regions upstream and downstream of the estimated processing sites, respectively.

Prediction of miRNA Expression with Linear Regression

The MPI and pri-miRNA expression levels were used as predictors for training a linear regression model to predict the logarithm of expression of the mature miRNA. The Pearson correlation between predicted and measured values was calculated in a 5-fold cross-validation setting and used as a measure of predicted accuracy.

Motif Enrichment

We divided expressed pre-miRNAs in two classes: efficiently processed and nonefficiently processed according to their MPI value and predicted the sequence features that are most likely to be associated with efficient processing using regularized logistic regression (see the Supplemental Experimental Procedures).

ACCESSION NUMBERS

Sequencing data were deposited to the NCBI Gene Expression Omnibus under accession number GSE56862.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.09.007.

AUTHOR CONTRIBUTIONS

T.C. performed all experimental work with the exception of validation assays in Figure 7, which were performed by M.G. A.M. performed all bioinformatics analyses. U.A.Ø. supervised research. T.C., A.M., M.G., and U.A.Ø. conceived the experiments, interpreted the data, and wrote the manuscript.

ACKNOWLEDGMENTS

We thank Bernd Timmermann and the MPI core sequencing facility for expert advice and help with sequencing and Stefan Haas for helpful discussion on the computational analysis. This work has been supported by DFG Grant SFB-TR84 (A.M.). The laboratory of U.A.Ø. is funded by the Federal Ministry for Education and Research in Germany through the Alexander von Humboldt Foundation Sofja Kovalevskaja Award (U.A.Ø.).

Received: April 11, 2014
Revised: July 2, 2014
Accepted: September 6, 2014
Published: October 9, 2014

REFERENCES

Allegra, D., and Mertens, D. (2011). In-vivo quantification of primary microRNA processing by Drosha with a luciferase based system. Biochem. Biophys. Res. Commun. *406*, 501–505.

Auyeung, V.C., Ulitsky, I., McGeary, S.E., and Bartel, D.P. (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. Cell *152*, 844–858.

Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell *150*, 279–290.

Chaulk, S.G., Thede, G.L., Kent, O.A., Xu, Z., Gesner, E.M., Veldhoen, R.A., Khanna, S.K., Goping, I.S., MacMillan, A.M., Mendell, J.T., et al. (2011). Role of pri-miRNA tertiary structure in miR-17∼92 miRNA biogenesis. RNA Biol. *8*, 1105–1114.

Cheng, T.L., Wang, Z., Liao, Q., Zhu, Y., Zhou, W.H., Xu, W., and Qiu, Z. (2014). MeCP2 suppresses nuclear microRNA processing and dendritic growth by regulating the DGCR8/Drosha complex. Dev. Cell *28*, 547–560.

Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E., et al. (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. *24*, 992–1009.

Choudhury, N.R., de Lima Alves, F., de Andrés-Aguayo, L., Graf, T., Cáceres, J.F., Rappsilber, J., and Michlewski, G. (2013). Tissue-specific control of brain-enriched miR-7 biogenesis. Genes Dev. *27*, 24–38.

Davis, B.N., Hilyard, A.C., Lagna, G., and Hata, A. (2008). SMAD proteins control DROSHA-mediated microRNA maturation. Nature *454*, 56–61.

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

Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature *505*, 696–700.

Fellmann, C., Zuber, J., McJunkin, K., Chang, K., Malone, C.D., Dickins, R.A., Xu, Q., Hengartner, M.O., Elledge, S.J., Hannon, G.J., and Lowe, S.W. (2011). Functional identification of optimized RNAi triggers using a massively parallel sensor assay. Mol. Cell *41*, 733–746.

Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. *40*, 37–52.

Friedländer, M.R., Lizano, E., Houben, A.J., Bezdan, D., Báñez-Coronel, M., Kudla, G., Mateu-Huertas, E., Kagerbauer, B., González, J., Chen, K.C., et al. (2014). Evidence for the biogenesis of more than 1,000 novel human microRNAs. Genome Biol. *15*, R57.

Fukuda, T., Yamagata, K., Fujiyama, S., Matsumoto, T., Koshida, I., Yoshimura, K., Mihara, M., Naitou, M., Endoh, H., Nakamura, T., et al. (2007). DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. Nat. Cell Biol. *9*, 604–611.

Gao, Y., Schug, J., McKenna, L.B., Le Lay, J., Kaestner, K.H., and Greenbaum, L.E. (2011). Tissue-specific regulation of mouse microRNA genes in endoderm-derived tissues. Nucleic Acids Res. *39*, 454–463.

Gregory, R.I., and Shiekhattar, R. (2005). MicroRNA biogenesis and cancer. Cancer Res. *65*, 3509–3512.

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

Guil, S., and Cáceres, J.F. (2007). The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. Nat. Struct. Mol. Biol. *14*, 591–596.

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

Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887–901.

Hatem, A., Bozdağ, D., Toland, A.E., and Çatalyürek, U.V. (2013). Benchmarking short sequence mapping tools. BMC Bioinformatics *14*, 184.

Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V.N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. Mol. Cell *32*, 276–284.

Kawai, S., and Amano, A. (2012). BRCA1 regulates microRNA biogenesis via the DROSHA microprocessor complex. J. Cell Biol. *197*, 201–208.

Kim, Y.K., and Kim, V.N. (2007). Processing of intronic microRNAs. EMBO J. *26*, 775–783.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol. *10*, 126–139.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M., et al. (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. Cell *129*, 1401–1414.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787–798.

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.

Miranda, K.C., Huynh, T., Tay, Y., Ang, Y.S., Tam, W.L., Thomson, A.M., Lim, B., and Rigoutsos, I. (2006). A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. Cell *126*, 1203–1217.

Mori, M., Triboulet, R., Mohseni, M., Schlegelmilch, K., Shrestha, K., Camargo, F.D., and Gregory, R.I. (2014). Hippo signaling regulates microprocessor and links cell-density-dependent miRNA biogenesis to cancer. Cell *156*, 893–906.

Morlando, M., Ballarino, M., Gromak, N., Pagano, F., Bozzoni, I., and Proudfoot, N.J. (2008). Primary microRNA transcripts are processed co-transcriptionally. Nat. Struct. Mol. Biol. *15*, 902–909.

Morlando, M., Dini Modigliani, S., Torrelli, G., Rosa, A., Di Carlo, V., Caffarelli, E., and Bozzoni, I. (2012). FUS stimulates microRNA biogenesis by facilitating co-transcriptional Drosha recruitment. EMBO J. *31*, 4502–4510.

Obad, S., dos Santos, C.O., Petri, A., Heidenblad, M., Broom, O., Ruse, C., Fu, C., Lindow, M., Stenvang, J., Straarup, E.M., et al. (2011). Silencing of microRNA families by seed-targeting tiny LNAs. Nat. Genet. *43*, 371–378.

Pawlicki, J.M., and Steitz, J.A. (2008). Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production. J. Cell Biol. *182*, 61–76.

Piskounova, E., Polytarchou, C., Thornton, J.E., LaPierre, R.J., Pothoulakis, C., Hagan, J.P., Iliopoulos, D., and Gregory, R.I. (2011). Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. Cell *147*, 1066–1079.

Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature *505*, 701–705.

Sakamoto, S., Aoki, K., Higuchi, T., Todaka, H., Morisawa, K., Tamaki, N., Hatano, E., Fukushima, A., Taniguchi, T., and Agata, Y. (2009). The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway. Mol. Cell. Biol. *29*, 3754–3769.

Siomi, H., and Siomi, M.C. (2010). Posttranscriptional regulation of microRNA biogenesis in animals. Mol. Cell *38*, 323–332.

Sood, P., Krek, A., Zavolan, M., Macino, G., and Rajewsky, N. (2006). Cell-type-specific signatures of microRNAs on target mRNA expression. Proc. Natl. Acad. Sci. USA *103*, 2746–2751.

Suzuki, H.I., Yamagata, K., Sugimoto, K., Iwamoto, T., Kato, S., and Miyazono, K. (2009). Modulation of microRNA processing by p53. Nature *460*, 529–533.

Tay, Y., Zhang, J., Thomson, A.M., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature *455*, 1124–1128.

Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., and Guigó, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. Genome Res. *22*, 1616–1625.

Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A.D., Filipowicz, W., Ramos, A., Gherzi, R., and Rosenfeld, M.G. (2009). The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. Nature *459*, 1010–1014.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E., and Chang, H.Y. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. Nature *505*, 706–709.

Wang, Y., Vogel, G., Yu, Z., and Richard, S. (2013). The QKI-5 and QKI-6 RNA binding proteins regulate the expression of microRNA 7 in glial cells. Mol. Cell. Biol. *33*, 1233–1243.

Xue, C., Li, F., He, T., Liu, G.P., Li, Y., and Zhang, X. (2005). Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. BMC Bioinformatics *6*, 310.

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

Kataoka, N., Fujita, M., and Ohno, M. (2009). Functional association of the Microprocessor complex with the spliceosome. Mol. Cell. Biol. *29*, 3243–3254.