Dicer-Processed Small RNAs: Rules and Exceptions

David Langenberger\textsuperscript{a,b}, M. Volkan Çakir\textsuperscript{b}, Steve Hoffmann\textsuperscript{a,b}, Peter F. Stadler\textsuperscript{a,b,c,d,e,f,g,h,i}

\textsuperscript{a}LIFE, Leipzig Research Center for Civilization Diseases, University Leipzig, Philipp-Rosenthal-Strasse 27, D-04107 Leipzig, Germany
\textsuperscript{b}Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107, Leipzig, Germany
\textsuperscript{c}Bioinformatics Group, Department of Computer Science, Härtelstraße 16-18, D-04107, Leipzig, Germany
\textsuperscript{d}studla@bioinf.uni-leipzig.de
\textsuperscript{e}Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany
\textsuperscript{f}RNomics Group, Fraunhofer Institut für Zelltherapie und Immunologie, Deutscher Platz 5e, D-04103 Leipzig, Germany
\textsuperscript{g}Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria
\textsuperscript{h}Center for non-coding RNA in Technology and Health, University of Copenhagen, Grønnegårdsvej 3, DK-1870 Frederiksberg, Denmark
\textsuperscript{i}Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA

Abstract

Canonical microRNAs are excised from their hairpin-shaped precursors by Dicer. In order to find possible exceptions to this rule and to identify additional substrates for Dicer processing we re-evaluate the small RNA sequencing data of the Dicer knockdown experiment in MCF-7 cells originally published by Friedländer et al. [Nucleic Acids Res. 40:37-52 (2012)]. While the well-known non-Dicer mir-451 is not sufficiently expressed in these experiments, there are several additional Dicer-independent microRNAs, among them the important tumor suppressor mir-663a. We recover previously described examples of non-miRNA Dicer substrates such as tRNA-Gln and several snoRNAs. Interestingly, sdRNAs derived from box C/D snoRNAs are Dicer-independent, while those derived from box H/ACA snoRNAs are often Dicer dependent. Several pol-III transcripts, in particular the vault RNAs and the great ape specific snaRs are processed by Dicer, while the small RNAs originating from Y RNAs seem to be Dicer independent.

Keywords: sdRNAs, snoRNAs, Y RNAs, vault RNAs, snaRs

Preprint submitted to Manuscript December 16, 2013
1. Introduction

Canonical microRNAs are processed from a primary pol II transcript by means of the Drosha-dependent microprocessor complex (Gregory et al., 2004), resulting in a characteristic hairpin of length 60–120 nucleotides. This pre-microRNA is then transported by Exportin-5 to the cytoplasm (Lund et al., 2004), where the hairpin is cut by Dicer into a double stranded RNA about 22nt in length with a 2nt 3’-overhang (Murchison and Hannon, 2004). Several alternative pathways that bypass Drosha have been reported. The most prominent example are mirtrons (Okamura et al., 2007; Ruby et al., 2007), whose precursor hairpins are produced by splicing. A related, mirtron-like source of small RNAs requires both splicing and exosome-mediated trimming to extract the pre-microRNA hairpin (Flynt et al., 2010; Chong et al., 2010). More recently, it was shown that a few microRNAs, in particular mir-451, are matured without the help of Dicer (Cheloufi et al., 2010; Cifuentes et al., 2010). For a recent review of the many alternative pathways for the biogenesis of microRNAs and other, microRNA-like small RNA species see e.g. (Yang and Lai, 2011).

Apparently, Dicer is not only involved in microRNA biogenesis, but appears to be involved also in the processing of other small RNA species. Short, microRNA-like small RNA species see e.g. (Yang and Lai, 2011).

2. Methods

2.1. Data and Mapping

We downloaded a previously published sequencing data set series (GSE31069, (Friedländer et al., 2012)) from the Gene Expression Omnibus (GEO) database (Edgar et al., 2002). The data consists of four different samples, two containing short reads from the total cell content and two containing reads from the cytoplasmic fraction only. Both pairs contrast small RNA expression before and after Dicer knock-down in a MCF-7 cell line. The analysis reported here uses only the cytoplasmic sample pair (GSM769509 and GSM769511). Since short RNA processing takes place in this compartment we expect to reduce the noise from the nucleus.

All the adapter-free reads were mapped against the human genome (NCBI36.50 Release of July 2008) using āgēmēl (Hoffmann et al., 2009): we activated the poly-A clipping, required small RNAs to map with an accuracy of at least 90% and selected the “best scoring hit strategy”. With these settings we mapped 8,743,377 of 15,493,265 reads (56%) of the control sample and 5,471,242 of 9,237,490 reads (59%) of the Dicer knockdown sample. The resulting sam files were converted to bam format, using samtools (Li et al., 2009) and subsequently translated to bigWig files using a custom perl script. The read density at each position in the bigWig files was normalized by the number of multiple hits of each read and the absolute number of mapped reads of each experiment (RPM) in order to make the two experiments comparable. We provide custom tracks for the UCSC Genome Browser (Kent et al., 2002) to make the mapping results publicly available.

2.2. Expressed Sites and Annotation

In order to identify previously un-annotated loci with small RNA expression we created sorted bed files and then used blockbuster (Langenberger et al., 2009) with default parameters to identify regions showing accumulations of at least 50 reads in at least one of control or Dicer knock-down data. We used āgēmēl from Bedtools (Quinlan and Hall, 2010) to obtain the final
list of expressed regions of interest (1,946 for control and 1,798 for the knock-down set), which we call “sites” from now on.

We downloaded the latest annotations from different sources (1523 microRNA loci from miRBase v18 (Griffiths-Jones, 2004); 631 tRNA loci from gtRNAdb (Chan and Lowe, 2009); 402 snoRNA loci as well as 4528 other RNAs from UCSC annotation (Karolchik et al., 2004)). This combined annotation track comprising 7,084 annotated ncRNA loci was compared with our list of sites using intersectBed (Quinlan and Hall, 2010).

Furthermore, all reads were overlapped with the UCSC repeat masker track (Jurka et al., 2000) and as soon as one read was mapped to a repeat associated region, all multiple hits of it were flagged with the type of repeat. If more than 50% of the expression of one site is caused by reads which are flagged as repeat associated, the whole site was flagged accordingly. In order to remove low-complexity sequences, which have a high probability of being random matches in short read data, we discarded all sites with a Shannon entropy of less than 1.6 bit.

2.3. Expression Levels

The expression level of each site, expressed in reads per kilobase of locus per million mapped reads (RPKM) was computed using the UCSC tool bigWigAverageOverBed (Kent et al., 2002). From these values we derived, for each site, the log₂-fold change $\lambda$ between the Dicer knock-down sample and the RPKM of the control sample. All sites with $\lambda < 0$ are interpreted as Dicer processed. All sites, together with their annotations, their expression values, their $\lambda$ and a link to the UCSC Genome Browser can be found at the supplement page http://www.bioinf.uni-leipzig.de/supplements/12-005.

2.4. Processing Pattern

Cleavage of a nearly double-stranded RNA by Dicer leads to a characteristic 2 nt overhang at the 3’end, see e.g. (Ji, 2008). In order to assess how important the thermodynamic stability of the precursor structure is for processing, we computed for a pair of putative single-stranded cleavage products, the following stability measure: RNAcofold (Bernhart et al., 2006) is used to compute the energy of the duplex with the constraint that the joint structure exhibits the 2nt overhang at the 3’ends. Then the inner part of both sequences was shuffled 100 times so that the dinucleotide composition is preserved, while the terminal base pairs and overhanging nucleotides were left untouched. The resulting z-score of the co-folding energies is recorded. For each site we considered the two consecutive tags with the largest expression as candidates for Dicer processing.

In order to assess the overall similarity of a site with canonical microRNAs we use RNAmicro (Hertel and Stadler, 2006). This tool evaluates structural features as well as the pattern sequence conservation. We retrieved alignments of all sites with 20nt flanking sequence on both sides from the 8way-multiZ alignment (human, chimp, orangutan, rhesus macaque, marmoset, mouse, opossum, platypus) (Blankenberg et al., 2011). We extracted sequences from 8way-multiZ file, re-alignes them using clustalw (Larkin et al., 2007) and used it to run RNAmicro. Then, the RNAmicro decision value (decV) was used to rate the sites, if they microRNA-like structures and conservations.

Dicer is well known to generate products in the narrow length range 21–28 nt, see e.g. (Starega-Roslan et al., 2011). We therefore recorded the distribution of read lengths for each locus. In addition, we determined the lengths of blocks of reads blockbuster (Langenberger et al., 2009) with default parameters. Read blocks summarize groups of reads that overlap nearly perfectly, hence its lengths is typically larger than that of individual reads.

3. Results

3.1. Identification of Dicer-dependent small RNAs

The Dicer knock-down (GSM769509) and control (GSM769511) datasets (Friedländer et al., 2012) together identify 2,115 expressed sites. Of these, 1,048 overlap with the 7,084 annotated ncRNAs and 1,067 remain unannotated. After filtering out the low-complexity sites, we retain 1,002 annotated and 539 unknown sites for further analysis.

Table 1: Fraction of Dicer processed sites among the annotated ncRNAs.

<table>
<thead>
<tr>
<th>type</th>
<th>Dicer processed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes</td>
</tr>
<tr>
<td>miRNA</td>
<td>255</td>
</tr>
<tr>
<td>tRNA</td>
<td>32</td>
</tr>
<tr>
<td>H/ACA snoRNA</td>
<td>8</td>
</tr>
<tr>
<td>C/D snoRNA</td>
<td>0</td>
</tr>
<tr>
<td>misc RNA</td>
<td>3</td>
</tr>
<tr>
<td>snRNA</td>
<td>2</td>
</tr>
<tr>
<td>scRNA</td>
<td>10</td>
</tr>
<tr>
<td>rRNA</td>
<td>28</td>
</tr>
</tbody>
</table>
Fig. 1 summarizes the response of the small RNA sites to Dicer knockdown. The log₂-fold change λ exhibits the expected bi-modal distribution separating in particular microRNAs from other small RNA products. Consistent with the original analysis of these datasets (Friedländer et al., 2012), microRNAs are strongly reduced upon reduction of Dicer activity. A closer inspection, however, shows a more differentiated picture.

On the one hand, a small subgroup of microRNAs does not respond to the knockdown of Dicer. On the other hand, a sizable number of unannotated sites (some of which might constitute previously undescribed microRNAs) are associated with well-known structured RNAs exhibiting large negative values of λ, see Table 1.

A substantial fraction of sites expressing small RNAs are annotated repetitive elements, Table 2. Disregarding a moderate number of simple repeats and low complexity regions, which cannot be unambiguously distinguished
Table 2: Fraction of Dicer processed sites among the NUMTs and repeat associated regions.

<table>
<thead>
<tr>
<th>type</th>
<th>Dicer processed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes</td>
</tr>
<tr>
<td>NUMT / chrM</td>
<td>1</td>
</tr>
<tr>
<td>SINE</td>
<td>126</td>
</tr>
<tr>
<td>LINE</td>
<td>141</td>
</tr>
<tr>
<td>LTR</td>
<td>81</td>
</tr>
<tr>
<td>DNA</td>
<td>27</td>
</tr>
<tr>
<td>Simple repeat</td>
<td>18</td>
</tr>
<tr>
<td>Low complexity</td>
<td>15</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
</tr>
<tr>
<td>Satellite</td>
<td>1</td>
</tr>
<tr>
<td>RNA</td>
<td>1</td>
</tr>
<tr>
<td>tRNA</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 2: Correlation of Dicer-response $\lambda$ with sequence-derived descriptors on the entire data set: (a) Free energy $z$-scores of constrained duplex structures with the 2-nt overhangs conforming to the canonical Dicer processing pattern. (b) Decision value of RNAmicro, a SVM-based machine learning tool trained to recognized canonical microRNAs. Its decision value combines the stability of the hairpin structure with patterns of sequence conservation but is agnostic about the location of the small RNA products. Only sites that form a hairpin structure can be scored by this method. (c) Reads with arising from Dicer processed regions, i.e. those with small values of $\lambda$, have a length within the size range between 20 and 25nt typical for microRNAs. The length distribution of reads without evidence for Dicer-processing is much broader and most reads fall outside the expected size range. (d) The reads originating from Dicer-responsive RNA form shorter, more coherent read blocks.

3.2. Characterization of Dicer-processed sites

Dicer-processed small RNAs typically derive from helical regions that are significantly more stable than the precursor secondary structures of Dicer-independent small RNAs. Fig. 2a shows that in particular putative precursor structures that give rise to the typical processing patterns with 2nt overhangs are substantially stabilized Dicer-responsive small RNAs.

Canonical microRNAs also exhibit a characteristic pattern of sequence conservation that can help to distinguish them from other, similar, sources of small RNAs and from hairpin-like structures that are not processed into small RNAs, see e.g. (Lai et al., 2003). RNAmicro (Hertel and Stadler, 2006) implements such a classifier based on a Support Vector Machine taking only a small number of structural and conservation based descriptors as input. Only sites that form hairpin structures can be scored by RNAmicro's SVM. We have previously used RNAmicro to distinguish microRNA-like from snoRNA-like small RNA sites (Langenberger et al., 2011). Fig. 2b shows that the RNAmicro decision value is also correlated with $\lambda$. With few exceptions, large decision values are limited to Dicer responsive sites.

Fig 2c summarizes the distribution of read lengths. As expected, nearly all reads arising from sites with $\lambda < -0.2$ have a lengths between 20 and 25nt, consistent with Dicer processing (Starega-Roslan et al., 2011). In contrast, short reads from sites with $\lambda > 0.2$, i.e., those that are clearly not resulting from Dicer cleavage, are typically longer and show a flat distribution. We also observe a difference in the length of read blocks as determined by blockbuster (Langenberger et al., 2009). Sites with $\lambda < 0$ have on average much shorter block sizes, often consisting only of a single block of microRNAs, Fig. 2d. Since the start and end position of mature microRNAs can vary by a couple of nucleotides (Ebbhardt et al., 2010) such that overlapping microRNAs read blocks have a length of around 30 nt.
3.3. Structured regions processed by Dicer indicate potential microRNA candidates

The data set used here has been generated specifically for the purpose of detecting novel microRNAs (Friedländer et al., 2012). Among the un-annotated, non-repetitive sites with $\lambda < 0$ six additional structured regions (Figure 3) were found. One of them (Figure 3a) is located in an intergenic region Figure 3a, is located in intergenic region far away from any annotation. Three of four intronic sequences (Figure 3d-f) fold into hairpins and the short reads map to the stem positions expected for mirtrons (Okamura et al., 2007; Ruby et al., 2007). The 3’ end of a candidate located in a SYT12-intron is determined by the splice acceptor (Figure 3b). Given its stable hairpin structure these findings suggest that it belongs to the recently described class of “semi-mirtrons” that require both splicing and exosome-mediated trimming for maturation (Flynt et al., 2010; Chong et al., 2010). The structure and the positions of mapped reads of the remaining candidate (Figure 3c) do not conform to a typical microRNA. Nevertheless, the ten-fold reduction of the read coverage in the Dicer knockdown experiment indicates Dicer-processing. Since these reads perfectly but not uniquely map to the intronic region, this candidate is of particular interest for further analysis.

3.4. Dicer-processed non-microRNAs

Surprisingly, there is a large number of well-known structured non-coding RNAs from which Dicer-sensitive small RNAs are produced.

A prominent example are the vault RNAs. The largest response is observed for vtRNA2-1 with $\lambda = -2.12$. This locus was originally classified as hsa-mir-886 but later recognized as a polymerase-III transcript (Canella et al., 2010) and vault RNA paralog (Nandy et al., 2009; Stadler et al., 2009). The other three vault RNA loci also give rise to short RNAs (Persson et al., 2009) and respond negatively to the Dicer depletion: $\lambda(vtRNA1-1) = -0.14$, $\lambda(vtRNA1-2) = -0.76$. The vtRNA1-3 locus is not sufficiently expressed.

The snaR ncRNAs (Parrott and Mathews, 2007) are pol-III transcripts that emerged in the ancestor of the African Great Apes from an Alu-derived precursor (Raha et al., 2010; Parrott et al., 2011). Fig. 4 shows that microRNA-like small RNAs are processed from the lower end of the stem-loop structure, which resembles a canonical pre-microRNA hairpin except for its length of more than 100nt. The snaR-derived small RNAs show the typical 2 nt 3’ overhangs. Their expression depends very strongly on the Dicer concentration.

The situation is more complex for tRNAs and snoRNAs. While many of them give rise to small RNA products, the majority is not influenced by the Dicer knockdown. A small subset of tRNAs, on the other hand is clearly subject to Dicer processing. These include in particular tRNA-Gln-CTG with $\lambda = -2.05$ as noted already previously by Cole et al. (2009). Other tRNAs with a clear Dicer signature are tRNA-Asn-GTT ($\lambda = -1.47$), tRNA-Asn-ATT ($\lambda = -0.83$), tRNA-Ala-CGC ($\lambda = -1.28$), tRNA-Ile-TAT ($\lambda = -1.19$), tRNA-Glu-TTC ($\lambda = -0.79$). None of the four mirbase “microRNAs” that are derived from tRNAs (mir-1274/tRNA-Lys, mir-1280/tRNA-Leu, mir-720/tRNA-Thr, mir-1308/tRNA-Gly) are expressed at sufficiently high levels to estimate $\lambda$.

Small nucleolar RNAs can share several characteristics with microRNAs, including similar components in their processing, see (Scott and Ono, 2011) for a recent review. The structural similarities between H/ACA snoRNAs and microRNAs are most obvious and have been noticed in several computational studies. Scott et al. (2009), for instance, report twenty miRNA precursors that show significant similarity to H/ACA snoRNAs; of these miR-151, miR-605, miR-664 = SNORA36B, miR-215, and miR-140 even bind to dyskerin, a component of the H/ACA snoRNP. On the other hand, Dicer processing has been demonstrated previously for SNORA45 (Ende et al., 2008). Consistently, we find $\lambda$(SNORA45) = $-1.55$. Of the 12 H/ACA snoRNAs with sufficient expression 8 have $\lambda < 0$ (Table 3), indicating that short reads from H/ACA snoRNAs are typically a product of Dicer processing. Interestingly, two H/ACA snoRNAs were classified as novel microRNAs by mirdeep2 (Friedländer et al., 2012): SNORA36A ($\lambda = -1.33$) and SNORA33 $\lambda = 0.15$. We emphasize, however, that only a small minority of H/ACA snoRNAs leads to abundant processing products. In addition these small RNAs are independent of Drosha (Ende et al., 2008; Taft et al., 2009; Braumeier et al., 2011), and in some cases Drosha even inhibits sdRNA formation (Taft et al., 2009), emphasizing that the snoRNAs and (canonical) microRNAs are in general clearly distinguished entities.

A quite different picture emerges for box C/D snoRNAs. Although small RNAs are abundantly produced from box C/D snoRNAs in our data set, Tab. 1, there is no indication that any of them is a Dicer substrate. The box C/D snoRNAs that are discussed as possibly microRNA-like in (Langenberger et al., 2011) show only marginal expression levels and no indication for Dicer processing. On the other hand, of the five microRNAs that resemble box C/D snoRNAs (having C and D boxes
in close proximity in the precursor and binding to fibrillarin) (Ono et al., 2011), four are Dicer substrates (miR-27b $\lambda = -0.90$, miR-16-1 $\lambda = -0.40$, mir-28 $\lambda = -0.95$, and let-7g $\lambda = -1.16$) and the fifth (mir-31) is not sufficiently expressed in MCF-7 cells. It appears, thus, that Dicer-processing clearly distinguished between bona fide microRNAs and small RNAs derived from box C/D snoRNAs.

Y RNAs are small pol-III transcripts that originate from RNA component of the Ro RNP particle and have a role in DNA replication (Christov et al., 2006). The four paralogous human Y RNAs form a cluster on Chr.7(148M) (Mosig et al., 2007; Perreault et al., 2007). The canonical loci show no evidence of Dicer processing hY3 $\lambda = 0.12$, hY4 $\lambda = 1.25$, hY1 $\lambda = 1.70$, hY5 $\lambda = 1.74$. We note that fragments from hY5 have also been annotated as mir-1975.

In addition to the canonical Y RNA cluster, however, there are more than a thousand Y RNA pseudogenes scattered across the genome (Perreault et al., 2007). The deep sequencing data shows that several of these loci form a source of short reads. A few of the Y4-derived loci sites have negative values of $\lambda$. We note, however, these have relatively low expression levels and might be confounded by mapping artefacts. In total, 11 sites that are derived from Y RNA sequences are classified as microRNAs by RNAmicro, six of which have moderate negative values of $\lambda$.

### 3.5. MicroRNA not processed by Dicer

The best-studied microRNA that is not processed by Dicer is mir-451. Unfortunately this site is not significantly expressed in MCF-7 cells, so that we cannot use it as a control. There are ten additional microRNAs with $\lambda > 0$. Six of them (mir-30a, mir-143, mir-374a, mir-379, mir-381, and mir-134) derive from precursor hairpins that are recognized by RNAmicro. Two of these, mir-30a and mir-374a, exhibit exceptionally high levels of expression and feature short RNAs derived from both sides of the precursor stem, Figure 7. We suspect that they are exceptionally good substrates for Dicer so that their maturation is least affected by Dicer concentrations. The evolutionarily ancient mir-125b-2 also exhibits both a canonical read pattern and a canonical pattern of sequence conservation. Nevertheless, it shows no reaction to Dicer knockdown, $\lambda = 0.03$.

For mir-143, mir-381, mir-134, mir-4417, and mir-4516 no mir$^*$ reads were detectable. Mir-4417 is present in monkeys only (Supplemental Material), and no homologs are detectable for hsa-mir-4516, precluding the analysis of patterns of sequence conservation for these two microRNAs.

The entire precursor hairpin of mir-3676 is covered by small RNA sequences. A closer inspection shows, however, that mir-3676 coincides with tRNAThr-AGT and is thus clearly an erroneous annotation. The mis-annotated

---

**Table 3: Box H/ACA snoRNAs processed by Dicer.** SNORA36B (Ender et al., 2008) (also annotated as mir-664) does not reach a sufficient expression level in MCF-7 cells.

<table>
<thead>
<tr>
<th>snoRNA</th>
<th>$\lambda$</th>
<th>snoRNA</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNORA45</td>
<td>-1.55</td>
<td>SNORA36B</td>
<td>-1.00</td>
</tr>
<tr>
<td>SNORA51</td>
<td>-2.42</td>
<td>SNORA46</td>
<td>-1.00</td>
</tr>
<tr>
<td>SNORA36A</td>
<td>-1.33</td>
<td>SNORA56</td>
<td>-0.93</td>
</tr>
<tr>
<td>SNORA17</td>
<td>-1.19</td>
<td>SNORA7B</td>
<td>-0.66</td>
</tr>
<tr>
<td>SCARNA3</td>
<td>-1.13</td>
<td>SNORA7A</td>
<td>-0.28</td>
</tr>
</tbody>
</table>

Figure 5: Several tRNAs are processed into small RNAs by Dicer. The processing patterns shown some similarities, in particular a tendency to have large short read coverage on the 3' side of the tRNA clover leaf. With the exception of tRNA-Ile-TAT the small RNAs are derived from within the mature tRNA.
“mir-3195”, furthermore, corresponds to a GC-rich low-complexity region located with the first exon of the TAF4 gene.

The sequence of mir-663a is very GU-rich and does not meet our exclusion criterion for low-complexity sequences. We retained it in our data set because it is well document as an important tumor suppressor (Pan et al., 2010; Yi et al., 2012). In contrast to canonical microRNAs, its primary sequence is quite poorly conserved although it can be found throughout the major eutherian groups. Its read pattern also strongly deviates from the expectation for microRNAs. Similar to mir-451, the precursor is covered with a background of short reads, as also seen in the cumulative read patterns provided by the MicroRNA Registry, albeit there is dominating, most frequently produced “mature microRNAs”.

4. Discussion

A rapidly expanding zoo of diverse small RNA species has emerged following the discovery of RNA interference (Fire et al., 1998) and microRNAs (Lee et al., 1993) almost two decades ago. With the rapid increase of high throughput sequencing data the boundaries between the different subdivisions of small RNAs have become more and more blurry.

Here we have focussed on the generation of small RNAs from their double-stranded precursors. Making use of a publicly available dataset (Friedländer et al., 2012) we find, consistent with the well-established knowledge, that the overwhelming majority of miRBase microRNAs is processed by Dicer. There are, however, several notable exceptions. Cole et al. (2009) argue that Dicer knockdown with siRNAs for a short period of time sometimes does not result in a significant change in the miRNA steady state level due to slow microRNA turnover. At least some of the Dicer-unresponsive miRNAs, however, exhibit unusual structural features and/or read patterns that deviate substantially from canonical microRNAs. While $\lambda > 0$ in itself is course not sufficient proof for Dicer-independence, it is at least a strong indication and helps to identify candidates for further analysis.

Dicer-processing is not limited to microRNAs. Several polymerase-III transcripts are prolific Dicer substrates, including human vault RNAs, the great ape specific small nuclear RNAs, and a small set of about a dozen tRNAs. While the vault RNAs products function like microRNAs, small RNAs derived from RNA-Gln-CTG do not function in this way: they do not associate with argonaute presumably due to the fact that these small RNAs are just too small (Cole et al., 2009). Despite their similarity with vault RNAs, including a secondary structure with a long terminal stem, there is no evidence that the abundant small RNAs deriving from Y RNAs are produced by Dicer cleavage. Both main classes of small nuclear RNAs are sources of abundant small RNAs. While all of the highly expressed box C/D snoRNAs are processed independently of Dicer, the situation is different for H/ACA snoRNAs. Most box H/ACA snoRNAs are a source of small RNAs, but in most cases the expression levels are small, at least in the investigated MCF-7 libraries. Among the highly expressed ones, however, the majority clearly is a Dicer substrate.

In summary, there does not seem to be a clear separation between processing pathways resulting in
small RNAs. Instead, the picture of an intricate network of interlaced alternatives emerges, in which the individual processing steps can be freely combined. As a consequence, it appears that a particular sequence of processing steps is neither a sufficient nor a necessary condition for a particular role. Small RNA sequencing data such as the ones analyzed here reveal only the end points of a likely more complex processing cascade. Longer potential intermediates, such as pre-microRNA hairpins or the parts of tRNAs resulting from stress-related cleavage in the anticodon loop (Jöchl et al., 2008; Thompson et al., 2008) are invisible here. It will be an interesting topic for future research to investigate if and how the generation of small RNAs is linked to other RNA processing mechanisms.

Acknowledgments

This work is supported in part by the Deutsche Forschungsgemeinschaft (SPP-1258 “Sensory and Regulatory RNAs in Prokaryotes”: STA 850/7-2 to PFS). LIFE – Leipzig Research Center for Civilization Diseases, Universität Leipzig is funded by means of the European Social Fund and the Free State of Saxony.

Conflict of interest

The authors declare no conflict of interest.

References


Figure 3: Six un-annotated non-repetitive loci that are processed by Dicer. (a) intergenic chr2:81,100,049-81,100,134(+); (b) a “semi-mirtron” in an intron of SYT12 chr11:66,569,729-66,569,790(+); (c) a source in an intron of SLC4A2 chr7:150,394,782-150,394,835(+); (d) FLNA chrX:153,235,873-153,235,943(-); (e) MAP3K4 three mirtrons: (c) SLC4A2 chr7:150,394,782-150,394,835(+); (f) TRIM28 chr19:63,753,464-63,753,555(+). The color scale represents the coverage on a logarithmic scale.

Figure 4: smaRs are processed by Dicer. Highlighted are the tags showing the highest expression.