Chapter 11
Different Types of Small RNAs in Protozoa

Ling-Ling Zheng, Jian-You Liao, Yan-Zi Wen, Geoff Hide,
Liang-Hu Qu and Zhao-Rong Lun

Abstract The origin and evolution of small RNA molecules is a long-standing mystery. The protozoa represent the earliest eukaryotes and encompass rich diverse genetic resources. However, there are many contradictions and disputes prevalent in studies on protozoan microRNAs. In this review, we summarize the research on the analysis of small RNA transcriptome data on two representative protozoans, *Trypanosoma brucei* and *Giardia lamblia*, as generated by high throughput sequencing. The results show that these protozoans do not have canonical miRNAs. Unexpectedly, there are many types of endogenous small interfering RNAs (endo-siRNAs). In addition, stress induced tRNA derived small RNAs (sitRNAs) were observed in different life cycle stages of these parasites. In total, there are six kinds of sitRNAs in *G. lamblia*. These small RNAs are then found to be involved in the differentiation processes of primitive eukaryotes. These results showed that systematic analysis of small RNAs in the protozoa revealed a more complex picture than previously thought.

L.-L. Zheng · Y.-Z. Wen · L.-H. Qu (✉) · Z.-R. Lun (✉)
Key Laboratory of Gene Engineering of the Ministry of Education, State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, People's Republic of China
e-mail: lssqlh@mail.sysu.edu.cn

Z.-R. Lun
e-mail: lsslzr@mail.sysu.edu.cn

J.-Y. Liao
Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Sun Yat-Sen Memorial Hospital, Research Center of Medicine, Sun Yat-Sen University, Guangzhou 510120, People's Republic of China

Y.-Z. Wen · Z.-R. Lun
Key Laboratory of Tropical Diseases and Control of the Ministry of Education, Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou 510080, People's Republic of China

G. Hide · Z.-R. Lun
Ecosystems and Environment Research Centre/Biomedical Research Centre, School of Environment and Life Sciences, University of Salford, Salford M5 4WT, UK

11.1   Introduction to Small RNAs in Protozoa

Small RNAs are defined by their length (20–30 nucleotides) and play important roles in post-transcriptional regulation (Ha and Kim 2014). There are two main types of well-studied small RNAs: miRNAs (microRNAs) and endo-siRNAs (endogenous small interfering RNAs). Although, these two types of small RNAs are all involved in RNAi (RNA interference) pathways, their source, processing mechanisms and functions are very distinct.

   MicroRNAs have been well studied in the past two decades since their first identification in *Caenorhabditis elegens*. They were later found in a diverse range of animals and these microRNAs appear to be widely expressed in animal somatic cells. They probably drive many of the epigenetic developmental processes that transform embryonic stem cells to terminally differentiated cells. Different tissues or cell lines contain their own specific profile of miRNAs, which regulate the expression of genes that are essential for cell-fate determination in cell differentiation (Sood et al. 2006; Li and Belmonte 2015) and hematopoietic development (Aravin and Hannon 2008). In the genome, microRNAs are self-governed genes which exist in inter-gene and intron regions. In animals, microRNA genes are typically transcribed by RNA polymerase II into long precursors pri-miRNAs (Li and Rana 2014). They are then processed by the Drosha enzyme into stem loop structure pre-miRNAs in the nucleus. The pre-miRNAs are then transported into cytoplasm, where pre-miRNAs are cleaved by the protein Dicer to generate a pair of 18–22 nt small RNAs—mature miRNA and miRNA star. The mature miRNA are loaded into the Argonaute complex and guide the complex to the 3′ UTR of mRNAs through base pair complementarity in the seed region (2–8 nt) of the microRNA. The expression of protein-coding genes are then suppressed at the post-transcriptional level. In general, the miRNA star is degraded in miRNA biogenesis, however, in some cases, these miRNA stars can also target mRNAs and have biological functions in development of diseases such as cancer (Yang et al. 2013; Shan et al. 2013; Wu and Arora 2014).

   Endo-siRNAs are derived from long double-strand RNAs (dsRNAs), and then processed into 21–26 nt small RNA clusters. These genes were first found in *Caenorhabditis elegans* and can silence the expression of genes by means of perfect complementarity with the transcripts (Wightman et al. 1993). Endo-siRNAs are mainly derived from Transponsable Elements (TEs) and are believed to be the main mechanism for retrotransposon element silencing. There are many types of endo-siRNAs in plants (Nilsen 2008), however, it is difficult to ascertain their characteristics in animals (Okamura and Lai 2008). The pioneer work in the investigation of endo-siRNAs was performed by two groups and both of them identified a substantial number of endo-siRNAs in mouse oocytes (Tam et al. 2008; Watanabe et al. 2008). A range of studies, conducted at a similar time, used deep-sequencing technology to reveal that there were also an abundant repertoire of endo-siRNAs in *Drosophila melanogaster* (Ghildiyal et al. 2008; Okamura et al. 2008a, b; Kawamura et al. 2008; Czech et al. 2008; Chung et al. 2008).

MicroRNAs have been discovered in sponges and in *Chlamydomonas rein-hardtii*, which are the lowest multicellular animals and plants, respectively. But homologous sequences and the characteristic microRNAs were not found between these two taxa (Wheeler et al. 2009). This suggested that microRNAs in plants and animals seem to have originated independently from different ancestors. A few studies have attempted to analyze the small RNAs in protozoans such as try-panosomes (Wen et al. 2011; Mallick et al. 2008; Tschudi et al. 2012a; Zheng et al. 2013), trichomonads (Chen et al. 2009; Lin et al. 2009; Huang et al. 2012), giardias (Chen et al. 2009; Liao et al. 2014; Saraiya and Wang 2008; Huang et al. 2012), amoeba (De et al. 2006) and toxoplasma (Braun et al. 2010). Some of them have reported that microRNAs demonstrate homology with those found in higher spe-cies. This implied that microRNAs might have originated from a common ancestor in the eukaryotes. Later other studies questioned the microRNAs identified in protists. Tarver et al. used critical criteria to re-evaluate these microRNAs and concluded that most of the identified microRNAs in protists were not bona fide microRNA (Tarver et al. 2012). These results bring more questions about the origin of microRNAs. Do microRNAs exist in ancient eukaryotes? Whether the features of these so called "microRNAs" are in common with/or specific from those that have been demonstrated in the plants and animals? If the molecules previously reported are not microRNAs, do they perform other specific functions? To answer these questions, a systematic analysis of small RNAs in protists is required. However, it is difficult work because of the following problems.

11.2 Current Problems Associated with the Identification of Small RNAs, and Particularly Authentic microRNAs, from Protozoan Organisms

There are at least four difficulties in small RNA identification in protozoan species:

1. The protozoans are a very diverse group of unicellular eukaryotic organisms that may well have had extensive polyphyletic origins (Cavalier-Smith 2003). The characteristics of these organisms are very broad, some are animal-like, some are plant-like and others also like fungi. Thus, there are a wide range of potential model protozoa organisms but may not be representative of all (Montagnes et al. 2012). Therefore, there are no common characteristics of small RNAs in pro-tozoan organisms which makes it very difficult to investigate them systematically.

2. Most of the parasitic or symbiotic protozoan species have complicated life-cycles, due to the fact that they can infect or inhabit multiple types of hosts (including a wide range of plants and animals). To adapt to the very different environments within different hosts, protozoans have to change their morphol-ogy frequently. These changes are dependent on the expression of genes

(Lun et al. 2015). Gene expression and regulation in protozoan organisms is, therefore, potentially very complicated.

3. The current analysis software for microRNA identification has been specifically designed for use with higher animals and plants. However, the small RNAs in protozoan species are very different from those found in higher organisms. Thus, this software cannot be used directly for the investigation of protozoan small RNAs.

4. The next generation sequencing (deep-sequencing) technology has improved the convenience and accuracy of microRNA identification. However, compared with higher organisms, there is very limited small RNA transcriptome data for protozoan species lodged in databases. This is especially true for comparative data between different life cycle stages.

11.3 Computational Methods for Genome-Wide Analysis of Small RNAs in Animal-Like Protozoans

To specifically investigate small RNAs in protozoans, we developed two computational methods to analyze endo-siRNAs and microRNAs from deep-sequencing results which could be used for animal-like protozoans. The origins of siRNAs have been found to be from longer double stranded RNA which have been cleaved by the protein Dicer. Thus siRNAs can be distinguished from random degradation or small RNA production by pretreating extracted protozoan RNA with Dicer (Fig. 11.1). The families of fragments are then subjected to deep sequencing using standard approaches. Next the 3′ adaptors were removed from the illumina-generated sequence reads and then the remaining 18–30 nt reads were mapped to the desired protozoan genome database(s). For reads that mapped to multiple loci, they were excluded if there were greater than nine hits on the genome. Figure 11.1 shows the flowchart for siRNA cluster identification. We first counted the numbers (N) of small RNAs that mapped to both strands of the genome to generate two numbers: one for the sense and the other for the antisense strand. These small RNAs are linked into a cluster if they have at least one base that is overlapping.The length of the longest cluster is then calculated (L). These numbers were then divided by the length of the cluster (L) to obtain the densities (N/L) of small RNA enrichment. We chose the density of 0.04 as a cutoff in order to select candidate NATs with high potential to produce NAT derived siRNAs. This threshold is the same as used in the study by Guo et al. on trans-NATs in rice (Guo et al. 2009). Identification of microRNA was carried out as shown in Fig. 11.2. These are usually found in a single location and have no adjacent clusters. They are recognized by regions of reverse complementarity enabling the formation of hairpin secondary structure.
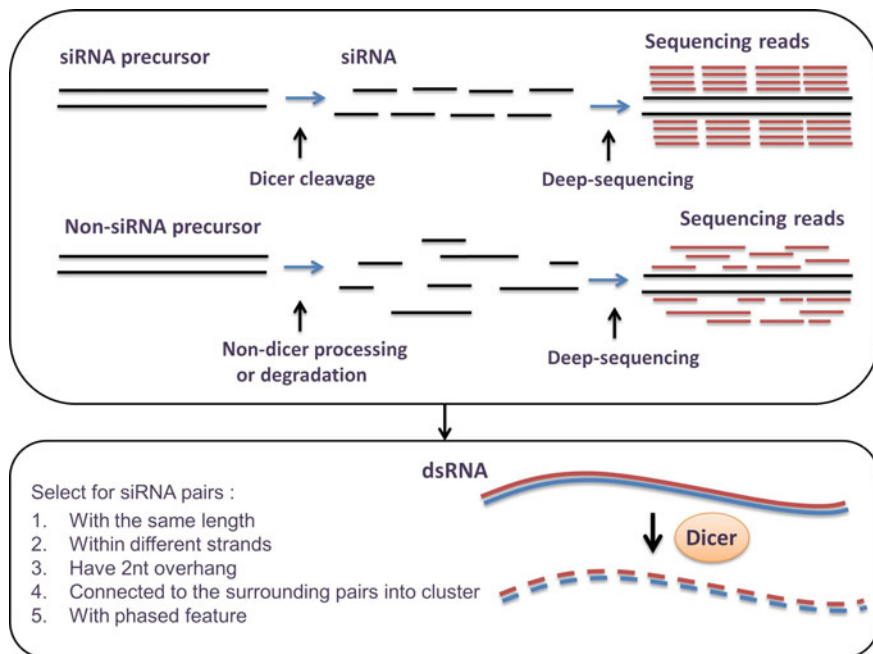
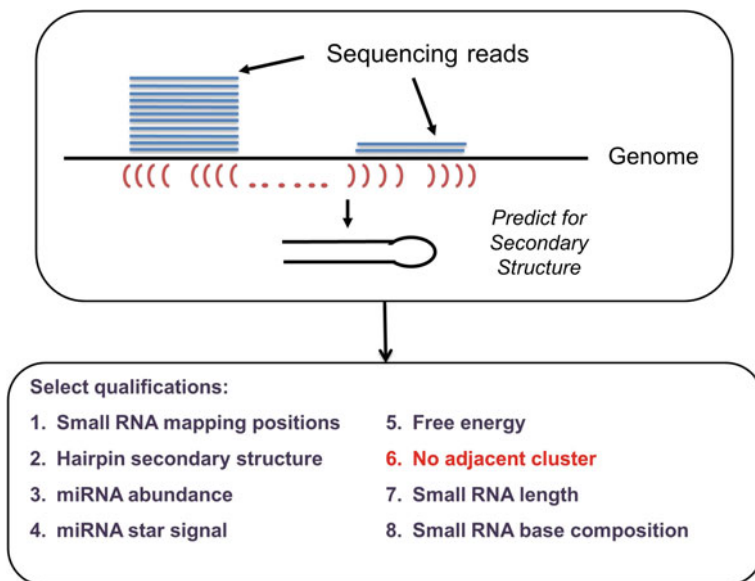**Fig. 11.1** siRNA cluster identification

**Fig. 11.2** microRNA identification

## 11.4   Small RNA Investigation in *Trypanosoma brucei*

We combined computational analysis and high-throughput sequencing to study siRNAs and microRNAs in *T. brucei*, produced during the two main life cycle stages, bloodstream form (BF) and procyclic form (PF). We focused specifically on Natural Antisense Transcript (NAT) pairs derived siRNAs in order to gain further insight into the extent and functional role of these RNAs in protozoan organisms. We identified hundreds of NATs on the annotated transcript-scale and whole genome-wide analyses (Tschudi et al. 2012a). We identified the majority of these small RNAs as siRNAs, clearly indicating the prevalence of this mechanism and its potential importance in protozoan parasites.

11.4.1   cis- *and* trans-*NATs Were Identified Within the* T. brucei *Genome*

Putative cis and trans-NATs identification was based on the annotation of *T. brucei* transcripts (Fig. 11.3). To identify the potential cis-NAT pairs, we compared the genomic loci of all annotated transcripts to search for gene pairs that in the same loci but from opposite strands. Eleven enclosed pairs of transcript elements were identified by overlapping reverse-complementary sequences and all of which were formed by protein coding genes and non-coding genes (snoRNAs and snRNA).

**Fig. 11.3** Genome-wide identification of cis- and trans-NATs
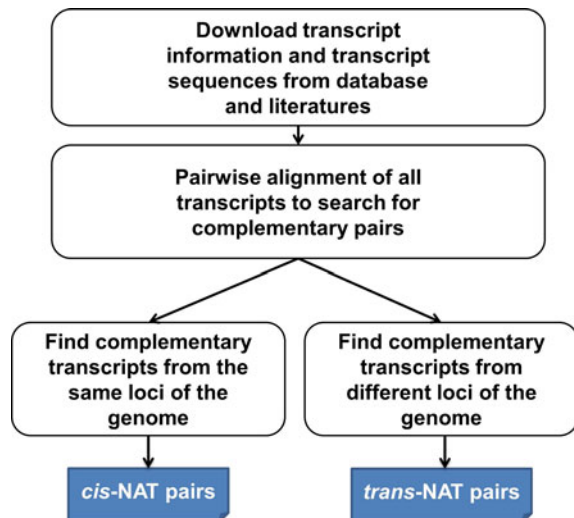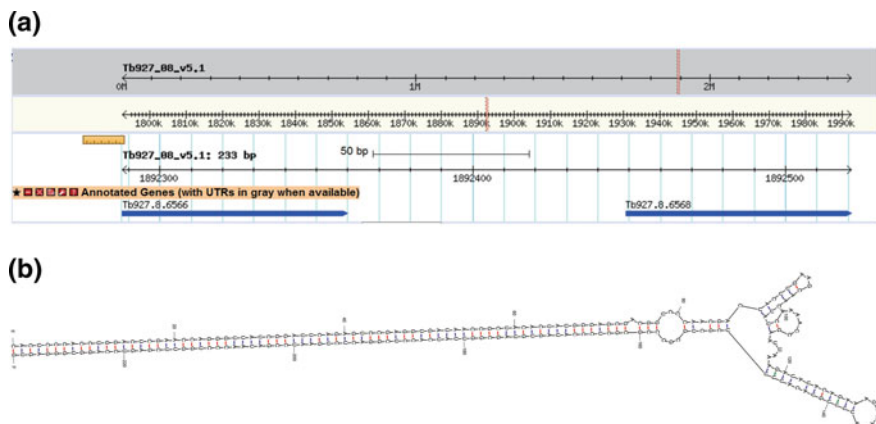
Fig. 11.4 Location and structure of two adjacent tRNAs which can form into NAT pairs. **a** Two Glutamine tRNAs (Tb927.8.6566 and Tb927.8.6568) at the adjacent locus of Tb927_08_v4 from 1892288 to 1892520. **b** Predicted structures that would be formed by base-pair interactions of the region. Secondary structures are folded using Mfold (Zuker 1989)

We carried out a pairwise alignment on all annotated transcript sequences to find *trans*-NAT pairs and identified 209 pairs of *trans*-NAT candidates. Specially, two tRNAs in adjacent locations could form into hairpin structures. Both of the two tRNAs were glutamine tRNA, locating in adjacent loci in plus strand of Tb927_08_v4 (Fig. 11.4a). The two inverted tRNAs were composed of perfect palindromes with an 89 nt internal sequence and were able to form highly stable hairpin loops (Fig. 11.4b).

We did not find any NATs in the mitochondrial genome. This was in accordance with the results in *P. falciparum* (Gunasekera et al. 2004) but inconsistant with the results in humans, mice and worms in which the abundance of antisense transcripts were low in the nuclear genome but high in abundance in the mitochondrial genome (Zhang et al. 2006). This might be an interesting subject for further study.

11.4.2 NAT-Derived Small RNAs Show Features of siRNAs

The NATs in both life stages were distributed across all of the 12 chromosomes of *T. brucei* TREU927 strain (Fig. 11.5). Chromosome sequences of *T. brucei* were downloaded from the TriTrypDB (release-2.1) component of EuPathDB (http://tritrypdb.org/tritrypdb/) (Aslett et al. 2010). Some NATs exhibit specific clustering across particular regions of the chromosomes (Fig. 11.5). Furthermore, we realized that these clusters were the regions where duplication events occurred frequently. For example, the longest NAT extended 8436 bp and spanned a cluster of
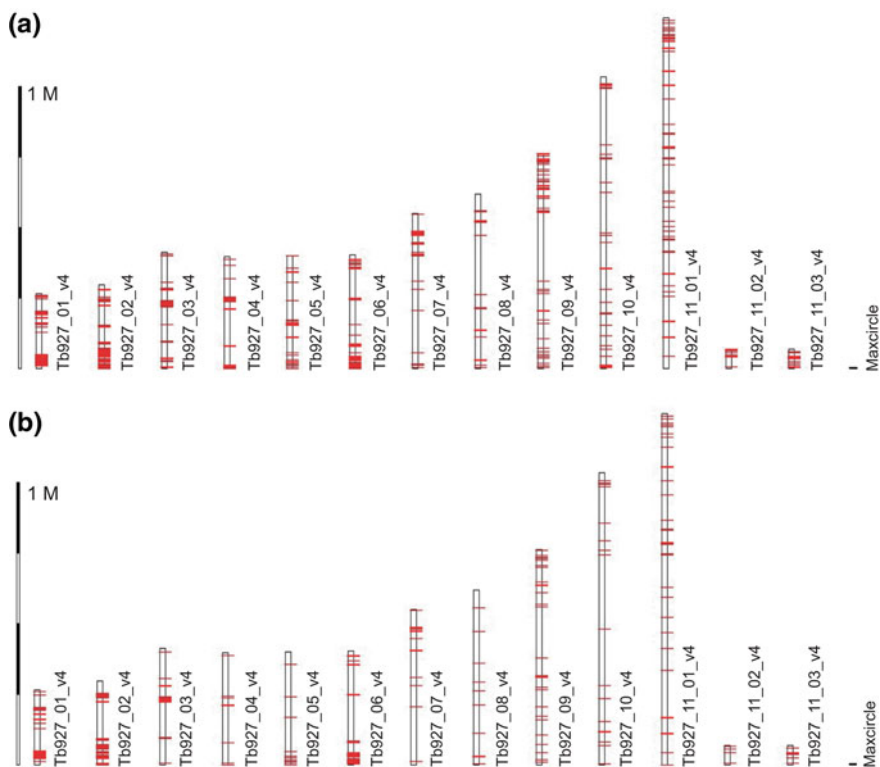
Fig. 11.5 Chromosome map of NATs. All of the NATs are mapped by there location on the chromosome in BF stage (**a**) and PF stage (**b**)

transcripts which were mainly transcribed from retrotransposon hot spot protein (RHS) pseudogenes. As shown in Fig. 11.6a, the NAT on Tb927_01_v4 from 126434 to 134869, where 6 transcripts were crossed, which included four RHS pseudogenes and two protein-coding genes with unknown function.

For each NAT type (i.e. cis-NATs or trans-NATs), we computed the densities of small RNA loci in the overlapping region. Small RNAs which derived from sense and antisense of the genome region are widely distributed in the bloodstream form and procyclic form (Fig. 11.6b). We then calculated the features of these small RNAs. In total, we got 259433 and 84339 NAT-derived small RNAs in bloodstream form and procyclic form, respectively. Most of these small RNAs were in the length region of 23–26 nt and had a characteristic "U" at the 5′-terminal (Fig. 11.7). These features correspond to the known siRNAs in *T. brucei* (Djikeng et al. 2001). Our further experiments showed that these small RNAs are indeed siRNAs and dependent on dicer processing (Zheng et al. 2013).
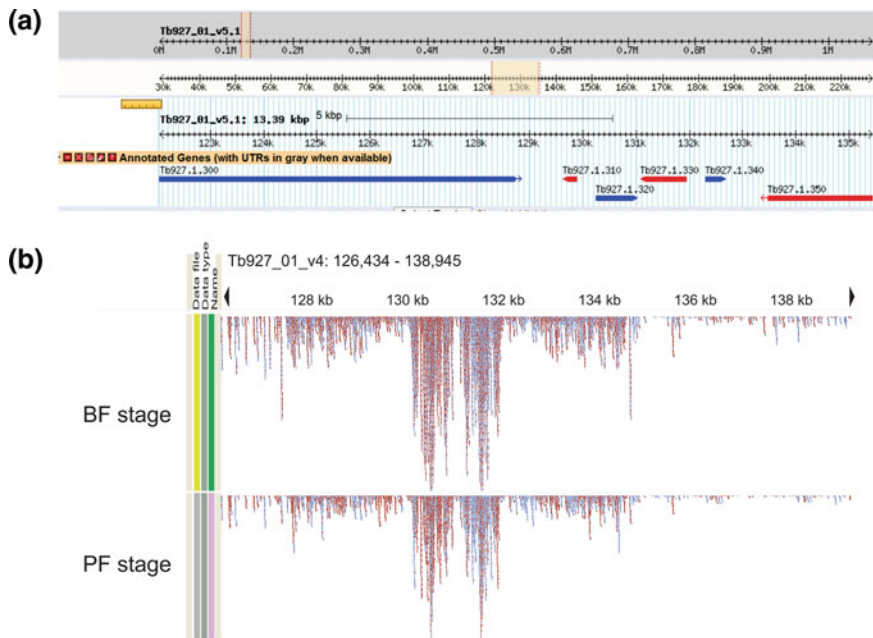
Fig. 11.6 Example of small RNAs derived from NAT. **a** is the screenshot of a region of T. brucei genome browser from TriTrypDB (Aslett et al. 2010) which show transcripts position information across the genome. *Arrow* indicates the transcription direction (5′ to 3′). *Color* also indicates the transcription direction: *blue* for sense transcripts and *red* for antisense transcripts. **b** Is the screenshot of a region of T. brucei genome browser from IGV (Thorvaldsdottir et al. 2013) The *small lines* with different *color* indicate the small RNAs originated position the *color* indicate their transcription direction (*blue* for sense *red* for antisense). *Top* is BF stage and *bottom* is PF stage

11.4.3 Endo-siRNAs Have Multiple Sources and Differential Expression in Two Life Cycle Stages

In addition to NATs, it has been shown that there are many regions in the genome that could generate siRNAs (Djikeng et al. 2001; Zheng et al. 2013; Tschudi et al. 2012a). These sources are listed in Table 11.1. Stage-specific siRNAs have been proved to be induced to respond to special environmental stress (Borsani et al. 2005; Hao et al. 2010). But what about the siRNAs in *T. brucei*? In our results, 78.6 % of siRNAs were only identified in the bloodsteam slender form, whereas only 9.2 % were specifically expressed in the insect procyclic form. Interestingly, the stage-specific siRNAs tended to be transcribed from clustered regions. For example, the region of Tb927_07_v4 (317344–621335) is a source of siRNAs. Surprisingly, although plenty of siRNAs were derived from this region in the slender form, little was found in the procyclic form (Fig. 11.8). Next, we analyzed the expression patterns of siRNAs in both stages and found 534 siRNAs exhibited
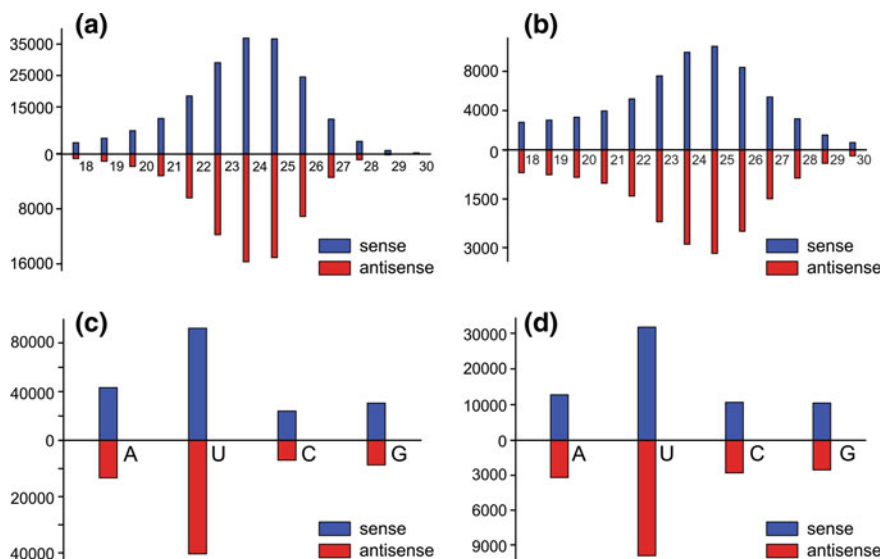
Fig. 11.7 Feature distribution of BTR-derived small RNAs. The length distribution of small RNAs derived from BTRs in BF stage (**a**) and PF stage (**b**). The 5′ end nucleotide distribution of small RNAs derived from BTRs in BF stage (**c**) and PF stage (**d**)

Table 11.1 The sources of siRNAs in *T. brucei*

Abb	Full Name	References
CIR147	Chromosome inteRNAl repeats	Obado et al. (2005)
SLACS	Spliced leader-associated conserved sequence	Djikeng et al. (2001)
Ingi	Ingi	Djikeng et al. (2001)
IR	Long inverted repeats	Tschudi et al. (2012a)
CTU	Convergent transcription unit	Tschudi et al. (2012a)
miscRNA	cMiscellaneous,	Tschudi et al. (2012a)
NAT	Nature antisense transcripts	Zheng et al. (2013)

significant expression differences (p < 0.001), 79 % are expressed significantly higher in the BF while only 21 % are expressed significantly higher in the PF stage.

11.4.4 Non-existence of Canonical microRNA in Selected Stages of T. brucei

In 2008, Mallick et al. reported that they used a bioinformatic approach to identify microRNA genes in *Trypanosoma brucei* and to predict their functions
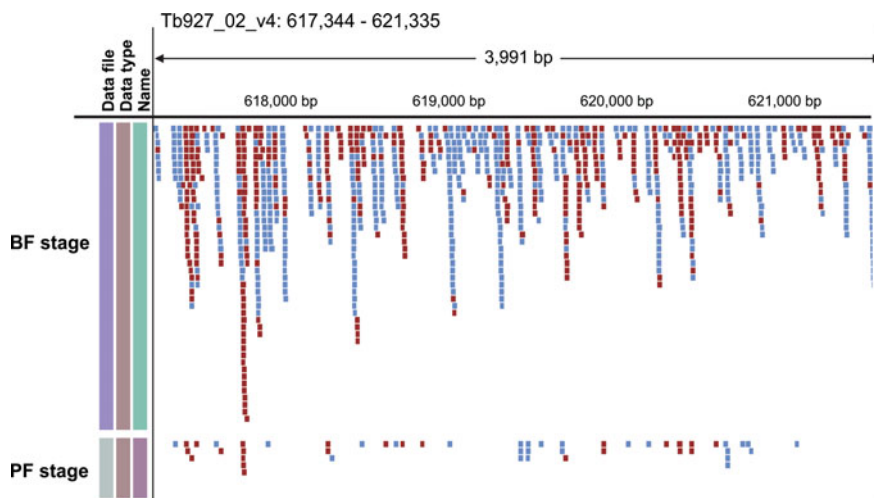
Fig. 11.8 Is the screenshot of a region of *T. brucei* genome browser from IGV(Thorvaldsdottir et al. 2013). The *small lines* with different *color* indicate the small RNAs original position and the *color* indicates their transcriptional direction (*blue* for sense *red* for antisense). *Top* is BF stage and *bottom* is PF stage

(Mallick et al. 2008). In total, they identified 1162 microRNA candidates and most of them targeted to the variant surface glycoprotein (VSG) in this parasite protozoan. They suggested that the microRNAs in *T. brucei* might play roles as genetic switches in modulating host-parasite interaction.

However, our results from the high-throughput sequencing on *T. brucei* did not identify any microRNA molecules from this parasite. Combined with the previous studies (Tschudi et al. 2012b; Kolev et al. 2011) we did not support the existence of microRNA in this parasite, at least in these two stages. We do not know the reasons behind this different result but it might be related to the microRNA genes which are expressed only in different stages of the life cycle of *T. brucei*. More studies need to be carried out to investigate this interesting topic.

No matter whether microRNAs exist in *T. brucei* or not, our results have shown that siRNAs as well as other small RNAs, like tRNA-derived small RNAs and rRNA derived small RNAs, are important regulators in the differentiation of *T. brucei* (Zheng et al. 2013). By comparison with the procyclic form, we found that some genes in the VSG family were specifically regulated by siRNAs in the slender form of *T. brucei*. VSG is an important antigen in the slender form of *T. brucei*. The extensive variety of this antigen allows the parasite to escape from the immunological defense of the mammalian hosts. Therefore, we propose that the siRNAs correspondingly evolve to eliminate the obsolete VSG transcripts as rapidly as possible, so that the novel VSG can be effective in a timely way. Taken together, our results provide genome-wide identification of small RNAs in *T. brucei*. These results will facilitate functional studies of siRNAs in this model

protozoan, as well as in other protozoan species, and help to unravel complex gene and non-coding RNA regulatory networks in eukaryotes.

11.5 Investigation into Small RNAs in *Giardia lamblia*

Giardia lamblia is one of the most primitive eukaryotes, and the study of its sRNA transcriptome could shed light on the origin and evolution of various sRNAs. We found that the *G. lamblia* sRNA transcriptome mainly consists of two kinds of sRNAs: endo-siRNA and tRNA derived sRNA (Liao et al. 2014). Both of these two kinds of sRNAs are also found in higher eukaryotic genomes, such as humans, suggesting they might be involved in the regulation of basic biological processes in eukaryotes.

11.5.1 Retrotransposon and eSGR Derived Endo-siRNAs

G. lamblia endo-siRNAs were primarily derived from two sources, retrotransposon and endo-siRNA generating regions (eSGR). The *G. lamblia* genome has three retrotransposons: two of these (GilT and GilM) are located in the telomeric regions, while the third retrotransposon (GilD) is dead (Arkhipova and Morrison 2001). Both functional retrotransposons, GilT and GilM, generate large number of endo-siRNAs. The dead GilD does not appear to generate any sRNAs. The endo-siRNA generating region (eSGR) is a region in the *G. lamblia* genome that seems to be specifically used to produce endo-siRNAs (Liao et al. 2014) and encodes three eSGRs (Fig. 11.9), named eSGRI, eSGRII and eSGRIII respectively. The eSGRI region is the longest and generates most of the eSGR derived endo-siRNAs (87.1 %). It spans
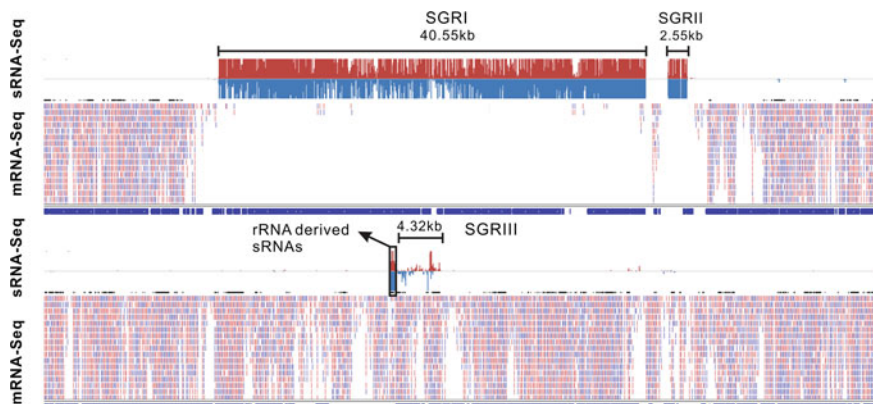
Fig. 11.9 The mRNA and sRNA expression statuses of three SGRs

40.55 kb in length and contains 34 protein-coding genes. Intriguingly, all these 34 protein-coding genes in eSGRI are not expressed. It is possible, therefore, that their expression is blocked by eSGRI derived endo-siRNA. In contrast, eSGRII and eSGRIII produce much less endo-siRNAs than eSGRI and they occupy shorter regions, 2.55 kb for eSGRII and 4.32 kb for eSGRIII (Fig. 11.9). These two regions also contain some protein coding genes, but these mRNAs are being expressed (Fig. 11.9). This might be due to the low expression levels of eSGRII and eSGRIII derived endo-siRNAs. The eSGR derived endo-siRNAs have the same characteristics as retrotransposon derived endo-siRNAs, which may be caused by them sharing an identical biogenesis pathway.

G. lamblia endo-siRNAs share very similar characteristics with endo-siRNAs of other eukaryotes including *T. brucei*. For instance, like endo-siRNAs in other eukaryotes, the *G. lamblia* endo-siRNAs have a length of between 20–30 nt, they are derived from both strands of the endo-siRNA generating regions and have one or more post-transcriptionally added 3′ untemplated nucleotides. The conservation of these characteristics suggests that they may be important for the function of endo-siRNAs. Unexpectedly, the 5′ end nucleotide of *G. lamblia* endo-siRNAs does not appear to be "U" like the endo-siRNAs of higher eukaryotes. It's unclear yet how and when endo-siRNA are processed to acquire their 5′ end "U". This is an interesting question worth pursuing in the future.

11.5.2 The Function of the G. lamblia *Endo-siRNA Pathway*

It's hard to clarify the detailed function of each endo-siRNA, because of the huge number of endo-siRNAs. We have identified more than a hundred thousand endo-siRNAs in the *G. lamblia* trophozoite, and the expression of each of these endo-siRNAs is low. A good way to investigate endo-siRNA function is to block the key protein in the endo-siRNA biogenesis pathway. Through knock-down of the *G. lamblia* Dicer gene (GLDICER), which is responsible for the slicing of the endo-siRNA precursor to generate mature endo-siRNAs (Macrae et al. 2006), we significantly reduced the expression level of all endo-siRNAs and observed that the ability of *G. lamblia* to differentiate was impaired. This result indicated that the endo-siRNA pathway might be involved in the regulation of *G. lamblia* differentiation. Another function of the *G. lamblia* endo-siRNA is in regulating antigenic variation (Prucca et al. 2008). This was found by Prucca et al. through knock-down of two proteins GLDICER and *G. lamblia* RNA dependent RNA Polymerase (GLRDRP). GLRDRP is a protein used to generate appropriate RNA for GLDICER slicing. Both the silencing of GLDICER or GLRDRP lead to the failure of mechanisms of *G. lamblia* antigenic variation. It's unclear yet how endo-siRNAs regulates differentiation and antigenic variation. The detailed study of the regulation mechanisms of endo-siRNAs is key to an understanding of the functional evolution of endo-siRNA in eukaryotes.

### 11.5.3 The G. lamblia *Genome Encodes Six Kinds of tRNA Derived sRNAs*

Recent studies have shown that tRNA derived sRNA (tsRNAs) is involved in the regulation of translation and various pathological processes like cancer (Goodarzi et al. 2015) and sperm tsRNAs from father may mediate intergenerational inheritance of diet-induced metabolic disorders (Chen et al. 2016; Sharma et al. 2016). We found that tRNA derived sRNAs are conserved in *G. lamblia* and exhibited changes in expression profiles between different life stages (Liao et al. 2014; Li et al. 2008), suggesting that the role of tRNA derived sRNA in cells might be more important than previous expected. In total, six kinds of tRNA derived sRNAs were found in *G. lamblia*, they were named as 5tasRNA, actasRNA, 3tasRNA, 5sitRNA, 3sitRNA and 5EsRNA respectively. The 5tasRNA, actasRNA, 3tasRNA members were derived from the 5′ end, middle region (most of the anticodon stem loop) and 3′ end of the mature tRNA, with lengths of around 20–22, 24–26, 28–30 nt respectively. Similarly 5tasRNA and 3tasRNA, 5sitRNA and 3sitRNA were also processed from the 5′ and 3′ ends of mature tRNA respectively. However these had much longer lengths, 50 nt for 5sitRNA and 46 nt for 3sitRNA. 5EsRNA was derived from the 5′ end of pseudotRNA-Gln (TTT), with a length of 36 nt.

11.5.4 The Biogenesis and Potential Function of *G. lamblia* tRNA Derived sRNAs

With the exception of 5EsRNA, all five remaining kinds of tRNA derived sRNAs were generated from various mature tRNAs. We identified two cleavage sites, Cleavage Site I and II (CSI and CSII), in the clover leaf structure of various tRNAs (Fig. 11.10). We found that 5tasRNA, actasRNA, 3tasRNA, and 5sitRNA, 3sitRNA might be all generated from cleavage at these two sites by an, as yet,

Fig. 11.10 Five kinds of sRNAs derived from mature tRNAs. CS refers to cleavage site
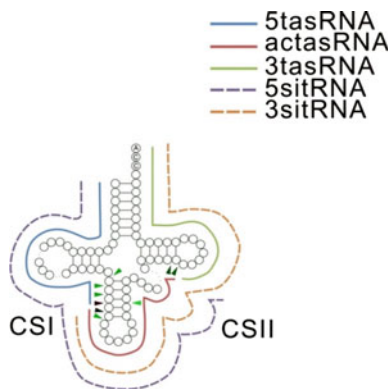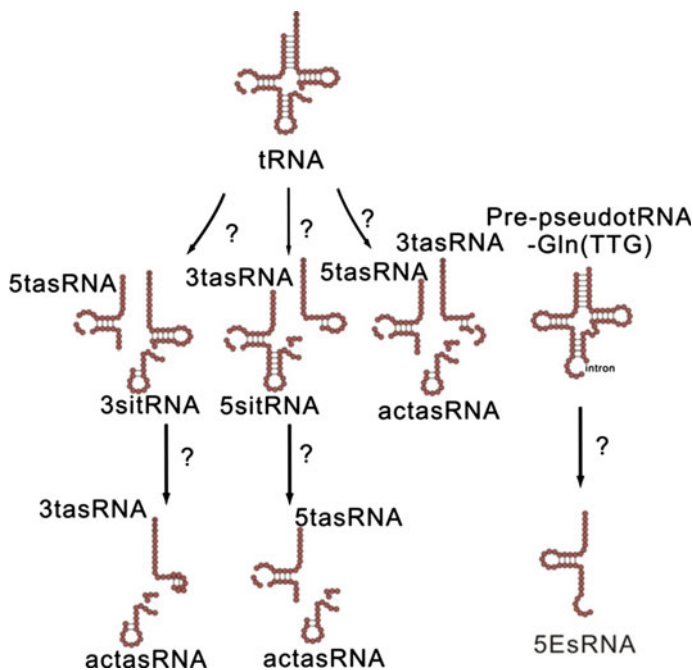
Fig. 11.11 The potential biogenesis pathway of six kinds of *G. lamblia* tRNA derived sRNAs

unknown RNA endonuclease. Considering that the 5tasRNA and 3sitRNA can be integrated into an intact tRNA, as does 3tasRNA and 5sitRNA, and also 5tasRNA, actasRNA and 3tasRNA, it's possible that the cleavage of mature tRNAs may generate one of these three groups of products (Fig. 11.11). The sequence of 5EsRNA covers the whole 5′-end exon of pseudotRNA-Gln(TTG), its biogenesis might borrow the pecursor process machinery from pseudotRNA-Gln(TTT).

The detailed functions of G. lamblia tRNA derived sRNAs are as yet unclear. 3sitRNA was the first identified *G. lamblia* tRNA derived sitRNA (Li et al. 2008). It is not expressed in trophozoites, but can be induced by several stress factors, thus connecting them with a potential stress resistant related function. Similarly, 3sitRNA, 5sitRNA, 5tasRNA, actasRNA and 3tasRNA are not expressed in the trophozoite stage either. However, all of these five kinds of sRNAs are expressed when the trophozoite starts to differentiate into cysts. They are also expressed at a high level in cysts, indicating that they might be involved in the differentiation process in *Giardia*. It's difficult to investigate the function of tRNA derived sRNAs as yet, because we currently lack appropriate methods to manipulate their gene expression. However, the identification of endonucleases responsible for the biogenesis of these sRNAs may overcome this obstacle.

11.5.5 *Does the* G. lamblia *Genome Possess microRNA-Encoding Genes?*

Because miRNAs share the same biogenesis machinery and effectors with endo-siRNA, it allows the possibility that the genome of *G. lamblia* may encode miRNA genes. Indeed, many studies have identified 166 potential miRNA in *G. lamblia* (Saraiya et al. 2011; Huang et al. 2012; Zhang et al. 2009; Chen et al. 2009; Saraiya and Wang 2008; Li et al. 2011, 2012). However, 76 % (126) of these miRNAs could not be detected by our very high depth sRNA sequencing. Among the 40 miRNAs that were revealed in our sequencing data, 12.5 % (5) were derived from ncRNAs, 50 % (20) were derived from an eSGR region, and the expression level of the remaining 37.5 % (15) miRNA is very low and only accounts for 0.0048 % of all reads in the trophozoite sRNA library. Therefore, there is no conclusive evidence, yet, that the *G. lamblia* genome encodes any miRNAs.

11.6 Concluding Remarks

In this review, we summarize the research on the analysis of small RNA transcriptome data on *Trypanosoma brucei* and *Giardia lamblia*. The results show that most of the protozoans do not have canonical miRNAs. However, there are many types of endogenous small interfering RNAs (endo-siRNAs) derived from pseudogenes, transcriptional elements, bidirectional transcription and NATs (natural antisense transcripts). In addition, widespread variation among the tRNA derived small RNAs were also observed in different stages of life cycle in these parasites. These results shed light on questions regarding the potential origins and evolution of mechanisms of small RNA-based control of gene expression.

References

Aravin AA, Hannon GJ (2008) Small RNA silencing pathways in germ and stem cells. In Cold Spring Harbor symposia on quantitative biology 73:283–290. doi:10.1101/sqb.2008.73.058

Arkhipova IR, Morrison HG (2001) Three retrotransposon families in the genome of *Giardia lamblia*: two telomeric, one dead. Proc Natl Acad Sci USA 98(25):14497–14502. doi:10.1073/pnas.231494798

Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, Gardner MJ, Gingle A, Grant G, Harb OS, Heiges M, Hertz-Fowler C, Houston R, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Logan FJ, Miller JA, Mitra S, Myler PJ, Nayak V, Pennington C, Phan I, Pinney DF, Ramasamy G, Rogers MB, Roos DS, Ross C, Sivam D, Smith DF, Srinivasamoorthy G, Stoeckert CJ, Jr., Subramanian S, Thibodeau R, Tivey A, Treatman C, Velarde G, Wang H (2010) TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res 38 (Database issue): D457-462. doi:10.1093/nar/gkp851

Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell 123(7):1279–1291. doi:10.1016/j.cell.2005.11.035

Braun L, Cannella D, Ortet P, Barakat M, Sautel CF, Kieffer S, Garin J, Bastien O, Voinnet O, Hakimi M-A (2010) A complex small RNA repertoire is generated by a plant/fungal-like machinery and effected by a metazoan-like Argonaute in the single-cell human parasite *Toxoplasma gondii*. PLoS Pathog 6:e1000920. doi:10.1371/journal.ppat.1000920

Cavalier-Smith T (2003) Protist phylogeny and the high-level classification of Protozoa. Eur J Protistol 39(4):338–348. doi:10.1078/0932-4739-00002

Chen Q, Yan M, Cao Z, Li X, Zhang Y, Shi J, Feng GH, Peng H, Zhang X, Zhang Y, Qian J, Duan E, Zhai Q, Zhou Q (2016) Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. Science 351(6271):397–400. doi:10.1126/science.aad7977

Chen XS, Collins LJ, Biggs PJ, Penny D (2009) High throughput genome-wide survey of small RNAs from the parasitic protists *Giardia intestinalis* and *Trichomonas vaginalis*. Genome Biol Evol 1:165–175. doi:10.1093/gbe/evp017

Chung WJ, Okamura K, Martin R, Lai EC (2008) Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. Curr Biol CB 18(11):795–802. doi:10.1016/j.cub.2008.05.006

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, Hannon GJ, Brennecke J (2008) An endogenous small interfering RNA pathway in *Drosophila*. Nature 453(7196):798–802. doi:10.1038/nature07007

De S, Pal D, Ghosh SK (2006) Entamoeba histolytica: computational identification of putative microRNA candidates. Exp Parasitol 113(4):239–243. doi:10.1016/j.exppara.2006.01.009

Djikeng A, Shi HF, Tschudi C, Ullu E (2001) RNA interference in *Trypanosoma brucei*: Cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. RNA 7(11):1522–1530

Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, Zamore PD (2008) Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. Science 320(5879):1077–1081. doi:10.1126/science.1157396

Goodarzi H, Liu X, Nguyen HC, Zhang S, Fish L, Tavazoie SF (2015) Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. Cell 161(4):790–802. doi:10.1016/j.cell.2015.02.053

Gunasekera AM, Patankar S, Schug J, Eisen G, Kissinger J, Roos D, Wirth DF (2004) Widespread distribution of antisense transcripts in the *Plasmodium falciparum* genome. Mol Biochem Parasit 136(1):35–42. doi:10.1016/j.molbiopara.2004.02.007

Guo X, Zhang Z, Gerstein MB, Zheng D (2009) Small RNAs originated from pseudogenes: cis- or trans-acting? PLoS Comput Biol 5(7):e1000449. doi:10.1371/journal.pcbi.1000449

Ha M, Kim VN (2014) Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol 15(8):509–524. doi:10.1038/nrm3838

Hao L, Cai P, Jiang N, Wang H, Chen Q (2010) Identification and characterization of microRNAs and endogenous siRNAs in *Schistosoma japonicum*. BMC Genom 11:55. doi:10.1186/1471-2164-11-55

Huang PJ, Lin WC, Chen SC, Lin YH, Sun CH, Lyu PC, Tang P (2012) Identification of putative miRNAs from the deep-branching unicellular flagellates. Genomics 99(2):101–107. doi:10.1016/j.ygeno.2011.11.002

Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H (2008) *Drosophila endogenous* small RNAs bind to Argonaute 2 in somatic cells. Nature 453 (7196):793–797. doi:10.1038/nature06938

Kolev NG, Tschudi C, Ullu E (2011) RNA interference in protozoan parasites: achievements and challenges. Eukaryot Cell 10(9):1156–1163. doi:10.1128/EC.05114-11

Li M, Belmonte JCI (2015) Roles for noncoding RNAs in cell-fate determination and regeneration. Nat Struct Mol Biol 22(1):2–4

Li ZH, Rana TM (2014) Therapeutic targeting of microRNAs: current status and future challenges. Nat Rev Drug Discovery 13(8):622–638. doi:10.1038/nrd4359

Li Y, Luo J, Zhou H, Liao JY, Ma LM, Chen YQ, Qu LH (2008) Stress-induced tRNA-derived RNAs: a novel class of small RNAs in the primitive eukaryote *Giardia lamblia*. Nucleic Acids Res 36(19):6048–6055. doi:10.1093/nar/gkn596

Li W, Saraiya AA, Wang CC (2011) Gene regulation in *Giardia lamblia* involves a putative microRNA derived from a small nucleolar RNA. PLoS Negl Trop Dis 5(10):e1338. doi:10.1371/journal.pntd.0001338

Li W, Saraiya AA, Wang CC (2012) The profile of snoRNA-derived microRNAs that regulate expression of variant surface proteins in *Giardia lamblia*. Cell Microbiol 14(9):1455–1473. doi:10.1111/j.1462-5822.2012.01811.x

Liao JY, Guo YH, Zheng LL, Li Y, Xu WL, Zhang YC, Zhou H, Lun ZR, Ayala FJ, Qu LH (2014) Both endo-siRNAs and tRNA-derived small RNAs are involved in the differentiation of primitive eukaryote *Giardia lamblia*. Proc Natl Acad Sci USA 111(39):14159–14164. doi:10.1073/pnas.1414394111

Lin WC, Li SC, Lin WC, Shin JW, Hu SN, Yu XM, Huang TY, Chen SC, Chen HC, Chen SJ, Huang PJ, Gan RRC, Chiu CH, Tang P (2009) Identification of microRNA in the protist *Trichomonas vaginalis*. Genomics 93(5):487–493. doi:10.1016/j.ygeno.2009.01.004

Lun ZR, Lai DH, Wen YZ, Zheng LL, Shen JL, Yang TB, Zhou WL, Qu LH, Hide G, Ayala FJ (2015) Cancer in the parasitic protozoans *Trypanosoma brucei* and *Toxoplasma gondii*. Proc Natl Acad Sci USA 112(29):8835–8842. doi:10.1073/pnas.1502599112

Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA (2006) Structural basis for double-stranded RNA processing by Dicer. Science 311(5758):195–198. doi:10.1126/science.1121638

Mallick B, Ghosh Z, Chakrabarti J (2008) MicroRNA switches in *Trypanosoma brucei*. Biochem Biophys Res Commun 372(3):459–463. doi:10.1016/j.bbrc.2008.05.084

Montagnes D, Roberts E, Lukes J, Lowe C (2012) The rise of model protozoa. Trends Microbiol 20(4):184–191. doi:10.1016/j.tim.2012.01.007

Nilsen TW (2008) Endo-siRNAs: yet another layer of complexity in RNA silencing. Nat Struct Mol Biol 15(6):546–548. doi:10.1038/nsmb0608-546

Obado SO, Taylor MC, Wilkinson SR, Bromley EV, Kelly JM (2005) Functional mapping of a *Trypanosome* centromere by chromosome fragmentation identifies a 16-kb GC-rich transcriptional "strand-switch" domain as a major feature. Genome Res 15(1):36–43. doi:10.1101/gr.2895105

Okamura K, Lai EC (2008) Endogenous small interfering RNAs in animals. Nat Rev Mol Cell Biol 9(9):673–678. doi:10.1038/nrm2479

Okamura K, Balla S, Martin R, Liu N, Lai EC (2008a) Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. Nat Struct Mol Biol 15(6):581–590. doi:10.1038/nsmb.1438

Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, Lai EC (2008b) The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. Nature 453(7196):803–806. doi:10.1038/nature07015

Prucca CG, Slavin I, Quiroga R, Elias EV, Rivero FD, Saura A, Carranza PG, Lujan HD (2008) Antigenic variation in *Giardia lamblia* is regulated by RNA interference. Nature 456 (7223):750–754. doi:10.1038/nature07585

Saraiya AA, Wang CC (2008) snoRNA, a novel precursor of microRNA in *Giardia lamblia*. PLoS Pathog 4(11):e1000224. doi:10.1371/journal.ppat.1000224

Saraiya AA, Li W, Wang CC (2011) A microRNA derived from an apparent canonical biogenesis pathway regulates variant surface protein gene expression in *Giardia lamblia*. RNA 17 (12):2152–2164. doi:10.1261/rna.028118.111

Shan SW, Fang L, Shatseva T, Rutnam ZJ, Yang X, Du W, Lu WY, Xuan JW, Deng Z, Yang BB (2013) Mature miR-17-5p and passenger miR-17-3p induce hepatocellular carcinoma by targeting PTEN, GalNT7 and vimentin in different signal pathways. J Cell Sci 126(Pt 6):1517–1530. doi:10.1242/jcs.122895

Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, Belleannee C, Kucukural A, Serra RW, Sun F, Song L, Carone BR, Ricci EP, Li XZ, Fauquier L, Moore MJ, Sullivan R, Mello CC, Garber M, Rando OJ (2016) Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. Science 351(6271):391–396. doi:10.1126/science.aad6780

Sood P, Krek A, Zavolan M, Macino G, Rajewsky N (2006) Cell-type-specific signatures of microRNAs on target mRNA expression. Proc Natl Acad Sci USA 103(8):2746–2751. doi:10.1073/pnas.0511045103

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, Hannon GJ (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature 453(7194):534–538. doi:10.1038/nature06904

Tarver JE, Donoghue PC, Peterson KJ (2012) Do miRNAs have a deep evolutionary history? BioEssays News Rev Mol Cell Dev Biol 34(10):857–866. doi:10.1002/bies.201200055

Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefs Bioinform 14(2):178–192. doi:10.1093/bib/bbs017

Tschudi C, Shi H, Franklin JB, Ullu E (2012a) Small interfering RNA-producing loci in the ancient parasitic eukaryote *Trypanosoma brucei*. BMC Genom 13:427. doi:10.1186/1471-2164-13-427

Tschudi C, Shi H, Franklin JB, Ullu E (2012b) Small interfering RNA-producing loci in the ancient parasitic eukaryote *Trypanosoma brucei*. BMC Genom 13:427. doi:10.1186/1471-2164-13-427

Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, Surani MA, Sakaki Y, Sasaki H (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature 453(7194):539–543. doi:10.1038/nature06908

Wen YZ, Zheng LL, Liao JY, Wang MH, Wei Y, Guo XM, Qu LH, Ayala FJ, Lun ZR (2011) Pseudogene-derived small interference RNAs regulate gene expression in African *Trypanosoma brucei*. Proc Natl Acad Sci USA 108(20):8345–8350. doi:10.1073/pnas.1103894108

Wheeler BM, Heimberg AM, Moy VN, Sperling EA, Holstein TW, Heber S, Peterson KJ (2009) The deep evolution of metazoan microRNAs. Evol Dev 11(1):50–68. doi:10.1111/j.1525-142X.2008.00302.x

Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. Cell 75(5):855–862

Wu C, Arora P (2014) MicroRNA passenger strand: orchestral symphony of paracrine signaling. Circ Cardiovas Genet 7(4):567–568. doi:10.1161/CIRCGENETICS.114.000805

Yang X, Du WW, Li H, Liu F, Khorshidi A, Rutnam ZJ, Yang BB (2013) Both mature miR-17-5p and passenger strand miR-17-3p target TIMP3 and induce prostate tumor growth and invasion. Nucleic Acids Res 41(21):9688–9704. doi:10.1093/nar/gkt680

Zhang Y, Liu XS, Liu QR, Wei LP (2006) Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. Nucleic Acids Res 34(12):3465–3475. doi:10.1093/nar/gkl473

Zhang YQ, Chen DL, Tian HF, Zhang BH, Wen JF (2009) Genome-wide computational identification of microRNAs and their targets in the deep-branching eukaryote *Giardia lamblia*. Comput Biol Chem 33(5):391–396. doi:10.1016/j.compbiolchem.2009.07.013

Zheng LL, Wen YZ, Yang JH, Liao JY, Shao P, Xu H, Zhou H, Wen JZ, Lun ZR, Ayala FJ, Qu LH (2013) Comparative transcriptome analysis of small noncoding RNAs in different stages of *Trypanosoma brucei*. RNA 19(7):863–875. doi:10.1261/rna.035683.112

Zuker M (1989) On finding all suboptimal foldings of an RNA molecule. Science 244(4900): 48–52