Chapter 9
Differential Expression of *Toxoplasma gondii* MicroRNAs in Murine and Human Hosts

Müşerref Duygu Saçar Demirci, Caner Bağcı and Jens Allmer

Abstract MicroRNAs are short RNA sequences involved in post-transcriptional gene regulation. MicroRNAs are known for a wide variety of species ranging from bacteria to plants. It has become clear that some cross-kingdom regulation is possible especially between viruses and their hosts. We hypothesized that intracellular parasites, like *Toxoplasma gondii*, similar to viruses would be able to modulate their host's gene expression. We were able to show that *T. gondii* produces many putative pre-miRNAs which are actually transcribed. Furthermore, some of these expressed pre-miRNAs have a striking resemblance to host mature miRNAs. Previous studies indicated that *T. gondii* infection coincides with increased abundance of some miRNAs. Here we were able to show that many of these miRNAs have close relatives in *T. gondii* which may not be distinguishable using PCR. Taken together, the similarity to host miRNAs, their confirmed expression, and their upregulation during infection, it suggests that *T. gondii* actively transfers miRNAs to regulate its host. We conclude, that this type of cross-kingdom regulation may be possible, but that targeted analysis is necessary to consolidate our computational findings.

9.1 Introduction

Mature microRNAs (miRNAs) are short (18–24 nt) nucleotide sequences which act as a recognition key via base-pairing with their target mRNAs involved in post-transcriptional gene regulation (Erson-Bensan 2014). Mature miRNAs are cut from larger stem-loop structures (Dicer for animals and DCL1 in the nucleus for

M.D. Saçar Demirci · J. Allmer (✉)
Molecular Biology and Genetics, Izmir Institute of Technology, 35430 Urla, Izmir, Turkey
e-mail: jens@allmer.de

C. Bağcı
Biotechnology, Izmir Institute of Technology, 35430 Urla, Izmir, Turkey

J. Allmer
Bionia Incorporated, IZTEKGEB A8, 35430 Urla, Izmir, Turkey

A.L. Leitão and F.J. Enguita (eds.), *Non-coding RNAs and Inter-kingdom Communication*, DOI 10.1007/978-3-319-39496-1_9

plants) and incorporated into protein complexes in the cytoplasm which then perform the regulatory action. The stem-loop structures are cut from pre-miRNAs (hairpins) within the nucleus (Drosha for animals and DCL1 for plants) and then exported into the cytosol. The pre-miRNAs are excised from pri-miRNA during transcription and they may derive from any part of a genome (Kim et al. 2009; Rodriguez et al. 2004) transcribed by either RNA-Polymerase II or III. The existence of miRNAs has been shown for a wide range of organisms ranging from protists (Gottesman 2005) and sponges (Kim et al. 2009) to plants (Xie et al. 2015) and animals (Okamura 2012). Even viruses contain miRNAs (Grey 2015); likely to also influence the host's gene expression (Li et al. 2014; Skalsky and Cullen 2010).

For *Toxoplasma gondii*, it has been shown that miRNAs and the necessary machinery for genesis and targeting exist (Braun et al. 2010) and that it is important during development (Hakimi and Menard 2010). Interestingly, the miRNA genesis pathway of *T. gondii* seems related to plant and fungal ones (Braun et al. 2010) which is striking since the plant miRNA genesis pathway may have evolved independently from the metazoan one (Chapman and Carrington 2007). It is even more curious that the targeting machinery in *T. gondii* seems related to the metazoan one (Braun et al. 2010). Around the same time with these findings, we predicted potential miRNA regulatory networks in *T. gondii* under the assumption of similarity to metazoan miRNAs (Cakir and Allmer 2010). This analysis may have to be redone using a mixture model although it led to the interesting finding that there are perhaps as many hub miRNAs in the *T. gondii* genome as there are major hosts. In summary, it has been established that *T. gondii* uses miRNA for its post transcriptional gene regulation.

Like viruses, obligatory intracellular parasites may benefit from influencing their hosts' gene expression. In this regard, it has been shown that *T. gondii* increases the amount of key miRNAs (miR-17 ~ 92 and miR-106b ~ 25), implicated in numerous hyperproliferative diseases, during infection (Zeiner et al. 2010). Zeiner et al. further demonstrated that this process is tied to *T. gondii* and that at least one other apicomplexan (*Neospora caninum*) does not impact the level of the selected miRNAs. Fast miRNA evolution has been reported before (Liang and Li 2009; Lu et al. 2008) and this seems to be further supported with the study by Zeiner et al. supporting their implicit idea that such regulation may be parasite-host specific. Other studies support this finding and show *T. gondii* dependent dysregulation of miRNAs (Cai et al. 2014; Cannella et al. 2014; Thirugnanam et al. 2013; Xiao et al. 2014).

We pondered how such dysregulation may be caused and hypothesized that a possible path would be for *T. gondii* to export miRNAs into its host. We then analyzed the miRNAs of *T. gondii* and concluded that communication between parasite and host via miRNAs is theoretically possible (Saçar et al. 2014) which is further supported by studies which show that secretion is important for parasite host communication (Boothroyd and Dubremetz 2008; Luder et al. 2009); for a recent

review see (English et al. 2015). Since *T. gondii* is present in a large number of hosts, we wondered whether some putative miRNAs are expressed host-specifically or whether some are highly expressed in multiple hosts. To investigate this, we used all human and mouse miRNAs available on miRBase (Griffiths-Jones et al. 2008) to establish machine learned models and applied them to all hairpins of the *T. gondii* ME49 genome. We acquired all available next generation sequencing datasets where any toxoplasma strain was sequenced within a host cell or singly from SRA (Leinonen et al. 2011) and evaluated the expression profiles of the predicted miRNAs. Clearly, the results can only guide further research since it is necessary to have evidence on the protein level for confident miRNA assignment as described in (Saçar and Allmer 2013c) something which can be achieved using mass spectrometry with quantitation (Allmer 2010). Mass spectrometric experiments can be guided by providing expected proteotypic peptide masses of proteins expected to be miRNA targets using multiple reaction monitoring type experiments (Kondrat et al. 1978).

T. gondii is a unicellular protist and as such is not expected to have many miRNAs with clear human or mouse like features. Yet, we found many such examples (2761066 and 2299419 hairpins passed human and mouse models at a prediction cutoff of 0.99) and they are conserved (5278 and 8190 hairpins showed high similarity to human and mouse mature microRNAs with a matchScore above or equal to 35) or differentially expressed (57789 and 64598 hairpins are differentially expressed in human and murine hosts with an absolute log2 fold change equal to or larger than 1 and a Benjamini Hochberg (BH) adjusted p-value equal to or smaller than 0.01) in mouse and human hosts. Since *T. gondii*'s targeting machinery is metazoan like, there are two modes of directly influencing the host's post transcriptional gene regulation. We suggest that either *T. gondii* transports naked stem-loop structures into the host which are then processed by the host machinery, or it can export loaded TgAgo (Al Riyahi et al. 2006) which may be directly functional in the host.

Host like miRNAs which are highly expressed by *T. gondii* in both hosts may have indispensable function for *T. gondii*; and gene enrichment using Reactome (Croft et al. 2014) indicates that this is very likely true for the miRNAs identified in this study. MicroRNAs that are differentially expressed between hosts may modulate *T. gondii*'s interference with host gene expression to specific hosts. Since *T. gondii* replicates in many hosts, and since it seems to be able to distinguish between primary and secondary hosts, there must be a mechanism allowing this distinction and miRNAs may play a role in communicating it. Both types of identified miRNAs can aid in developing new diagnostic biomarkers and may serve as drug targets (Hoy et al. 2014; Manzano-Roman and Siles-Lucas 2012). Although drugs against *T. gondii* are currently available, the need for novel ones is evident (Blader and Saeij 2009).

9.2 Materials and Methods

9.2.1 *Hairpin Extraction from the* T. gondii *Genome*

The genome assembly of *T. gondii* ME49 was downloaded from ToxoDB (Gajria et al. 2008) Release 25 (http://toxodb.org/common/downloads/release-25/TgondiiME49/fasta/data/). With an in-house script we divided the genome into 500 nt long sequences with 250 nt overlap for both strands. These sequence lengths were chosen since there is no human or murine hairpin in miRBase longer than 250 nt; so all potential pre-miRNAs in the *T. gondii* genome can be completely captured within one 500 nt fragment. The resulting 263964 fragments were then folded using RNAFold (Hofacker 2003). The structure provided by RNAFold was parsed using an in-house script which extracted all hairpin-like structures only requiring a stem with at least three consecutive bonds somewhere in the stem and a terminal loop with at least 3 nucleotides. All extracted hairpins ($\sim$4.8 million) were then examined with human and murine models established via machine learning.

9.2.2 *MicroRNA Feature Calculation*

There are no miRNA examples available on miRBase for *T. gondii* or other Apicomplexa and therefore targeted feature selection is not possible. Since the interest of the present study is on how *T. gondii* may regulate its host gene expression using miRNAs, this is not a problem and the study can draw from a larger body of knowledge describing human miRNA prediction. About a dozen studies performing ab initio pre-miRNA prediction have been published (Batuwita and Palade 2009; Bentwich 2008; Bentwich et al. 2005; Ding et al. 2010; Gao et al. 2012; Gudys et al. 2013; Jiang et al. 2007; Lopes Ide et al. 2014; Ng and Mishra 2007; Ritchie et al. 2012; van der Burgt et al. 2009; Xu et al. 2008; Xue et al. 2005). All features and their natural extensions and normalizations were implemented in JAVA™ and since calculations for the analysis of all putative hairpins extracted from the *T. gondii* genome were time consuming, HTCondor (Thain et al. 2005) was used for parallel computation.

9.2.3 *Machine Learning*

MicroRNA detection is quite involved experimentally and, therefore, computational methods are indispensable (Ng and Mishra 2007). Numerous methods for computational detection of pre-miRNAs have been developed (Allmer 2014; Allmer and Yousef 2012; Saçar et al. 2014). Most approaches are based in machine learning, specifically two-class classification although some use one-class

classification (Yousef et al. 2008, 2015). For successful machine learning, the pre-miRNAs need to be parameterized and many features to describe hairpins have been proposed (Saçar and Allmer 2013b, 2013c). The most important factor for effective machine learning is, however, data with positive and negative examples correctly assigned (Saçar and Allmer 2013b). We took into account problems with available data (Saçar et al. 2013) and class-imbalance (Saçar and Allmer 2013b) while training models using KNIME (Berthold et al. 2009). We previously showed that as little as 50 features may be enough for training a classifier and, therefore, we performed feature selection prior to training of the classifier (Fig. 9.1). We first clustered the about 700 features using *k*-Means clustering into 100 clusters. Then 50 clusters were selected based on calculation speed and information gain and from each cluster, the feature with highest information gain was included into our final feature set (Supplementary Table 2). In previous studies we have not seen a great impact of the classifier on classification efficiency (Saçar and Allmer 2013a) and, therefore, chose to use decision tree as our model since it is very fast and allows serialization of the model in KNIME.

For positive data we used all pre-miRNAs from miRBase for *Homo sapiens* (hsa) and *Mus musculus* (mmu) and as negative data we used the pseudo dataset (Ng and Mishra 2007). For the training of the classifier thousand fold Monte Carlo (MC) cross validation (Xu and Liang 2001) was performed and for each fold randomly sampled 70 % of the positive data for training and 30 % of it for testing. Negative data was randomly sampled from the pseudo dataset such the number of examples matched the number of positive examples. The models trained for hsa and mmu were then applied to all hairpins that could be extracted from the *T. gondii*

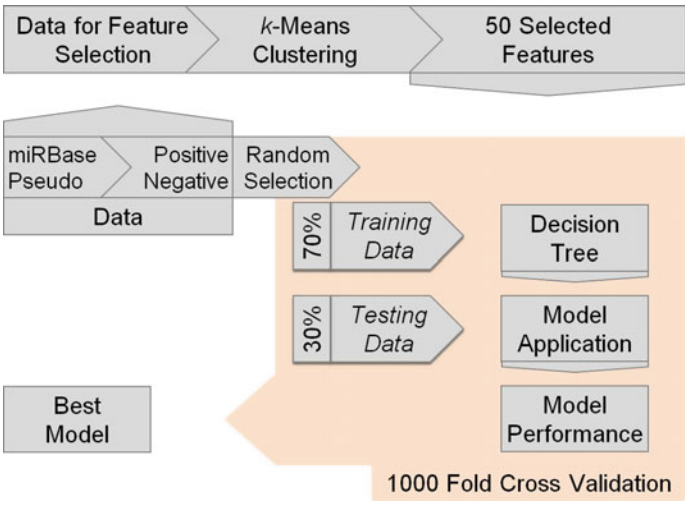

Fig. 9.1 Machine learning regime for establishing models that describe pre-miRNAs for different organisms. Feature selection was performed on a per species basis

ME49 genome assembly. Commonly used performance measures like specificity, sensitivity and accuracy were calculated.

9.2.4 MicroRNA Expression Analysis

All *T. gondii* sequencing datasets (74) available on the small sequencing read archive (Leinonen et al. 2011) were downloaded and analyzed using FastQC (Andrews 2010). 14 datasets were finally accepted for further analysis (Supplementary Table 1). Supplementary Table 1 also details why some of the datasets did not fit the present study and why they were excluded with the major reasons being low base-calling quality and overrepresentation of a dataset. The datasets were from parasite cultures (3), human foreskin fibroblast cells (3), murine bone-marrow macrophages (3) and mouse peritoneal exudate (5).

All accepted datasets were trimmed from adapter contaminations using an in-house script to determine adapter sequences and cut adapt (Martin 2010) in default settings to remove them. Remaining sequences were further trimmed from low-quality and ambiguous base callings with sickle (http://omictools.com/sickle-s714.html) using a quality threshold of 30 (implying a 0.001 base-calling error probability). Reads shorter than 20 nt after the two step trimming processes were discarded (Supplementary Table 1).

The ToxoDB Release 25 genome was indexed using Bowtie 2 (Langmead and Salzberg 2012) and trimmed reads were aligned to the genome using Tophat 2 (Kim et al. 2013) with default settings which include a maximum edit distance of 2 and take splicing into account. Reads that were ambiguously mapped to multiple locations on the genome and reads that did not map in respect to proper pairs for the paired-end sequencing datasets (mate pairs mapped on different chromosomes or wrong orientation) were excluded from further analysis.

Read counts were established using featureCounts (Liao et al. 2014) from the Subread package (http://subread.sourceforge.net) for both annotated *Toxoplasma gondii* genes from ToxoDB Release 25 and for the predicted hairpins that were mapped back to the genome by BLAST (Altschul and Gish 1996). All blast hits were recorded that pass two criteria: (1) predicted hairpin being longer than 40 nucleotides and (2) the hairpin matching to the genome perfectly. Reads spanning two or more genes/hairpins were assigned to the gene/hairpin with the largest overlap. For paired-end sequencing, fragments (read pairs) were counted instead of reads.

We previously established a similarity measure to measure the similarity of mature sequences among selected microRNAs and predicted hairpins (Saçar et al. 2014). Aligning them using BLAST with ‘blastn-short‘ option and disallowing any mismatches from negative strand of both the mature sequences and the hairpins. The resulting matchScore was assigned to each hit using the following formula:

$$matchScore + (SM \cdot 5) + M - (SMM \cdot 5) - MM$$

where SM is the number of matches in the seed region (first 8 nucleotides), M is the number of matches in the non-seed region, SMM is the number of mismatches and gaps in the seed region, MM is the number of mismatches and gaps in the non-seed region. Some factors were tested for weighting of the seed region and 5 was arbitrarily chosen from a number of promising candidates.

9.3 Results and Discussion

9.3.1 MicroRNA Detection Model Training

For human and murine miRNA detection models were trained. 50 selected features (Supplementary Table 2) were calculated for positive and negative data and 1000-fold MC cross validation was performed. Accuracy for both models was always above 75 % and never reached 90 % (Fig. 9.2). The median accuracy for human was slightly above 83 % and thus about 1 % better than the murine one which was about 82 %. Both range and interquartile range were slightly larger for the generated murine model.

The models selected for miRNA detection for human and murine achieved an accuracy of 0.875 and 0.865, respectively. This accuracy is slightly better than the best accuracy achieved in our previous study (Saçar et al. 2014), but expectedly lower when compared to accuracies achieved earlier using feature sets from

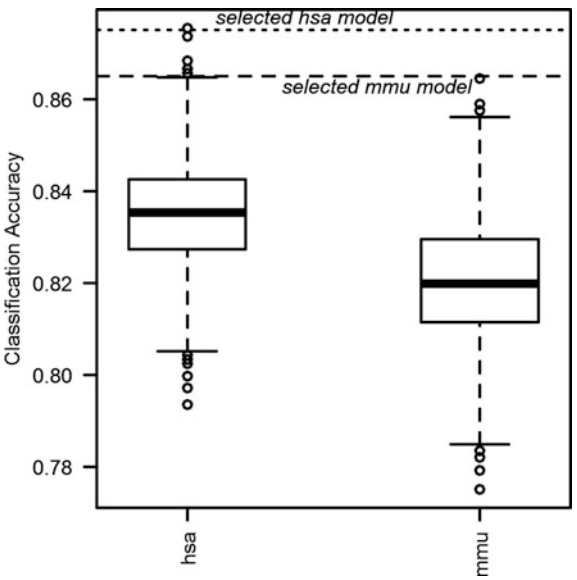

Fig. 9.2 Model accuracy for 1000-fold Monte Carlo cross validation during training of the human and murine classifiers. The selected human model (0.875) and the selected murine model (0.865) are displayed as *dotted* and *dashed lines*, respectively

ab initio studies trained with randomly generated negative data (Saçar and Allmer 2013b). Thus these models seemed applicable for the analysis of putative hairpins extracted from the *T. gondii* genome.

9.3.2 Hairpin Extraction and Analysis

Both strands of the genome were fragmented into 500 nt pieces with 250 nt overlap. The secondary structures for all fragments were predicted using RNAFold. About 2.4 million hairpin-like structures were found in both strands and two examples are given in Fig. 9.3.

Only hairpins with a minimum of three consecutive bonds in the stem (Fig. 9.3; A: 1-7) were extracted from the secondary structures of the folded 500-mers. The murine and human machine learned models were applied to all ~4.8 million hairpins extracted from the genome (Fig. 9.4). While forward and reverse strands show similar prediction score distributions for hsa and mmu, hsa and murine display distinct prediction score distributions. For all further analyses, we determined a prediction score threshold at 0.99 prediction score (Fig. 9.4).

Of the ~4.8 million extracted hairpins approximately 2.3 and 2.8 million passed the models at 99 % prediction score cutoff for murine and human, respectively. In our previous study (Saçar et al. 2014), far less miRNAs passed the threshold which may occur because we didn't use hairpin length or stem length cutoffs. Human and murine predictions for the forward strand share about half of the predicted hairpins at a cutoff of 99 % model score (Fig. 9.5) and the same is true for the reverse strand (data not shown).

Clearly, not all of these hairpins can represent true pre-miRNAs. The number may also be inflated by a factor of up to two due to the creation of overlapping

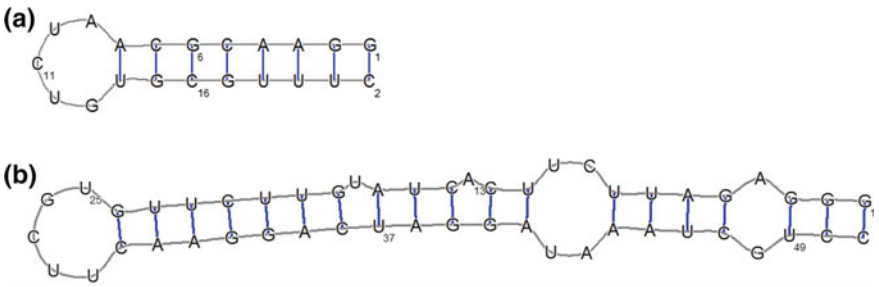

Fig. 9.3 The first two examples for stemloop structures from the forward strand of the *T. gondii* genome. The structures were drawn with RNAShapes (Steffen et al. 2006). Structure A is too short to be considered for further analysis whereas B fulfills expectations for further analysis

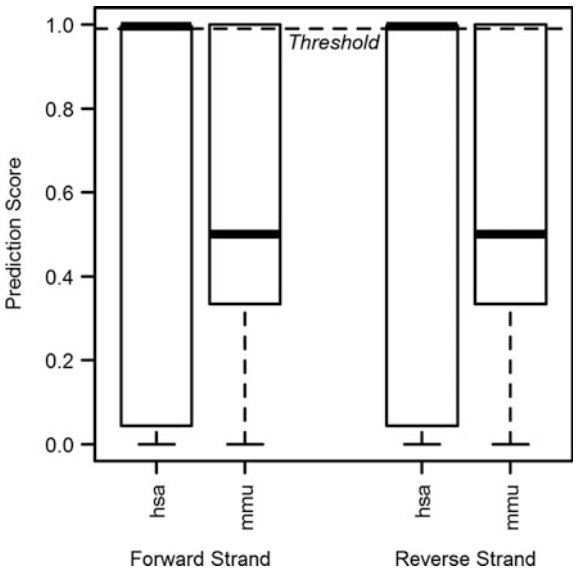

Fig. 9.4 Prediction score distribution for application of human (hsa) and mouse (mmu) models to putative hairpins extracted from the *T. gondii* genome. The 4.8 million hairpins were randomly sampled down to 100.000 to enable plotting. The *dashed line* indicates the 99 % score cutoff

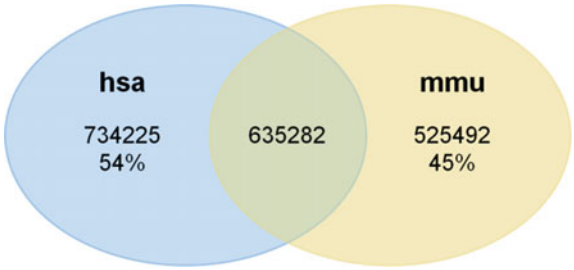

Fig. 9.5 Number of hairpins (coding strand only) passing the human (hsa) and mouse (mmu) models at a score cutoff of 0.99. The percentages are in respect to the organism i.e.: 46 % of predicted *T. gondii* hairpins that pass the hsa model also pass the mmu model whereas, when viewed from the mmu perspective it amounts to 55 %

fragments which was not accounted for, since folds are quite different and hairpins that are found at the same locus may still be quite different in consecutive overlapping fragments. Another convoluting factor is that the models achieved an accuracy of around 87 % which means that at least 13 % of hairpins that pass these models at a cutoff of 99 % are still wrongly predicted. A good way to filter such a large number of results is to check whether they are actually expressed.

9.3.3 Hairpin Expression Analysis

Many parts of a genome can be folded into structures strongly resembling pre-miRNAs. However, only if the hairpins are expressed can they perform any function and therefore expression was analyzed using RPKM (Mortazavi et al. 2008) as the measure (Fig. 9.6). We further analyzed whether there is a relationship between prediction score and hairpin expression abundance but were not able to find any (Fig. 9.1 in Supplementary File 1).

Most of the predicted hairpins were not expressed in at least one sample and, therefore, were not further considered. As can be expected, with increasing RPKM less hairpins fall into the ranges (Fig. 9.6). It is of note, that some of the expression values of the predicted hairpins for *T. gondii* are above the median expression of hairpins (hsa) provided in miRBase (~1000 RPKM).

Even more instructive than the pure expression of a predicted hairpin is whether it is shared among multiple organisms (Fig. 9.7). With the exception of Kunming mice which share only few highly expressed predicted pre-miRNAs with other organisms (perhaps due to the sequencing strategy), the larger part of highly expressed pre-miRNAs (~60 %) are shared among species (Fig. 9.7). 4071 predicted *T. gondii* pre-miRNAs are shared among the murine, human, and parasite

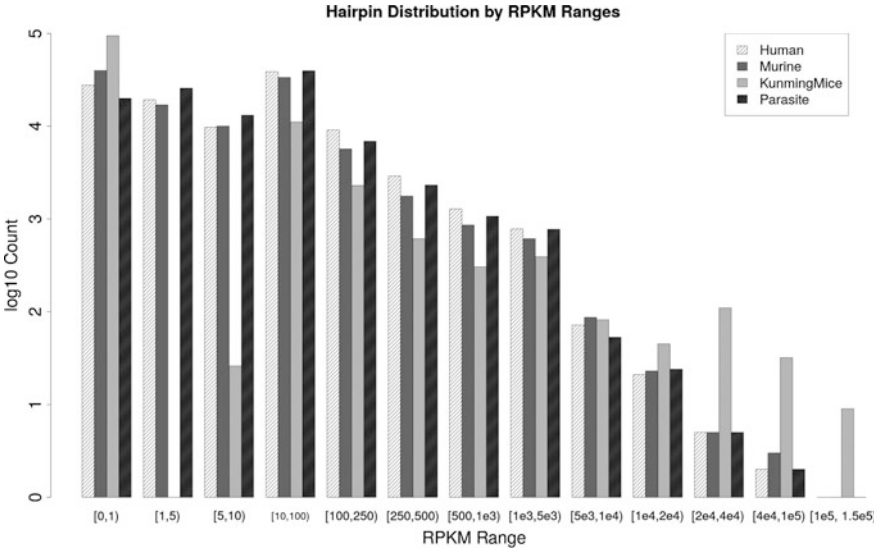

Fig. 9.6 Only hairpins that were expressed in at least one sample were considered for expression analysis. The predicted *T. gondii* hairpins were grouped by their RPKM and the counts are provided for the samples (pooled by species). The groups were chosen arbitrarily such that a small number of groups can capture the complete range of RPKM values

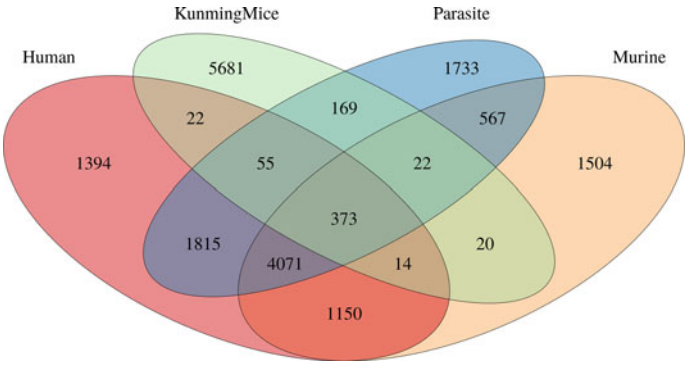

Fig. 9.7 Predicted *T. gondii* hairpins that are expressed above 1000 RPKM in the samples and how they are shared among species that were investigated

samples. Murine and parasite share an additional 5033, human and parasite 6314, and human and murine 5608 of the highly expressed predicted *T. gondii* pre-miRNAs. The number of predicted hairpins which are uniquely expressed in one of the species is, therefore, much lower than the miRNAs whose expression is shared among samples from different species (human: 1394, murine: 1504, and parasite: 1733). MicroRNAs modulate gene expression and this modulation can be systemic or specific to the environment. It seems likely that the predicted hairpins perform a fundamental function whereas the ones that are highly expressed in only one or two samples perform a more specific function.

The shared predicted hairpins discussed above are expressed at similar rates as hairpins typically presented on miRBase (http://www.mirbase.org/cgi-bin/miRNA_summary.pl?org=hsa) which for human range between 0.1 and 135000 with a median of 97 and an average of 1100 RPM. Unfortunately, these values cannot be translated into absolute concentrations which would be essential for drawing conclusions about the abundance of miRNAs in the different samples.

It is of interest whether some of the miRNAs previously found to be dysregulated in human or mouse hosts are expressed in *T. gondii*. For this we matched the predicted hairpins to human and mouse mature miRNAs using the matchScore with a threshold of 35. The hairpins were further required to be expressed in at least one sample.

It has been shown that miR-17 ~ 92, miR-106b ~ 25, and miR-106a ~ 363 increase in abundance within the host upon infection with *T. gondii* (Cai et al. 2014; Zeiner et al. 2010). Interestingly, predicted *T. gondii* hairpins similar to miR-18a, miR-20b, miR-92, and miR-363 were expressed in the human and murine samples (Supplementary Table 3). Another study showed that mmu-miR-712-3p, mmu-miR-511-5p and mmu-miR-217-5p are indicative for *T. gondii* infection in mice. Very similar hairpins to the former two miRNAs are also predicted for *T. gondii* in this study and they are expressed at relatively high levels (Supplementary Table 3). Xiao et al. reported the upregulation of miR-30c-1, miR-125b-2,

miR-132, mir-23b, and miR-17 ~ 92 (Xiao et al. 2014). Except for miR125-b-2 and miR-132 we also found similar *T. gondii* hairpins expressed in human and murine samples (Supplementary Table 3). Wang et al. analyzed *T. gondii* in respect to its miRNAs and found 17 conserved miRNAs 2 of which were related to metazoan miRNA families (Wang et al. 2012). Here, similar *T. gondii* pre-miRNAs to mmu-miR-466i-5p, mmu-miR-574-5p, and has-miR-574-5p are also found. Toxoplasma persistence in human brain seems to be associated with miR-146a, miR-155, and miR-1246 (Cannella et al. 2014). While we did not find miR-146a, we found similar *T. gondii* hairpins to miR-155 and miR-1246 to be expressed in human and/or murine samples (Supplementary Table 3). In summary, for many of the upregulated miRNAs indicative of toxoplasmosis there exist highly similar hairpins in *T. gondii*. Often PCR is performed to validate the existence of miRNAs in the referenced studies, but it is not clear whether the host miRNAs can be distinguished from the very similar *T. gondii* hairpins.

The targets of the miRNAs which have previously been reported to be upregulated in the host were submitted to Reactome analysis (D'Eustachio 2011). For all predicted *T. gondii* miRNAs similar to the previously reported miRNAs (both human and murine) that were expressed in both human and murine samples the gene targets are always signifying increased metabolism and infection (Supplementary Table 4).

9.3.4 Known Toxoplasma Gondii *Pre-microRNAs*

A recent study identified 339 novel miRNAs in Toxoplasma and compared expression between RH and ME49 strains (Wang et al. 2012). In order to check whether the miRNAs were also among the hairpins identified in this study, we acquired their data. Only perfect complete matches of their mature miRNA sequences to our putative pre-miRNAs were accepted using blastn-short (Supplementary Table 5). 48 of the mature sequence were not part of our predicted hairpins. 43 out of these were found in the sequence assembly we used, but were not part of hairpins that passed the human or mouse model. Finally, the remaining 4 mature sequences were not found in the *T. gondii* genome assembly we used in this study. Interestingly, the average expression of matches in their data is ~2080 and ~2270 while the average expression for the reported mature sequences without counterpart in our data is 1.4 and 4.3. All matching sequences also pass the human model at a prediction score cutoff of 0.99 which indicates that the identified miRNAs by Wang and colleagues (Wang et al. 2012) are very similar to human type pre-miRNAs and thus confirms the claim that the miRNAs are metazoan like (Braun et al. 2010).

9.4 Conclusion

MicroRNAs in *T. gondii* have metazoan like features in respect to targeting (Braun et al. 2010). Therefore, predictive models based on human and murine miRNAs seem to be applicable and the models trained here achieved around 87 % accuracy (Fig. 9.2). Due to the similarity and the fact that transfer from *T. gondii* to the host cell is possible, it can be reasoned that *T. gondii* hairpins may perform regulatory function in the host. Some of the predicted hairpins are expressed in various hosts and of these a large number is shared among them. Additionally, some predicted hairpins are highly similar to host miRNAs and are relatively abundant. Many of the miRNAs that have previously been shown to be upregulated in diseases have highly expressed close *T. gondii* homologs. Together, these clues are suggestive, but not fully conclusive, for the regulatory mechanism where *T. gondii* actively transfers miRNAs into the host to modulate gene expression.

T. gondii miRNAs, different from host miRNAs, which are highly expressed in all hosts would be good disease marker candidates. Differentially expressed miRNAs that are similar to host miRNAs may provide therapeutic leads for treatment of Toxoplasmosis. Unfortunately, sequencing data was only available for human and murine hosts. We hope that in the future we will be able to acquire data from more hosts; especially from cats.

Acknowledgments This work was supported by the Scientific and Technological Research Council of Turkey (Grant No. 113E326) awarded to JA.

Supplementary Materials Supplementary material is available at the following URL: http://bioinformatics.iyte.edu.tr/supplements/ncRNA2016

References

Al Riyahi A, Al-Anouti F, Al-Rayes M, Ananvoranich S (2006) Single Argonaute protein from *Toxoplasma gondii* is involved in the double-stranded RNA induced gene silencing. Int J Parasitol 36(9):1003–1014. doi:10.1016/j.ijpara.2006.04.014

Allmer J (2010) Label-free quantitation, an extension to 2DB. Amino Acids 38(4):1075–1087. doi:10.1007/s00726-009-0317-9

Allmer J (2014) Computational and bioinformatics methods for microRNA gene prediction. Methods Mol Biol 1107:157–175. doi:10.1007/978-1-62703-748-8_9

Allmer J, Yousef M (2012) Computational methods for ab initio detection of microRNAs. Front Genet 3:209. doi:10.3389/fgene.2012.00209

Altschul SF, Gish W (1996) Local alignment statistics. Methods Enzymol 266:460–480

Andrews S (2010) FASTQC: a quality control tool for high throughput sequence data. Babraham. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Batuwita R, Palade V (2009) microPred: effective classification of pre-miRNAs for human miRNA gene prediction. Bioinformatics 25(8):989–995. doi:10.1093/bioinformatics/btp107

Bentwich I (2008) Identifying human microRNAs. Curr Top Microbiol Immunol 320:257–269

Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E, Sharon E, Spector Y, Bentwich Z (2005) Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet 37(7):766–770. doi:10.1038/ng1590

Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, Ohl P, Thiel K, Wiswedel B (2009) KNIME—The Konstanz Information Miner

Blader IJ, Saeij JP (2009) Communication between *Toxoplasma gondii* and its host: impact on parasite growth, development, immune evasion, and virulence. APMIS 117(5–6):458–476. doi:10.1111/j.1600-0463.2009.02453.x

Boothroyd JC, Dubremetz JF (2008) Kiss and spit: the dual roles of *Toxoplasma* rhoptries. Nat Rev Microbiol 6(1):79–88. doi:10.1038/nrmicro1800

Braun L, Cannella D, Ortet P, Barakat M, Sautel CF, Kieffer S, Garin J, Bastien O, Voinnet O, Hakimi MA (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(5):e1000920. doi:10.1371/journal.ppat.1000920

Cai Y, Chen H, Mo X, Tang Y, Xu X, Zhang A, Lun Z, Lu F, Wang Y, Shen J (2014) *Toxoplasma gondii* inhibits apoptosis via a novel STAT3-miR-17-92-Bim pathway in macrophages. Cell Signal 26(6):1204–1212. doi:10.1016/j.cellsig.2014.02.013

Cakir MV, Allmer J (2010) Systematic computational analysis of potential RNAi regulation in *Toxoplasma gondii*. In: 2010 5th international symposium on health informatics and bioinformatics. IEEE, Ankara, Turkey, pp 31–38

Cannella D, Brenier-Pinchart MP, Braun L, van Rooyen JM, Bougdour A, Bastien O, Behnke MS, Curt RL, Curt A, Saeij JP, Sibley LD, Pelloux H, Hakimi MA (2014) miR-146a and miR-155 delineate a MicroRNA fingerprint associated with *Toxoplasma* persistence in the host brain. Cell Rep 6(5):928–937. doi:10.1016/j.celrep.2014.02.002

Chapman EJ, Carrington JC (2007) Specialization and evolution of endogenous small RNA pathways. Nat Rev Genet 8(11):884–896. doi:10.1038/nrg2179

Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR, Jassal B, Jupe S, Matthews L, May B, Palatnik S, Rothfels K, Shamovsky V, Song H, Williams M, Birney E, Hermjakob H, Stein L, D'Eustachio P (2014) The Reactome pathway knowledgebase. Nucleic Acids Res 42(Database issue):D472–477. doi:10.1093/nar/gkt1102

D'Eustachio P (2011) Reactome knowledgebase of human biological pathways and processes. Methods Mol Biol 694:49–61. doi:10.1007/978-1-60761-977-2_4

Ding J, Zhou S, Guan J (2010) MiRenSVM: towards better prediction of microRNA precursors using an ensemble SVM classifier with multi-loop features. BMC Bioinformatics 11(Suppl 11): S11. doi:10.1186/1471-2105-11-S11-S11

English ED, Adomako-Ankomah Y, Boyle JP (2015) Secreted effectors in *Toxoplasma gondii* and related species: determinants of host range and pathogenesis? Parasite Immunol 37 (3):127–140. doi:10.1111/pim.12166

Erson-Bensan AE (2014) Introduction to microRNAs in biological systems. Methods Mol Biol 1107:1–14. doi:10.1007/978-1-62703-748-8_1

Gajria B, Bahl A, Brestelli J, Dommer J, Fischer S, Gao X, Heiges M, Iodice J, Kissinger JC, Mackey AJ, Pinney DF, Roos DS, Stoeckert CJ, Jr., Wang H, Brunk BP (2008) ToxoDB: an integrated *Toxoplasma gondii* database resource. Nucleic Acids Res 36(Database issue):D553–556. doi:10.1093/nar/gkm981

Gao Z, Luo X, Shi T, Cai B, Zhang Z, Cheng Z, Zhuang W (2012) Identification and validation of potential conserved microRNAs and their targets in peach (Prunus persica). Mol Cells 34 (3):239–249. doi:10.1007/s10059-012-0004-7

Gottesman S (2005) Micros for microbes: non-coding regulatory RNAs in bacteria. Trends Genet 21:399–404. doi:10.1016/j.tig.2005.05.008

Grey F (2015) Role of microRNAs in herpesvirus latency and persistence. J Gen Virol 96(Pt 4):739–751. doi:10.1099/vir.0.070862-0

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36(Database issue):D154–158. doi:10.1093/nar/gkm952

Gudys A, Szczesniak MW, Sikora M, Makalowska I (2013) HuntMi: an efficient and taxon-specific approach in pre-miRNA identification. BMC Bioinform 14:83. doi:10.1186/1471-2105-14-83

Hakimi MA, Menard R (2010) Do apicomplexan parasites hijack the host cell microRNA pathway for their intracellular development? F1000 Biol Rep 2. doi:10.3410/B2-42

Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31:3429–3431

Hoy AM, Lundie RJ, Ivens A, Quintana JF, Nausch N, Forster T, Jones F, Kabatereine NB, Dunne DW, Mutapi F, Macdonald AS, Buck AH (2014) Parasite-derived microRNAs in host serum as novel biomarkers of helminth infection. PLoS Negl Trop Dis 8 (2):e2701. doi:10.1371/journal.pntd.0002701

Jiang P, Wu H, Wang W, Ma W, Sun X, Lu Z (2007) MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. Nucleic Acids Res 35(Web Server issue):W339–344. doi:10.1093/nar/gkm368

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14(4):R36. doi:10.1186/gb-2013-14-4-r36

Kim VN, Han J, Siomi MC (2009) Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol 10(2):126–139. doi:10.1038/nrm2632

Kondrat RW, McClusky GA, Cooks RG (1978) Multiple reaction monitoring in mass spectrometry/mass spectrometry for direct analysis of complex mixtures. Anal Chem 50:2017–2012. doi:10.1021/ac50036a020

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9 (4):357–359. doi:10.1038/nmeth.1923

Leinonen R, Sugawara H, Shumway M (2011) The sequence read archive. Nucleic Acids Res 39 (Database issue):D19–21. doi:10.1093/nar/gkq1019

Li C, Hu J, Hao J, Zhao B, Wu B, Sun L, Peng S, Gao GF, Meng S (2014) Competitive virus and host RNAs: the interplay of a hidden virus and host interaction. Protein Cell 5(5):348–356. doi:10.1007/s13238-014-0039-y

Liang H, Li WH (2009) Lowly expressed human microRNA genes evolve rapidly. Mol Biol Evol 26(6):1195–1198. doi:10.1093/molbev/msp053

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30(7):923–930. doi:10.1093/bioinformatics/btt656

Lopes Ide O, Schliep A, de Carvalho AC (2014) The discriminant power of RNA features for pre-miRNA recognition. BMC Bioinformatics 15:124. doi:10.1186/1471-2105-15-124

Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu CI (2008) The birth and death of microRNA genes in *Drosophila*. Nat Genet 40(3):351–355. doi:10.1038/ng.73

Luder CG, Stanway RR, Chaussepied M, Langsley G, Heussler VT (2009) Intracellular survival of apicomplexan parasites and host cell modification. Int J Parasitol 39(2):163–173. doi:10.1016/j.ijpara.2008.09.013

Manzano-Roman R, Siles-Lucas M (2012) MicroRNAs in parasitic diseases: potential for diagnosis and targeting. Mol Biochem Parasitol 186(2):81–86. doi:10.1016/j.molbiopara.2012.10.001

Martin M (2010) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5(7):621–628. doi:10.1038/nmeth.1226

Ng KL, Mishra SK (2007) De novo SVM classification of precursor microRNAs from genomic pseudo hairpins using global and intrinsic folding measures. Bioinformatics 23(11):1321–1330. doi:10.1093/bioinformatics/btm026

Okamura K (2012) Diversity of animal small RNA pathways and their biological utility. Wiley Interdiscip Rev RNA 3(3):351–368. doi:10.1002/wrna.113

Ritchie W, Gao D, Rasko JE (2012) Defining and providing robust controls for microRNA prediction. Bioinformatics 28(8):1058–1061. doi:10.1093/bioinformatics/bts114

Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A (2004) Identification of mammalian microRNA host genes and transcription units. Genome Res 14(10A):1902–1910. doi:10.1101/gr.2722704

Saçar MD, Allmer J (2013a) Comparison of four Ab initio MicroRNA prediction tools. In: Proceedings of the international conference on bioinformatics models, methods and algorithms. SciTePress—Science and and Technology Publications, Barcelona, Spain, pp 190–195

Saçar MD, Allmer J (2013b) Current limitations for computational analysis of miRNAs in cancer. Pakistan J Clin Biomed Res 1:3–5

Saçar MD, Allmer J (2013c) Data mining for microrna gene prediction: On the impact of class imbalance and feature number for microRNA gene prediction. In: 2013 8th international symposium on health informatics and bioinformatics. IEEE, Ankara, Turkey

Saçar MD, Hamzeiy H, Allmer J (2013) Can MiRBase provide positive data for machine learning for the detection of MiRNA hairpins? J Integr Bioinform 10(2):215. doi:10.2390/biecoll-jib-2013-215

Saçar MD, Bagci C, Allmer J (2014) Computational prediction of microRNAs from *Toxoplasma gondii* potentially regulating the hosts' gene expression. Genomics Proteomics Bioinformatics 12(5):228–238. doi:10.1016/j.gpb.2014.09.002

Skalsky RL, Cullen BR (2010) Viruses, microRNAs, and host interactions. Annu Rev Microbiol 64:123–141. doi:10.1146/annurev.micro.112408.134243

Steffen P, Voss B, Rehmsmeier M, Reeder J, Giegerich R (2006) RNAshapes: an integrated RNA analysis package based on abstract shapes. Bioinformatics 22:500–3. doi:10.1093/bioinformatics/btk010

Thain D, Tannenbaum T, Livny M (2005) Distributed computing in practice: the condor experience. Concurr Comput Pract Exp 17:2–4

Thirugnanam S, Rout N, Gnanasekar M (2013) Possible role of *Toxoplasma gondii* in brain cancer through modulation of host microRNAs. Infect Agent Cancer 8(1):8. doi:10.1186/1750-9378-8-8

van der Burgt A, Fiers MW, Nap JP, van Ham RC (2009) In silico miRNA prediction in metazoan genomes: balancing between sensitivity and specificity. BMC Genomics 10:204. doi:10.1186/1471-2164-10-204

Wang J, Liu X, Jia B, Lu H, Peng S, Piao X, Hou N, Cai P, Yin J, Jiang N, Chen Q (2012) A comparative study of small RNAs in *Toxoplasma gondii* of distinct genotypes. Parasit Vectors 5:186. doi:10.1186/1756-3305-5-186

Xiao J, Li Y, Prandovszky E, Karuppagounder SS, Talbot CC, Jr., Dawson VL, Dawson TM, Yolken RH (2014) MicroRNA-132 dysregulation in *Toxoplasma gondii* infection has implications for dopamine signaling pathway. Neuroscience 268:128–138. doi:10.1016/j.neuroscience.2014.03.015

Xie M, Zhang S, Yu B (2015) microRNA biogenesis, degradation and activity in plants. Cell Mol Life Sci 72 (1):87–99. doi:10.1007/s00018-014-1728-7

Xu Q-S, Liang Y-Z (2001) Monte Carlo cross validation. Chemom Intell Lab Syst 56:1–11. doi:10.1016/S0169-7439(00)00122-2

Xu Y, Zhou X, Zhang W (2008) MicroRNA prediction with a novel ranking algorithm based on random walks. Bioinformatics 24 (13):i50–58. doi:10.1093/bioinformatics/btn175

Xue C, Li F, He T, Liu GP, Li Y, Zhang X (2005) Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. BMC Bioinf 6:310. doi:10.1186/1471-2105-6-310

Yousef M, Jung S, Showe LC, Showe MK (2008) Learning from positive examples when the negative class is undetermined–microRNA gene identification. Algorithms Mol Biol 3:2. doi:10.1186/1748-7188-3-2

Yousef M, Allmer J, Khalifa W (2015) Sequence motif-based one-class classifiers can achieve comparable accuracy to two-class learners for plant microRNA detection. J Biomed Sci Eng 8:684–694. doi:10.4236/jbise.2015.810065

Zeiner GM, Norman KL, Thomson JM, Hammond SM, Boothroyd JC (2010) Toxoplasma gondii infection specifically increases the levels of key host microRNAs. PLoS One 5(1):e8742. doi:10.1371/journal.pone.0008742