Entropy-Based Model for MiRNA Isoform Analysis

Shengqin Wang[1,2], Jing Tu[1], Lei Wang[1], Zuhong Lu[1,3]*

1 State Key Lab of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing, 210096, China, **2** Zhejiang Provincial Key Laboratory for Water Environment and Marine Biological Resources Protection, College of Life and Environmental Science, Wenzhou University, Wenzhou, 325035, China, **3** Department of Biomedical Engineering, College of Engineering, Peking University, Beijing, 100781, China

* zhlu@seu.edu.cn

Abstract

MiRNAs have been widely studied due to their important post-transcriptional regulatory roles in gene expression. Many reports have demonstrated the evidence of miRNA isoform products (isomiRs) in high-throughput small RNA sequencing data. However, the biological function involved in these molecules is still not well investigated. Here, we developed a Shannon entropy-based model to estimate isomiR expression profiles of high-throughput small RNA sequencing data extracted from miRBase webserver. By using the Kolmogorov-Smirnov statistical test (KS test), we demonstrated that the 5p and 3p miRNAs present more variants than the single arm miRNAs. We also found that the isomiR variant, except the 3' isomiR variant, is strongly correlated with Minimum Free Energy (MFE) of pre-miRNA, suggesting the intrinsic feature of pre-miRNA should be one of the important factors for the miRNA regulation. The functional enrichment analysis showed that the miRNAs with high variation, particularly the 5' end variation, are enriched in a set of critical functions, supporting these molecules should not be randomly produced. Our results provide a probabilistic framework for miRNA isoforms analysis, and give functional insights into pre-miRNA processing.

Introduction

MiRNAs are ~22 nt endogenous small non-coding RNAs, mediating the translation repression or trigger degradation by paring with target mRNAs in post-translational regulation to control gene expression [1,2]. Advances in next-generation sequencing (NGS) technology are giving rise to a fast accumulation of known miRNAs. In the lasted miRBase version, the human genome encodes for over 1,500 miRNAs [3].

Typically, a mature miRNA commences from the genome as a primary miRNA transcript (pri-miRNA) via RNA polymerase II-mediated transcription. Together with DGCR8, the nuclear RNase III-type protein Drosha cleaves the pri-miRNA to release the precursor miRNA (pre-miRNA), a hairpin-like secondary structure. With the exportin 5-dependent pathway, the

pre-miRNA is then exported to the cytoplasm, where it is processed into a short double-stranded RNA (dsRNA) duplex by the enzyme Dicer [4,5]. One or both strands of the duplex may serve as the functional mature miRNA, and anneal to target mRNA that have complementary target sequence with the guide of the RNA-induced silencing complex (RISC) [5,6]. The imprecise precursor cropping or dicing can change the Drosha and Dicer cleavage sites and generate miRNA isoform products, which make variations in their 5' and/or 3' end positions compared with canonical miRNAs [7].

Many high-throughput small RNA sequencing projects have demonstrated the existence of isomiR variants [8–11]. The frequency of variations at same sites is seen repeatedly and unlikely attribute to degradation or sequencing error, and some of them have been proved to play an important biological role in the control of miRNA-mediated gene expression [12–16]. Variant in the 5' end position of miRNA is supposed to alter the seed region, which is supposed to be very important for target recognition [17–19], thereby reshuffling the target region and affecting the related biological pathway [20–22]. And adding specific nucleotides to the 3' end can modify the stability of miRNA and/or the efficiency of target repression[23–25].

To our knowledge, the isomiR profile can be attributable to three main factors: Drosha and Dicer cleavage, nucleotide addition, and nucleotide substitution. The template nucleotide addition can be the result of the imprecise cleavage by Drosha and Dicer, which has been reported to be more frequent than the non-template nucleotide addition [26,27]. The non-template nucleotide addition can be originated in nucleotide addition [23] or nucleotide substitution by post-transcriptional modifications [28]. Most of non-template nucleotide additions are located at 3' end of miRNAs, and the frequency of them is quite low based on the pervious transcriptome data analysis [29].

Despite the distribution of isomiRs is unlikely to be random, the biological relevance of these molecules has been overlooked in previous studies[7]. Here, we developed a Shannon entropy-based model to measure the isomiR expression profiles from high-throughput small RNA sequencing data, and to find the candidate functional role of these molecules.

Materials and Methods

Data sources

We fetched the high-throughput small RNA sequencing data for multiple alignment format employed in miRBase webserver [3], including 81 Homo sapiens related experiments collected from five recently published papers [30–34]. These experiments included miRNAs from different developmental stages of different tissues and cell lines, and the multiple alignment data pooled these miRNAs together. Corresponding pre-miRNAs and their Minimum Free Energy (MFE) information were also retrieved. Since too few sequences will result in a systematic underestimation of isomiR variants, as well as too many sequences may be contributed by PCR amplification bias, our analysis only included miRNAs with number of sequences more than 50 and less than 10000 (S1 Fig.).

Shannon entropy calculation

At first, we defined the isomiRs as sequences that matched the known pre-miRNA in the canonical mature miRNA region ± 4 nt. To characterize the isomiR variants of a given miRNA, in the case of multiple alignments of miRNA sequences, we defined *MIH*, MiRNA Isoform entropy (abbreviated H), as the average of Shannon entropies of the observed symbol distribution

Fig 1. Region selection for *MIH* calculation. Sequences located in the dotted frame (canonical mature miRNA region ± 4 nt) are accepted for *MIH* calculation. Bases located in the left rectangle are used for *MIH*5 calculation, and those located in the right rectangle are used for *MIH*3 calculation (including the gap character "-").

doi:10.1371/journal.pone.0118856.g001

for each site in the related region:

$$MIH = -\frac{1}{L}\sum_{i}^{L}\sum_{j}^{N}p_j^i log_2 p_j^i \tag{1}$$

Here, p_j^i is the observed frequency which calculated as the frequency of the character j at particular sequence position i divided by the number of sequences in the alignment, and N is the number of distinct symbols for the given sequence type (five for RNA: A, U, G, C, and gap character "-"). For *MIH* calculation, L is equivalent to the length of canonical mature miRNA + 8, including the upstream 4 nt and downstream 4 nt. Since the variations differ greatly between 5' ends and the 3' ends, we calculated *MIH*5 for 5' end variations and *MIH*3 for 3' end variations, respectively. For both of them, L is equal to 9, including the end site and flanking region (± 4 nt). For example, as shown in Fig. 1, the *MIH* of has-miR-1 is defined as the average of entropies of the symbol distribution for all positions located in the green dotted frame, and corresponding *MIH*5 and *MIH*3 can be calculated from the left and right yellow rectangle region.
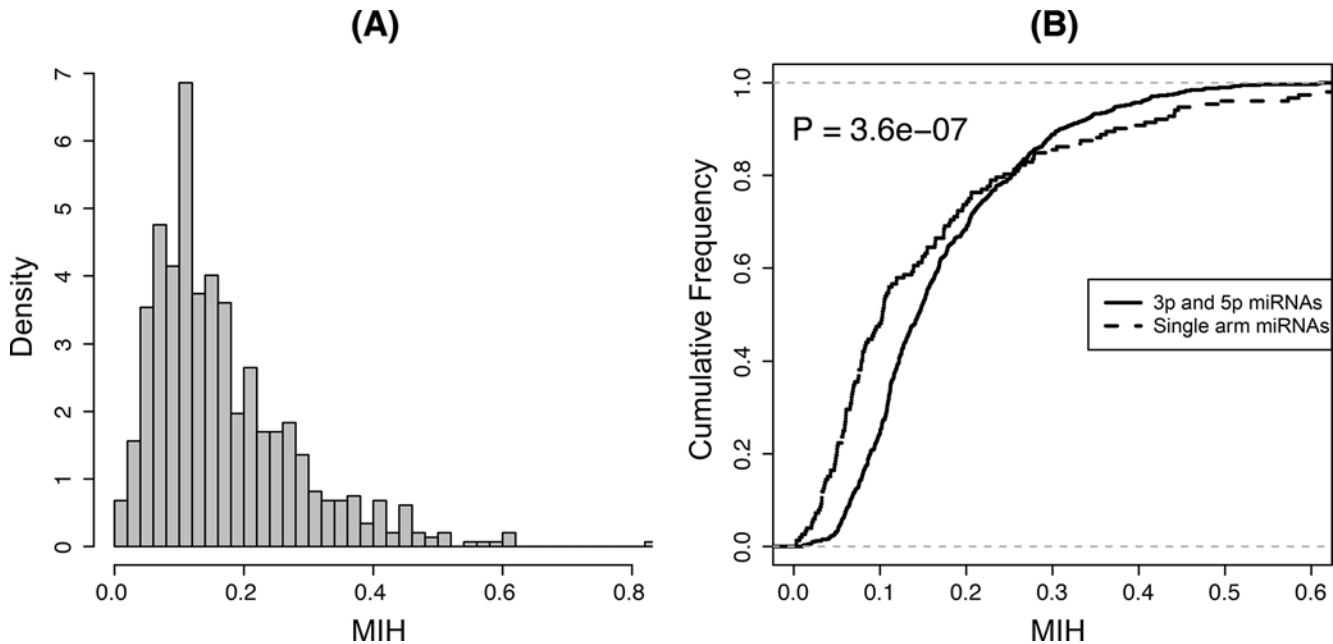
Fig 2. Distribution of *MIH* values. (A) *MIH* distribution of all mature miRNAs after filter. (B) KS-test comparison cumulative fraction plot. Solid line indicates the cumulative distribution of mature miRNA *MIH* values from both 5' and 3' arms of the hairpin precursors (the 3p and 5p miRNAs). Dot line shows the cumulative distribution of mature miRNA *MIH* values from only single arms (the single arm miRNAs).

Functional enrichment analysis

Here, we took TAM tool to perform enrichment analysis to get an insight into the functional role of the miRNAs with high isoform entropy values [35]. The TAM collected miRNAs and their related functions reported in the publications. After removing redundancy, the corresponding high isoform entropy precursors were used to perform functional enrichment analysis, and all precursors with characterized isomiR variants were selected as the reference set. We also used the DAVID Bioinformatics Tools[36,37] to estimate the pathway enrichment of the experimentally validated miRNA targets that were supported by assay or Western blot from miRTarBase[38] and of the predicted miRNA targets from TargetScan[39], respectively. The Enrichment Thresholds or EASE was set as 0.05.

Results

Summary of miRNA isoform entorpy values

In this study, we calculated *MIH* values using the small RNA sequences extracted from miR-Base website. After discarding miRNAs with the very large or very small numbers, we collected 736 mature miRNAs from 545 precursors with varied *MIH* range from 0.003 to 1.096 (S1 Table), which show the presence of sequence heterogeneity for all the analyzed miRNAs, supporting the complexity of Drosha and Dicer processing in the pre-miRNA transcription [40,41]. High *MIH* value demonstrates blurred patterns and broad diversity of miRNA processing, while low *MIH* value means the expression of isomiRs is dominant on some mature miRNA sequences and great uniformity along the pre-miRNA. Thus, *MIH* is suitable to compare the variant profiles across different miRNAs. From density plot Fig. 2A, we found most of *MIH* values are located in a region around 0.1, which means most of the miRNAs have low

isoform variations, coinciding with the result of previous study[42]. For example, the MIH score of miR-1 in Fig. 1 is 0.1956 (S1 Table).

These *MIH* values can be divided into two sets by the name of mature miRNAs, one set named with 5p or 3p means both arms of pre-miRNAs can generate mature miRNAs without regarding to the preferred strand (the 5p and 3p miRNAs), and the other set named without 5p or 3p means these miRNAs prefer one arm of pre-miRNAs (the single arm miRNAs). Here, we applied the KS test for determining whether isomiR distribution patterns are different between the 5p and 3p miRNAs and the single arm miRNAs. KS test can be used to check if two sets differ significantly under making no assumption about the distribution of data. Compared with the single arm miRNAs, the 3p and 5p miRNAs have greater *MIH* values (Fig. 2B), which means the 3p and 5p miRNAs have greater isoform variations, demonstrating the entropy of products from pre-miRNA (mature miRNA types) is consistent with the entropy of mature miRNA sequences (isomiR variants).

In order to find the relative contributions of Drosha and Dicer to the isomiR variants, we compared the *MIH5* and *MIH3* values of the paired 5p and 3p miRNAs, respectively (S1 Table. Dependent 2-group Wilcoxon Signed Rank Test). We found that the mean of the 3p miRNA *MIH3* values (Drosha cleavage, mean = 0.321) is less than the mean of the 5p miRNA *MIH3* values (Dicer cleavage, mean = 0.359), and the difference between them is significant (P value < 0.01). The mean of the 5p miRNA *MIH5* values (Drosha cleavage, mean = 0.146) is also less than the mean of the 3p miRNA *MIH5* values (Dicer cleavage, mean = 0.151), though the difference between them is not significant (P value > 0.1). Our result agreed with the report that Drosha is higher fidelity than Dicer, which is described in a recent study based on the dominant isomiR analysis in 17 mouse samples [27]. Comparing with their study, our analysis only included the miRNAs with high expression level (count >50), which might be one of the proposed reasons why we cannot detect the significant difference between the Drosha cleavage and Dicer cleavage at the 5' end.

Correlation analysis

Here, we investigated the MFE and length of pre-miRNA to figure out if the intrinsic feature contributes to the isomiR expression profiles. By comparing *MIH* and length with MFE extracted from miRBase, we found the MFE and length of pre-miRNAs should have a great contribution to the isomiR variants (Fig. 3A and B). It seems that the thermostability of folding secondary structures can result in the variation in the Drosha and Dicer cleavage sites. Interestingly, the *MIH5* was also detected to be correlated with the MFE and length of pre-miRNA, while the *MIH3* was not (Fig. 3C, D, E and F). We also compared *MIH* with the number of miRNA sequences to check if the abundance of expression level can affect the *MIH* values, whereas they are not significantly correlated (Fig. 3G), which means the sequence depth can not give a reason for the difference of *MIH* values between miRNAs.

Since *MIH* is positively correlated with MFE, the isomiR expression profiles of mature miRNAs transcribed from two arms of pre-miRNAs should be also correlated. For most pre-miRNAs, both arms can generate functional mature miRNAs, and show various isomiR expression profiles [43]. However, different pre-miRNAs show diversity dominant mature miRNAs between 5p arm and 3p arm, and present different *MIH* values between 5p and 3p. Some miRNAs have higher *MIH* at 5p arms and others at 3p arms. In order to verify whether the *MIH* values are consistent between the 5p and 3p miRNAs, here, we classified them into two sets (high *MIH* and low *MIH*) by simply comparing the 5p and 3p *MIH* values with each other, and calculated the relationship between them (Fig. 3H). Our result showed the *MIH* values are well correlated between 5p and 3p arms, indicating the isomiRs derived from the same
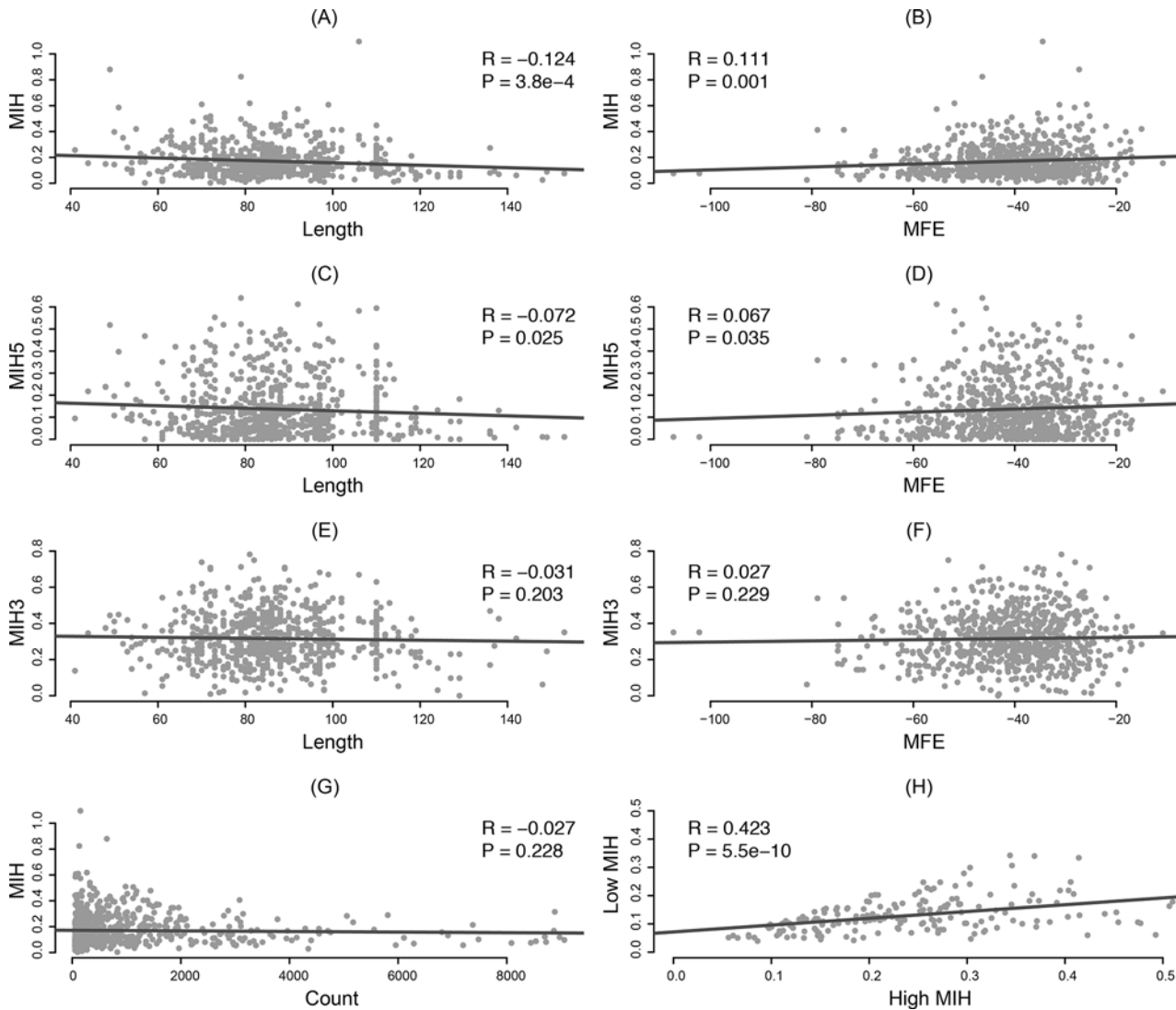
Fig 3. Pearson's correlation analysis. (A) *MIH* and length of pre-miRNA; (B) *MIH* and MFE; (C) *MIH5* and length of pre-miRNA; (D) *MIH5* and MFE; (E) *MIH3* and length of pre-miRNA; (F) *MIH3* and MFE; (G) *MIH* and count of reads; (H) High *MIH* and low *MIH*. High and low *MIH* values are calculated by comparing the *MIH* values between the 5p and 3p miRNAs from the same precursor.

doi:10.1371/journal.pone.0118856.g003

pre-miRNAs show consistent expression profile, which strongly implies that there could be some similar regulation mechanism controlling miRNA isoform production from two arms. It is also pointed out that not all of 5p and 3p arm miRNAs have consistent isomiR expression profiles, sometime extremely high *MIH* in one arm and extremely low *MIH* in another, which demonstrates there should be other more complexly regulate mechanisms existed.

Functional enrichment analysis

The isomiR expression profiles are non-random, indicating that these sequences could be regulated and therefore functional. In order to assess the biological relevance of these molecules, the TAM tool was utilized to perform functional enrichment analysis for miRNAs with high isoform entropy values. We found the *MHI5* have the greatest significantly enriched function items, while the *MIH3* have little contribution to the enrichment analysis (Fig. 4). Compared
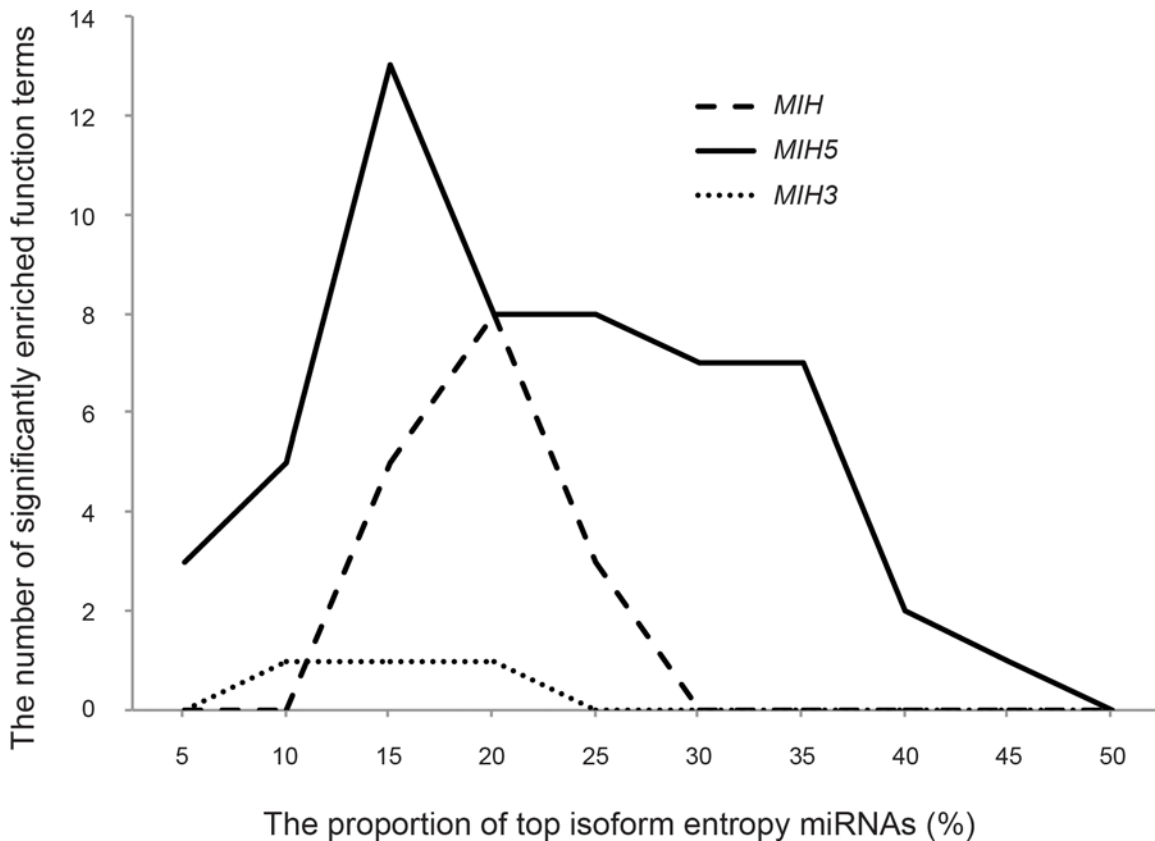
Fig 4. (A) The functional enrichment analysis for miRNA with high isoform entropy values by TAM tool. The X-axis displays the proportion of the top isoform entropy miRNAs. The Y-axis is the number of significantly enriched function terms.

doi:10.1371/journal.pone.0118856.g004

with the variations at the 3' end, the variations at the 5' end usually have lower values and exhibits higher fidelity, illustrating more biological function repression existed.

The *MIH5* have the largest number of significantly enriched function items when using the top 15% entropy values. After removing redundancy, 103 corresponding precursors were assigned to determine a *P* value for the functional enrichment of the target genes by compared with 545 precursors as a reference set. As shown in Table 1, many miRNAs with high *MIH5* values are enriched in cancer-related function, such as miRNA tumor suppressors, angiogenesis, cell cycle related, cell differentiation, etc. In order to further check the function of these miRNA, 729 experimentally validated targets extracted from miRTarBase and 1269 conserved target genes predicted by the TargetScan are used for pathway analysis. Of which, 703 experimentally validated targets and 1237 predicted targets are accepted for the DAVID Bioinformatics Tools to estimate the pathway enrichment, respectively. Based on the reported KEGG pathway terms, both of them are also enriched in cancer-related pathway (S2 and S3 Tables).

Discussion

In this study, we developed the Shannon entropy-based model *MIH* for the isomiR variant measurement. The variation in the 5' end and 3' end leads to a dominating effect on the entropy and gives a *MIH* value. Most of *MIH* values are low, which means the variation expression is still rare. We found that both MFE and length of pre-miRNA are strongly correlated with the variation of isomiRs (Fig. 3A and B), which gives a clue that the intrinsic feature of pre-miRNA

Table 1. Significantly enriched functions of the MiRNAs with the top 15% *MIH5* values.

Function term	Count	*P*-value	FDR
miRNA tumor suppressors	19	3.17E-06	3.56E-04
Cell cycle related	24	2.25E-05	1.44E-03
HIV latency	11	4.50E-05	2.53E-03
Hormones regulation	21	9.14E-05	4.10E-03
Apoptosis	17	5.75E-04	1.99E-02
Angiogenesis	11	8.13E-04	2.43E-02
Akt pathway	9	1.16E-03	2.91E-02
Immune response	16	1.44E-03	3.23E-02
Cell proliferation(Hwang etal BJC2006)	5	2.26E-03	3.76E-02
Folliculogenesis	5	2.26E-03	3.90E-02
Human embryonic stem cell(hESC) regulation	21	2.71E-03	4.32E-02
onco-miRNAs	12	3.43E-03	4.69E-02
Cell proliferation	11	3.70E-03	4.80E-02

doi:10.1371/journal.pone.0118856.t001

should be one of the important factors for the miRNA regulation. The intrinsic feature related miRNA regulation was also found in a recent report that the thermostability of seed binding is strongly correlated with the percentage of seed-pairing target sites[17]. The entropy of miRNA types produced from pre-miRNA should be consistent with the entropy of isomiRs. Then, more types of mature miRNAs are correlated with more isomiR variants, and the 5p and 3p miRNAs usually have larger *MIH* values than the single arm miRNAs (Fig. 2B). In addition, the variations at 5p arms are well correlated with the variations at 3p arms (Fig. 3H).

IsomiR variants, which are not randomly produced by the cleavage of Drosha and Dicer or other mechanisms, should be the characteristic of miRNA expression response under specific condition. Here, we detected the isomiR expression profile at both 5' and 3' end by integrating miRNAs under different developmental stages of different tissues and cell lines. The candidate functions of miRNAs with high isoform entropy are enriched in a set of critical functions, strongly implying that there could be important biological role involved, indicating that the complete repertoire of functional miRNAs is likely more complex than previously appreciated. The miRNAs with high isoform entropy at 5' end have the greatest significantly function items, most of which are cancer-related functions. Considering many of our input data come from cancer-related research, we suggest that the miRNA cannot only change the expression level but also evolve many isomiR variants to implement the function in a more effective way. Our results agree with previous reports that the isomiRs could play important roles in oncogenesis process [44–46]. The other enriched functions show that the regulation spectrum of the isomiR sequences is related to a very broad diversity of biological processes.

Although isomiRs are highly overlapped with each other and most of them should have similar function, variants in the 5' end position of miRNAs are expected to alter the canonical seed sequences, which usually complementary binding to the target mRNA genes, thereby considerable versatility in miRNA target selection [16,20–22,47]. Functional enrichment analysis of miRNAs with high *MIH*5 also supports the 5' end variation should play an important role. An earlier study reported that the biotin-labeled isomiRs can act cooperatively with canonical miRNAs to target functional related but different genes, suggesting biologically relevant and functionally cooperative about these molecules [9]. The existence of multiple 5' isomiRs could enable miRNA genes greatly expand the targeting potential with the utilization of length heterogeneity and implement the function in a more effective way for some specific networks. In

our result, the variation at the 3' end was not significantly correlated with the MFE and length of pre-miRNA, and the miRNAs with high isoform entropy at the 3' end can be hardly enriched in function terms. The 3' isomiRs, which is proved to be equally effective inhibitor to target gene as the canonical miRNA in recent study[48], should affect less than the 5' isomiRs to the post-transcriptional gene regulation.

Previous position shift based method, like the weighted average size of nucleotide variation (WAZNV), the location of reference miRNA sequence (defined by the most dominant sequence) is the most important factor used to calculate the relative distance of isomiR variants [27,46]. However, switching of the most dominant isomiRs found in several studies makes it hard to define the reference sequence for a specific miRNA [27,43,44], which leads to the difference between the detected reference sequence and the canonical mature miRNA [49]. Also, the KS test based method for isomiR distribution pattern comparison can be only used to examine individual miRNA isoform profile, and then check if the distribution patterns in two different conditions are unique or not[44]. One of the advantages of our strategy is less dependent on the exact location of the most dominant mature miRNA, and the wide range region (canonical mature miRNA ± 4 nt) should include most of the isomiR variants. Therefore, the location of reference miRNA should have limited contribution for the *MIH* value calculation.

Supporting Information

S1 Fig. Plot of the count of aligned sequences and *MIH* for each miRNA. Two lines on x-axis indicate the cut off of 50 and 10000. The solid line shows a lowess smooth of the plot. (DOC)

S1 Table. List of miRNAs and their *MIH* values.
(DOC)

S2 Table. The KEGG pathway enrichment of the experimentally validated miRNA target genes using the DAVID Bioinformatics Tools.
(DOC)

S3 Table. The KEGG pathway enrichment of the TargetScan predicted miRNA target genes using the DAVID Bioinformatics Tools.
(DOC)

Acknowledgments

Author Contributions

Conceived and designed the experiments: SW ZL. Performed the experiments: SW. Analyzed the data: SW JT. Contributed reagents/materials/analysis tools: JT LW. Wrote the paper: SW ZL.

References

1. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281–297. PMID: 14744438

2. Huntzinger E, Izaurralde E (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet 12: 99–110. doi: 10.1038/nrg2936 PMID: 21245828

3. Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42: D68–D73. doi: 10.1093/nar/gkt1181 PMID: 24275495

4. Gu S, Jin L, Zhang Y, Huang Y, Zhang F, Valdmanis PN, et al. (2012) The Loop Position of shRNAs and Pre-miRNAs Is Critical for the Accuracy of Dicer Processing In Vivo. Cell 151: 900–911. doi: 10.1016/j.cell.2012.09.042 PMID: 23141545

5. Kim VN, Han J, Siomi MC (2009) Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol 10: 126–139. doi: 10.1038/nrm2632 PMID: 19165215

6. Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 27: 91–105. PMID: 17612493

7. Neilsen CT, Goodall GJ, Bracken CP (2012) IsomiRs–the overlooked repertoire in the dynamic micro-RNAome. Trends Genet 28: 544–549. doi: 10.1016/j.tig.2012.07.005 PMID: 22883467

8. Guo L, Lu Z (2010) Global expression analysis of miRNA gene cluster and family based on isomiRs from deep sequencing data. Comput Biol Chem 34: 165–171. doi: 10.1016/j.compbiolchem.2010.06.001 PMID: 20619743

9. Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S, et al. (2011) MicroRNAs and their isomiRs function cooperatively to target common biological pathways. Genome Biol 12: R126. doi: 10.1186/gb-2011-12-12-r126 PMID: 22208850

10. Körbes AP, Machado RD, Guzman F, Almerão MP, de Oliveira LFV, Loss-Morais G, et al. (2012) Identifying conserved and novel microRNAs in developing seeds of Brassica napus using deep sequencing. PLoS ONE 7: e50663. doi: 10.1371/journal.pone.0050663 PMID: 23226347

11. Pritchard CC, Cheng HH, Tewari M (2012) MicroRNA profiling: approaches and considerations. Nat Rev Genet 13: 358–369. doi: 10.1038/nrg3198 PMID: 22510765

12. Vickers KC, Sethupathy P, Baran-Gale J, Remaley AT (2013) Complexity of microRNA function and the role of isomiRs in lipid homeostasis. J Lipid Res 54: 1182–1191. doi: 10.1194/jlr.R034801 PMID: 23505317

13. Ameres SL, Zamore PD (2013) Diversifying microRNA sequence and function. Nat Rev Mol Cell Biol 14: 475–488. doi: 10.1038/nrm3611 PMID: 23800994

14. Fernandez-Valverde SL, Taft RJ, Mattick JS (2010) Dynamic isomiR regulation in Drosophila development. RNA 16: 1881–1888. doi: 10.1261/rna.2379610 PMID: 20805289

15. Guo L, Zhao Y, Zhang H, Yang S, Chen F (2013) Close association between paralogous multiple isomiRs and paralogous/orthologues miRNA sequences implicates dominant sequence selection across various animal species. Gene 527: 624–629. doi: 10.1016/j.gene.2013.06.083 PMID: 23856130

16. Hinton A, Hunter SE, Afrikanova I, Jones GA, Lopez AD, Fogel GB, et al. (2014) sRNA-seq analysis of human embryonic stem cells and definitive endoderm reveal differentially expressed microRNAs and novel isomiRs with distinct targets. Stem Cells.

17. Wang X (2014) Composition of seed sequence is a major determinant of microRNA targeting patterns. Bioinformatics 30: 1377–1383. doi: 10.1093/bioinformatics/btu045 PMID: 24470575

18. wang S, Xu Y, Lu Z (2014) Genome-wide miRNA seeds prediction in Archaea. Archaea 2014: 671059. doi: 10.1155/2014/671059 PMID: 24948879

19. Friedman RC, Farh KK-H, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 19: 92–105. doi: 10.1101/gr.082701.108 PMID: 18955434

20. Bizuayehu TT, Lanes CFC, Furmanek T, Karlsen BO, Fernandes JMO, Johansen SD, et al. (2012) Differential expression patterns of conserved miRNAs and isomiRs during Atlantic halibut development. BMC Genomics 13: 11. doi: 10.1186/1471-2164-13-11 PMID: 22233483

21. Ebhardt HA, Fedynak A, Fahlman RP (2010) Naturally occurring variations in sequence length creates microRNA isoforms that differ in argonaute effector complex specificity. Silence 1: 12. doi: 10.1186/1758-907X-1-12 PMID: 20534119

22. Humphreys DT, Hynes CJ, Patel HR, Wei GH, Cannon L, Fatkin D, et al. (2012) Complexity of murine cardiomyocyte miRNA biogenesis, sequence variant expression and function. PLoS ONE 7: e30933. doi: 10.1371/journal.pone.0030933 PMID: 22319597

23. Wyman SK, Knouf EC, Parkin RK, Fritz BR, Lin DW, Dennis LM, et al. (2011) Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. Genome Research 21: 1450–1461. doi: 10.1101/gr.118059.110 PMID: 21813625

24. Katoh T, Sakaguchi Y, Miyauchi K, Suzuki T, Kashiwabara S-I, Baba T, et al. (2009) Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. Genes & Development 23: 433–438.

25. Jones MR, Quinton LJ, Blahna MT, Neilson JR, Fu S, Ivanov AR, et al. (2009) Zcchc11-dependent uridylation of microRNA directs cytokine expression. Nat Cell Biol 11: 1157–1163. doi: 10.1038/ncb1931 PMID: 19701194

26. Seitz H, Ghildiyal M, Zamore PD (2008) Argonaute loading improves the 5' precision of both Micro-RNAs and their miRNA* strands in flies. Curr Biol 18: 147–151. doi: 10.1016/j.cub.2007.12.049 PMID: 18207740

27. Zhou H, Arcila ML, Li Z, Lee EJ, Henzler C, Liu J, et al. (2012) Deep annotation of mouse iso-miR and iso-moR variation. Nucleic Acids Res 40: 5864–5875. doi: 10.1093/nar/gks247 PMID: 22434881

28. Alon S, Mor E, Vigneault F, Church GM, Locatelli F, Galeano F, et al. (2012) Systematic identification of edited microRNAs in the human brain. Genome Research 22: 1533–1540. doi: 10.1101/gr.131573.111 PMID: 22499667

29. Guo L, Yang Q, Lu J, Li H, Ge Q, Gu W, et al. (2011) A comprehensive survey of miRNA repertoire and 3' addition events in the placentas of patients with pre-eclampsia from high-throughput sequencing. PLoS ONE 6: e21072. doi: 10.1371/journal.pone.0021072 PMID: 21731650

30. Bar M, Wyman SK, Fritz BR, Qi J, Garg KS, Parkin RK, et al. (2008) MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. Stem Cells 26: 2496–2505. doi: 10.1634/stemcells.2008-0356 PMID: 18583537

31. Zhu JY, Pfuhl T, Motsch N, Barth S, Nicholls J, Grässer F, et al. (2009) Identification of novel Epstein-Barr virus microRNA genes from nasopharyngeal carcinomas. J Virol 83: 3333–3341. doi: 10.1128/JVI.01689-08 PMID: 19144710

32. Stark MS, Tyagi S, Nancarrow DJ, Boyle GM, Cook AL, Whiteman DC, et al. (2010) Characterization of the Melanoma miRNAome by Deep Sequencing. PLoS ONE 5: e9685. doi: 10.1371/journal.pone.0009685 PMID: 20300190

33. Witten D, Tibshirani R, Gu SG, Fire A, Lui W-O (2010) Ultra-high throughput sequencing-based small RNA discovery and discrete statistical biomarker analysis in a collection of cervical tumours and matched controls. BMC Biol 8: 58. doi: 10.1186/1741-7007-8-58 PMID: 20459774

34. Kuchen S, Resch W, Yamane A, Kuo N, Li Z, Chakraborty T, et al. (2010) Regulation of microRNA expression and abundance during lymphopoiesis. Immunity 32: 828–839. doi: 10.1016/j.immuni.2010.05.009 PMID: 20605486

35. Lu M, Shi B, Wang J, Cao Q, Cui Q (2010) TAM: a method for enrichment and depletion analysis of a microRNA category in a list of microRNAs. BMC Bioinformatics 11: 419. doi: 10.1186/1471-2105-11-419 PMID: 20696049

36. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44–57. doi: 10.1038/nprot.2008.211 PMID: 19131956

37. Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37: 1–13. doi: 10.1093/nar/gkn923 PMID: 19033363

38. Hsu S-D, Tseng Y-T, Shrestha S, Lin Y-L, Khaleel A, Chou CH, et al. (2014) miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Res 42: D78–D85. doi: 10.1093/nar/gkt1266 PMID: 24304892

39. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. (2011) Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nature Publishing Group 18: 1139–1146.

40. Lee LW, Zhang S, Etheridge A, Ma L, Martin D, Galas D, et al. (2010) Complexity of the microRNA repertoire revealed by next-generation sequencing. RNA 16: 2170–2180. doi: 10.1261/rna.2225110 PMID: 20876832

41. Ebhardt HA, Tsang HH, Dai DC, Liu Y, Bostan B, Fahlman RP. (2009) Meta-analysis of small RNA-sequencing errors reveals ubiquitous post-transcriptional RNA modifications. Nucleic Acids Res 37: 2461–2470. doi: 10.1093/nar/gkp093 PMID: 19255090

42. Langenberger D, Pundhir S, Ekstrøm CT, Stadler PF, Hoffmann S, Gorodkin J. (2012) deepBlockAlign: a tool for aligning RNA-seq profiles of read block patterns. Bioinformatics 28: 17–24. doi: 10.1093/bioinformatics/btr598 PMID: 22053076

43. Guo L, Zhang H, Zhao Y, Yang S, Chen F (2014) Selected isomiR expression profiles via arm switching? Gene 533: 149–155. doi: 10.1016/j.gene.2013.09.102 PMID: 24120620

44. Li S-C, Liao Y-L, Ho M-R, Tsai K-W, Lai C-H, Lin W-C. (2012) miRNA arm selection and isomiR distribution in gastric cancer. BMC Genomics 13 Suppl 1: S13. doi: 10.1186/1471-2164-13-S1-S13 PMID: 22369582

45. Saito K, Inagaki K, Kamimoto T, Ito Y, Sugita T, Nakajo S, et al. (2013) MicroRNA-196a is a putative diagnostic biomarker and therapeutic target for laryngeal cancer. PLoS ONE 8: e71480. doi: 10.1371/journal.pone.0071480 PMID: 23967217

46. Chang H-T, Li S-C, Ho M-R, Pan H-W, Ger L-P, Hu L-Y, et al. (2012) Comprehensive analysis of micro-RNAs in breast cancer. BMC Genomics 13 Suppl 7: S18. doi: 10.1186/1471-2164-13-S7-S18 PMID: 23281739

47. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, et al. (2010) Mammalian micro-RNAs: experimental evaluation of novel and previously annotated genes. Genes & Development 24: 992–1009. doi: 10.1101/gad.1884710 PMID: 20413612

48. Tan GC, Chan E, Molnar A, Sarkar R, Alexieva D, Isa IM, et al. (2014) 5' isomiR variation is of functional and evolutionary importance. Nucleic Acids Res.

49. Giles KM, Barker A, Zhang PM, Epis MR, Leedman PJ (2011) MicroRNA regulation of growth factor receptor signaling in human cancer cells. Methods Mol Biol 676: 147–163. doi: 10.1007/978-1-60761-863-8_11 PMID: 20931396