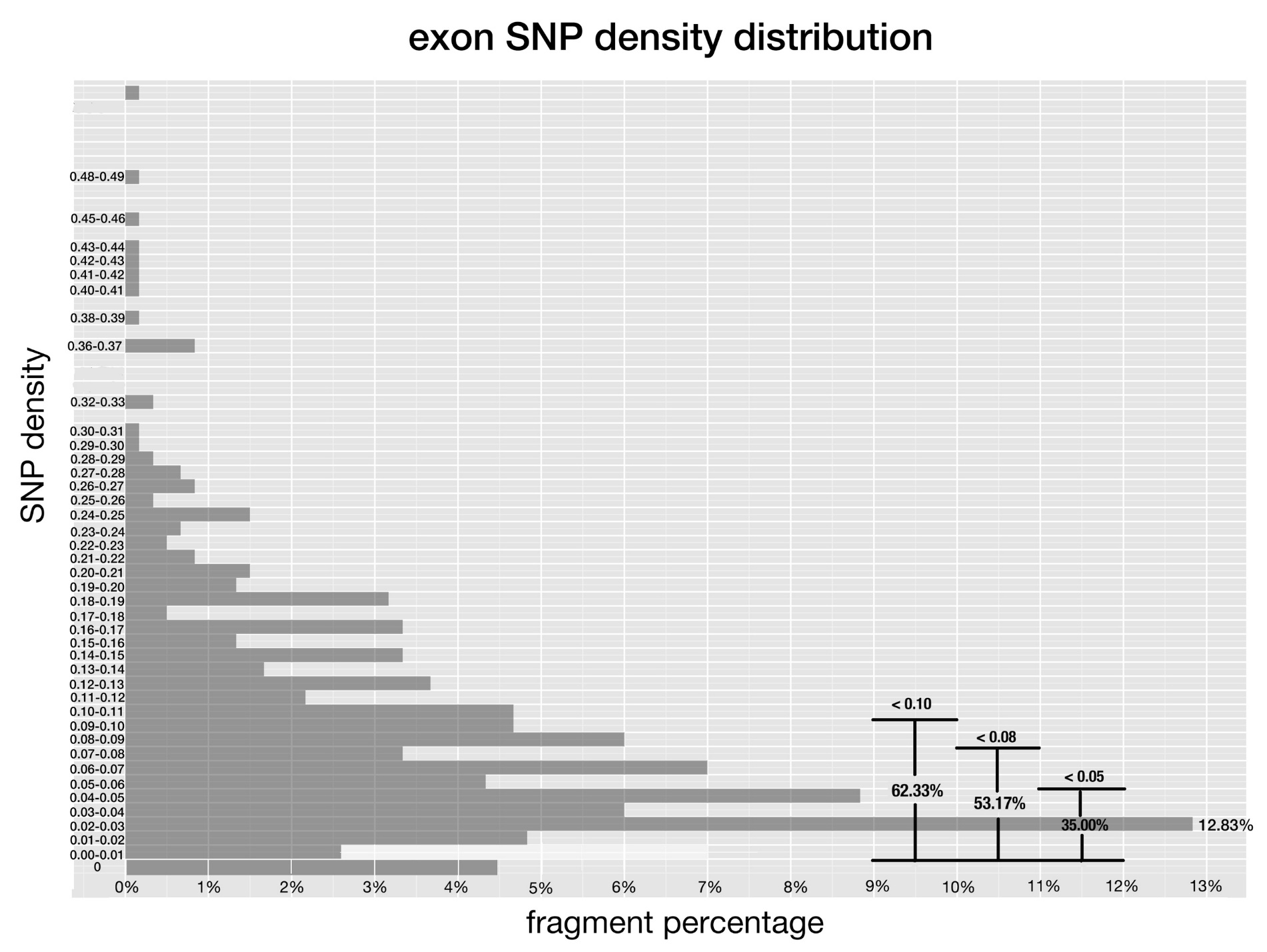
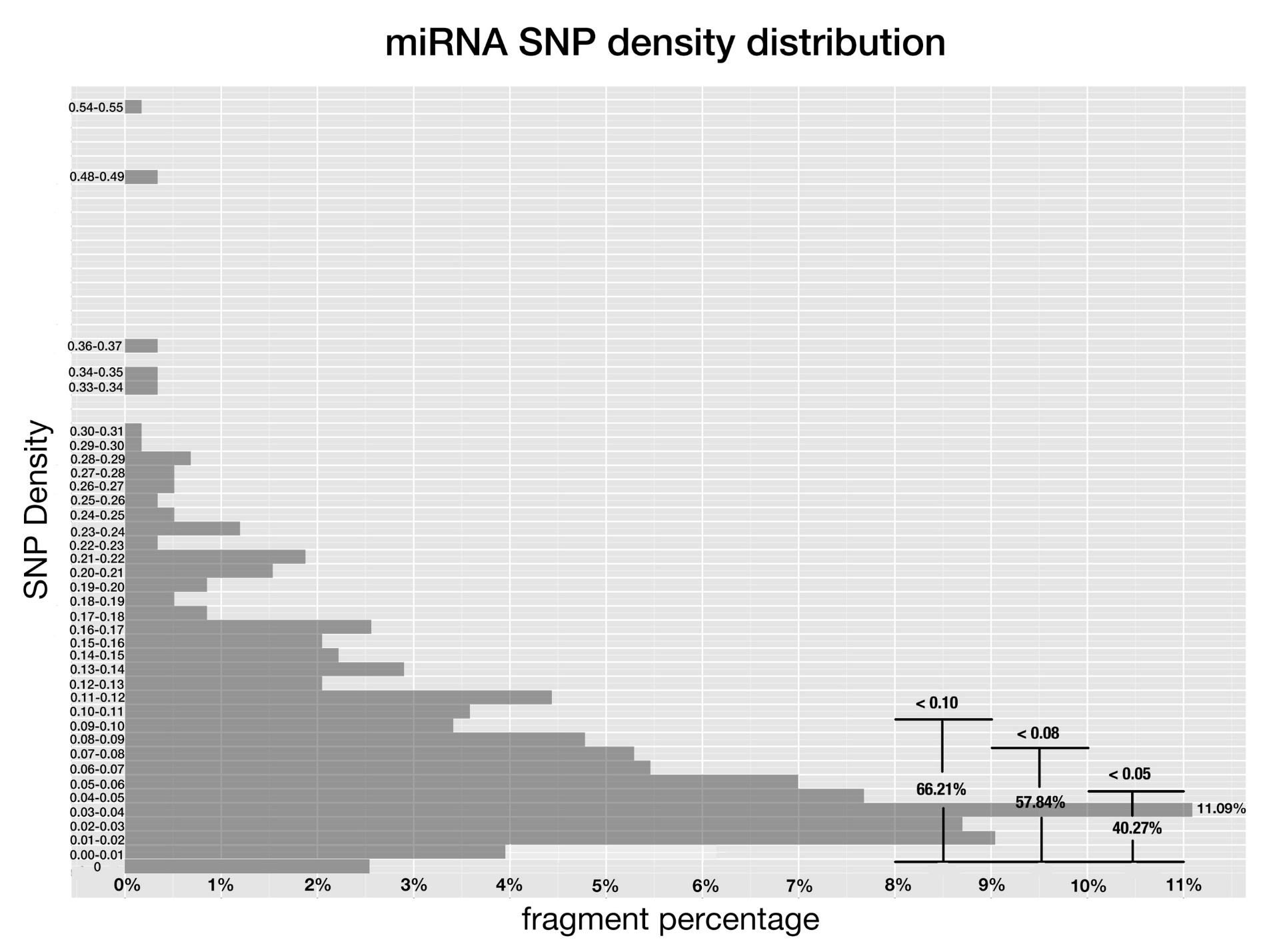
**Results**

**Overall SNP densities of rice miRNAs from 3K Rice Genome Project**

SNP density can reflect the selection pressure imposed on given genomic regions. The higher the pressure is; the lower the SNP density will be [1], and vice versa. Since miRNAs are functional genomic units and master regulators, they should have gone through different selection pressure compared with intergenic regions, and exons [2].To investigate this, were collected3 4chosen the, which is defined as the division of SNP number by the length of the genetic region, respectively

As expected, SNP density of pre-miRNAs and exons were found to be lower than that of intergenic regions. For pre-miRNAs, the percentage of genomics fragments kept

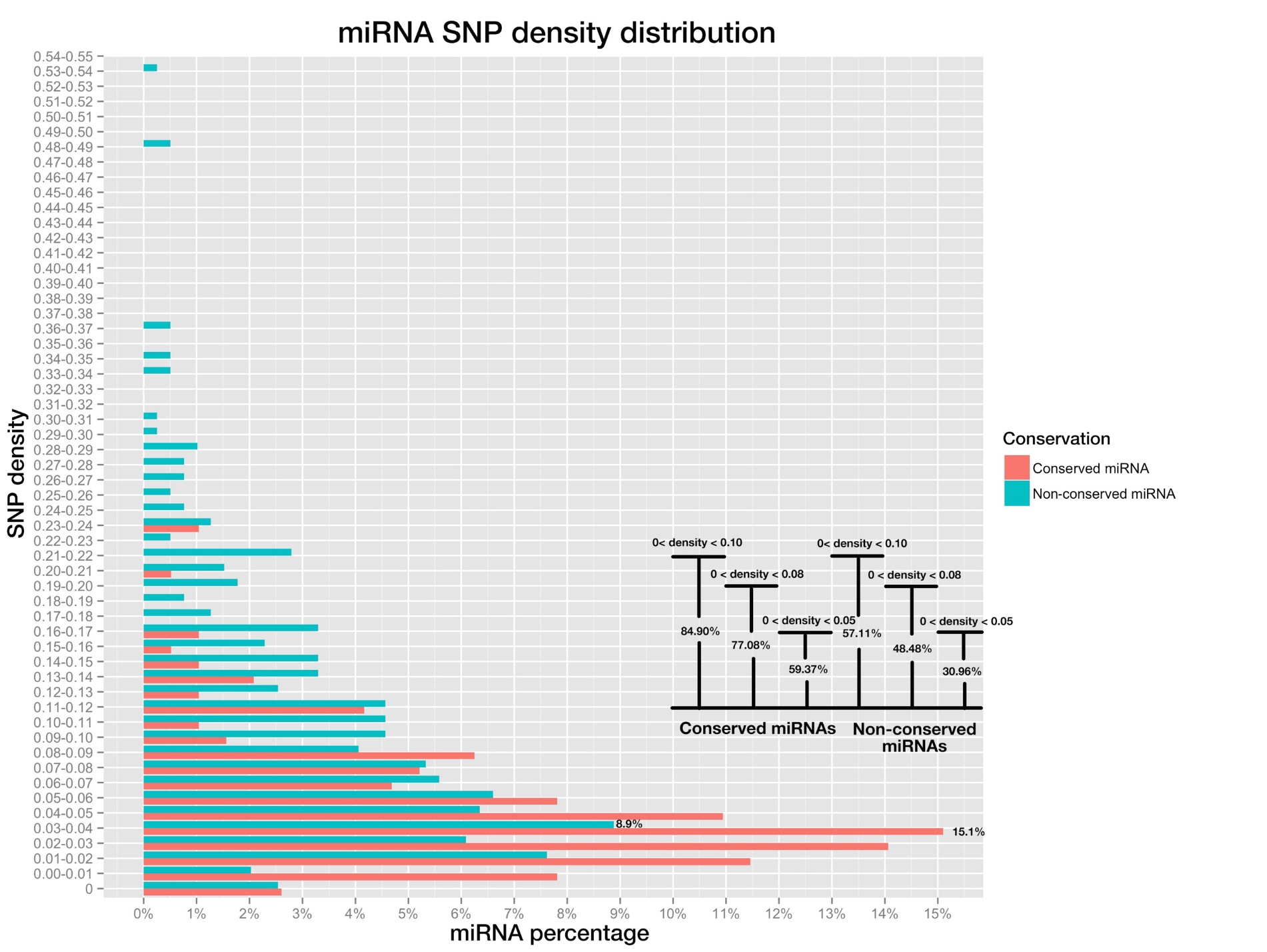
rising with the SNP density until it peaked at the range of 0.03-0.04. It then decreased gradually after 0.04 (Fig. 1).The same trend was observed for the overall SNP distribution in exons sampled in this study, except that the abundance of corresponding fragments peaked at the SNP density range of 0.02-0.03 with a percentage of 12.83 (Fig. 2). No such trend was seen for the intergenic regions analyzed (Fig. 3). Through comparing the percentage of fragments fell within the lower SNP density ranges of 0-0.10, 0-0.08 and 0-0.05 individually (Fig. 1, 2, 3), it is clear that most pre-miRNAs clustered at similar density as exons, where significantly less intergenic regions fell into the same ranges. This demonstrates that pre-miRNAs have gone through stricter evolutionary selections than both intergenic regions and exons, which is consistent with their established roles as master regulators in many genetic pathways.



**Fig 1-3. SNP density of pre-miRNAs, exon regions and intergenic regions in rice.**

the of the genetic regionthe same as Fig 1, and x-axis corresponds to the percentage of fragments that have SNP density at given range.

Due to their evolutionary conservation [9], conserved rice miRNAs should have lower SNP density than that of non-conserved ones. This is the case as shown in Figure 4, most conserved miRNAs clustered at lower SNP density ranges compared to non-conserved ones after comparing the percentage of fragments fell within the lower SNP density ranges of 0-0.10, 0-0.08 and 0-0.05, respectively.



**Fig 4. miRNA SNP density distribution of pre-miRNAs, both conserved miRNAs (in red color) and non-conserved miRNAs (blue color).**

Bar plot on the bottom right shows the percentage of miRNAs whose SNP densities are below 0.10, 0.08 and 0.05, for conserved and non-conserved miRNAs respectively.

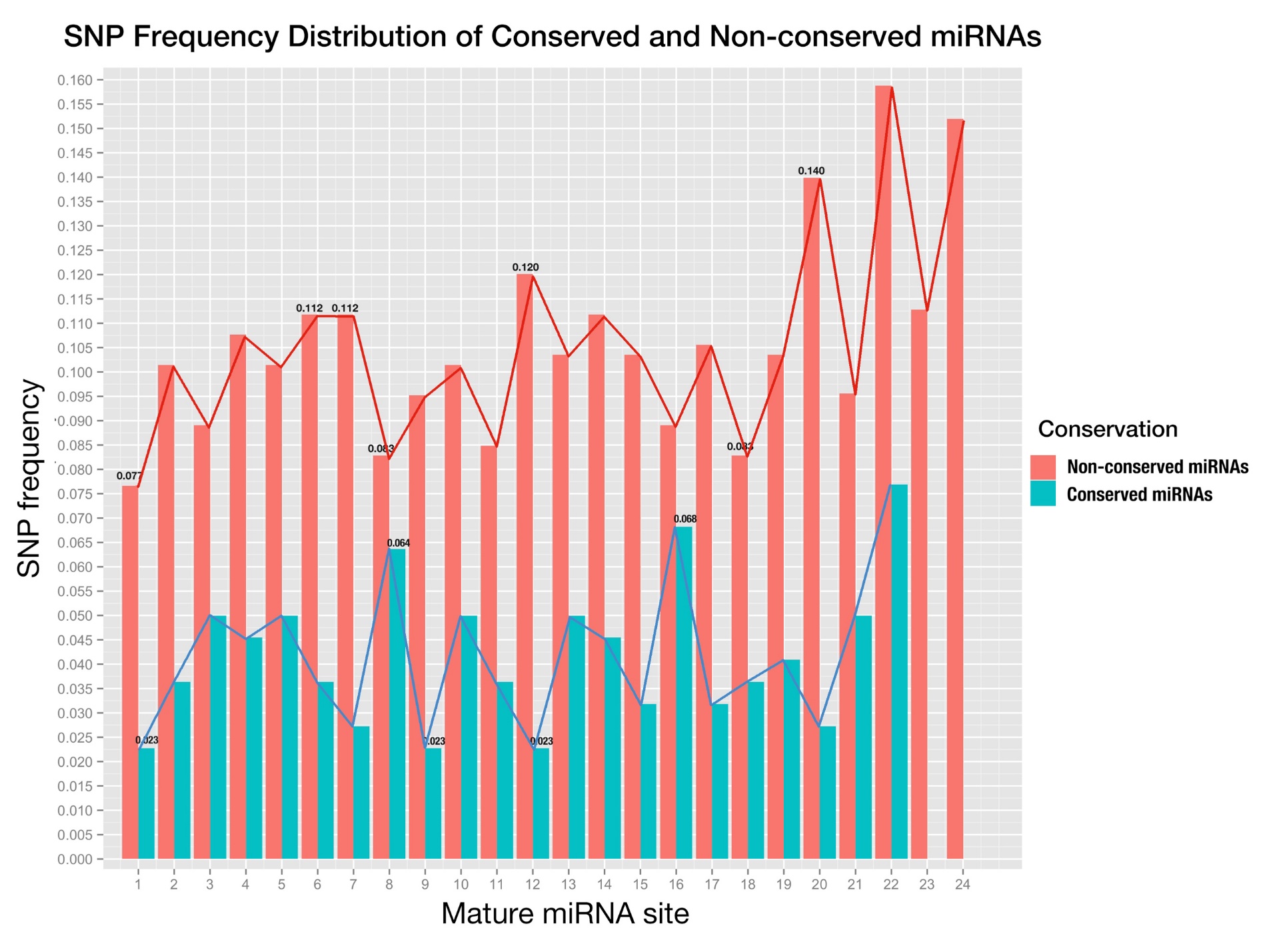
**Positional SNP distributions are differential in conserved and non-conserved miRNAs**

Plant miRNAs have variable efficacies toward different targets. This can be attributed to their complementarity patterns, which are determined by mature miRNAs and the cognate target sites co-jointly. Moreover, there is evidence suggesting that different positions along the 21 nt mature miRNA functionally diversify in target recognition and silencing [5], as mutations on certain positions cause complete abortion of silencing while others do not have obvious impact (Reference). This may also be reflected by the differential SNP frequency of each position, as lower frequency implies higher selection pressure. Therefore, collective analyses of SNP frequencies of individual positions along mature miRNAs may further reveal their differential functional importance in silencing.

Positional SNP frequencies of conserved and non-conserved mature miRNAs were calculated separately and compared. Since most miRNAs are 21-nt in length, this study focused only on position 1 to 21. As expected all positions on conserved miRNAs had SNP frequenciescorresponding positionson Despite so, it was expected that they would share similar rankings of SNP frequencies among different positions, given both are under the same selection pressure mechanistically. However, this doesn’t appear to be the case, except position one was the lowest SNP frequencies for both conserved and non-conserved miRNAs, which may be explained by its importance in the loading of miRNAs into the AGO proteins (Reference…).

Pairing at position 10 and 11 is thought to be critical for plant miRNA functioning through cleavage [6,7, 8], which adds to another level of restriction on the evolution of both positions and this may result in the lower SNP densities than other miRNA positions. Conserved miRNAs are functionally more important and conserved than non-conserved miRNAs, so, SNP frequencies of positions along conserved mature miRNAs are supposed to be lower than that of positions along non-conserved ones. To verify these guesses,

The positions with highest SNP frequencies were positions 20, 12 as well as positions 6 and 7 sharing the same frequency, while the positions with lowest SNP frequencies were positions 1, 8 and 18,; by contrast, for conserved miRNAs, the positions with highest SNP frequencies were positions 16, 8 and followed by 3, 6, 10, 13 sharing the same frequency, while the positions with lowest SNP frequencies were 1, 9 and 12 (Fig. 5). For both conserved miRNAs and non-conserved miRNAs, position 1 won out to be the lowest SNP frequency site, which may be explained by the fact that position 1 determines which Argonaut protein to load for mature miRNAs and this would make the site subject to high selection pressure. But position 10 and 11 are not among the lowest SNP frequency positions, not consistent with the empirical claims.



**Fig 5. Positional SNP distribution of conserved miRNAs (blue) and non-conserved miRNAs (red).**

X-axis is the sites in mature miRNA from 5’end - 3’end, and y-axis is SNP frequency which is calculated as number of SNPs at this site divided by number of miRNAs;

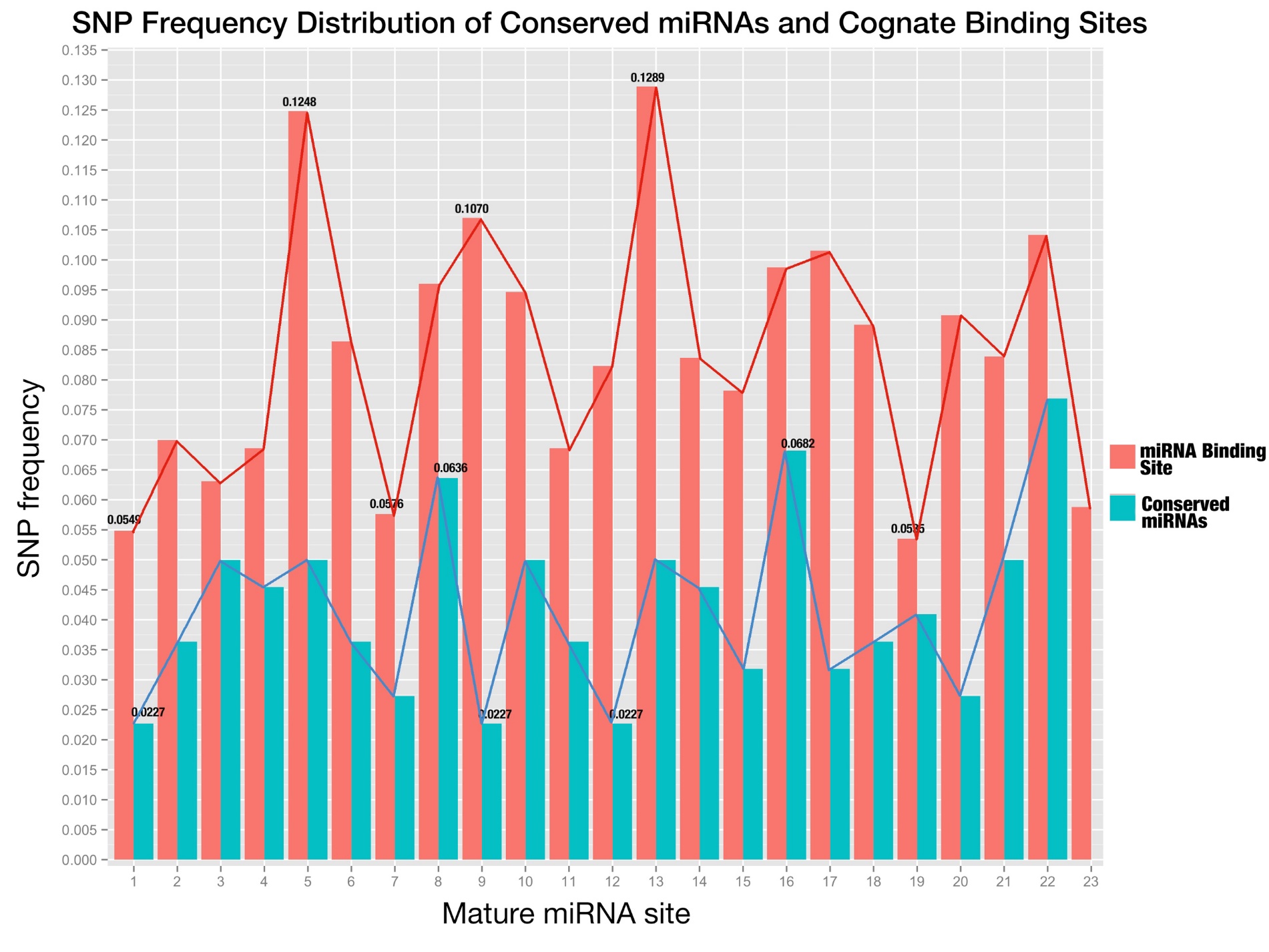
**Positive correlations between positional SNP frequencies of conserved miRNAs and cognate targets reveal co-evolving constrain**

To further utilize the 3K data to investigate the functional

Correlations between matures the corresponding positions onfunctional importance

Since conserved miRNAs have more identifiable targets than non-conserved miRNAs in plants [9] and are functionally more important, we only studied the binding sites of genes targeted by conserved miRNAs here. Also, the diversified importance of positions along mature miRNAs and binding sites may be reflected by the difference of SNP frequencies of varied positions. With the help of online miRNA target prediction tool, *psRNATarget* [10] and transcriptome-wide degradome validation of rice miRNA targets [11], a total number of 823 genes were found being targeted by conserved miRNAs. Then distribution of SNP frequencies of positions along both conserved miRNAs and their binding sites was shown side-by-side in a bar-plot (Fig. 6).

Compared with conserved mature miRNAs, SNP frequency of each position of miRNA binding site was higher, and this suggested that selection pressure upon conserved mature miRNAs is higher than that upon binding site on cognate targets. For plant miRNA regulates multiple targets with high complementarity, more restrictions would be added to mature miRNA and possibly cause the SNP frequency along mature miRNAs to be lower than that along miRNA binding sites. The positions with highest SNP frequencies were position 12, 5 and 9; while the positions with lowest SNP frequencies were position 19, 1 and 7. The highest SNP frequency (0.1289) of positions in miRNA binding site was more than doubled higher compared with the lowest SNP frequency (0.0535), meanwhile, the highest SNP frequency of positions in conserved miRNA (0.0682) was tripled higher compared with the lowest SNP frequency (0.0227). The great difference between SNP frequencies along both conserved miRNA and its binding site implies the importance of positions to the miRNA:target interaction is diversified. Furthermore, Pearson correlation coefficient between SNP frequencies of all positions along conserved miRNAs and that of positions along miRNA binding sites was calculated to be 0.5891, which indicated a moderate positive linear relationship, and the p-value was 0.002455 with null hypothesis that there was no correlation between them, which was less than 0.05 and meant a significant correlation between them. The Pearson correlation test shew a positive correlation between SNP frequencies of positions in conserved miRNA and its binding sites and suggested the co-evolution of miRNAs and its binding sites.



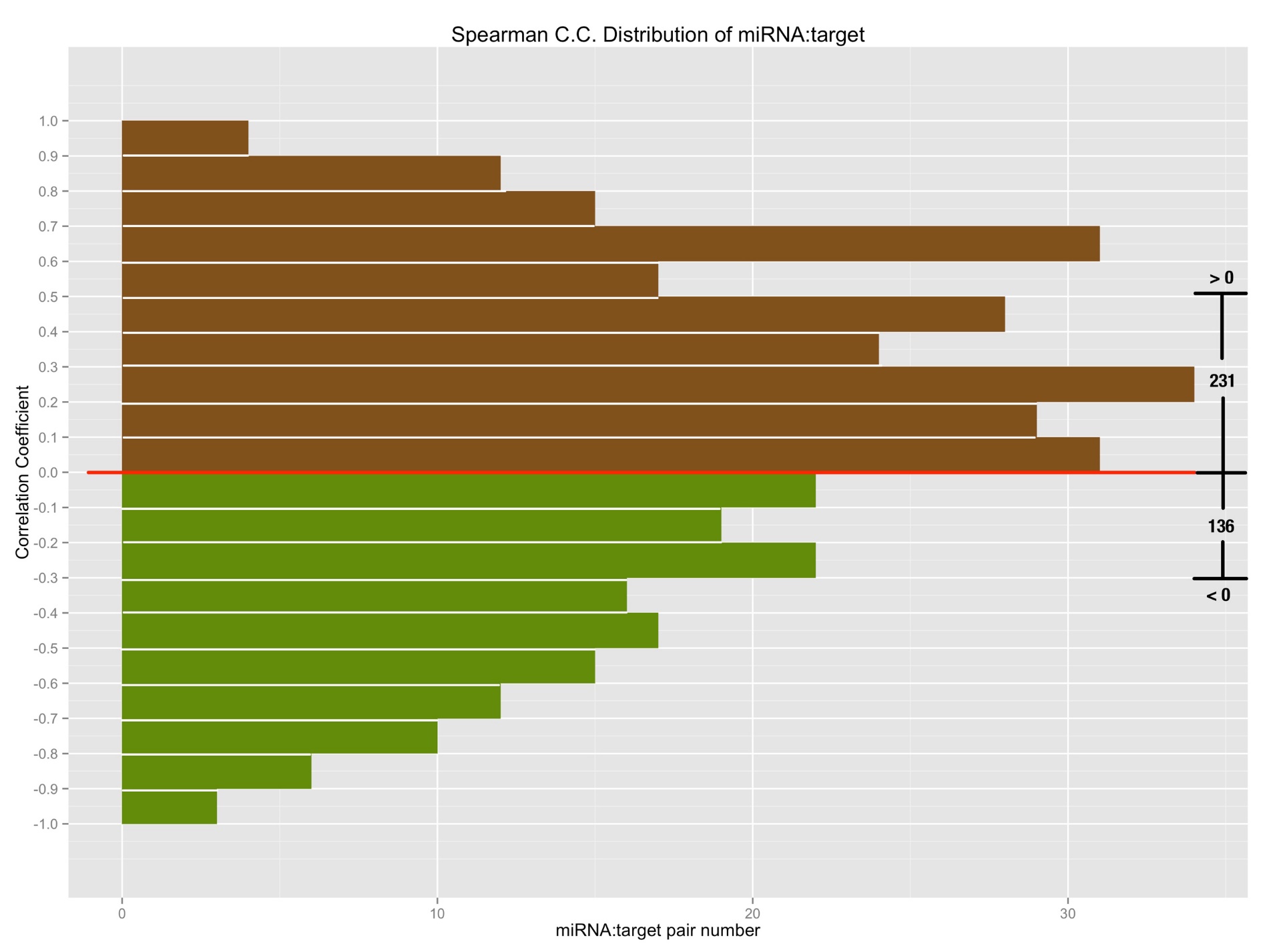
**Fig 6. SNP distribution of all sites along conserved mature miRNAs and their binding sites.**

The sites of miRNA binding site are placed in the same order as mature miRNAs (from 5’end to 3’end in the miRNA);

**Part III. No obvious correlation was found for the expression of confirmed miRNA:target pairs**

In contrast to animal miRNAs, plant miRNAs target genes with high degree of complementarity, which allows the confident prediction with bioinformatics tools. But how to filter those false positive predictions remains a headache. Under the general assumptions that complementarity is the sole determinant of silencing and that any given plant miRNA-loaded RNA-induced silencing complex (miRISC) is able to act independently [12, 13], target mRNAs are downregulated by corresponding miRNAs and the expression level of plant miRNA is negatively correlated with that of cognate target mRNAs, which was also supported by experiment in which five mRNAs encoding TCP transcription factors are downregulated in plants overexpressing miR319 [14]. To verify this hypothesis, expression data of rice miRNAs and genes was extracted from RiceFREND database [15], and correlation test was performed on miRNA:target pairs validated by degradome with 3-week old rice seedling samples [11].

Only 136 out of 367 miRNA:target pairs were examined to be negatively correlated on their expression level (Fig. 7) and this strongly objects the hypothesis that there is negative expression correlation between confirmed targets and their miRNAs. Furthermore, more than half (197 out of 367) pairs fell in the range of weak correlation (-0.4~0.4), which shows there may not be direct relationship between miRNAs and cognate target mRNAs concerning expression level.



**Fig 7. The spearman correlation coefficient of degradome validated miRNA:target relationships;**

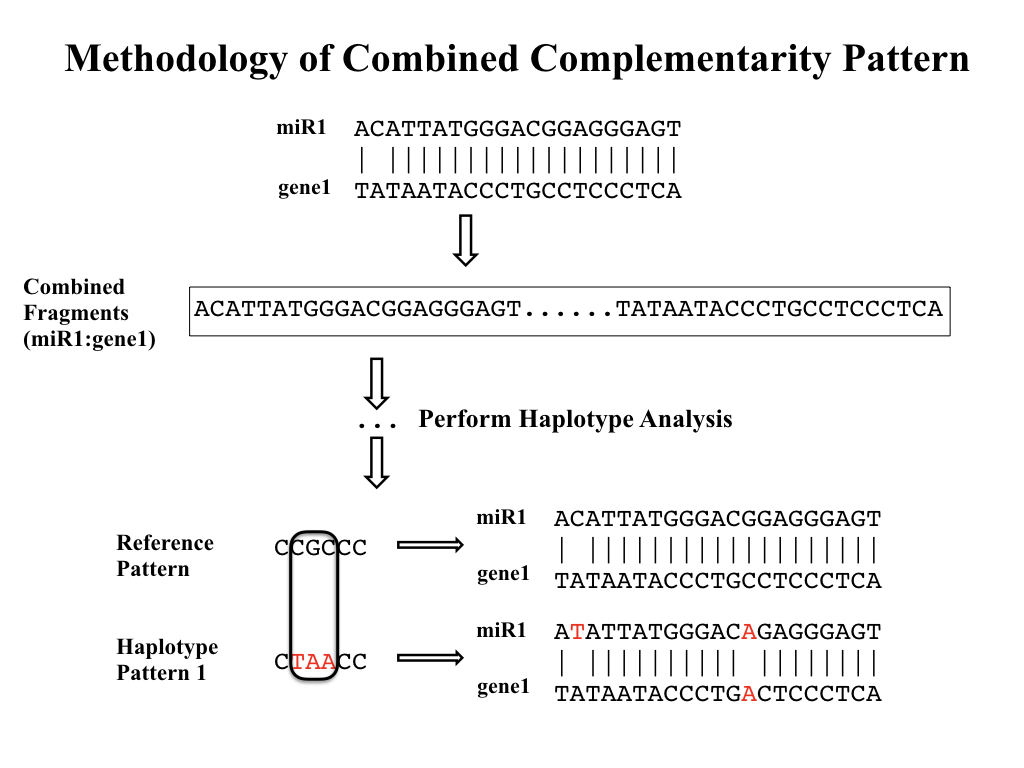
X-axis is the number of pairs that fall on the specified range of correlation coefficient; so bars with green color in the lower part denote the negatively correlated miRNA:target pairs, by contrast, bars with brown color in the upper part denote the positively correlated miRNA:target pairs.

**Part IV. Combined Complementarity Pattern Analysis (CCPA) discovered miR818 family had positions recovered their pairing with two SNPs on both miRNA and binding site**

Haplotype analysis is frequently used in population genetic analysis, and rice cultivars can be grouped together into different sets according to the haplotype patterns when studying the function of a specific genomic region. MiRNA function as a regulator through interaction with its target gene, to be specific, the binding site of cognate target gene. So, I extended the haplotype analysis into Combined Complementarity Pattern Analysis (CCPA) to work for both sequences on mature miRNA and miRNA binding site that are interacting while silencing.

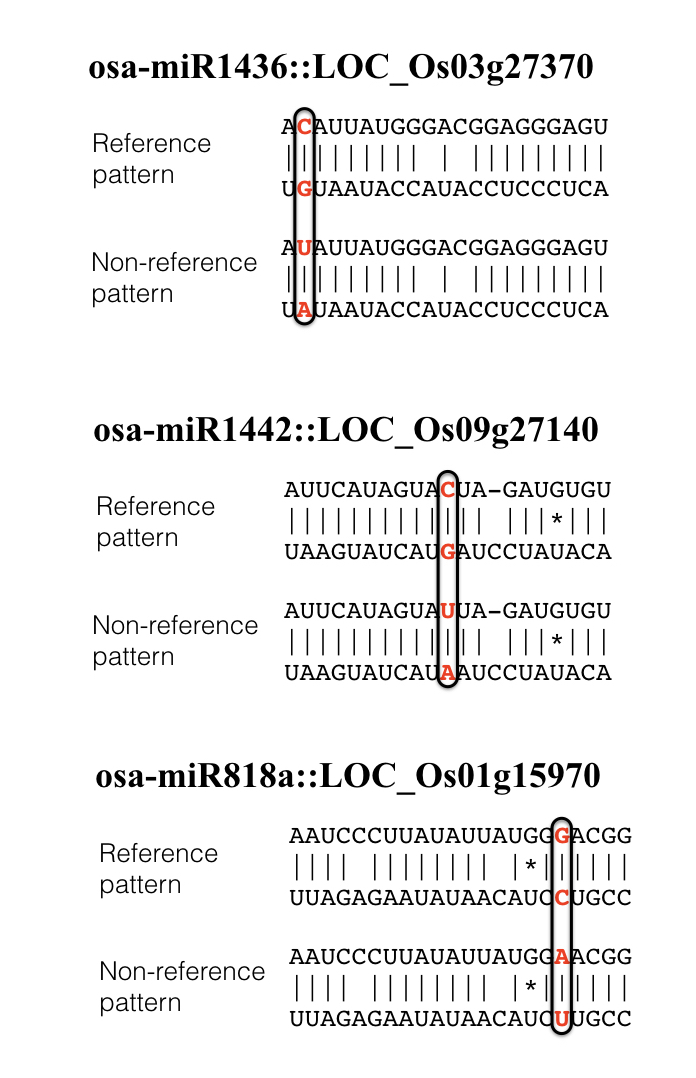
SNP was adopted as the sole biological marker in CCPA. All SNPs would be obtained from the database querying with the genomic region of both the mature miRNA and cognate miRNA binding on question, and then sorted in the ascending order of their id, which composed of 11 digits starting with number 1, followed by 2 digits representing rice genome chromosome id and then zero-filling 8 digits representing genomic position of the SNP. Each rice cultivar was denoted as a haplotype pattern which is a sequence of alleles at the sorted SNP positions, and one allele for one SNP position with upper letter for homozygotes and lower letter of the higher frequency allele for heterozygotes. After that, all 3,024 rice cultivars were divided into different sets according to their haplotype patterns.

Then, each haplotype pattern was mapped back to the sequences of mature miRNA and miRNA binding site to form the specified genotype of a rice cultivar set concerning the miRNA:target pair in question. Both sequences would then be aligned in the reverse order to form the complementarity pattern (and that’s why this method is called combined complementarity pattern analysis), for the orientation of mature miRNA is opposite to that of binding site of its target while binding, and the complementarity pattern would be compared with that of reference genome which is called the reference complementarity pattern to unfold the changes that SNPs do to the miRNA:target complementarity.



**Fig 8. Workflow of Combined Complementarity Pattern Analysis (CCPA)**

Generally, positions within complementarity pattern of miRNA:target pair could undergo four types of changes when SNPs take place in either mature miRNA or miRNA binding site (or both of them). The complementarity of the positions could switch from pairing to mismatch, from mismatch to pairing, from pairing to pairing (remain pairing) or from mismatch to mismatch (remain mismatch). Applying CCPA to all conserved miRNAs in rice as well as their target genes, a special type of complementarity change was found tha AQ t there are positions remained pairing after two SNPs were introduced to both sequences at the given position of the complementarity pattern. And this type of change took place in osa-miR818a-e, osa-miR1436, osa-miR1439, osa-miR1442, osa-miR1862b, osa-miR2275, osa-miR444a/b/d and their cognate targets. Osa-miR444 family was previously reported belonging to a unique class of miRNAs called natural antisense miRNAs that derive from the natural cis-antisense transcript pairs, and can generate mature miRNAs that perfectly match their targets [16]. The recovery of complementarity can be explained by the antisense property because a single SNP introduced to the rice genome would cause mutations in both miRNA and its target but they still remain perfectly matched. But for other miRNAs, two SNPs were needed to keep the position still pairing. More interestingly, except osa-miR2275, all other miRNAs belong to the same miRNA family of related hairpin sequences - miR818, according to *miRBase.org*.



**Fig 9. Examples of complementarity patterns that remained pairing after introducing 2 SNPs**

In the complementarity pattern, a vertical bar would be placed denoting match, a blank denotes mismatch, an asterisk denotes G:U pair.

**Part V. Phenotypes of miR156, miR172 showed no big difference compared with that of reference combined complementarity pattern**

**References:**

[1] Chen, K. & Rajewsky, N. Natural selection on human miRNA binding sites inferred from SNP data. *Nature Genet.* **38**, 1452–1456 (2006).

[2] Saunders, M. A., Liang, H. & Li, W. H. Human polymorphism at microRNAs and microRNA target sites. Proc. Natl Acad. Sci. USA 104, 3300–3305 (2007).

[3] Alexandrov, N. *et al.* SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res.* **43**, D1023–D1027 (2015).

[4] Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42, D68–D73 (2014).

[5] Liu Q, Wang F, Axtell M.J. Analysis of complementarity requirements for plant microRNA targeting using a Nicotiana benthamiana quantitative transient assay. Plant Cell 26: 741–753 (2014).

[6] Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M., and Weigel, D. (2005). Specific effects of microRNAs on the plant transcriptome. Dev. Cell 8: 517–527.

[7] Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., García, J.A., and Paz- Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. Nat. Genet. 39: 1033–1037.

[8] Todesco, M., Rubio-Somoza, I., Paz-Ares, J., and Weigel, D. (2010). A collection of target mimics for comprehensive analysis of microRNA function in Arabidopsis thaliana. PLoS Genet. 6: e1001031.

[9] Jones-Rhoades MW. 2012. Conservation and divergence in plant microRNAs. *Plant Mol. Biol.* 80:3–16

[10] Xinbin Dai and Patrick X. Zhao, psRNATarget: A Plant Small RNA Target Analysis Server, Nucleic Acids Research, 2011, W155-9. doi: 10.1093/nar/gkr319.

[11] Li YF, Zheng Y, Addo-Quaye C, Zhang L, Saini A, Jagadeeswaran G, Axtell MJ, Zhang W, Sunkar R. Transcriptome-wide identification of microRNA targets in rice. Plant J. 2010;62:742-759.

[12] Iwakawa, H. and Tomari, Y. (2013) Molecular insights into microRNA-  mediated translational repression in plants. Mol. Cell 52, 591–601

[13] Tang, G. et al. (2003) A biochemical framework for RNA silencing in

plants. Genes Dev. 17, 49–63

[14] Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, et al. 2003. Control of leaf mor- phogenesis by microRNAs. *Nature* 425:257–63

[15] Sato Y, Namiki N, Takehisa H, Kamatsuki K, Minami H, Ikawa H, Ohyanagi H, Sugimoto K, Itoh J, Antonio B, Nagamura Y (2013) RiceFREND: a platform for retrieving coexpressed gene networks in rice. Nucleic Acids Research 41:D1214-D1221

[16] Lu, C., et al. (2008). Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). Proc. Natl. Acad. Sci. USA 105: 4951–4956.