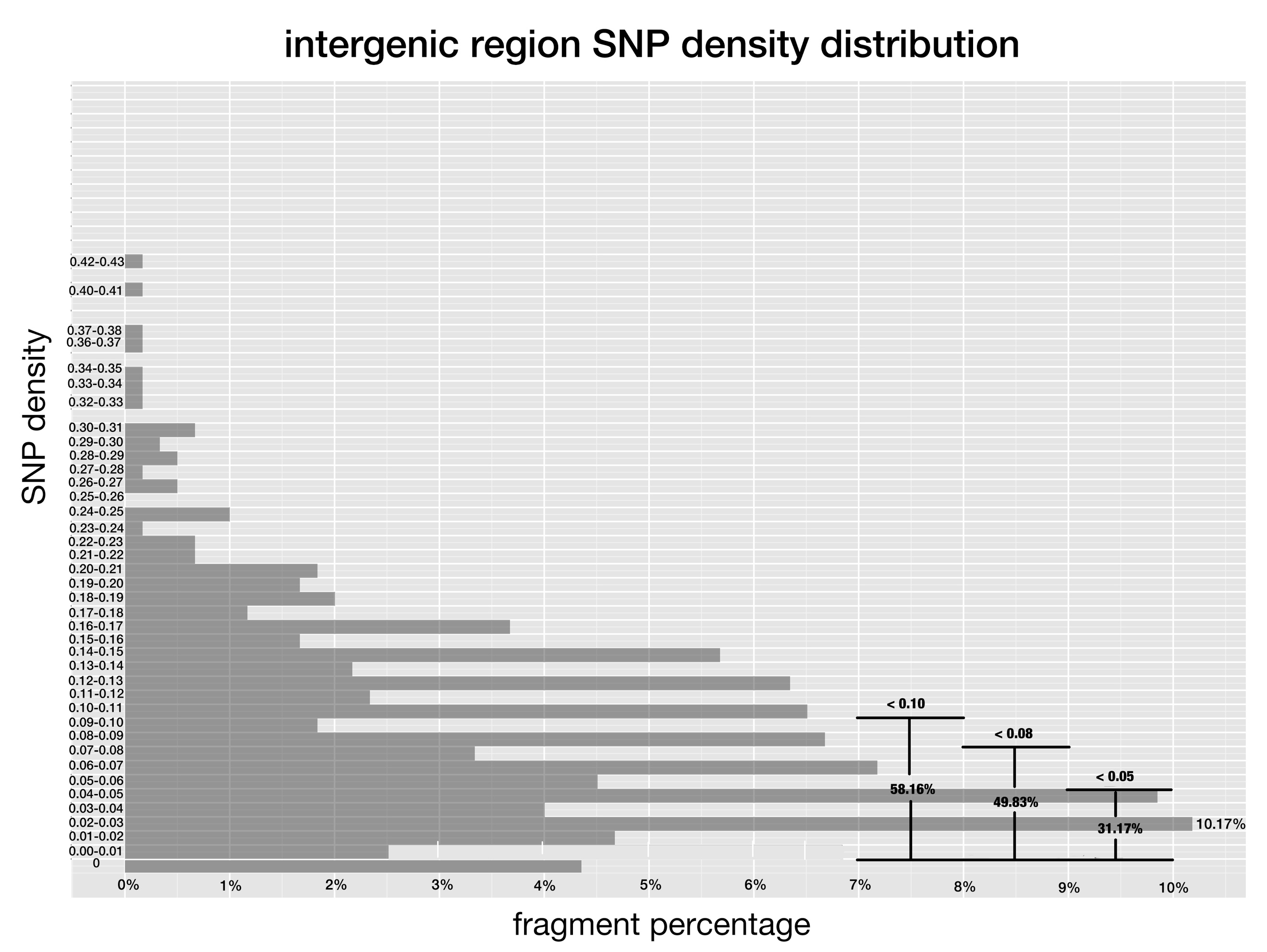
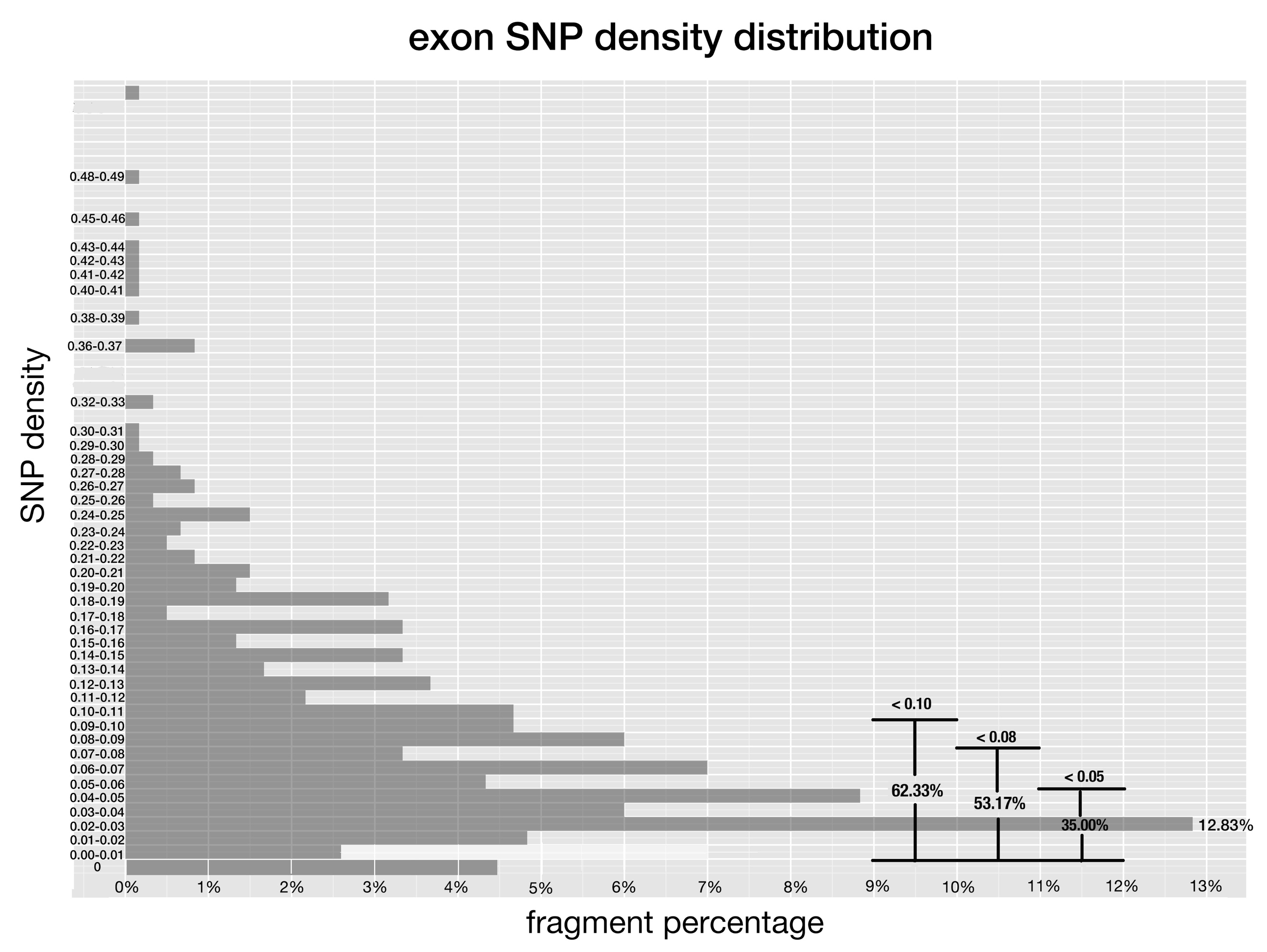
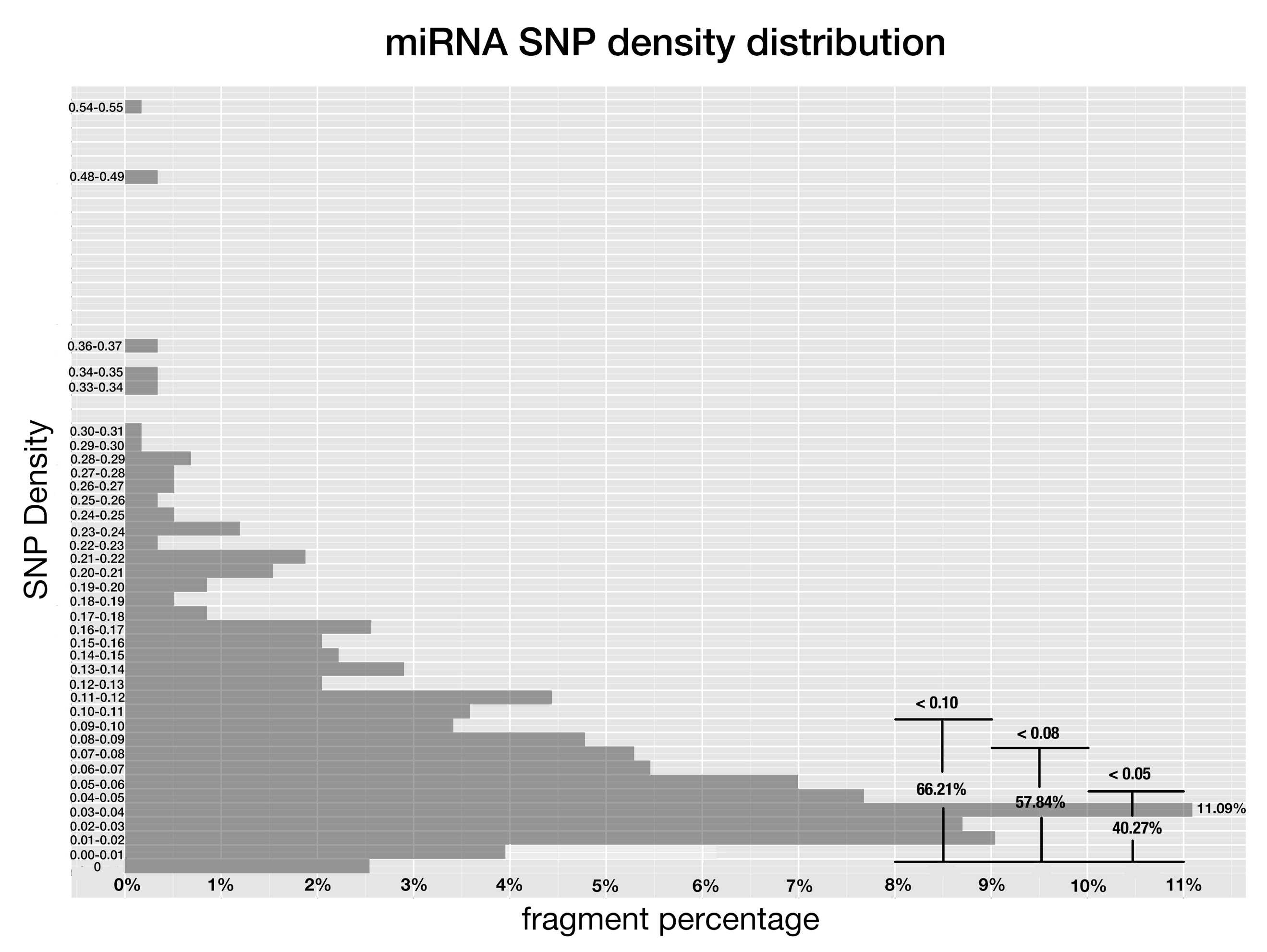
**Results**

**Part I. Overall SNP density of rice miRNAs**

SNP density can reflect the selection pressure imposed on given genomic regions. The higher the pressure is; the lower the SNP density will be. Since miRNAs are functional genomic units and master regulators, their SNP density may reflect different selection pressure compared with that of intergenic regions and exons [!!!Needs Quotation]. Compared with intergenic regions, SNP density of pre-miRNAs and exons are supposed to be lower. I obtained SNPs from Rice SNP-Seek Database (derived from 3,000 Rice Genome Project, *snp-seek.irri.org/*) [2], for all rice pre-miRNAs which were deposited in *miRBase.org/* (miRBase release v 20) [3] and randomly selected intergenic regions and exons across rice genome. Then SNP density distribution of rice pre-miRNAs, intergenic regions and exons were plotted and compared.

A general trend of SNP density distribution of pre-miRNAs was observed that fragment percentage increased from 0-0.03 and it peaked at range 0.03-0.04 with percentage of 11.09%, then fragment percentage kept dropping after 0.04 (Fig. 1). While the overall distribution of SNP density in exons sampled in this study was similar to that of pre-miRNAs except that the peak value fell within the range of 0.02-0.03 with percentage of 12.83% (Fig. 2). But as for the distribution of SNP density in intergenic regions sampled in this study, no obvious trend was found (Fig. 3). Through comparing the fragment percentage fell within ranges 0-0.10, 0-0.08 and 0-0.05 (Fig. 1, 2, 3), it was shown that SNPs of pre-miRNAs and exons clustered denser at low ranges than that of intergenic regions, which implied both pre-miRNAs and exons are under stricter evolutionary pressure than intergenic regions and furthermore this may be explained by the fact that pre-miRNAs and exons are functional more important. Compared with SNPs of exons, SNPs of pre-miRNAs clustered denser at low ranges, and this suggests pre-miRNAs are more strictly selected than exons.

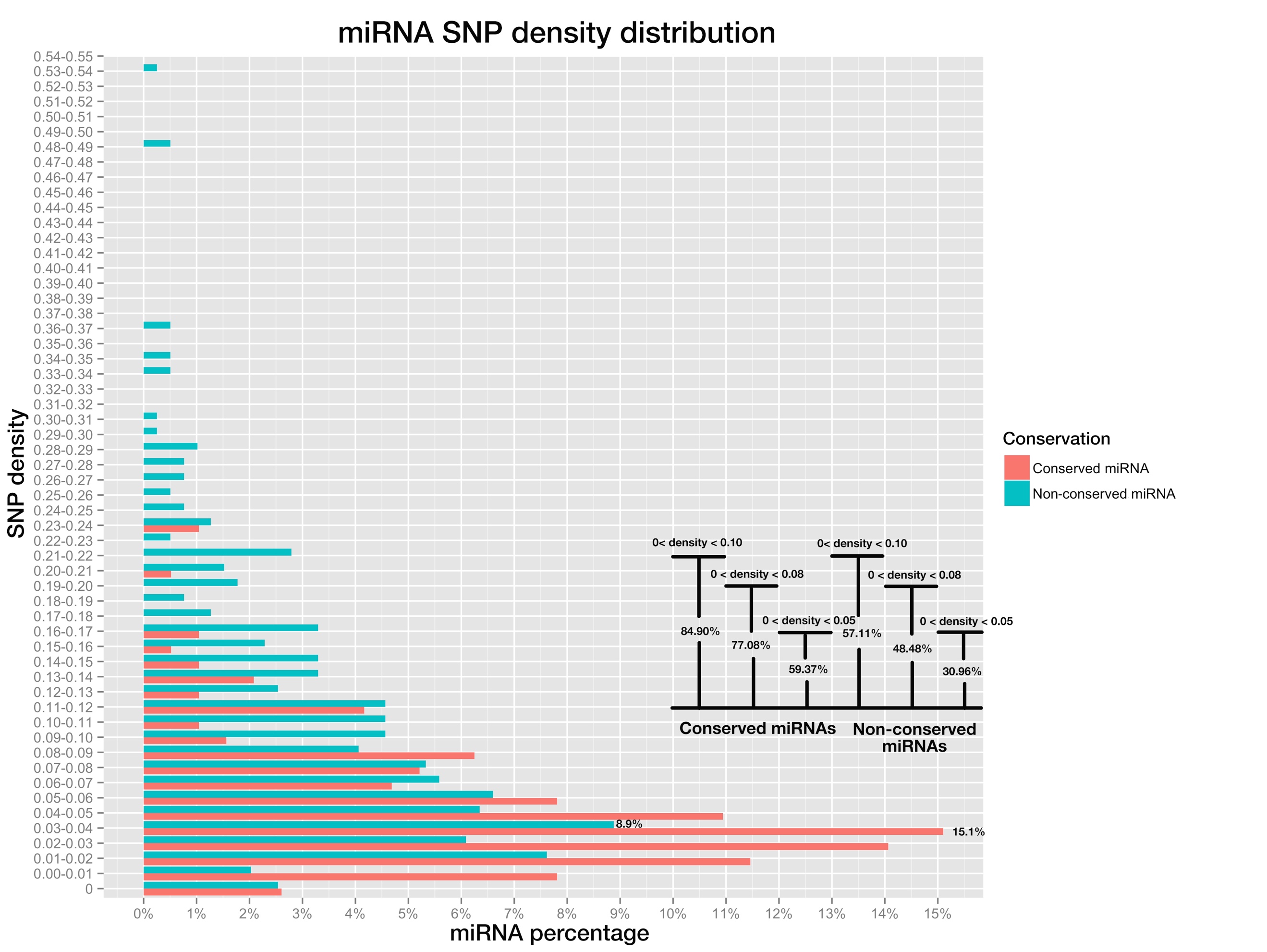


**Fig 1-3. SNP density of pre-miRNAs, exon regions and intergenic regions.**

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

Conserved miRNAs are evolutionarily more important with function than non-conserved ones in plants. Thus, it was appropriately supposed that SNPs of conserved miRNAs would cluster denser at low range than that of non-conserved ones. I classified pre-miRNAs into conserved and non-conserved ones, and drew histogram for them respectively.

The trend of SNP density distribution of both conserved miRNAs and non-conserved miRNAs was similar to that of all pre-miRNAs, except for different percentage values at the peak, which was 15.1% for conserved miRNAs and 8.9% for non-conserved miRNAs. SNPs of conserved miRNAs cluster denser at low ranges than that of non-conserved miRNAs, furthermore, the differences of the miRNA percentage that fell at the low range between conserved miRNAs and non-conserved ones are greater than the differences between pre-miRNAs and intergenic regions (Fig. 4), which implies the difference between evolutionary pressure imposed on conserved miRNAs and non-conserved miRNAs is even bigger than the evolutionary pressure difference between pre-miRNAs and intergenic regions.



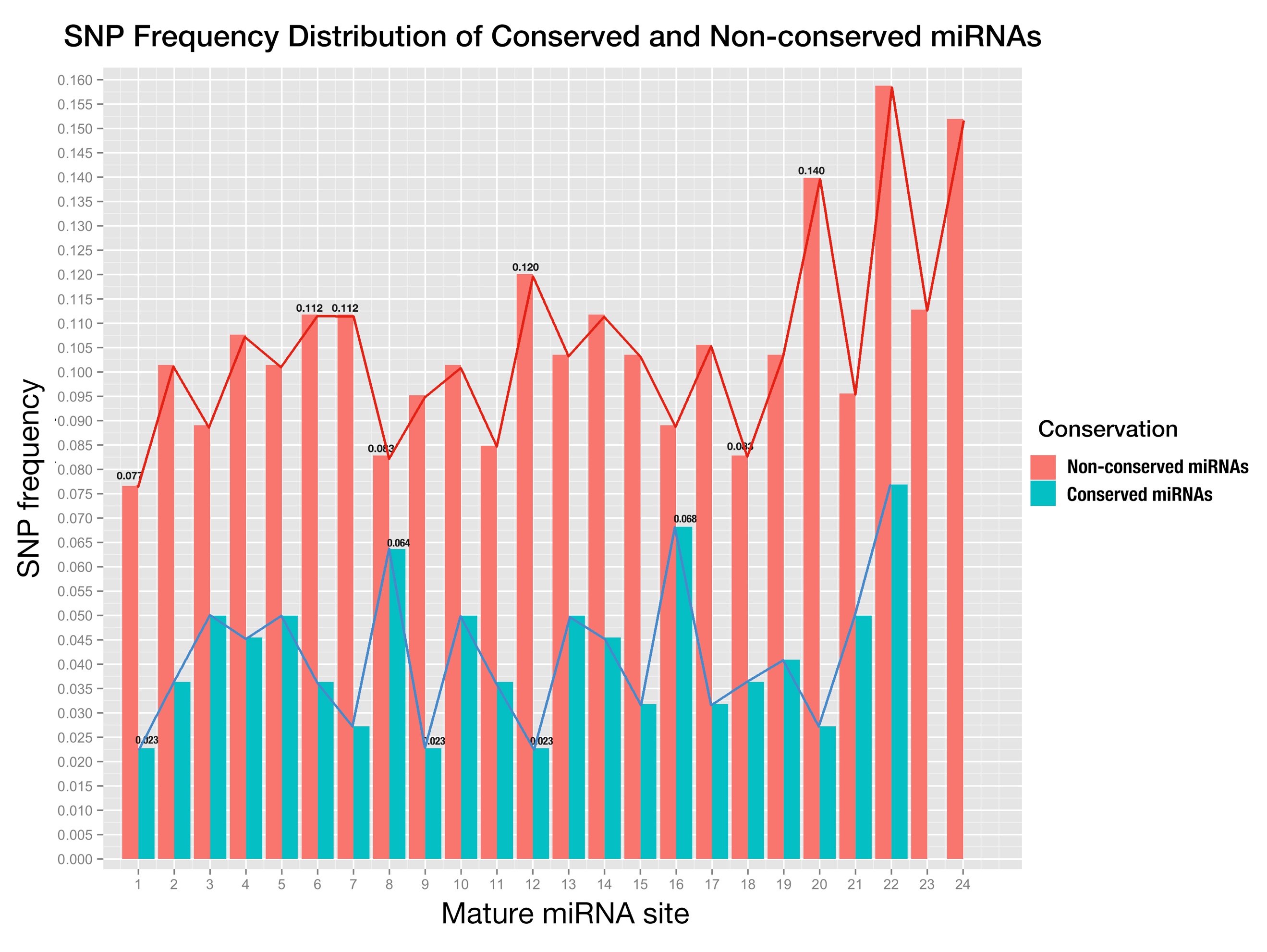
**Fig 4. miRNA SNP density distribution of pre-miRNAs, both conserved miRNAs (in red color) and non-conserved miRNAs (blue color).**

SNP density of a miRNA is the division of SNP number by miRNA length. Bar plot on 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.

**Part II. SNP distribution along mature miRNAs and miRNA binding sites**

Natural occurring sites of miRNAs have variable efficacies attributable to their complementarity patterns [4], which are determined by positions along the mature miRNAs and their binding sites of cognate targets co-jointly. Thus, different positions diversify in their importance in target silencing. The importance may be reflected by the SNP frequency of positions and lower frequency implies the given position is under stricter selection pressure and is more important for miRNA:target interaction. Pairing at position 10 and 11 is thought to be critical for plant miRNA functioning through cleavage [5,6,7], 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, SNP frequencies of all positions of both conserved and non-conserved miRNAs were calculated and put in a bar-plot for comparison in our study.

Overall, SNP frequency of conserved miRNA is lower than that of non-conserved miRNAs at each position as expected. As most miRNAs are 21-nt in length, our focus is on position 1-21.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, for non-conserved miRNAs; 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.

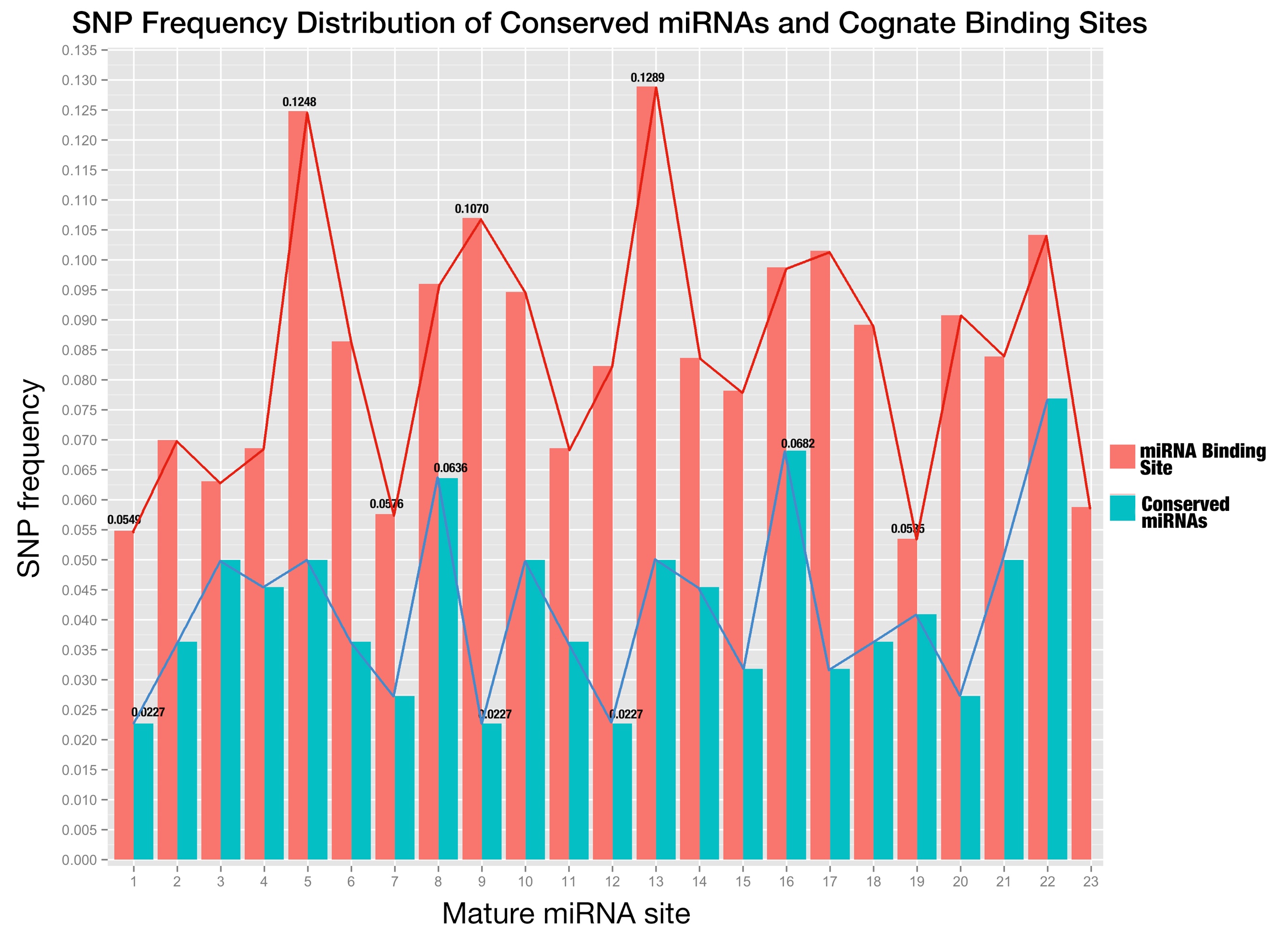


**Fig 5. SNP distribution of all sites along mature miRNA, both conserved miRNAs (blue) and non-conserved miRNAs (red).**

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;

Since conserved miRNAs have more identifiable targets than non-conserved miRNAs in plants [8] and are functionally more important, we studied only binding sites of genes targeted by conserved miRNAs. Plant miRNA regulates target mRNA through complementarity, which is determined by positions along miRNA and its binding site on the cognate target. And SNP frequencies of positions along rice miRNA and its binding sites may reflect the correlation between them. 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* [9] and transcriptome-wide degradome validation of rice miRNA targets [10], 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.



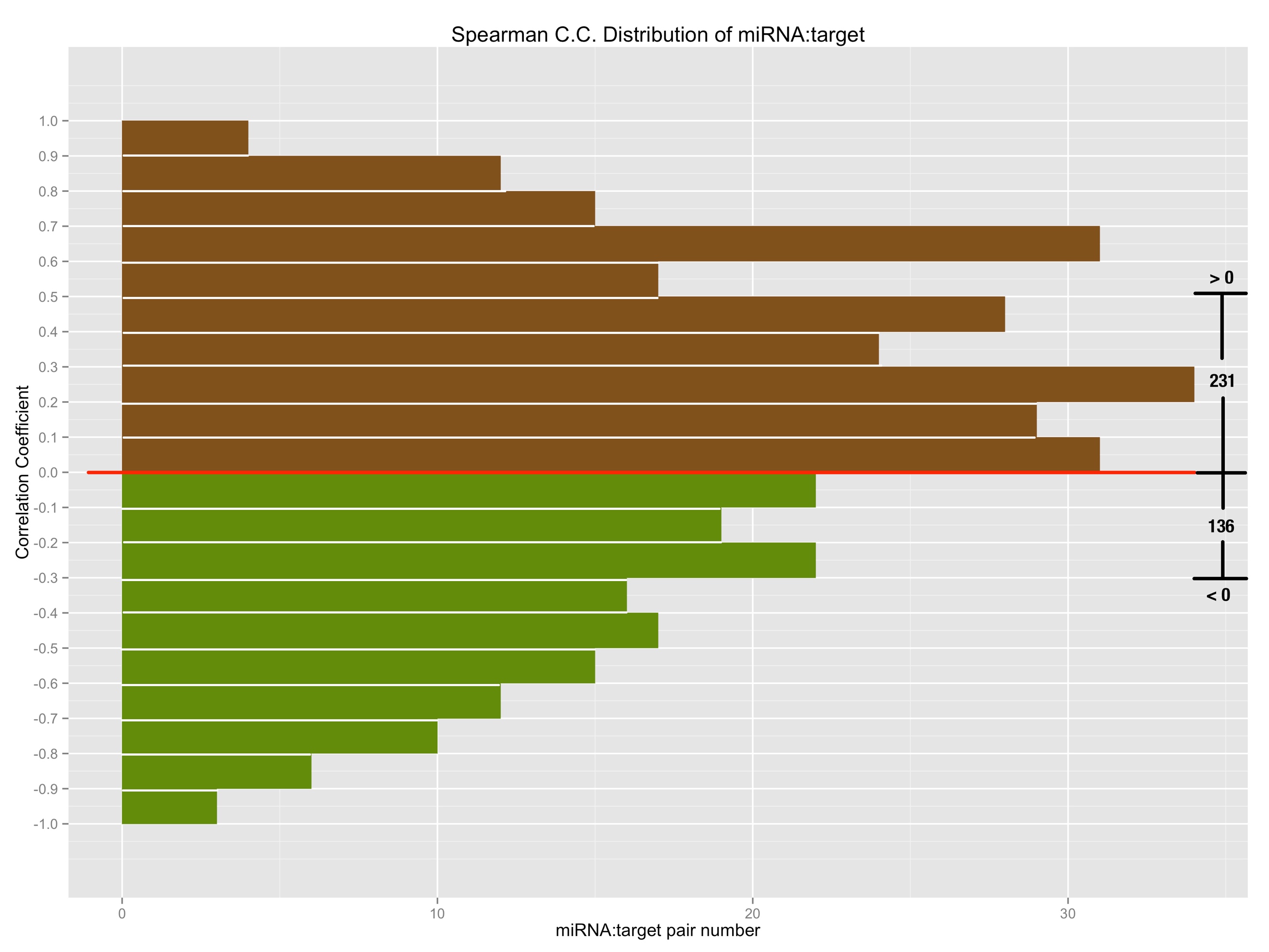
**Fig 6. SNP distribution of all sites along conserved mature miRNAs and their binding sites.**

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 [11, 12], 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 [13]. To verify this hypothesis, expression data of rice miRNAs and genes was extracted from RiceFREND database [14], and correlation test was performed on miRNA:target pairs validated by degradome with 3-week old rice seedling samples [10].

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;**

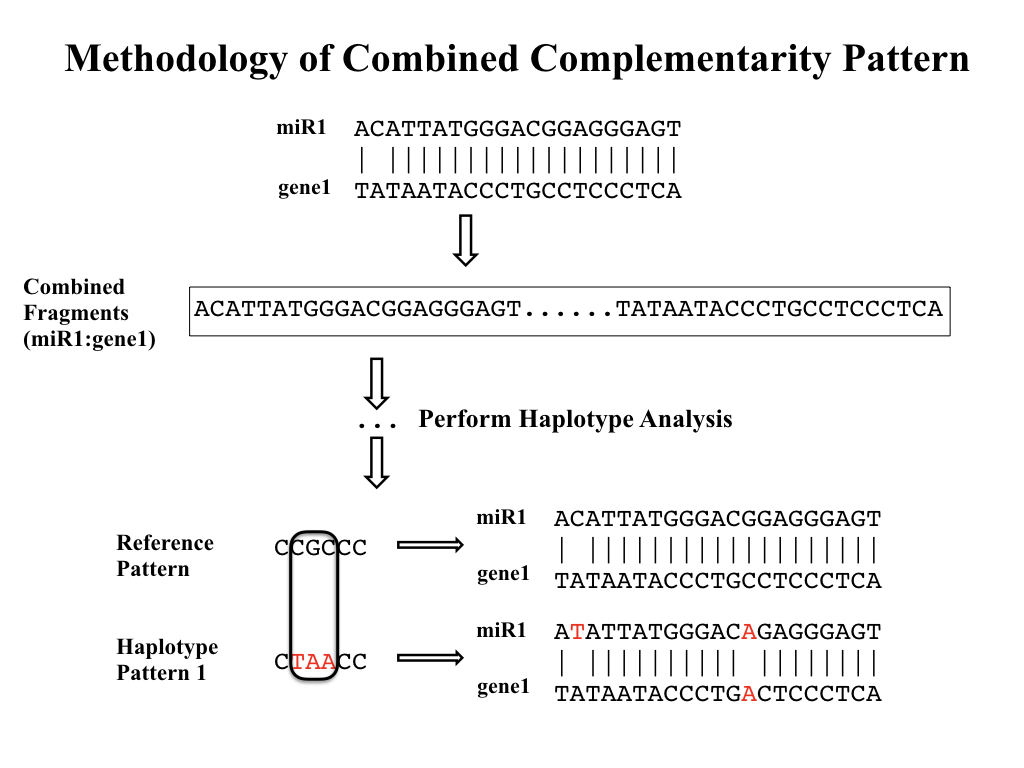
X-axis is the number of pairs that fall on the specified range of correlation coefficient; so bars with green color in the lower part denote the negatively correlated miRNA:target pairs, by contrast, bars with brown color in the upper part denote the positively correlated miRNA:target pairs.

**Part IV. Combined Complementarity Pattern Analysis Method**

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. Finally, the phenotypes or other features of different rice cultivar could be analyzed set-by-set with the different influences SNPs brought up to the miRNA functioning.



**Fig 8. Workflow of Combined Complementarity Pattern Analysis**

**Part V. Combined Complementarity Pattern Analysis Method**

Quotations:

[1] !!! Needs quotation

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

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

[4] 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).

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

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

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

plants. Genes Dev. 17, 49–63

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

[14] 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

[15]