

Role of SNPs in the biogenesis of mature miRNAs

Ying Wang

Shenyang Ligong University

Jidong Ru (✉ rujidong@126.com)

College of Light Industry and Textile, Qiqihar University

Meng Xianglian

Changzhou institute of Technology

Research article

Keywords: SNP, isomiRs, miRNA, splicing mechanism

Posted Date: May 26th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-29018/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BioMed Research International on June 17th, 2021. See the published version at <https://doi.org/10.1155/2021/2403418>.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Role of SNPs in the biogenesis of mature miRNAs

Ying Wang^{1,2}, Jidong Ru^{3,*}, Xianglian Meng⁴

¹ College of Equipment control, Shenyang Ligong University, No.6, nanping middle road, hunnan new district, Shenyang, Liaoning, 110159, China, ² College of Computer and Control Engineering, Qiqihar University, No.42, Wenhua Street, Qiqihar, Heilongjiang, 161006, China, ³ College of Light Industry and Textile, Qiqihar University, No.42, Wenhua Street, Qiqihar, Heilongjiang, 161006 China, ⁴ School of Computer Information & Engineering, Changzhou Institute of Technology, 213032, China

Email addresses of the authors

YW, wangying0129@126.com; RJD, rujidong@126.com; MXL, mengxl@czust.edu.cn.

#Correspondence address:

Jidong Ru, PhD
Professor
No.42, Wenhua Street
Qiqihar University
China
E-mail: rujidong@126.com

Keywords

SNP; isomiRs; miRNA; splicing mechanism;

26 **Abstract**

27 **Background**

28 SNPs within pre-miRNA regions play a significant role in miRNA generation, processing and
29 function by different molecular mechanisms and are thought to be major contributors to the
30 variations in phenotypes and diseases. Therefore, whole-genome analysis of how SNPs affect
31 mature miRNA biogenesis is important for precision medicine.

32 **Results**

33 In this study, aiming to analyze the role of SNPs in mature miRNA biogenesis genome-wide, we
34 constructed a SNP-pre-miRNA database, named miRSNPBase, consisting of 886 pre-miRNAs
35 and 2640 SNPs based on the latest data. Then, we identified 10574 SNP-pre-miRNAs based on
36 886 pre-miRNAs and their associated 2640 SNPs, and we performed genome-wide association
37 analyses to identify isoform miRNAs (isomiRs) based on miRFind that are associated with the
38 mechanism of SNPs affecting miRNA maturation. A total of 4235 nor-SNP-pre-miRNAs based
39 on 480 nor-pre-miRNAs and 1250 nor-SNPs were identified. We analyzed the effects of SNP
40 type, SNP location and SNP-mediated free energy change during mature miRNA biogenesis and
41 found that they are closely related to mature miRNA biogenesis. In addition, the MAF
42 distribution of the iso-pre-miRNAs and nor-SNPs was analyzed based on the 1000 Genomes
43 Project. The results demonstrated that individuals who contained the iso-SNPs were in
44 the minority, and those who contained the nor-SNPs were in the majority. Notably, to verify our
45 method and identify important biomarkers, we identified isomiRs and iso-SNPs in 18 GBR
46 individuals of European origin. In the results, 209 iso-pre-miRNA candidates and 71 verified iso-
47 pre-miRNAs of the 18 GBR samples were identified, and 2667 isomiRs of 209 pre-miRNAs
48 were verified by miRNA sequencing data.

49 **Conclusions**

50 Our results clearly indicated that SNPs that altered the mature miRNA splicing mechanism and
51 led to the production of isomiRs, were closely related to and affected normal life processes, and
52 led to epigenetic changes and diseases.

53

54 **Background**

55

56 MicroRNAs (miRNAs) are single-stranded small noncoding RNAs 20~24 nucleotides (nt) in
57 length that regulate gene expression at the posttranscriptional level. It is estimated that miRNAs
58 regulate approximately 60% of the transcription process in humans. The abnormal expression of
59 miRNAs is closely related to many diseases, and more than 50% of human miRNAs are located
60 in gene fragment sections and are related to cancers, including breast cancer, lung cancer,
61 colorectal cancer, skin cancer, nasopharyngeal carcinoma, ovarian cancer, and nerve cell
62 carcinoma. In addition, single nucleotide polymorphisms (SNPs) are DNA sequence
63 polymorphisms caused by a single individual diversity variation at the genome level. More than
64 38 million SNPs, including 140 million small insertion/deletion events and 14,000 structure
65 variations, exist in the human genome and contribute to human phenotypic differences and
66 diseases as molecular markers identified in different research fields [1]. GWAS have reported
67 1300 variations and 200 phenotypes [2, 3]. Remarkably, approximately 88% of the variations in
68 disease-related SNPs are the most abundant form of noncoding region variations in the human
69 genome, thousands of SNPs exist in the miRNA sequences and their upstream and downstream
70 flanking regions, and more than 40% of pre-miRNAs contain at least one SNP [4]. SNPs within
71 pre-miRNA regions may be responsible for several of the reported associations between SNPs,
72 miRNAs and complex human phenotypes and diseases.

73

74

75 MiRNAs are transcribed from genomic DNA and are spliced in several steps that heavily depend
76 on correct secondary and tertiary structures. The secondary structure is determined by the RNA
77 sequence, which is in turn determined by the genomic sequence. The nucleotide composition of
78 miRNA sequences plays important roles in the biogenesis of mature miRNAs, and variations
79 affect the molecular structure, thermodynamic stability and functional strand selection of miRNA
80 sequences. The sequence features of pre-miRNAs are arbitrary but follow some rules, such as the
81 striking frequency of GGAC in pre-miRNA sequences [5] and the strong preference for 'U' in
82 the first nucleotide position of the 5' arm of mature miRNAs [6].

83 The SNPs within pre-miRNA regions alter pre-miRNA sequences by adding, replacing, or
84 deleting nucleotides [7], and the structure and the minimum free energy of SNP-related miRNAs
85 are altered [8, 9]. For example, 11 SNPs located in the stem of pre-miRNAs reduce the number
86 of mature miRNA products, and the thermodynamic changes of more than 44% of pre-miRNAs
87 are greater than 2.0 kcal/mole [10].

88

89

90 SNPs within pre-miRNA regions affect the required secondary structure and thermodynamics
91 and alter the miRNA maturation process, including Drosha enzyme processing, Dicer enzyme
92 processing and functional strand choice [6, 7], which are closely related to a variety of
93 phenotypes and diseases [11, 12]. The processing of quite a few pre-miRNAs has been reported
94 to be altered by SNPs; for example, miR-125a, which contains a SNP in the seed region,
95 produces one mature miRNA in one arm, and the miRNA from the other arm is prevented from

96 undergoing Drosha enzyme processing[9]. Specifically, SNPs change the pre-miRNA processing
97 sites[13]; for example, miR-934, which contains rs73558572, generates five mature miRNAs that
98 are offset (1-2 nt) from the 3' arm reference mature miRNA[14]. As a result, SNPs lead to
99 imprecise precursor cropping or dicing and affect the expression level of miRNA[15]. For
100 example, when the terminal loop of pre-miRNA is changed, mature miRNA levels are reduced
101 by 60% [16]. Sun et al [14] demonstrated that this process is one of the mechanisms for isomiR
102 generation[17].

103

104

105

106 Although a multitude of SNPs have been found to play a significant role in miRNA generation,
107 processing and function by different molecular mechanisms and are closely related to various
108 diseases and phenotypes[14], exactly how SNPs affect mature miRNA biogenesis remains
109 unclear; therefore, studies on SNP-affected miRNA maturation mechanisms can provide
110 evidence for causal SNPs and contribute to precision medicine.

111

112 In this study, our aim was to perform a genome-wide analysis of the role of SNPs in the
113 biogenesis of mature miRNAs. We present a database, miRSNPBase, that provides
114 comprehensive information about SNPs and SNP-related miRNA loci. Then, all the sequences of
115 pre-miRNAs containing SNPs were surveyed using the position in the human genome as
116 the correlation factor. Four splice sites of each SNP-pre-miRNA were predicted based on
117 miRFind to further identify all the normal miRNAs and isomiRs. Based on the relationship with
118 the produced miRNAs, pre-miRNAs and SNPs were divided into iso-pre-miRNAs and iso-SNPs,

119 which lead to isomiR production, and nor-SNPs and nor-pre-miRNAs, which contribute to
120 normal miRNA production. Furthermore, the isomiRs and iso-SNP candidates were presented
121 for subsequent studies. Aiming to evaluate how the SNPs affect miRNA maturation mechanisms
122 and penetrance in individuals, the minor allele frequencies of the SNPs included in the normal
123 sequences and isomiRs were statistically analyzed. In addition, we analyzed the effects of SNP
124 type, SNP location and SNP-affected free energy change during the biogenesis of mature
125 miRNAs. Specifically, we verified the predicted isomiRs based on miRNA sequencing data for
126 18 GBR individuals. The schematic of the overall method is illustrated in Figure 1.

127 **Methods**

128

129 *Data*

130

131

132

133 *SNP-related dataset*

134 Information on SNPs was obtained from the 1000 Genomes Project ([ftp://ftp-
135 trace.ncbi.nih.gov/1000genomes/ftp/release/20130502/](ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/20130502/)) which was constructed based on 2,504
136 individuals from 26 populations and included over 84.7 million SNPs and >99% of SNPs with a
137 frequency of >1% for a variety of ancestries.

138

139 *MiRNASNP dataset*

140 The miRNASNP dataset (<http://bioinfo.life.hust.edu.cn/miRNASNP2/download.php>) is a
141 database that catalogs a total of 2257 SNPs in 1596 human pre-miRNAs based on miRBase
142 version 19 and the dbSNP database (version 137).

143

144 *Validation data*

145 The validation data include VCF files, the miRNA sequencing data of GBR (British in England
146 and Scotland) and the *Homo sapiens* GRCh37 reference sequence.

147

148 The 1000 Genomes project consists of a total of 26 subpopulations from five major populations
149 (Americans, Europeans, East Asians, South Asians and Africans). The VCF file contains the
150 final variant call set with phased genotypes for chr1-22, chrX and chrY consisting of 2504
151 individuals from 26 populations based on the phase3 analysis of the 1000 Genomes sequence
152 data. Eighteen GBR individuals of European origin were selected for our study. We downloaded
153 the VCF files of GBR population-specific SNP data of the 18 individuals from the ftp server of
154 the 1000 Genomes project (<ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>)[18, 19].

155

156 The miRNA sequencing data consist of the set of human lymphoblastic cell line samples from
157 the GBR population. is the dataset was downloaded from
158 <https://www.ebi.ac.uk/arrayexpress/experiments/E-GEUV-2/samples/>.

159

160 The *Homo sapiens* GRCh37 reference sequence was from 1000 Genomes project phase 2 and
161 was downloaded from
162 ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/human_g1k_v37.fasta.gz.

163

164 *Identification of SNP-related miRNAs (SNP-miRNAs)*

165 The attribute of miRNA genome position information is described as:

166

167 $Attr_{mir_pos}=(Hogg \text{ and } Harries);$

168 The attributes of the SNPs are listed as follows:

169

170 $Attr_{snp}=\{ \text{Chromosome, Coordinates, SNPID, Reference, Alter, Qual, Filter, Number of}$
171 $\text{allele, Frequency of allele, Amount of allele, Type of variation, Read of the variation } \};$

172

173 The position information of miRNAs and SNPs were mapped based on the hg38 and hg19
174 coordinates, and liftOver (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>) was used to convert the
175 coordinates from hg38 to hg19. Then, based on the coordinates, all SNPs were mapped to the
176 pre-miRNAs, and the SNP extraction conditions are described as $Attr_{snp}\{\text{Coordinates}\} \in$
177 $Attr_{mir_pos}\{\text{End} - \text{Start}\}$. All identified pre-miRNAs and SNPs were used to construct our
178 database miRSNPBase. The SNP-pre-miRNA information in miRSNPBase is defined as
179 $Attr_{SNP_pre}=\{ \text{name, plus/trans strand, SNP positions, start of miRNA, end of miRNA, reference}$
180 $\text{nucleotide, alter nucleotide, minor allele frequency}\}$, among them, minor allele frequency, MAF,
181 is the proportion of the SNP variable genes in the species.

182

183 Because our miRSNPBase was constructed based on the 1000 Genomes project, we expanded
184 our database with miRNASNP, which was developed on the basis of NCBI dbSNP (version 137).

185

186 *Identification of four splicing sites of SNP-pre-miRNA*

187 To construct the SNP-pre-miRNA sequences, SNPs were mapped to pre-miRNA sequences
188 based on coordinates. To avoid favoritism, SNP-pre-miRNA sequences of each SNP-associated

189 pre-miRNA were constructed based on the combination method, and the specific nucleotides of
190 corresponding positions in the pre-miRNAs were substituted by SNPs. The number of SNP-pre-
191 miRNAs is shown as:

$$192 \quad N_i = C_n^1 + C_n^2 + \dots + C_n^k + \dots + C_n^n \quad (1-1)$$

193 where N_i is the number of SNP-pre-miRNAs for the i th pre-miRNA and n is the number of SNPs
194 mapping to one pre-miRNA based on coordinates. An example of constructing a SNP-pre-
195 miRNA sequence is illustrated in Figure 2:

196
197 In this process, the SNP position is defined based on the coordinates of the plus or trans strand.
198 For two different situations, we used two different conversion methods. The calculation methods
199 for the variation position of the plus strand and trans strand are illustrated in Figure 3:

200
201 To systematically identify mature miRNAs of the SNP-pre-miRNAs, miRFind software
202 developed in our previous work [20] was used to identify the splicing sites. miRFind is software
203 aiming to identify mature miRNAs within pre-miRNA, and it provides five mature miRNA
204 candidates with an accuracy as high as 68%. We defined the start and end sites of the 5' arm
205 mature miRNA as P5_5 and P5_3, similarly, the start and end sites of the 3' arm mature miRNA
206 as P3_5 and P3_3. Based on the identified mature miRNAs, we extracted the normal miRNAs
207 (nor-miRNAs) and isomiRs. After counting the splicing site offsets between the five predicted
208 miRNA candidates and reference mature miRNAs, we defined the SNP-pre-miRNAs, pre-
209 miRNAs and SNPs whose splicing sites were changed as iso-SNP-pre-miRNAs, iso-pre-
210 miRNAs and iso-SNPs, respectively, and not changed as nor-SNP-pre-miRNAs, nor-pre-
211 miRNAs and nor-SNPs, respectively. Specifically, a nor-iso-pre-miRNA is defined as a pre-
212 miRNA with different SNPs that produces normal miRNAs and isomiRs at the same time.

213

214

215 *Effects of SNPs on the biogenesis of mature miRNAs*

216 To research the effects of SNP position, SNP type and SNP-affected free energy change on the
217 mature miRNA splicing mechanism, we investigated the enrichment of the iso-pre-miRNAs and
218 nor-pre-miRNAs in the SNP positions and the distribution of SNP types of iso-pre-miRNAs and
219 nor-pre-miRNAs. Moreover, we researched the enrichment of SNP-pre-miRNAs based on free
220 energy change by extracting the free energy changes between the iso-SNP-pre-miRNAs and
221 normal pre-miRNAs and between the nor-SNP-pre-miRNAs and normal pre-miRNAs. The free
222 energy of each normal pre-miRNA and SNP-pre-miRNA was calculated using RNAfold[21].

223

224

225 *The MAF of SNPs in the iso-pre-miRNA and nor-pre-miRNA*

226 The iso-SNPs that potentially affect mature miRNA biogenesis are functional SNPs. To research
227 the distribution of iso-SNPs and nor-SNPs in humans, we obtained the MAF of the nor-SNPs and
228 iso-SNPs to analyze the rules of SNP genes in the population based on the 1000 Genomes project.
229 Specifically, to make the investigation results more representative, we only focused on iso-SNPs
230 and nor-SNPs.

231

232 *Identification and verification of isomiRs based on a GBR population from 1000 Genomes*

233 To verify our method and identify important biomarkers, the VCF files of GBR population-
234 specific SNP data and related miRNA sequencing data were selected in our study. First, we
235 extracted the SNPs of each chromosome based on the VCF files. In this process,

236 GenomeAnalysisTK.jar was used to compare the samples with the human GRCh37 reference
237 sequence to extract chromosome data and VCF files of the 23 chromosomes. Second, we
238 integrated all the SNPs of the 23 chromosomes. Based on miRSNPBase, we obtained the pre-
239 miRNA-SNP dataset of each sample and then the pre-miRNA-SNP sequences of this sample we
240 constructed. Third, we identified four Dicer sites of the pre-miRNA-SNP sequences, and then the
241 canonical miRNAs and isomiR candidates were extracted. Finally, we aligned all the isomiRs
242 with miRNA sequencing data of the corresponding sample. The miRNAs that can be found in the
243 miRNA sequencing data are the verified isomiRs.

244 **Results**

246 *Establishment of the miRSNPBase database*

247 A total of 1881 human pre-miRNAs were extracted based on the miRNA genome position
248 information. On the basis of the coordinates in the genome, the SNPs were mapped to the pre-
249 miRNAs and flanking regions, and we found 2146 SNPs located in the pre-miRNAs. Among
250 them, 995 SNPs were found in the dbSNP, and 1151 SNPs were not in the dbSNP. The results
251 demonstrate that our method affords a great degree of data integrity. As such, the miRNASNP
252 data were integrated to construct our database, named miRSNPBase. In total, miRSNPBase
253 includes 886 pre-miRNAs and 2640 SNPs (Additional file 1: Table 1), among them, 551 pre-
254 miRNAs have mature miRNA in the 5' arm and 566 pre-miRNAs have mature miRNA in the 3'
255 arm (Additional file 2: Table 2).

256
257

258 To compare SNP enrichment among the different regions of the pre-miRNAs, we accounted for
259 the number of SNPs located in five regions, the 3' flanking region, miRNA-5P, terminal loop,

260 miRNA-3P, and 5' flanking region of the pre-miRNA. We further characterized the enrichment
261 of SNPs located in the mature miRNAs of each pre-miRNA. The distribution of SNPs in the pre-
262 miRNA sequences is shown in Figure 4:

263

264

265 As shown in Figure 4, the enrichment of SNPs in the miRNA-5P region was higher than that in
266 the other regions. The enrichment of SNPs in the terminal loop was the lowest. In mature
267 miRNAs, SNPs were the most enriched in the 13th nucleotide position. The lowest enrichment
268 was in the 23rd position; since the length of most mature miRNAs is 22 nt, the 23rd position was
269 not considered, and the lowest enrichment of SNPs in pre-miRNAs was found in the 1st, 5th and
270 17th positions. For most positions, the 13th, 14th and 15th positions had the highest SNP
271 enrichment.

272

273

274 *Identification of splicing sites of SNP-pre-miRNA*

275 Each pre-miRNA and related SNPs were used to construct SNP-pre-miRNAs based on
276 nucleotide substitution and the composition method. Taking hsa-mir-1181 as an example, there
277 are 6 SNPs within hsa-mir-1181, based on which we constructed 63 SNP-pre-miRNAs. By the
278 same token, we constructed 10574 SNP-pre-miRNAs using 886 pre-miRNAs and their
279 associated 2640 SNPs. When we identified the mature miRNAs of each SNP-pre-miRNA using
280 miRFind, the five mature miRNA candidates were considered to improve the identified accuracy.

281

282 Based on the miRFind results, the mature miRNAs were divided into nor-miRNAs and isomiRs;
283 correspondingly, the SNP-pre-miRNAs associated with nor-miRNAs and isomiRs were named

284 nor-SNP-pre-miRNAs and iso-SNP-per-miRNAs, respectively. The pre-miRNAs associated with
285 nor-miRNAs and iso-miRNAs were named nor-pre-miRNAs and iso-per-miRNAs, respectively.
286 The result of mature miRNA identification of SNP-pre-miRNAs is illustrated in Table 2:

287

288

289 As shown in Table 2, the enrichment of iso-SNP-pre-miRNAs, iso-pre-miRNAs and iso-SNPs
290 was lower than that of nor-SNP-pre-miRNAs, nor-pre-miRNAs and nor-SNPs in P5_5, and the
291 enrichment of iso-SNP-pre-miRNAs, iso-pre-miRNAs and iso-SNPs was higher than that of nor-
292 SNP-pre-miRNAs, nor-pre-miRNAs and nor-SNPs in other sites. Taking P5_5 as an example,
293 we identified 607 iso-SNP-pre-miRNAs that contained 143 iso-pre-miRNAs and 427 iso-SNPs.
294 Conversely, we identified 4235 nor-SNP-pre-miRNAs that contained 480 nor-pre-miRNAs and
295 1250 nor-SNPs. All the iso-pre-miRNAs, nor-pre-miRNAs, nor-SNPs and iso-SNPs associated
296 with the four splicing sites are shown in Additional file 3: Table 3.

297

298 The distribution of pre-miRNAs and SNPs associated with the normal miRNAs and isomiRs are
299 shown in Figure 5 and Additional file 4: Table 4:

300

301

302 The common nor-pre-miRNAs, nor-SNPs, iso-pre-miRNAs and iso-SNPs of the four splicing
303 sites were extracted, and the results suggested that these nor-pre-miRNAs tended to remain in the
304 normal pre-miRNA splicing process. The nor-SNPs tended to not affect the Drosha and Dicer
305 enzyme splicing of pre-miRNAs. The iso-pre-miRNAs tended to be spliced into isomiRs, and the

306 nor-SNPs tended to change the splicing sites and produce isomiRs. Remarkably, all these pre-
307 miRNAs and SNPs can be used as candidates for future studies.

308

309

310 *Analysis of the MAF of SNPs in the iso-pre-miRNAs and nor-pre-miRNAs*

311 The MAF of iso-SNPs and nor-SNPs is illustrated in Figure 6:

312

313

314 The highest MAF of nor-SNPs is '1', which means that a specific set of SNPs will not affect
315 mature miRNA biogenesis. The highest MAF of the iso-SNPs did not reach '1', which means
316 that there is a specific set of SNPs in the samples that does not affect mature miRNA biogenesis,
317 and a specific set of SNPs that always affects miRNA maturation does not exist.

318

319

320 As shown in Figure 6, we sorted the MAF based on value and compared the value of the bottom
321 75% and 50%. For all sites, the MAF of iso-SNPs of iso-SNPs was lower than that of nor-SNPs,
322 which means that the population with SNPs that can affect miRNA mature biogenesis is always
323 less than the population with SNPs that do not affect miRNA mature biogenesis.

324

325

326 Based on the MAF of iso-SNPs and nor-SNPs, most molecules with SNPs underwent the
327 traditional mature miRNA splicing process, which is consistent with the existence of
328 heterogeneous miRNAs and mature miRNAs. The results suggested that SNPs altered the

329 biogenesis of mature miRNAs, which would result in the production of many isomiRs thus
330 changing miRNA expression and affecting normal cellular processes, which can lead to more
331 diseases, but this scenario is obviously not what is actually observed. Therefore, our study
332 illustrates that the MAF of iso-SNPs is lower than that of nor-SNPs, proving that only a few
333 SNPs affect mature miRNA biogenesis.

334

335

336 *Effects of SNPs on the biogenesis of mature miRNAs*

337 The iso-pre-miRNAs, nor-pre-miRNA, iso-SNPs and nor-SNPs were analyzed to determine
338 which factors might play an important role in the biogenesis of mature miRNAs. We analyzed
339 the relationships between SNP position, SNP type, SNP-affected free energy change and splice
340 site variation of pre-miRNAs. Because isomiRs can be generated from mutations of all four
341 splicing sites, we analyzed all miRNA sites. The enrichment of pre-miRNAs based on SNP
342 positions is described in Figure 7:

343

344

345 Taking P5_5 as an example, the nor-pre-miRNAs had the highest enrichment and the lowest
346 enrichment when the SNPs were located in section 1 and section 3, respectively. Accordingly,
347 the iso-pre-miRNAs had the highest and lowest enrichment when the SNPs were located in
348 section 4 and section 3, respectively. When the SNPs were located in section 1, the iso-pre-
349 miRNAs had a higher enrichment in that region and the nor-pre-miRNAs had a lower enrichment,
350 and their enrichments were insignificantly different. Additionally, when the SNPs were located

351 in section 3, the iso-pre-miRNAs had a lower enrichment, the nor-pre-miRNAs had a higher
352 enrichment, and their enrichments were significantly different.

353

354

355 For P5_5-associated pre-miRNAs, when SNPs were located in section 1, the SNP-pre-miRNAs
356 tended to splice and produce isomiRs, and when SNPs were located in sections 2 and 3, the SNP-
357 pre-miRNAs tended to splice and produce normal miRNAs. In a similar way, for P5_3
358 associated pre-miRNAs, when SNPs were located in sections 2 and 5, the SNP-pre-miRNAs
359 tended to splice and produce normal miRNAs, and when SNPs were located in the other sections,
360 the SNP-pre-miRNAs tended to splice and produce isomiRs. For P3_5- and P3_3-associated pre-
361 miRNAs, when the SNPs were located in sections 1 and 2, the SNP-pre-miRNAs tended to
362 splice and produce normal miRNAs, and when SNPs were located in the other sections, the SNP-
363 pre-miRNAs tended to splice and produce isomiRs.

364

365

366

367 We analyzed the enrichment of pre-miRNAs in SNPs located in the mature miRNAs, as shown
368 in Figure 8.

369

370

371 We focused on three cases: (1) The enrichment of SNPs in iso-pre-miRNAs/nor-pre-miRNAs is
372 higher than that in nor-pre-miRNAs/iso-pre-miRNAs; (2) the enrichment of SNPs in one class of
373 pre-miRNAs has a large degree of change in specific sites of mature miRNAs; and (3) the

374 enrichments of SNPs in iso-pre-miRNAs and nor-pre-miRNAs have a slight change in specific
375 sites in mature miRNAs.

376

377

378 For P5_5 splicing-related mature miRNAs, when the SNPs were located in the 7 nt, 9 nt and 15
379 nt positions, the enrichment in nor-pre-miRNAs was higher and that in iso-pre-miRNAs was
380 lower, and their enrichments were insignificantly different. Additionally, when the SNPs were
381 located in the 6 nt, 8 nt and 13 nt positions, the iso-pre-miRNAs had a lower enrichment, the nor-
382 pre-miRNAs had a higher enrichment, and their enrichments were significantly different. When
383 the SNPs were located in 17-22 nt positions, the enrichment of nor-pre-miRNAs and iso-pre-
384 miRNAs was positive correlated. These findings suggest that pre-miRNAs containing SNPs in
385 the 7 nt, 9 nt and 15 nt positions tend to splice and produce normal miRNAs, pre-miRNAs with
386 SNPs located in the 6 nt, 8 nt and 13 nt positions tend to splice and produce isomiRs, and when
387 the SNPs are located in the 17-22 nt positions, they do not affect the pre-miRNA splicing process.

388

389

390 Similarly, for P5_3 splicing-related mature miRNAs, the results suggest that pre-miRNAs
391 containing SNPs in the 1 nt, 2 nt 4 nt, 6 nt, 11 nt 12 nt and 17 nt positions tend to splice and
392 produce normal miRNAs and tend to splice and produce isomiRs when SNPs are located in the 3
393 nt, 8 nt 13 nt and 14 nt positions, while the SNPs located in the 18-22 nt positions do not affect
394 the pre-miRNAs splicing process. For P3_5 splicing-related mature miRNAs, the data suggest
395 that pre-miRNAs containing SNPs in the 4 nt, 10 nt, 13 nt, 15 nt and 21 nt positions tend to
396 splice and produce normal miRNAs and tend to splice and produce isomiRs when SNPs are

397 located in the 2 nt, 9 nt and 22 nt positions, while the SNPs located in the 5 nt, 6 nt and 17 nt
398 positions do not affect the pre-miRNAs splicing process. For P3_5 splicing-related mature
399 miRNAs, the results suggest that pre-miRNAs containing SNPs in the 4 nt, 10 nt, 13 nt, 15 nt
400 and 21 nt positions tend to splice and produce normal miRNAs and tend to splice and produce
401 isomiRs when SNPs are located in the 2 nt, 9 nt and 22 nt positions, while the SNPs located in
402 the 5 nt, 6 nt and 17 nt positions do not affect the pre-miRNAs splicing process. For P3_3-
403 splicing related mature miRNAs, the data suggest that pre-miRNAs containing SNPs in the 2 nt,
404 4 nt, 15 nt and 22 nt positions tend to splice and produce normal miRNAs and tend to splice and
405 produce isomiRs when SNPs are located in the 3 nt, 4 nt, 16 nt and 21 nt positions, while the
406 SNPs located in the 5 nt, 10 nt, 11 nt, 18 nt, 19 nt and 20 nt positions do not affect the pre-
407 miRNAs splicing process.

408

409

410 In general, the enrichments of iso-pre-miRNAs and nor-pre-miRNAs are negatively correlated in
411 most sites of mature miRNAs, which is consistent with the actual observations; that is, as the
412 SNP-pre-miRNA counts to normal miRNA counts increase, the SNP-pre-miRNA counts to
413 isomiRs counts decrease.

414

415

416 The enrichment of pre-miRNAs and SNPs based on SNP type is described in Figure 9:

417

418

419 For P5_5 splicing-related mature miRNAs, when the SNP types are 'C' and 'G', the percent of
420 iso-pre-miRNAs is higher; otherwise, when the SNP types are 'A' and 'U', the percent of nor-
421 pre-miRNAs is higher. Additionally, as the SNP type is 'A', the difference in the percent
422 between iso-pre-miRNAs and nor-pre-miRNAs is largest. These observations suggest that when
423 SNP types are 'C' and 'G', the pre-miRNAs tend to splice and produce isomiRs, while when
424 SNP types are 'A' and 'U', pre-miRNAs tend to splice and produce normal miRNAs.

425

426

427 Similarly, for P5_3 splicing-related mature miRNAs, the results suggest that when SNP types are
428 'C' and 'G', the pre-miRNAs tend to splice and produce isomiRs, while they tend to splice and
429 produce normal miRNAs when SNP types are 'A' and 'U'. For P3_5 splicing-related mature
430 miRNAs, the results suggest that when SNP types are 'C', 'G' and 'U', the pre-miRNAs tend to
431 splice and produce isomiRs, while they tend to splice and produce normal miRNAs when the
432 SNP type is 'A'. For P3_3 splicing-related mature miRNAs, the data suggest that when SNP
433 types are 'G' and 'U', the pre-miRNAs tend to splice and produce isomiRs, while they tend to
434 splice and produce normal miRNAs when the SNP type is 'A'.

435

436

437

438 The enrichment of SNP-pre-miRNAs based on free energy change is described in Figure 10:

439

440 The change in free energy mostly focuses on 0-8 kcal/mol. As free energy increases greater than
441 4 kcal/mol, SNP-pre-miRNAs tend to shear and produce isomiRs. When the free energy change

442 decreased by more than 2 kcal/mol, for P5_5 and P3_5 mature miRNA loci, the percentage of
443 iso-SNP-pre-miRNAs is higher than that of nor-SNP-pre-miRNAs, and for P5_3 and P3_3
444 mature miRNA loci, the percentage of nor-SNP-pre-miRNAs is higher than that of iso-SNP-pre-
445 miRNAs. These findings suggest that the decrease in the free energy tends to alter the splicing
446 sites of the 5' end of mature miRNAs and is less affected by the 3' ends of mature miRNAs.

447

448

449 On the basis of the free energy change, the enrichment distribution of iso-SNP-pre-miRNAs and
450 nor-SNP-pre-miRNAs associated with the 5' end (P5_5 and P5_3) and 3' end (P5_3 and P3_3)
451 sites of mature miRNAs has similar characteristics, indicating that the effects of the free energy
452 change on 5' end and 3' end mature miRNA biogenesis are largely consistent.

453

454 For P3_3 mature miRNAs, when the free energy increases, SNP-pre-miRNAs tend to maintain
455 normal mature miRNA biogenesis, and when the free energy decreases, SNP-pre-miRNAs tend
456 to splice and produce normal mature miRNAs.

457

458 *Identification and verification of isomiRs based on a GBR population from 1000 Genomes*

459 Based on our method, we identified the isomiRs and iso-SNPs of 18 GBR individuals of
460 European origin. Taking HG00097 as an example, because the sample VCF data were provided
461 by karyotype, we extracted the variation information of HG00097 from 23 chromosome files and
462 integrated all the variation information. All SNPs were mapped to pre-miRNAs to construct the
463 SNP-pre-miRNAs for HG00097. Furthermore, we identified the mature miRNA sequences at the
464 four sites. As a result, we predicted 695 isomiRs of 92 pre-miRNAs with 94 SNPs in a different

465 guide strand with the incorporation of variations in its sequence. The pre-miRNAs, iso-SNPs,
466 and isomiRs of HG00097 are summarized in Additional file 5: Table 5. The results suggest that
467 all SNPs within pre-miRNAs could have a potential impact on miRNA biogenesis and function.

468

469 Furthermore, we identified the isomiRs and iso-SNPs of 18 GBR individuals of European origin,
470 and 209 iso-pre-miRNA candidates and 71 verified iso-pre-miRNAs of the 18 GBR samples are
471 shown in Additional file 6: Table 6. In addition, 2667 isomiRs of 209 pre-miRNAs were verified
472 by miRNA sequencing data, and the isomiRs and iso-SNPs of the 18 GBR individuals are shown
473 in Table 3 and are detailed in Additional file 7: Table 7.

474

475

476 We validated them with the miRNA sequencing data. The 158 verified isomiRs of the 18 GBR
477 samples are shown in Table 4, and the details are shown in Additional file 8: Table 8:

478

479 **Discussion**

480 As an important molecular mechanism by which SNPs significantly contribute to miRNA
481 generation mechanisms and functions, experiments have shown that the molecular structure,
482 thermodynamic stability and functional strand selection are affected by SNPs located in pre-
483 miRNAs. As a result, SNPs influenced the selection of Drosha enzyme processing, Dicer enzyme
484 processing and functional strands in the splicing process of pre-miRNAs and altered the
485 expression levels of miRNAs, which is closely related to various phenotypes and diseases.
486 However, because the expression levels of iso-miRNAs are low, iso-miRNAs are hard to detect,

487 and the mechanism of SNP-affected mature miRNA biogenesis remains unclear. Therefore, we
488 systematically studied the role of SNPs in the biogenesis of mature miRNAs.

489

490

491 We constructed a SNP-associated pre-miRNA database based on the latest data from miRBase
492 and the 1000 Genomes project, and by integrating the miRNASNP database, we obtained our
493 database, named miRSNPBase.

494

495

496 We performed an analysis of the relationships between SNP type, SNP location and SNP-
497 affected free energy change and the biogenesis of mature miRNAs. The results showed that these
498 three factors played important roles in the mature miRNA generation mechanism. On the other
499 hand, we analyzed the penetrance of nor-SNPs and iso-SNPs based on the MAF. The results
500 proved that most SNPs do not alter normal miRNA production, and most individuals contain
501 these SNPs. Moreover, only a few SNPs affect mature miRNA biogenesis, and only a few
502 individuals contain such SNPs.

503

504 We identified isomiRs and iso-SNPs in 18 GBR individuals; specifically, these isomiRs were
505 verified by miRNA sequencing data of 18 GBR samples. As a result, we obtained epigenetic-
506 associated isomiRs and SNPs; furthermore, by comparison with the miRNA sequencing data, we
507 found and verified the presence of isomiRs.

508

509

510 Conclusions

511

512 Overall, our study suggested that SNPs affect the biological characteristics and lead to changes
513 in the Dicer sites of mature miRNAs undergoing the maturation process, therefore leading to the
514 generation of isomiRs. The distribution of the MAF of SNPs showed that only a few SNPs affect
515 the splicing mechanism of mature miRNAs, which is consistent with reported facts. In addition,
516 some isomiRs were verified based on miRNA sequencing data of 18 GBR individuals.
517 Identification of isomiRs in miRNA sequencing data also indicates that our method is effective.
518 In conclusion, the results suggest that the SNPs play an important role in the biogenesis of
519 mature miRNAs.

520

521

522 **Declarations**

523 **Ethics approval and consent to participate**

524 Not applicable

525

526 **Consent for publication**

527 Not applicable

528

529 **Availability of data and materials**

530 The miRNA genome position information used in this paper can be downloaded from
531 <ftp://mirbase.org/pub/mirbase/>. The information of SNPs described in this manuscript was
532 obtained from the 1000 Genomes Project ([23](ftp://ftp-</p></div><div data-bbox=)

533 trace.ncbi.nih.gov/1000genomes/ftp/release/20130502/). MiRNASNP dataset can be downloaded
534 from <http://bioinfo.life.hust.edu.cn/miRNASNP2/download.php>. The VCF files used in this
535 paper can be downloaded from <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>. And
536 the miRNA sequencing data described in this manuscript can be obtained from
537 <https://www.ebi.ac.uk/arrayexpress/experiments/E-GEUV-2/samples/>. The Homo sapiens
538 GRCh37 reference sequence used in this paper can be downloaded from
539 ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/human_g1k_v37.fasta.gz.

540

541 **Competing interests**

542 The authors declare that they do not have any competing commercial interests in relation to the
543 submitted work.

544 **Funding**

545

546 Publication charges for this article have been funded by the Youth Science Fund of Heilongjiang
547 Province of China (no. QC2017079), the science and technology project of Qiqihar (GYGG-
548 201915, GYGG-201709), the Basic Business of Educational Commission of Heilongjiang
549 Province of China (135109246), Humanities and Social Science Fund of Ministry of Education
550 of China (19YJCZH120), and National Natural Science Foundation (NNSF) of China
551 (61901063).

552

553 **Authors' contributions**

554 YW designed the work program, drafted the manuscript and wrote the code and implemented the
555 analysis. MXL constructed the database. RJD participated in the writing of the manuscript and
556 the interpretation of the results.

557

558

559 **References**

- 560 1. Shastry BS: **SNPs: impact on gene function and phenotype.** *Methods Mol Biol* 2009,
561 **578**:3-22.
- 562 2. Lander ES: **Initial impact of the sequencing of the human genome.** *Nature* 2011,
563 **470**(7333):187-197.
- 564 3. Ryan BM, Robles AI, Harris CC: **Genetic variation in microRNA networks: the**
565 **implications for cancer research.** *Nat Rev Cancer* 2010, **10**(6):389-402.
- 566 4. Jin Y, Lee C: **Single Nucleotide Polymorphisms Associated with MicroRNA**
567 **Regulation.** *Biomolecules* 2013, **3**(2):287-302.
- 568 5. Alarcon CR, Lee H, Goodarzi H, Halberg N, Tavazoie SF: **N6-methyladenosine marks**
569 **primary microRNAs for processing.** *Nature* 2015, **519**(7544):482-485.
- 570 6. Hu HY, Yan Z, Xu Y, Hu H, Menzel C, Zhou YH, Chen W, Khaitovich P: **Sequence**
571 **features associated with microRNA strand selection in humans and flies.** *BMC*
572 *genomics* 2009, **10**:413.
- 573 7. Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim
574 VN: **Molecular basis for the recognition of primary microRNAs by the Drosha-**
575 **DGCR8 complex.** *Cell* 2006, **125**(5):887-901.
- 576 8. Hiard S, Charlier C, Coppieters W, Georges M, Baurain D: **Patrocles: a database of**
577 **polymorphic miRNA-mediated gene regulation in vertebrates.** *Nucleic Acids Res*
578 2010, **38**(Database issue):D640-651.
- 579 9. Duan R, Pak C, Jin P: **Single nucleotide polymorphism associated with mature miR-**
580 **125a alters the processing of pri-miRNA.** *Hum Mol Genet* 2007, **16**(9):1124-1131.
- 581 10. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, Sun J, Guo AY: **Genome-wide**
582 **identification of SNPs in microRNA genes and the SNP effects on microRNA target**
583 **binding and biogenesis.** *Human mutation* 2012, **33**(1):254-263.
- 584 11. Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RK, Hua Y,
585 Gueroussov S, Najafabadi HS, Hughes TR *et al*: **RNA splicing. The human splicing**
586 **code reveals new insights into the genetic determinants of disease.** *Science* 2015,
587 **347**(6218):1254806.
- 588 12. Jin Y, Lee CG: **Single Nucleotide Polymorphisms Associated with MicroRNA**
589 **Regulation.** *Biomolecules* 2013, **3**(2):287-302.
- 590 13. Hogg DR, Harries LW: **Human genetic variation and its effect on miRNA biogenesis,**
591 **activity and function.** *Biochemical Society transactions* 2014, **42**(4):1184-1189.
- 592 14. Sun G, Yan J, Noltner K, Feng J, Li H, Sarkis DA, Sommer SS, Rossi JJ: **SNPs in**
593 **human miRNA genes affect biogenesis and function.** *Rna* 2009, **15**(9):1640-1651.

- 594 15. Li MJ, Yan B, Sham PC, Wang J: **Exploring the function of genetic variants in the**
595 **non-coding genomic regions: approaches for identifying human regulatory variants**
596 **affecting gene expression.** *Briefings in bioinformatics* 2015, **16**(3):393-412.
- 597 16. Zhang X, Zeng Y: **The terminal loop region controls microRNA processing by**
598 **Drosha and Dicer.** *Nucleic Acids Res* 2010, **38**(21):7689-7697.
- 599 17. Guo L, Chen F: **A challenge for miRNA: multiple isomiRs in miRNAomics.** *Gene*
600 2014, **544**(1):1-7.
- 601 18. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs
602 RA, Hurles ME, McVean GA: **A map of human genome variation from population-**
603 **scale sequencing.** *Nature* 2010, **467**(7319):1061-1073.
- 604 19. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM,
605 Handsaker RE, Kang HM, Marth GT, McVean GA: **An integrated map of genetic**
606 **variation from 1,092 human genomes.** *Nature* 2012, **491**(7422):56-65.
- 607 20. Wang Y, Li X, Tao B: **Improving classification of mature microRNA by solving class**
608 **imbalance problem.** *Scientific reports* 2016, **6**:25941.
- 609 21. Auyeung VC, Ulitsky I, McGeary SE, Bartel DP: **Beyond secondary structure:**
610 **primary-sequence determinants license pri-miRNA hairpins for processing.** *Cell*
611 2013, **152**(4):844-858.
612

613 **Figure Legends**

614

615 **Figure 1: The schematics of the overall method**

616 **Figure 2: The example of constructing the SNP-pre-miRNA sequence**

617 **Figure 3: The calculation method for variation position of plus-strand and trans-strand**

618 **Figure 4: The distribution of SNPs in the pre-miRNA sequences**

619 **Figure 5: The distribution of pre-miRNAs and SNPs associated with the normal and**

620 **isomiRs**

621 **Figure 6: The MAF of iso-SNPs and nor-SNPs**

622 **Figure 7: The enrichment of pre-miRNAs based on SNPs positions**

623 **Figure 8: The enrichment of pre-miRNAs as SNPs located in the mature miRNAs**

624 **Figure 9: The enrichment of pre-miRNAs and SNPs based on SNPs type**

625 **Figure 10: The enrichment of SNP-pre-miRNAs based on free energy change**

626 **Table 1.** Example of the SNPs information

Table	Parameter	Table	Parameter	Table	Parameter
chrom	1	AC	1	AFR_AF	0
pos	55285	AF	0.000199681	EUR_AF	0
ID	rs532608387	AN	5008	SAS_AF	0.001
REF	T	NS	2504	AA	t
ALT	C	DP	18296	VT	SNP
QUAL	100	EAS_AF	0		
FILTER	PASS	AMR_AF	0		

627

628

629

630

631 **Table 2.** The result of mature miRNA sits identification of SNP-pre-miRNAs

632

633

634

sit	nor-SNP-pre-miRNAs	nor-pre-miRNAs	nor-SNPs	iso-SNP-pre-miRNA	iso-pre-miRNA	iso-SNPs
P5_5	4235	480	1250	607	143	427
P5_3	1825	218	569	3385	405	1065
P3_5	2089	264	707	3435	363	1005
P3_3	2026	267	721	3779	395	1081

635

636

637 **Table 3.** The isomiRs and iso-snp of 18 GBR populations.

Category	sample	isomiR	SNP	sample	isomiR	SNP
Number	HG00096	697	109	HG00108	719	69
Number	HG00097	601	105	HG00109	718	112
Number	HG00099	426	94	HG00110	663	128

Number	HG00100	710	119	HG00111	711	123
Number	HG00101	646	104	HG00112	463	84
Number	HG00102	657	113	HG00114	665	103
Number	HG00105	683	104	HG00115	682	117
Number	HG00106	732	128	HG00116	714	110
Number	HG00107	315	71	HG00117	639	126

638

639 **Table 4.** The verified isomiRs of 18 GBR.

Category	sample	Veirified- isomiR	sample	Veirified- isomiR
Number	HG00096	96	HG00108	37
Number	HG00097	19	HG00109	19
Number	HG00099	19	HG00110	36
Number	HG00100	67	HG00111	27
Number	HG00101	36	HG00112	27
Number	HG00102	32	HG00114	16
Number	HG00105	23	HG00115	45
Number	HG00106	23	HG00116	33
Number	HG00107	7	HG00117	55

640 **Additional data files**

641

642 **Additional file 1: Table S1:** The database miRSNPBase. (xls)

643 **Additional file 1: Table S2:** The pre-miRNAs which have mature miRNA in the 5' arm and
644 3'arm. (xls)

645 **Additional file 1: Table S3:** All the iso-pre-miRNAs, nor-pre-miRNAs, nor-SNPs and iso-SNPs
646 associated with four splicing sites. (xls)

647 **Additional file 1: Table S4:** The pre-miRNAs and SNPs associated with the normal and
648 isomiRs. (xls)

649 **Additional file 1: Table S5:** The pre-miRNAs, iso-SNPs, and isomiRs of HG00097. (xls)

650 **Additional file 1: Table S6:** The isomiRs and iso-snp of 18 GBR populations. (xls)

651 **Additional file 1: Table S7:** The verified isomiRs of 18 GBR. (xls)

652 **Additional file 1: Table S8:** The iso-pre-miRNA candidates and the verified iso-pre-miRNAs of
653 18 GBR samples. (xls)

654

Figures

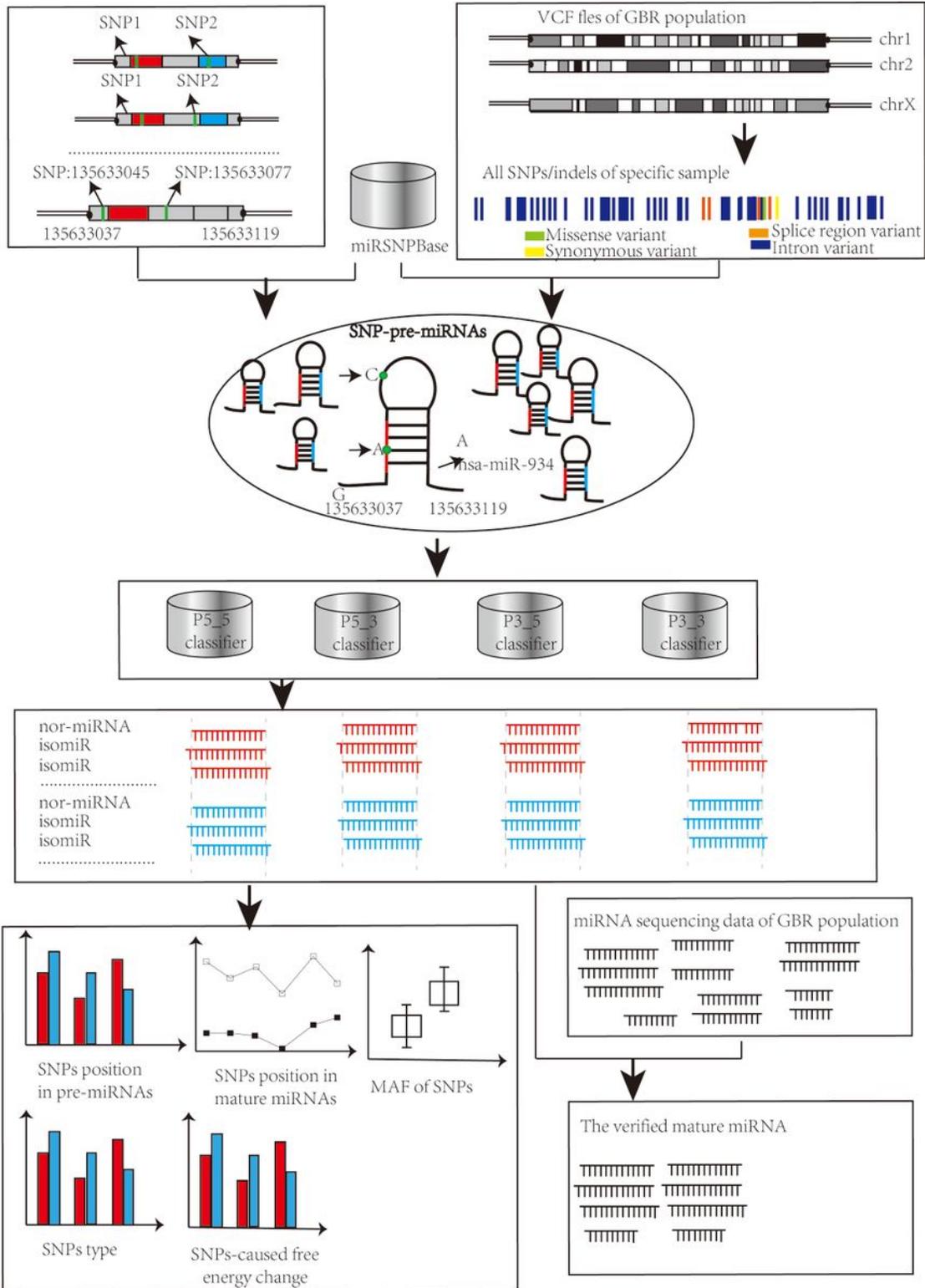


Figure 1

The schematics of the overall method

>hsa-mir-187

GGUCGGGCUCACCAUGACACAGUGUGAGACCUCGGGCUACAACACAGGACCCGGGCGCUGCUCUGACCCCUCGUG
UCUUGUGUUGCAGCCGGAGGGACGCAGGUCCGCA

名称	染色体	变异位置	miRNA起始	miRNA终止	参考	变异	变异名称	...
hsa-mir-187	chr18	33484783	33484781	33484889	G	A	rs114964240	...
hsa-mir-187	chr18	33484792	33484781	33484889	C	U	rs41274312	...
hsa-mir-187	chr18	33484837	33484781	33484889	G	A	rs375688661	...

SNP-pre-miRNA

>hsa-mir-187 rs114964240

GGUCGGGCUCACCAUGACACAGUGUGAGACCUCGGGCUACAACACAGGACCCGGGCGCUGCUCUGACCCCUCGUG
UCUUGUGUUGCAGCCGGAGGGACGCAGGUCC**A**CA

>hsa-mir-187 rs41274312

GGUCGGGCUCACCAUGACACAGUGUGAGACCUCGGGCUACAACACAGGACCCGGGCGCUGCUCUGACCCCUCGUG
UCUUGUGUUGCAGCCGGAGGGA**U**GCAGGUCCGCA

>hsa-mir-187 rs375688661

GGUCGGGCUCACCAUGACACAGUGUGAGACCUCGGGCUACAACACAGGACCC**A**GGCGCUGCUCUGACCCCUCGUG
UCUUGUGUUGCAGCCGGAGGGACGCAGGUCC**A**

>hsa-mir-187 rs114964240 rs41274312

GGUCGGGCUCACCAUGACACAGUGUGAGACCUCGGGCUACAACACAGGACCCGGGCGCUGCUCUGACCCCUCGUG
UCUUGUGUUGCAGCCGGAGGGA**U**GCAGGUCC**A**CA

>hsa-mir-187 rs114964240 rs375688661

GGUCGGGCUCACCAUGACACAGUGUGAGACCUCGGGCUACAACACAGGACCC**A**GGCGCUGCUCUGACCCCUCGUG
UCUUGUGUUGCAGCCGGAGGGACGCAGGUCC**A**CA

>hsa-mir-187 rs41274312 rs375688661

GGUCGGGCUCACCAUGACACAGUGUGAGACCUCGGGCUACAACACAGGACCC**A**GGCGCUGCUCUGACCCCUCGUG
UCUUGUGUUGCAGCCGGAGGGA**U**GCAGGUCCGCA

>hsa-mir-187 rs114964240 rs41274312 rs375688661

GGUCGGGCUCACCAUGACACAGUGUGAGACCUCGGGCUACAACACAGGACCC**A**GGCGCUGCUCUGACCCCUCGUG
UCUUGUGUUGCAGCCGGAGGGA**U**GCAGGUCC**A**CA

Figure 2

The example of constructing the SNP-pre-miRNA sequence

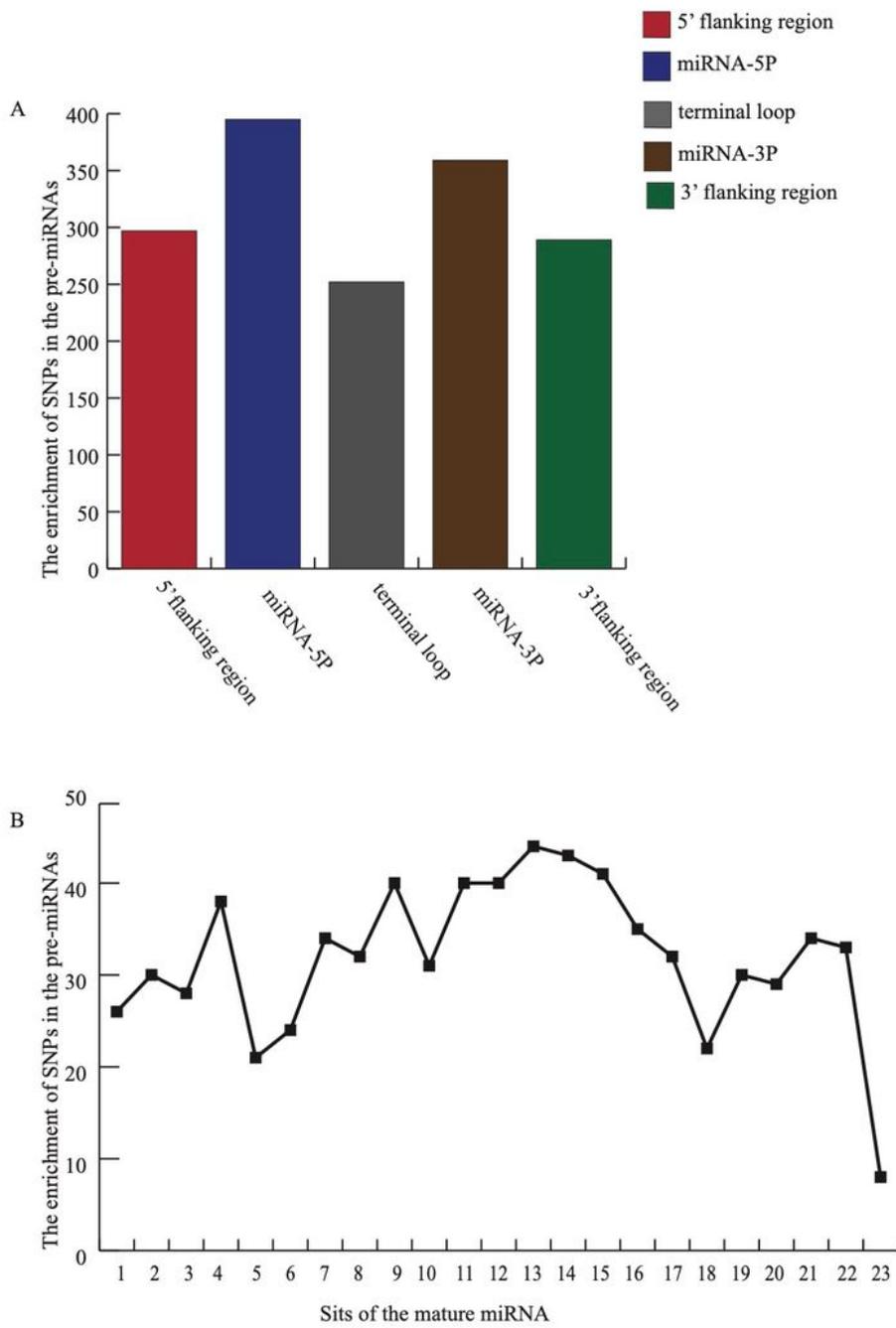
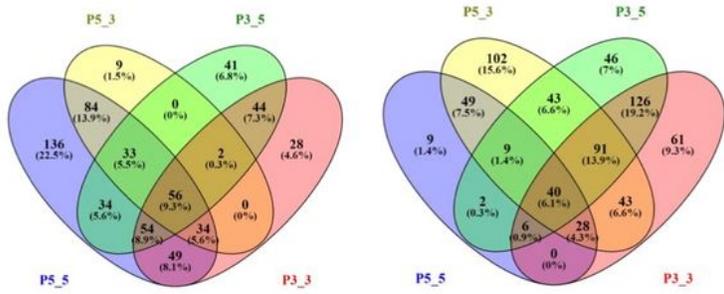


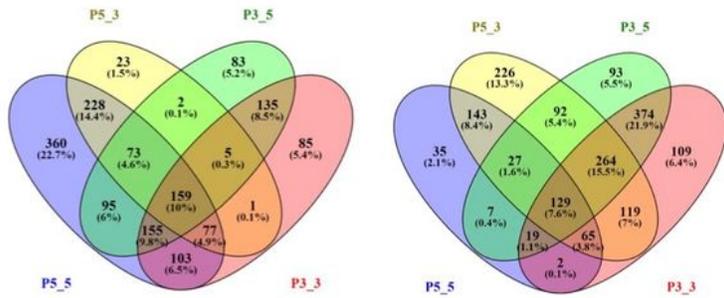
Figure 4

The distribution of SNPs in the pre-miRNA sequences



A The nor-pre-miRNAs

B The iso-pre-miRNAs



C The nor-SNPs

D The iso-SNPs

Figure6 The distribution of pre-miRNAs and SNPs associated with the normal and isoform miRNAs

Figure 5

The distribution of pre-miRNAs and SNPs associated with the normal and isomiRs

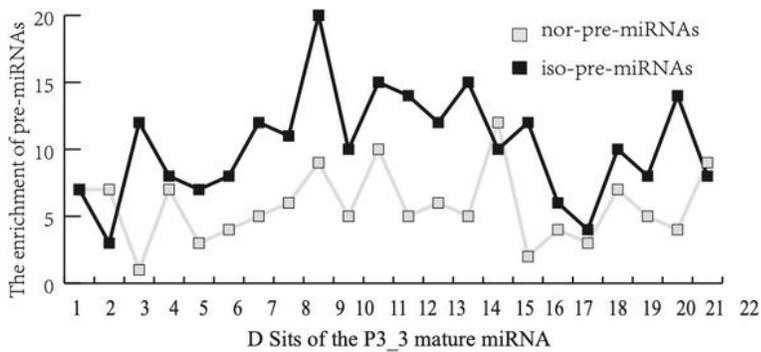
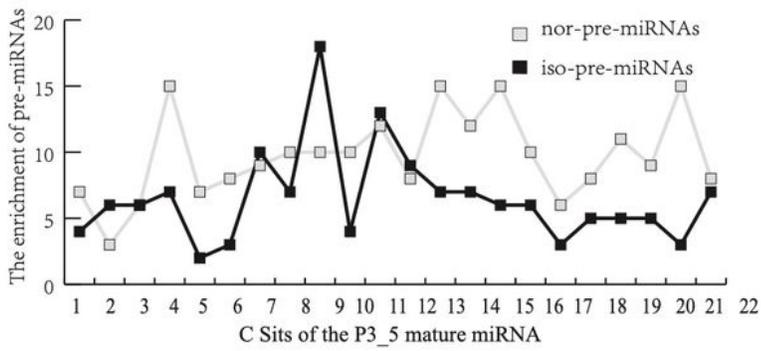
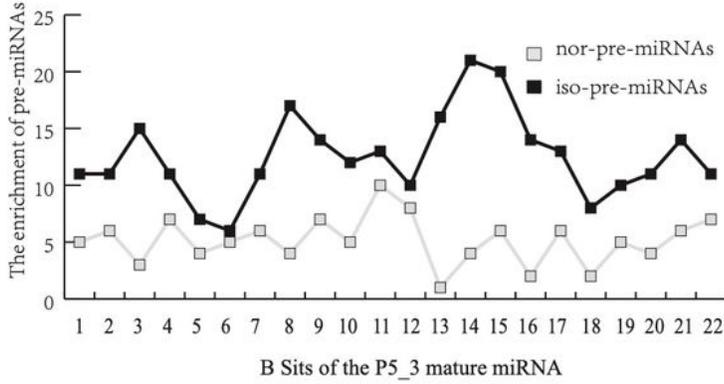
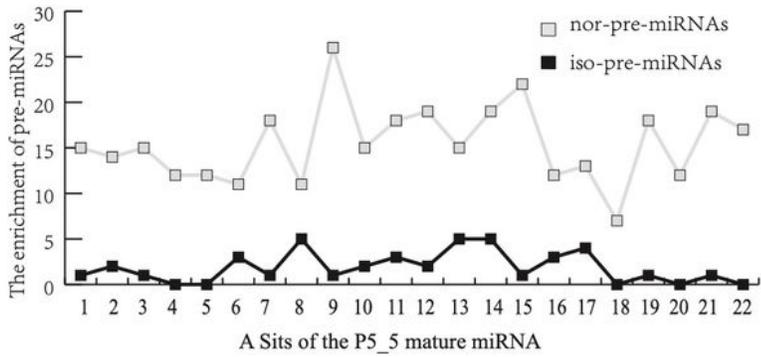


Figure 6

The MAF of iso-SNPs and nor-SNPs

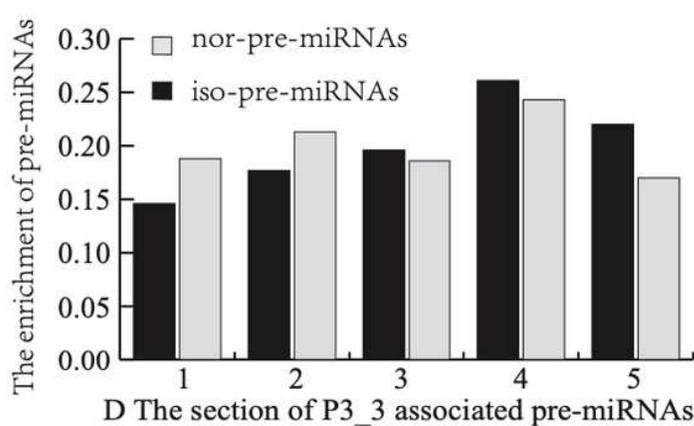
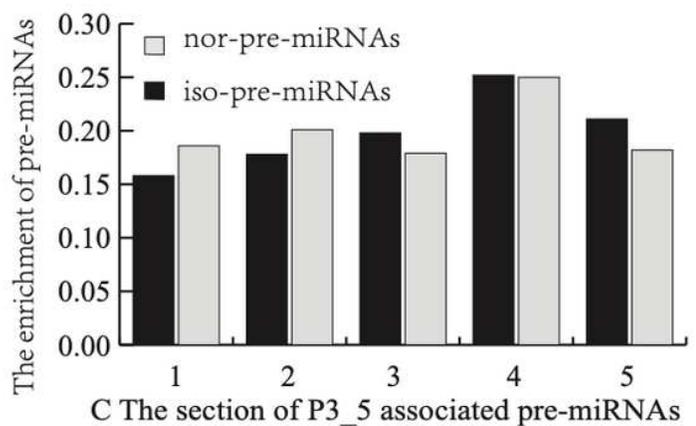
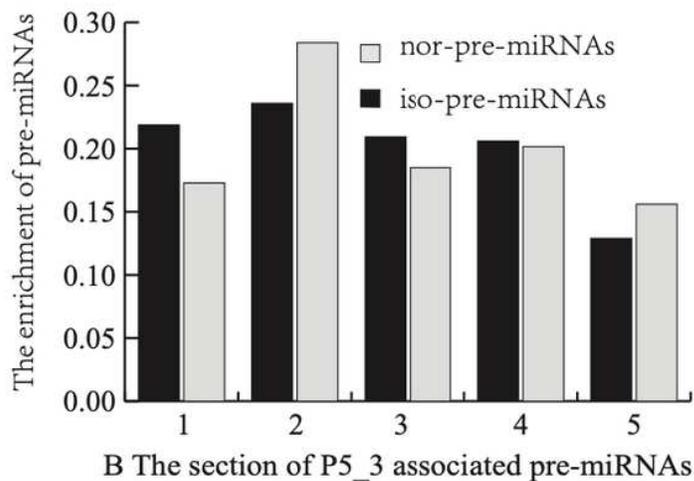
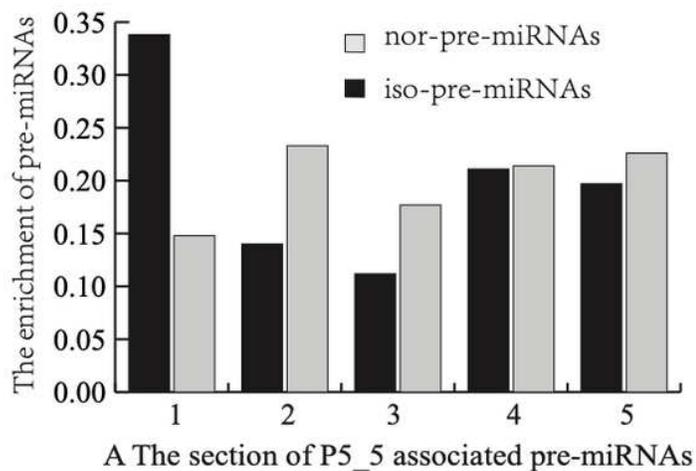
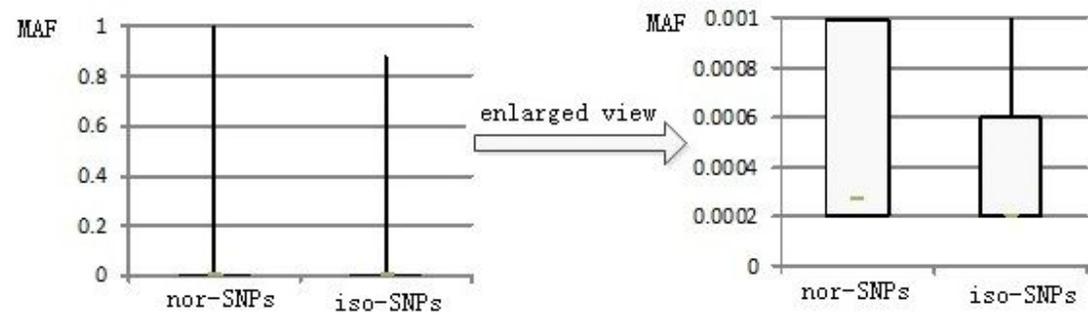
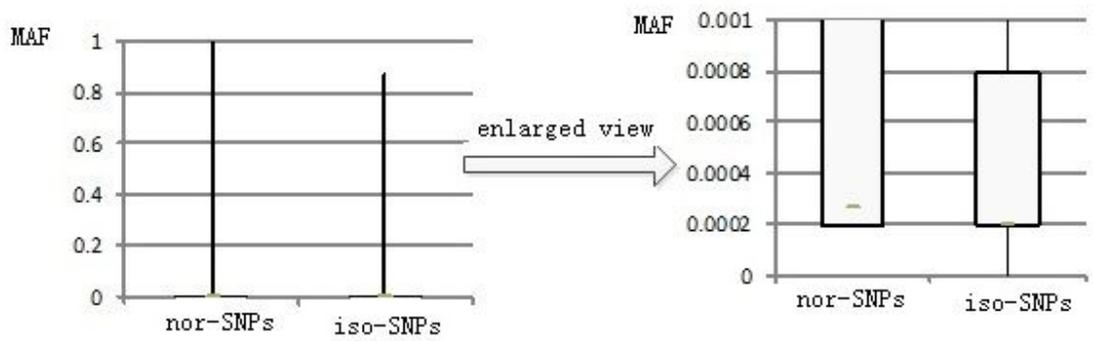


Figure 7

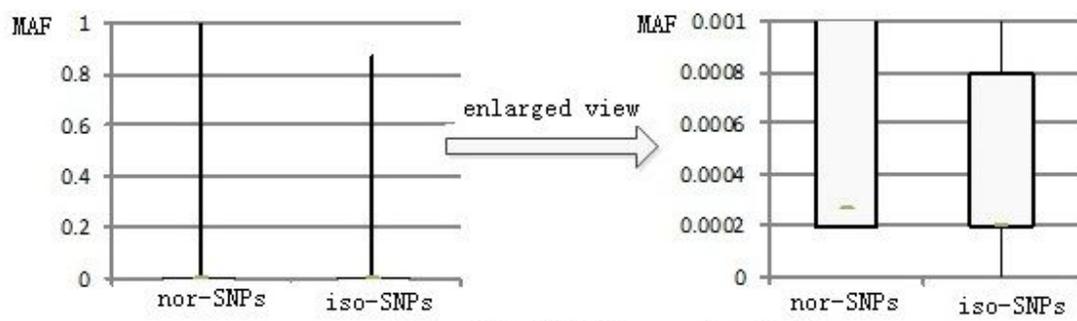
The enrichment of pre-miRNAs based on SNPs positions



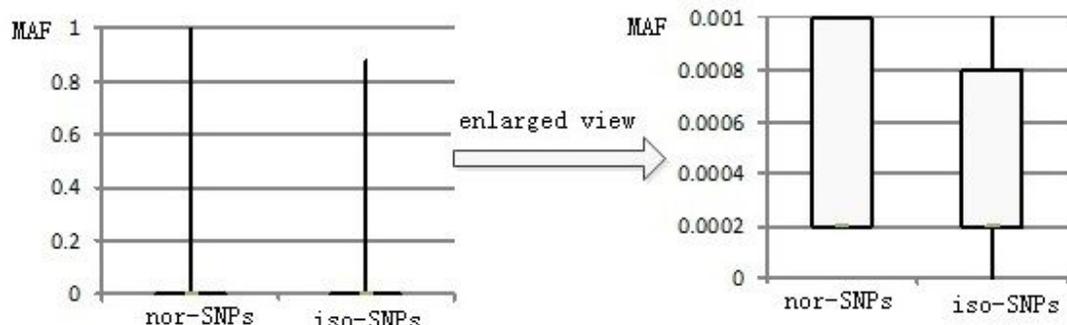
A The MAF of P5_5 associated SNPs



B The MAF of P5_3 associated SNPs



C The MAF of P3_5 associated SNPs



D The MAF of P3_3 associated SNPs

Figure 8

The enrichment of pre-miRNAs as SNPs located in the mature miRNAs

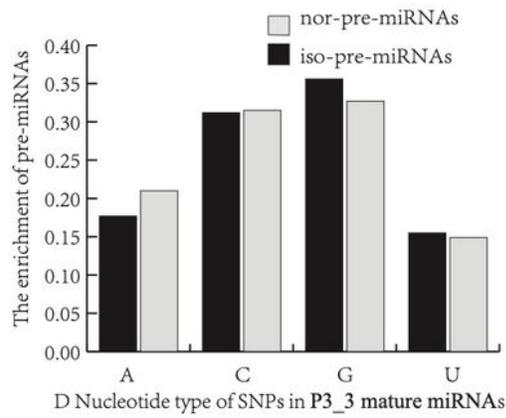
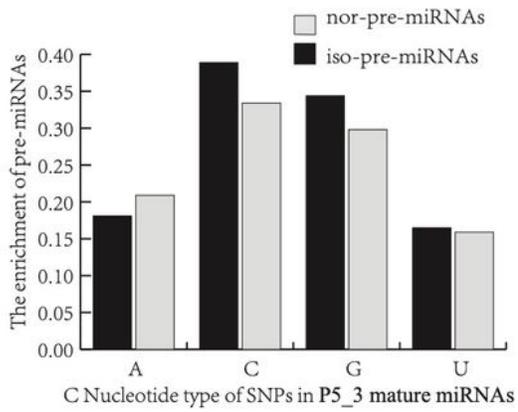
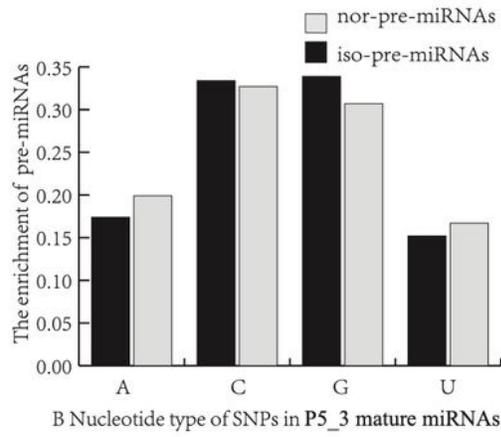
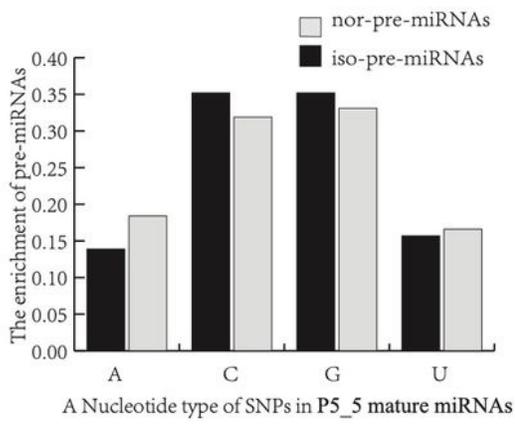


Figure 9

The enrichment of pre-miRNAs and SNPs based on SNPs type

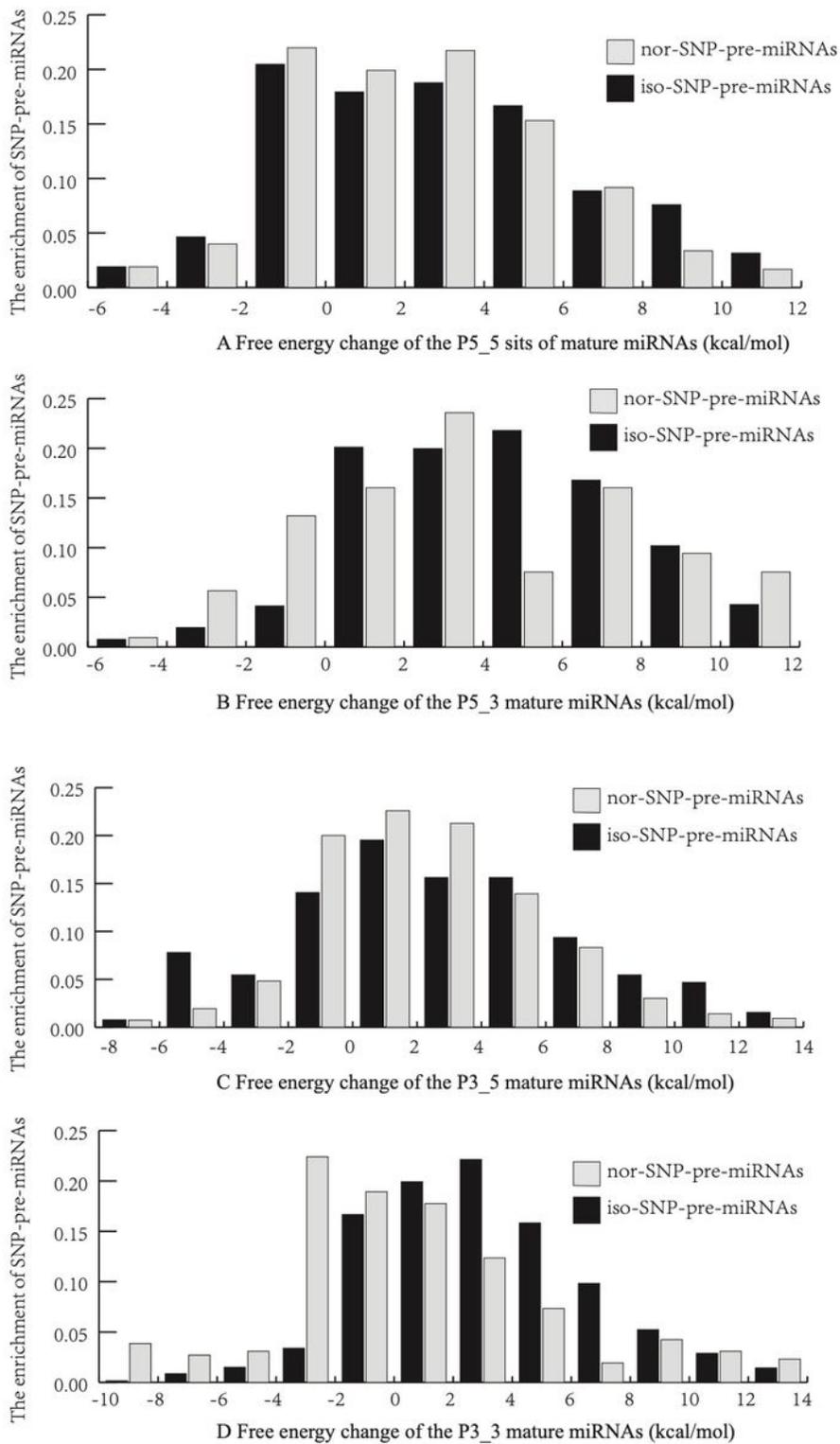


Figure 10

The enrichment of SNP-pre-miRNAs based on free energy change

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SuppTab8.xlsx
- SuppTab1.xlsx
- SuppTab4.xlsx
- SuppTab2.xlsx
- SuppTab6.xlsx
- SuppTab5.xlsx
- SuppTab3.xlsx
- SuppTab7.xlsx