SCIENTIFIC REPORTS

natureresearch

OPEN

Genome-Wide Analysis of MicroRNA-related Single Nucleotide Polymorphisms (SNPs) in Mouse Genome

Gideon Omariba[1,3], Fuyi Xu[1,2,3], Maochun Wang [1], Kai Li[1], Yuxun Zhou[1] & Junhua Xiao[1*]

MicroRNAs are widely referred to as gene expression regulators for different diseases. The integration between single nucleotide polymorphisms (SNPs) and miRNAs has been associated with both human and animal diseases. In order to gain new insights on the effects of SNPs on miRNA and their related sequences, we steadily characterized a whole mouse genome miRNA related SNPs, analyzed their effects on the miRNA structural stability and target alteration. In this study, we collected 73643859 SNPs across the mouse genome, analyzed 1187 pre-miRNAs and 2027 mature miRNAs. Upon mapping the SNPs, 1700 of them were identified in 702 pre-miRNAs and 609 SNPs in mature miRNAs. We also discovered that SNP densities of the pre-miRNA and mature miRNAs are lower than the adjacent flanking regions. Also the flanking regions far away from miRNAs appeared to have higher SNP density. In addition, we also found that transitions were more frequent than transversions in miRNAs. Notably, 841 SNPs could change their corresponding miRNA's secondary structure from stable to unstable. We also performed target gain and loss analysis of 163 miRNAs and our results showed that few miRNAs remained unchanged and many miRNAs from wild mice gained target site. These results outline the first case of SNP variations in the mouse whole genome scale. Those miRNAs with changes in structure or target could be of interest for further studies.

MicroRNAs (miRNAs) are non-coding RNA molecules with a length of about 18–24 nucleotides[1]. Initially, miR-NAs are transcribed from primary (pri-) miRNA transcript by the aid of a microprocessor complex called Drosha, to form the precursor (pre-miRNA) in the nucleus[2]. Then pre-miRNAs are transported to the cytoplasm for more processing into miRNA duplexes. The duplex is then weighed down onto Argonaut (Ago) protein of the RNA induced silencing complex (RISC) producing a mature single stranded miRNA[3,4]. It is estimated that untranslated region (UTR) is the potential region for post transcriptional regulation of mRNA[5]. Currently several miRNAs have been identified in various species including human and mice[6]. It is believed that miRNAs are involved in various biological processes by targeting mRNA induces translational inhibition of target gene[7]. Since the miRNA gene regulatory capacity was discovered[8,9], researchers have developed mechanisms of finding out the potential of miRNAs as causative and therapeutic agent for diseases like cancer and heart failure[10,11].

Single nucleotide polymorphism (SNP) is one of the most common genetic variations in a genome[12]. SNPs have been reported to have an important role on gene function and regulation[13]. With the advance of high throughput technology, increasing number of research has revealed that SNPs have profound influence in miRNA function, stability and targeting[14]. The SNPs are commonly identified in the miRNA genes, or in the binding site of mRNA of miRNAs. SNPs in the miRNA binding site are capable of modifying miRNA by generating or removing a miRNA binding site in the target mRNA[15,16]. These SNPs can effactually cause target alterations and gene expression interference[17]. To date, it has been reported by a number studies that SNPs in miRNAs genes or in the target site are involved in complex traits and diseases[18–25]. For example pre miR-146a- SNP (rs2910164) was reported to be associated with the reduced expression of the pre-miR-146a[19,21]. Also the association of mir-499-SNP (rs3746444) with various immunity diseases was reported[22,24,25]. Zhou et al., have also identified common polymorphism pre-miRNA associated with cardiomyopathy[23].

[1]College of Chemistry, Chemical Engineering, and Biotechnology, Donghua University, Shanghai, 201620, China. [2]Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN, 38163, United States. [3]These authors contributed equally: Gideon Omariba and Fuyi Xu. *email: xiaojunhua@dhu.edu.cn

1

Term	Number
SNPs in dbSNP146	67850467
SNPs in MGP(v5)	59241109
SNPs in C1SLs	3727081
Total SNPs	73643859
Total pre-miRNAs	1187
Total miRNA	2027
pre-miRNA containing SNPs	702
SNPs in pre-miRNA	1700
mature miRNA containing SNPs	464
SNPs in miRNA	609
seed region containing SNPs	163
Numberof SNPs in seed region	175

Table 1. Summary of data collection and miRNA related statistics.

So far, a number of lines of inbred mice with several isolated mutations have been established. This has enriched genetic diversity and alterations giving light to mammalian biology and human disease at large[26,27]. The mouse is one of the potential and extraordinary organisms towards our understanding of human gene function[28]. About 3,587 mouse genotypes, modeling human diseases have been reported, with at least one or more mutated genes. Around 700 inbred mice lines have been developed and are used for genetic mapping studies[29]. The variations in these inbred lines have enabled the identification of genes and variations contributing to complex traits, quantitative trait locus (QTLs) and diseases like cancers, infectious diseases[30] physiology behavior[31] and reproduction.

Despite of fierce research done on human genome and other species, very little research has been done on miRNA-related SNPs in mouse genome. In this study, we steadily characterize whole mice genome miRNA-SNPs, analyze their effects on the miRNA secondary structure stability and target alteration (target gain and loss). We discovered that SNP densities of the pre-miRNA and mature miRNAs are lower than the adjacent flanking regions. We also found that, 841 miRNAs could change their secondary structure from stable to unstable. The results also depict a significant change in miRNA target interaction in the miRNA seed region. These results outline the first case of SNP variations in the whole mice genome, which creates a better platform for further analysis and applications on human disease.

Data collection and characteristics of miRNA with SNPs

In this study, we totally collected 73643859 unique SNPs across the mouse genome from three databases, in which, 68 million SNPs from dbSNP146 of NCBI, 59 million SNPs from MGP(v5), and 3.7 million SNPs from C1SLs (Table 1). There are 1187 pre-miRNAs and 2027 mature miRNAs were deposited in miRBase. After mapping the 7.4 million SNPs onto the mouse pre-miRNAs, 1700 SNPs were identified in 702 pre-miRNAs, with 1–14 SNPs in every pre-miRNA, approximately. We also discovered that out of the 1700 SNPs, 609 of them are located in 4624 mature miRNAs while 173 of them are located in 163 seed regions of the miRNA (Table 1).

As shown in Fig. 1A, Among the 702 pre-miRNAs with SNPs, 279 (40%) pre-miRNA genes contain one SNPs, 176 (25%) contain two SNPs, and 105 (15%) have three SNPs. The total number of pre-miRNA genes have more than 3 SNPs are 142 which account for 20% of the pre-miRNAs, with only 3 pre-miRNAs having more than 10 SNPs. The highest number of SNPs is recorded mmu-miR-6238 (14), which is followed by mmu-miR-713 (12) and finally mmu-miR-6339 (11). For mature miRNA, nearly 77% (1565) do not contain any SNPs, while for the 462 mature miRNAs with SNPs, most of them are only contain one (343, 74% or two (96, 21%) SNPs. The rest of them containing 3~6 SNPs only account for 4% (23), with mmu-miR-1931 having the most SNPs (6), followed by mmu-miR- 3473b (5), mmu-miR-3060-3p (4), and mmu-miR-7037-3p(4). We also analyzed the SNP distribution per base along the mature miRNA. Results showed that sites including 1, 3, 6, 13, and 15 especially for site 1 tend to have accumulated less SNPs, while sites 2 and 14 tend to containing more SNPs (Fig. 1B).

Data has shown in mean + SE. We also compared various SNP densities in different regions, characterized them according to their flanking regions and upstream or downstream regions. The results show that mature miRNAs have the lowest SNP density (11 SNPs per kb, SD = 20, SE = 0.45). Pre-miRNAs also shows lower SNPs density with 17.5 SNPs per kb (SD = 21.5, SE = 0.62). Moreover, the flanking regions that are more far away from miRNAs genes tend to have higher SNP density. As a comparison, SNP densities of the pre or mature miRNAs are significantly lower than its adjacent flanking regions (Fig. 1D).

Moreover, we additionally found that transitions (71.3%) were undeniably more continuous than transversions (28.7%) in the miRNAs. When compared with genome wide, transitions in miRNAs are higher than in genome wide (71.3% vs 66.0%). For transition, A-C, A-T, and G-T substitutions are lower in miRNAs especially for A-T (5.3% vs. 9.0%), while C-G substitutions in miRNAs are slightly higher than those in genome wide (7.3% vs. 6%) (Fig. 1C).
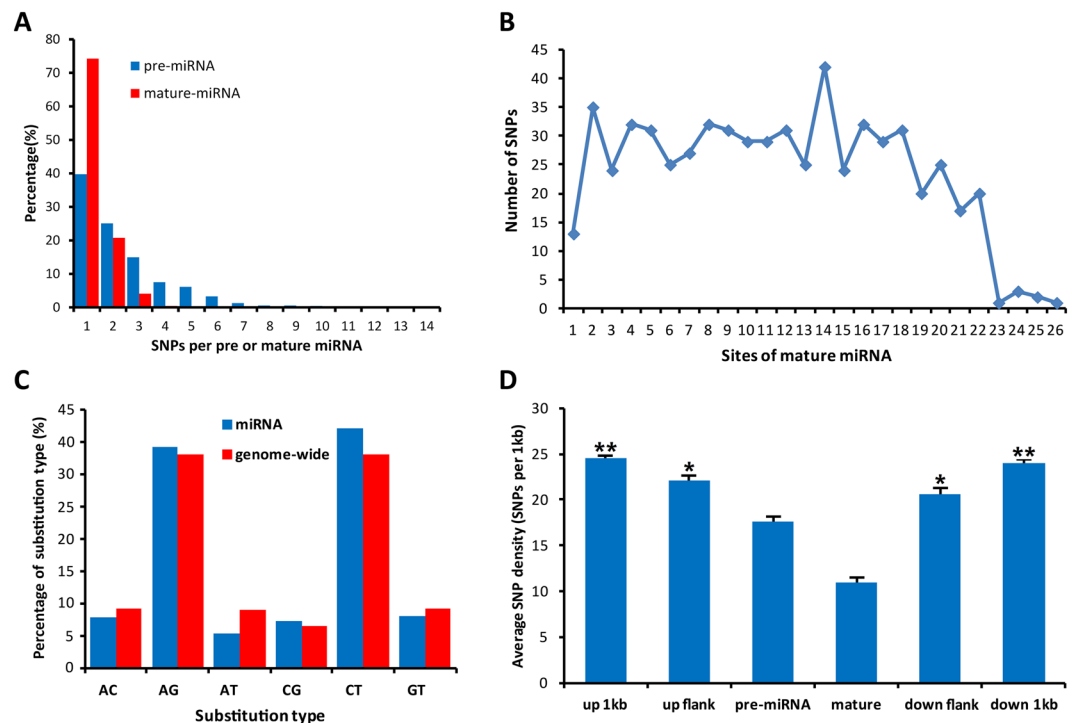
Figure 1. Characteristics of miRNA genes with SNPs. (**A**) Frequency distributions for various SNP numbers in pre-miRNAs and mature miRNAs. (**B**) SNP distribution per base along the mature miRNA. (**C**) Types of substitutions identified in mouse pre-miRNAs. (**D**) SNP density of pre-miRNAs, mature-miRNAs, and their flanking regions. The up or down flanking regions are equal to the pre-miRNAs length, they are in the adjacent location to the corresponding pre-miRNAs. The evaluation of these regional differences was done by T test (* represents $p < 0.1$ and ** represent $p < 0.01$).
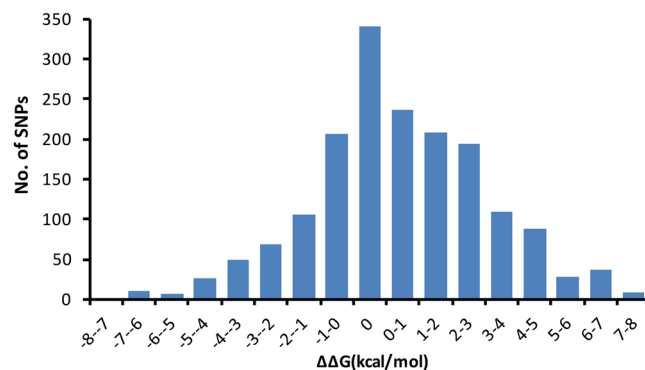
Figure 2. Frequency distribution of energy change ($\Delta\Delta G$, kcal/mol) for the secondary structure of miRNAs caused by SNPs.

Secondary structural stability of miRNA with SNPs

So as to explore the impact of SNPs in dependability on miRNAs secondary structure, we utilized program RNAfold for secondary structure prediction and looked at the vitality change between wild and mutant type of each miRNAs. As shown in Fig. 2, 1387 SNPs in 626 miRNAs were changed MFE with $|\Delta\Delta G|$ ranging from 0.1 to 7.2 kcal/mol (Supplementary Table 1). The average free energy change caused by SNP on secondary structure is 2.17 kcal/mol4 which is similar with the one in the human genome (2.1 kcal/mol). Previous study showed that the base energy for maintaining the miRNAs secondary structure stability would be 0.3 kcal/mol[32]. We found 841 SNPs in 476 miRNAs could change their secondary structure from stable to unstable according to this criterion (Supplementary Table 1). For example, mmu-miR-3060, where a nucleotide G changed to A, caused a small loop structure with 6.7 kcal/mol MEF change (Fig. 3). On the other hand, 398 SNPs in 294 miRNAs increased the stability of these miRNAs (Supplementary Table 1) such as mmu-miR-5709, where a A-C SNP seems to stabilize the structure of its hairpin more at the stem part with −7.1 kcol/mol MEF change (Fig. 3).
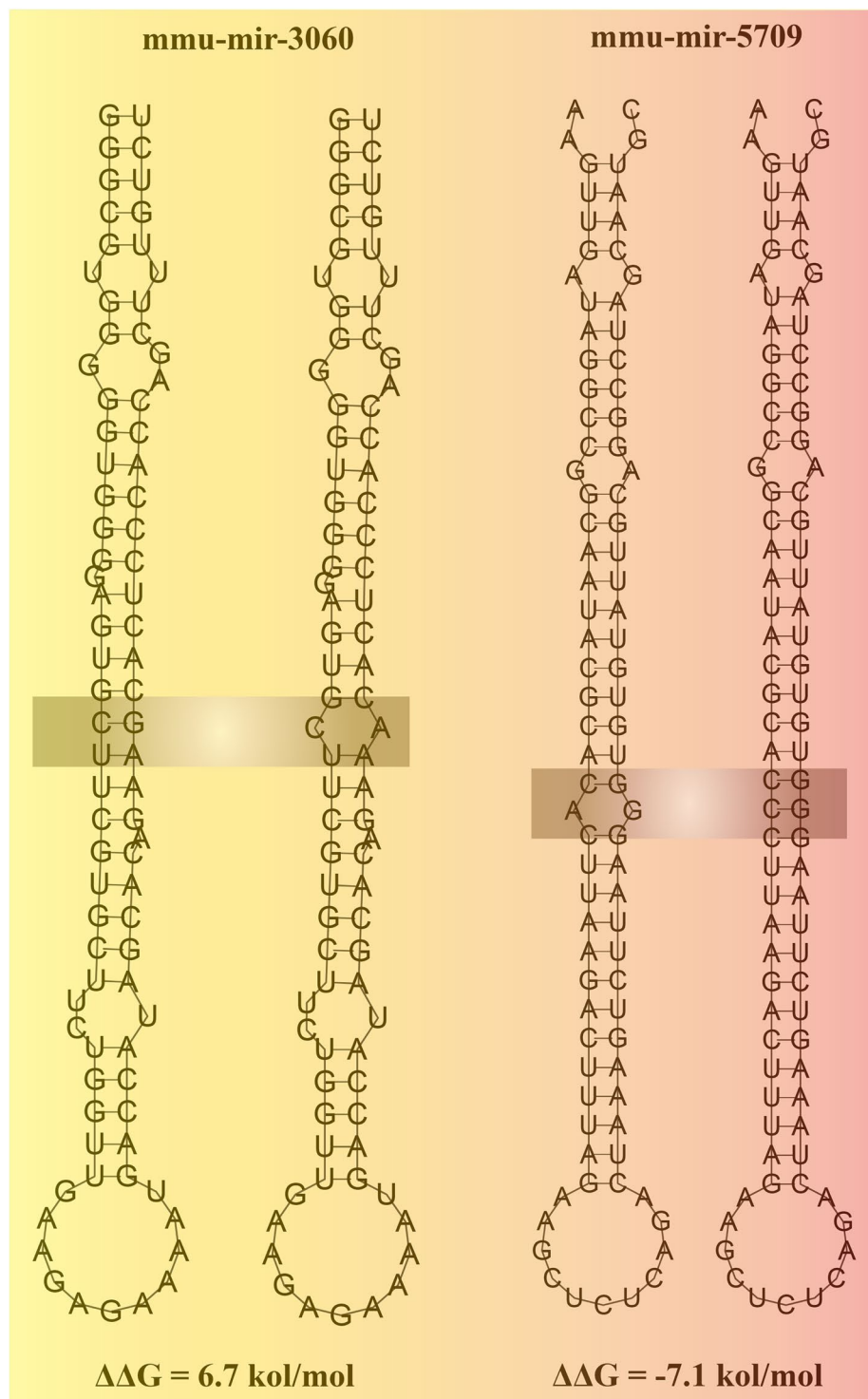
Figure 3. Illustration of the secondary structure of miRNAs with SNPs. For mmu-mir-3060, nucleotide G changed to A causing a small loop structure with a 6.7 kcal/mol MEF change, while mmu-mir-5709 A-to-C SNP in the stem region made its hairpin structure more stable with a −7.1 kcal/mol change.

Although the above analysis for miRNAs can evaluate single SNP effect on its secondary structure, it could not reflect the strain-specific miRNA secondary structure change. In other words, for a certain strain, it may contain more than one SNP in one miRNA sequence. Therefore, the secondary structure caused by single SNP or by combining all SNPs would result in to a very different result. We next performed the strain-specific miRNAs secondary structure analysis in 36 MGP and 18 C1SLs mice which we had their exact genotype information. As shown in Fig. 4A, on average, each strain in MGP contain 89 miRNAs (SD = 92.3) genome wide with C57BL/10 J mouse contain only 5 and SPRET/EiJ mouse with 502. While for C1SLs, each line has 19.8 miRNAs (SD = 3.5) in
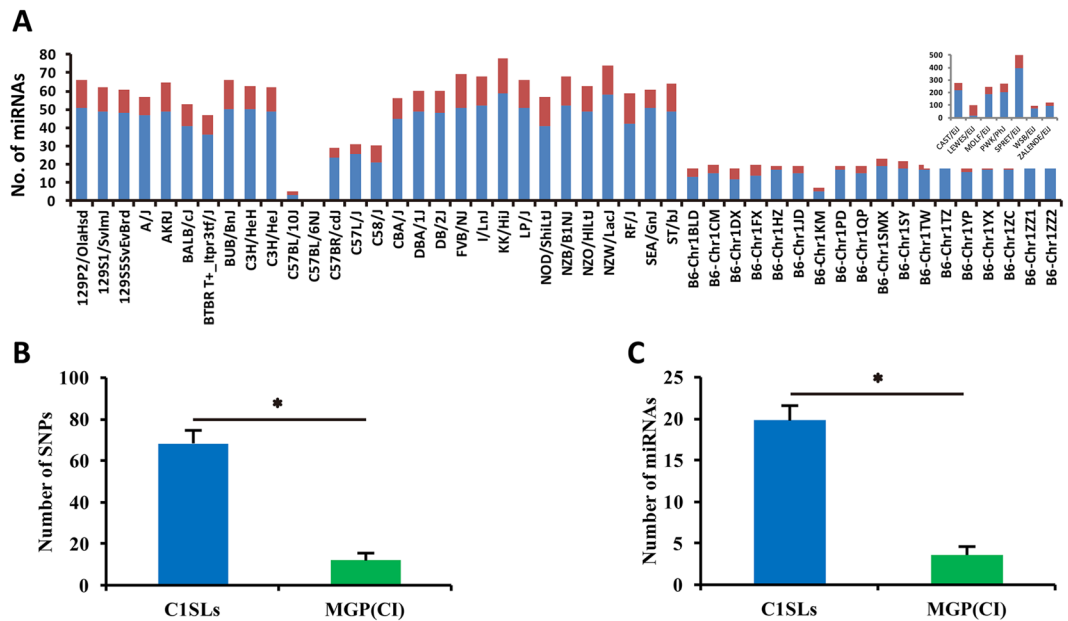
Figure 4. Number of miRNAs containing SNPs in 36 MGP and 18 C1SLs mice. (**A**) The number of miRNAs containing SNPs in 36 MGP and 18 C1SLs. The upper right shows 7 wild-derived strains. Blue bar represents $|\Delta\Delta G| > 0.3$ kcal/mol of the secondry structure of miRNAs, while red bar represents $|\Delta\Delta G| <= 0.3$ kcal/mol. Upper right: seven wild-derived mice. (**B**) Shows the number of miRNA SNPs in C1SLs mice and MGP strains. (**C**) Number of miRNAs in C1SLs and MGP mice.

chromosome 1 which is significantly higher than in MGP sequenced 29 classical inbred (CI) strains (excluding seven wild-derived strains) (Fig. 4B,C). Next, we performed the secondary structure analysis, results showed that the overall proportion of $\Delta\Delta G > 0.3$ kcal/mol for strain-specific was account for 76.4% (SD = 11.6). In addition, we found 9 miRNA with $|\Delta\Delta G|$ over 10 kcol/mol in 24 strains (Table 2). The detailed information for line-specific secondary change was listed in Supplementary Table 2.

Target loss and gain analysis

The miRNA's target loss and gain is greatly affected by presence of SNPs in the seed regions of the miRNA, thus affecting the miRNA function dramatically. In this study, we obtained our potential miRNA target loss and gain using the comparison difference of wild type targets and SNP- miRNA targets. According to the results, we discovered that few miRNAs remained unchanged, while many of them gained and lost target site respectively (Fig. 5). Interestingly, many miRNAs gained target site happen to be in the wild mice category.

Discussion

In this study, we steadily characterize a whole mouse genome miRNA SNPs, analyze their effects on miRNA secondary structure stability and target alteration (target gain and loss). Totally, we collected 73643859 SNPs across the mouse genome. When we mapped the SNPs, 1700 SNPs were identified in 702 pre-miRNAs, 609 SNPs were located in mature miRNAs and 173 SNPs in 163 miRNA seed regions. The SNP distribution per base analysis showed different SNP accumulation along the mature miRNA. We discovered that SNP densities of the pre-miRNA and mature miRNAs are lower than the adjacent flanking regions. The flanking regions that are far away from miRNA genes were found to have higher SNP density than the closer ones. We also found that the secondary structure of 476 miRNAs could change from stable to unstable. The results also depict a significant change in miRNA target interaction in the seed regions of the miRNA. In our target gain and loss analysis, results showed that few of the miRNAs remained unchanged. Most of the miRNAs that gained target belong to wild mice category. These results outline the first case of SNP variations in the whole mouse genome, which creates a better platform for further analysis and applications on human disease. The wild type miRNAs that gained target site also create an interesting incite for further studies.

It is worthy to note that most of our findings are consistent with other similar studies done on different species. In our SNP density analysis, we discovered that pre-miRNAs and mature miRNAs have lower SNP densities than the flanking regions which agree with most studies. For instance, a study done by Gong *et al.*, on human genome discovered that SNP densities of the flanking regions were higher than SNPs of pre-miRNAs and MIR seed region[33]. In another study, Liu *et al.*, found that most research done on rice had indicated that pre-miRNAs contained higher SNP density compared to flanking regions which was contrary to their findings. In their results, they unexpectedly discovered that the SNP densities in the mature miRNAs and pre-miRNAs were higher than SNPs in the flanking regions, which could be as a result of hasty evolutions on miRNAs in specific species of rice[34]. It is also worthy to note that in another cucumber study[35], they discovered that mature sequences contained less SNP than the binding sites. These divergent results in rice need an in-depth analysis to find out on why some

miRNA	Wild MFE	Mutant MFE	ΔΔG	Strain	No. of SNPs
mmu-mir-3060	42.4	27.7	−14.7	BUB/BnJ	4
mmu-mir-3060	42.4	29.9	−12.5	CAST/EiJ	4
mmu-mir-3060	42.4	27.7	−14.7	LEWES/EiJ	4
mmu-mir-3060	42.4	27.7	−14.7	NOD/ShiLtJ	4
mmu-mir-3060	42.4	27.7	−14.7	NZO/HlLtJ	4
mmu-mir-3060	42.4	27.7	−14.7	NZW/LacJ	4
mmu-mir-3060	42.4	27.7	−14.7	ST/bJ	4
mmu-mir-3060	42.4	27.7	−14.7	WSB/EiJ	4
mmu-mir-3470b	47.1	35	−12.1	CAST/EiJ	8
mmu-mir-3470b	47.1	35	−12.1	MOLF/EiJ	8
mmu-mir-3473b	42.5	16.6	−25.9	LEWES/EiJ	6
mmu-mir-3473b	42.5	24.5	−18	PWK/PhJ	3
mmu-mir-3473b	42.5	24.5	−18	SPRET/EiJ	3
mmu-mir-3473b	42.5	16.6	−25.9	WSB/EiJ	6
mmu-mir-3473b	42.5	16.6	−25.9	ZALENDE/EiJ	6
mmu-mir-3473c	41.4	29.5	−11.9	SPRET/EiJ	2
mmu-mir-3473f	69.5	58.1	−11.4	B6-Chr1BLD	3
mmu-mir-3473f	69.5	58.1	−11.4	B6-Chr1CM	3
mmu-mir-3473f	69.5	56.9	−12.6	B6-Chr1HZ	4
mmu-mir-3473f	69.5	58.1	−11.4	B6-Chr1JD	3
mmu-mir-3473f	69.5	58.1	−11.4	B6-Chr1SMX	3
mmu-mir-3473f	69.5	58.1	−11.4	B6-Chr1SY	3
mmu-mir-3473f	69.5	58.1	−11.4	B6-Chr1TW	3
mmu-mir-3473f	69.5	56.9	−12.6	B6-Chr1TZ	4
mmu-mir-3473f	69.5	58.1	−11.4	B6-Chr1YX	3
mmu-mir-3473f	69.5	58.1	−11.4	B6-Chr1ZC	3
mmu-mir-3473f	69.5	58.1	−11.4	B6-Chr1ZZ1	3
mmu-mir-3473f	69.5	58.1	−11.4	B6-Chr1ZZ2	3
mmu-mir-5135	44.1	31.1	−13	MOLF/EiJ	3
mmu-mir-5135	44.1	31.1	−13	PWK/PhJ	3

Table 2. Lists of miRNAs with $|\Delta\Delta G| > 10$ kcol/mol.
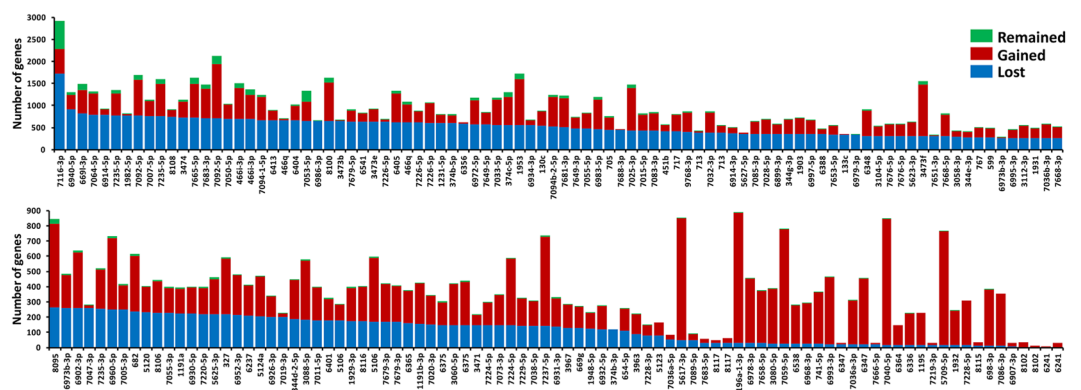
Figure 5. Target alteration analysis of miRNAs in regard with their seed regions. The green bar represents the number of target sites that neither gained nor lost. The red and blue bar represent the number of target sites that gained and lost, respectively.

few species have different SNP densities from the rest. We also realized that the far the flanking regions are from miRNA genes, the higher the SNP density they have. Our target loss and gain analysis results showed that most miRNAs either lost or gained targets, which concurs with Gong et al.[33], who discovered that miRNAs SNPs in the seed regions could highly lead to target gain and loss in humans. They also experimentally validated the target loss of miR-124. It is also important to highlight that Harnprasopwat et al., found out that an A/G SNP in mature miR-26 could destabilize the secondary structure of pri-miRNA126 and prevent its processing to mature miRNA[36]. On the other hand, SNP G/A at the stem region of miR510 led to mature miRNA production[37]. In our study, we

equally agree with others that SNPs could cause unstable hairpin structures in the stem regions hence affect the production of mature miRNAs and affect their stability.

Despite of the mouse being one of the potential and extraordinary organisms towards our understanding of human gene function[28], very few researches on miRNA-related SNPs in mouse have been done. Due to establishment of a number of inbred mice, isolation of several impulsive mutations has been achieved. This has enriched genetic diversity and alterations giving light to mammalian biology and human disease at large[26,27]. The fact that model animals can be kept in favorable controlled environments; inbred mice stand a better chance to give a highly reliable phenotypic data and information for genetic analysis[32], About 3,587 mouse genotypes, modeling human diseases have been reported, with at least one or more mutated genes. Around 700 inbred mice lines have been developed and are used for genetic mapping studies[29]. The variations in these inbred lines have enabled the identification of genes and variations contributing to complex traits (QTLs) and diseases like cancers, infectious diseases[30] physiology behavior[31] and reproduction. Through the aid of knockout and mutant mice, a number of studies on miRNA expression comparison profiles have been achieved[38]. In a given study, significant differences on miRNA expression profiling of mice retina, showed that miRNAs are actively involved in retinal disease[39]. As a result of over expression of miR-17-92 on transgenic mice's lymphocytes, Xiao et al. discovered premature death and lymphorproliferative disease in the mice[40]. miRNAs have been reported to be involved in mouse and tumor development[6]. For example in a given study several miRNAs were identified close to tumor loci in mice[41]. This clearly indicates that miRNAs processes in mice play an important role in genetics and other biological functions.

Recently, wild mice have emerged as one of the important resources for identification of genes and fine mapping. This is because, wild mice is capable of harboring larger genetic variations and have a wide historical recombination than laboratory mice[42]. Taking advantage of this capability of wild mice, Xiao et al., were able to develop a novel strategy fro genetic dissection of complex traits using the wild mice. as part of current research group conducted a study on complex traits using specific chromosome substitution strains of wild mice[43]. The mammalian biology has made tremendous steps due to various insights from wild mice studies[44]. In recent studies, researchers have also pointed out that wild mice has gained more preference in studying complex traits because they have been found to be easily cross bred with classical laboratory mice strains hence producing offspring with natural polymorphisms[45]. A recent study by Fuyi et al., also who is part of this research study team, developed an important resource for studying genetic complex traits using chromosome 1 substitution line that was a derivative of wild mice[46]. This also validates the wide and important use of wild mice in genetic research. Our current results of target gain and loss analysis showed many of the miRNAs that gained target site belong to wild mice category. This creates a big interest on these miRNAs further studies and explorations. Another study also revealed that miRNA-3473b from wild mice is directly involved in expression control of neuroinflammation in cerebral is chemia[47]. This wild type miRNA is also identified in our wild mice miRNA list having aggressive characteristics especially in controlling gene expression and can easily gain target site. In another interesting study, Tiwari et al., found out that mmu-mir-3473f, which belongs to wild mice category, is capable of regulating different biological processes associated with macrophage dysfunction in parasites[48]. The wild mice miRNAs seem to have a wide scoop for more analysis and various applications in genetics.

Up to date most studies on miRNA SNP analysis have been done on human and other species. We introduce the first miRNA SNP analysis with interesting results, especially our target gain and loss analysis results which showed that many of the miRNAs that gained target site belong to wild mice hence creating the arch for further studies on the same. These results outline the first case of SNP variations in the whole mouse genome, which creates a better platform for further analysis and applications on human disease.

Materials and Methods

Data collection.   MiRNA data was retrieved from miRBase database (http://www.mirbase.org/) (Release 21)[49,50]. The SNP information was collected from the following three data sets: (1) the NCBI dbSNP146 for mouse[51]; (2) the Sanger Institute (http://www.sanger.ac.uk/science/data/mouse-genomes-project)[52]; and (3) SNPs identified on C1SLs derived from Chinese wild mice[46].

Identification of miRNAs related SNPs.   We compared both coordinates of SNPs and the related miRNAs to identify SNPs located in mouse pre-miRNAs genes and those in their adjacent regions (flanking or 1 kb upstream/downstream regions). Intersect subcommand applied in bedtools[53] was used to perform this analysis.

Characterizing SNPs in mature miRNA and other sequence regions.   We wrote a in home script for SNP density calculation. The assessment of SNP density differences in mature miRNAs and other sequence regions (pre-miRNA, flanking regions and up or downstream regions) was done by T test. The transition/tranversion ratio was calculated with vcftools[53,54].

Analysis of miRNA secondary structure.   The program RNAfold implemented in ViennaRNA (version 2.2.5) package[55] was employed to analyze miRNA secondary structure. This calculates minimum free energy (MFE), secondary structure and prints the MFE structure in bracket notation. We used RNAplot for pre-miRNAs secondary structure display[56].

Analysis of miRNA target loss or gain.   Target gain and loss was analyzed by TargetScan (http://www.TargetScan.org/) (Release 7.2)[57] and miRanda (http://www.microrna.org) (version 3.3a)[58].

Data availability

References

1. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509–524 (2014).
2. Burke, J. M., Kelenis, D. P., Kincaid, R. P. & Sullivan, C. S. A central role for the primary microRNA stem in guiding the position and efficiency of Drosha processing of a viral pri-miRNA. *RNA* **20**, 1068–1077 (2014).
3. Faller, M. *et al.* DGCR8 recognizes primary transcripts of microRNAs through highly cooperative binding and formation of higher-order structures. *RNA* **16**, 1570–1583 (2010).
4. Ma, H., Wu, Y., Choi, J.-G. & Wu, H. Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. *Proc. Natl. Acad. Sci.* **110**, 20687–20692 (2013).
5. Yi, R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**, 3011–3016 (2003).
6. Bernstein, E. *et al.* Dicer is essential for mouse development. *Nat. Genet.* **35**, 215–217 (2003).
7. Wilson, R. C. *et al.* Dicer-TRBP Complex Formation Ensures Accurate Mammalian MicroRNA Biogenesis. *Mol. Cell* **57**, 397–407 (2015).
8. Catalanotto, C., Cogoni, C. & Zardo, G. MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions. *Int. J. Mol. Sci.* **17**, 1712 (2016).
9. Valinezhad Orang, A., Safaralizadeh, R. & Kazemzadeh-Bavili, M. Mechanisms of miRNA-Mediated Gene Regulation from Common Downregulation to mRNA-Specific Upregulation. *Int. J. Genomics* **2014**, 1–15 (2014).
10. Bertino, J. R., Banerjee, D. & Mishra, P. J. Pharmacogenomics of microRNA: a miRSNP towards individualized therapy. *Pharmacogenomics* **8**, 1625–1627 (2007).
11. Ha, T.-Y. MicroRNAs in Human Diseases: From Cancer to Cardiovascular Disease. *Immune Netw.* **11**, 135 (2011).
12. Suh, Y. & Vijg, J. SNP discovery in associating genetic variation with human disease phenotypes. *Mutat. Res. Mol. Mech. Mutagen.* **573**, 41–53 (2005).
13. Lohrer, H. D. & Tangen, U. Investigations into the Molecular Effects of Single Nucleotide Polymorphism. *Pathobiology* **68**, 283–290 (2000).
14. Li, L., Meng, T., Jia, Z., Zhu, G. & Shi, B. Single nucleotide polymorphism associated with nonsyndromic cleft palate influences the processing of miR-140. *Am. J. Med. Genet. Part A* **152A**, 856–862 (2010).
15. Sethupathy, P. & Collins, F. S. MicroRNA target site polymorphisms and human disease. *Trends Genet.* **24**, 489–497 (2008).
16. Zorc, M. *et al.* Catalog of MicroRNA Seed Polymorphisms in Vertebrates. *PLoS One* **7**, e30737 (2012).
17. Ryan, B. M., Robles, A. I. & Harris, C. C. Genetic variation in microRNA networks: the implications for cancer research. *Nat. Rev. Cancer* **10**, 389–402 (2010).
18. Bao, L. *et al.* PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits. *Nucleic Acids Res.* **35**, D51–D54 (2007).
19. Shen, H., Hu, Z., Chen, J. & Tian, T. Genetic variants of miRNA sequences and non small cell lung cancer survival. *Eur. J. Cancer Suppl.* **6**, 10–11 (2008).
20. Jazdzewski, K. *et al.* Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. *Proc. Natl. Acad. Sci.* **105**, 7269–7274 (2008).
21. Shen, J. *et al.* A functional polymorphism in the miR-146a gene and age of familial breast/ovarian cancer diagnosis. *Carcinogenesis* **29**, 1963–1966 (2008).
22. Liu, Z. *et al.* Genetic variants in selected pre-microRNA genes and the risk of squamous cell carcinoma of the head and neck. *Cancer* **116**, 4753–4760 (2010).
23. Zhou, B. *et al.* Common genetic polymorphisms in pre-microRNAs were associated with increased risk of dilated cardiomyopathy. *Clin. Chim. Acta* **411**, 1287–1290 (2010).
24. George, G. P., Gangwar, R., Mandal, R. K., Sankhwar, S. N. & Mittal, R. D. Genetic variation in microRNA genes and prostate cancer risk in North Indian population. *Mol. Biol. Rep.* **38**, 1609–1615 (2011).
25. Wang, Y., Yang, B. & Ren, X. Hsa-miR-499 polymorphism (rs3746444) and cancer risk: A meta-analysis of 17 case–control studies. *Gene* **509**, 267–272 (2012).
26. MGI-Mouse Genome Informatics-The international database resource for the laboratory mouse. Available at, http://www.informatics.jax.org/. (Accessed: 23rd January 2019).
27. Smith, C. L., Blake, J. A., Kadin, J. A., Richardson, J. E. & Bult, C. J. Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. *Nucleic Acids Res.* **46**, D836–D842 (2018).
28. Bult, C. J. From information to understanding: the role of model organism databases in comparative and functional genomics. *Anim. Genet.* **37**, 28–40 (2006).
29. Callaway, E. How to build a better mouse. *Nature* **475**, 279–279 (2011).
30. Durrant, C. *et al.* Collaborative Cross mice and their power to map host susceptibility to Aspergillus fumigatus infection. *Genome Res.* **21**, 1239–48 (2011).
31. Philip, V. M. *et al.* Genetic analysis in the Collaborative Cross breeding population. *Genome Res.* **21**, 1223–1238 (2011).
32. Hunter, K. W. & Crawford, N. P. S. The Future of Mouse QTL Mapping to Diagnose Disease in Mice in the Age of Whole-Genome Association Studies. *Annu. Rev. Genet.* **42**, 131–141 (2008).
33. Gong, J. *et al.* Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. *Hum. Mutat.* **33**, 254–263 (2012).
34. Liu, Q., Wang, H., Zhu, L., Hu, H. & Sun, Y. Genome-wide identification and analysis of miRNA-related single nucleotide polymorphisms (SNPs) in rice. *Rice* **6**, 10 (2013).
35. Ling, J. *et al.* Genome-wide analysis of microRNA targeting impacted by SNPs in cucumber genome. *BMC Genomics* **18**, 275 (2017).
36. Harnprasopwat, R. *et al.* Alteration of processing induced by a single nucleotide polymorphism in pri-miR-126. *Biochem. Biophys. Res. Commun.* **399**, 117–122 (2010).
37. Sun, G. *et al.* SNPs in human miRNA genes affect biogenesis and function. *RNA* **15**, 1640–1651 (2009).
38. van Rooij, E. *et al.* Control of Stress-Dependent Cardiac Growth and Gene Expression by a MicroRNA. *Science (80-.).* **316**, 575–579 (2007).

39. Loscher, C. J. *et al.* Altered retinal microRNA expression profile in a mouse model of retinitis pigmentosa. *Genome Biol.* **8**, R248 (2007).
40. Xiao, C. *et al.* Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. *Nat. Immunol.* **9**, 405–414 (2008).
41. Sevignani, C. *et al.* MicroRNA genes are frequently located near mouse cancer susceptibility loci. *Proc. Natl. Acad. Sci.* **104**, 8017–8022 (2007).
42. Salcedo, T., Geraldes, A. & Nachman, M. W. Nucleotide Variation in Wild and Inbred Mice. *Genetics* **177**, 2277–2291 (2007).
43. Xiao, J. *et al.* A novel strategy for genetic dissection of complex traits: the population of specific chromosome substitution strains from laboratory and wild mice. *Mamm. Genome* **21**, 370–376 (2010).
44. Phiferrixey, M. & Nachman, M. W. The Natural History of Model Organisms: Insights into mammalian biology from the wild house mouse Mus musculus. *Elife* (2015).
45. Guénet, J.-L. & Bonhomme, F. Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet.* **19**, 24–31 (2003).
46. Xu, F. *et al.* Genome Sequencing of Chromosome 1 Substitution Lines Derived from Chinese Wild Mice Revealed a Unique Resource for Genetic Studies of Complex Traits. G3 Genes|Genomes|Genetics, https://doi.org/10.1534/g3.116.033902 (2016).
47. Wang, X. *et al.* miRNA-3473b contributes to neuroinflammation following cerebral ischemia. *Cell Death Dis.* **9**, 11 (2018).
48. Tiwari, N. *et al.* Identification and Characterization of miRNAs in Response to Leishmania donovani Infection: Delineation of Their Roles in Macrophage Dysfunction. *Front. Microbiol.* **8** (2017).
49. Feero, W. G., Guttmacher, A. E. & Collins, F. S. Genomic Medicine — An Updated Primer. *The New England journal of medicine* **362** (2010).
50. Griffiths-Jones, S. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34**, D140–D144 (2006).
51. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
52. Daneck, P. *et al.* High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome Biol.* **13**, R26 (2012).
53. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
54. Betel, D., Wilson, M., Gabow, A., Marks, D. S. & Sander, C. The microRNA.org resource: targets and expression. *Nucleic Acids Res.* **36**, D149–D153 (2007).
55. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
56. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
57. Auber, D., Delest, M., Domenger, J.-P. & Dulucq, S. Efficient drawing of RNA secondary structure. *J. Graph Algorithms Appl.* **10**, 329–351 (2006).
58. Friedman, R. C., Farh, K. K.-H., Burge, C. B. & Bartel, D. *Genome research. Cold Spring Harbor Laboratory Press* (Cold Spring Harbor Laboratory Press, 2008).
59. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26.6**, 841–842 (2010).

Acknowledgements

Author contributions

G.O. and F.X. wrote the main text and performed the analysis. M.W., K.L. and Y.Z. reviewed the manuscript. J.X. designed the study and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-62588-6.

Correspondence and requests for materials should be addressed to J.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.