Non-coding RNA Research 6 (2021) 123–129

Contents lists available at ScienceDirect

Non-coding RNA Research

journal homepage: www.sciencedirect.com/journal/non-coding-rna-research

Characteristics of miRNA-SNPs in healthy Japanese subjects and non-small cell lung cancer, colorectal cancer, and soft tissue sarcoma patients

Koki Katayama [a,1], Shimon Nakashima [a,1], Hiroo Ishida [b], Yutaro Kubota [b], Masataka Nakano [a,c], Tatsuki Fukami [a,c], Yasutsuna Sasaki [b], Ken-ichi Fujita [d], Miki Nakajima [a,c,*]

[a] Drug Metabolism and Toxicology, Faculty of Pharmaceutical Sciences, Kanazawa University, Kakuma-machi, Kanazawa, 920-1192, Japan
[b] Division of Medical Oncology, Department of Medicine, Showa University School of Medicine, 1-5-8 Hatanodai, Shinagawa-ku, Tokyo, 142-8555, Japan
[c] WPI Nano Life Science Institute (WPI-NanoLSI), Kanazawa University, Kakuma-machi, Kanazawa, 920-1192, Japan
[d] Division of Cancer Genome and Pharmacotherapy, Department of Clinical Pharmacy, Showa University School of Pharmacy, 1-5-8 Hatanodai, Shinagawa-ku, Tokyo, 142-8555, Japan

ARTICLE INFO

Keywords:
Polymorphisms
NGS
Biomarker
microRNA

ABSTRACT

Single nucleotide polymorphisms in genes encoding microRNAs (miRNA-SNPs) may affect the maturation steps of miRNAs or target mRNA recognition, leading to changes in the expression of target mRNAs to cause gain- or loss-of-function changes. Several miRNA-SNPs are known to be associated with the risk of diseases such as cancer. The purpose of this study was to comprehensively determine the miRNA-SNPs in Japanese individuals to evaluate the differences in allele frequencies between ethnicities by comparing data from the global population in the 1000 Genomes Project and differences between healthy subjects and cancer patients. We performed next-generation sequencing targeting genes encoding 1809 pre-miRNAs. As a result, 403 miRNA-SNPs (146 miRNA-SNPs per subject on average) were identified in 28 healthy Japanese subjects. We observed significant differences in the allele frequencies between ethnicities in 33 of the 403 miRNA-SNPs. The numbers of miRNA-SNPs per subject in 44 non-small cell lung cancer (NSCLC), 33 colorectal cancer (CRC), and 15 soft tissue sarcoma (STS) patients were almost equal to those in healthy subjects. Significant differences in allele frequencies were observed for 14, 11, and 9 miRNA-SNPs in NSCLC, CRC, and STS patients compared with the frequencies in healthy subjects, suggesting that these SNPs can be biomarkers of risk for each type of cancer assessed. In summary, we comprehensively characterized miRNA-SNPs in Japanese individuals and found differences in allele frequencies of several miRNA-SNPs between ethnicities and between healthy subjects and cancer patients. Studies investigating a larger number of subjects should be performed to confirm the potential of miRNA-SNPs as biomarkers of cancer risk.

1. Introduction

MicroRNAs (miRNAs) are endogenous noncoding RNAs 19–25 nucleotides in length. miRNAs are first transcribed as long transcripts called primary miRNAs (pri-miRNAs) by RNA polymerase II, processed into precursor miRNAs (pre-miRNAs) having a stem-loop structure 70–100 nucleotides in length, and then processed into single-stranded mature miRNAs [1]. To date, 1881 pre-miRNAs and 2588 mature miRNAs have been identified in humans (miRBase 21). Mature miRNAs bind to target mRNAs to downregulate gene expression via translational repression or mRNA degradation [2,3]. For target recognition by miRNAs, complete sequence matching with the target mRNA is not required, but the complementarity of the seed sequence, nucleotides 2–8 from the 5' end of the miRNA, is critical. miRNA-mediated gene regulation is involved in every biological process, such as cell differentiation, proliferation, development, and apoptosis [3,4]. Dysregulation of miRNA expression is relevant to the onset or progression of various diseases, including cancer [5,6].

Abbreviations: 3′-UTR, 3′-untranslated region; BWA, Burrows-Wheeler Aligner; CI, confidence interval; CRC, colorectal cancer; GATK, Genome Analysis Toolkit; InDel, insertion or deletion mutation; miRNA, microRNA; mOR, modified odds ratio; NGS, next-generation sequencing; NSCLC, non-small cell lung cancer; pre-miRNA, precursor miRNA; SNP, single nucleotide polymorphism; STS, soft tissue sarcoma.

* Corresponding author. Drug Metabolism and Toxicology, Faculty of Pharmaceutical Sciences, Kanazawa University, Kakuma-machi, Kanazawa, 920-1192, Japan.
E-mail address: nmiki@p.kanazawa-u.ac.jp (M. Nakajima).

[1] These authors equally contributed to this work.

https://doi.org/10.1016/j.ncrna.2021.06.002
Received 1 April 2021; Received in revised form 24 June 2021; Accepted 24 June 2021
Available online 27 June 2021

Single nucleotide polymorphisms (SNPs) are a major type of genetic variation. Since SNPs are distributed throughout the whole genomic region, they exist in genes encoding proteins and genes encoding miRNAs (miRNA-SNPs). SNPs located in pre-miRNAs might affect their processing to cause a change in the expression of mature miRNAs. SNPs located in mature miRNA, especially on the seed sequence, might change target mRNA recognition. Accordingly, the miRNA-SNPs may cause functional alterations. Accumulating evidence shows that functional alterations caused by miRNA-SNPs are associated with certain diseases. For example, it has been reported that a SNP in pre-miR-27a (rs11671784) resulted in decreased expression of mature miR-27a and increased expression of a target gene, runt-related transcription factor 1, in bladder cancer cells [7]. SNPs in miR-196a-2 and miR-146a have been reported to be involved in the pathogenesis of Behcet's disease [8] and diabetic neuropathy [9], respectively. Therefore, miRNA-SNPs have the potential to be biomarkers to predict the risk of diseases and to be used to clarify the mechanisms of pathogenesis or progression of diseases.

The analysis of miRNA-SNPs is progressing in the cancer research field. Jia et al. (2014) reported that the frequency of a SNP on miR-146a (rs2910164) was significantly higher in NSCLC patients than in healthy controls in a Chinese population [10]. Qian et al. (2016) reported that several variants in miR-339–3p were associated with breast cancer risk in women of African ancestry [11]. In earlier studies, SNPs of specific miRNAs were analyzed to focus on diseases of interest. To the best of our knowledge, there are no comprehensive studies on miRNA-SNPs. In the present study, we sought to comprehensively determine miRNA-SNPs in Japanese individuals to clarify ethnic differences in allele frequencies and to examine whether allele frequencies are different between healthy subjects and patients suffering from non-small cell lung cancer (NSCLC), colorectal cancer (CRC), or soft tissue sarcoma (STS).

2. Materials and methods

2.1. Chemicals and reagents

A KAPA Library Prep Kit, SeqCap Pure Capture Bead Kit, SeqCap EZ Hybridization and Wash Kit, SeqCap EZ Accessory Kit v2, SeqCap Adapter Kit A/B and SeqCap HE-Oligo Kit A/B were obtained from Roche (Basel, Switzerland). A Qubit dsDNA HS Assay Kit was obtained from Thermo Fisher Scientific (Waltham, MA). An Agilent DNA 1000 Kit was obtained from Agilent Technologies (California, USA). MiSeq Reagent Nano Kit v2 (300 cycles), MiSeq Reagent Micro Kit v2 (300 cycles), MiSeq Reagent Kit v3 (150 cycles), and PhiX control kit v3 were obtained from Illumina (California, USA). All other chemicals and solvents were of the highest grade commercially available.

2.2. Subjects and genomic DNA preparation

This study was approved by the Ethics Committees of Kanazawa University (Kanazawa, Japan), Showa University (Tokyo, Japan), and Nagoya City University (Nagoya, Japan). We recruited 28 healthy Japanese subjects, 44 NSCLC patients, 33 CRC patients, and 15 soft tissue sarcoma (STS) patients (Table 1). Written informed consent was obtained from all subjects. Blood samples were collected from cubital veins. Genomic DNA was extracted from peripheral lymphocytes using a Gentra Puregene Blood Kit (Gentra Systems, Minneapolis, MN, USA).

Table 1
Characteristics of subjects.

Variables	Healthy subjects	NSCLC patients	CRC patients	STS patients
Number of subjects	28	44	33	15
Age Median (range)	34 (27–74)	70 (45–89)	68 (41–78)	56 (33–80)
Gender Male (%)	19 (67.9)	28 (63.6)	14 (42.4)	11 (73.3)
Female (%)	9 (32.1)	16 (36.4)	19 (57.6)	4 (26.7)

NSCLC: non-small cell lung cancer. CRC: colorectal cancer. STS: soft tumor sarcoma.

2.3. Library design for hybrid selection

For DNA capture, a custom NimbleGen SeqCap EZ Choice Library (Roche, Basel, Switzerland) was designed to target the genomic sequence of 1881 human pre-miRNAs registered in the miRBase 21 database. DNA baits were selected using stringent settings for probe design (uniqueness was tested by sequence search and alignment was determined by the hashing algorithm). After removing target sequences predicted to provide no increased depth of coverage, 1809 (96.1%) pre-miRNAs were covered with capture baits for a final target region of 0.2 Mbp.

2.4. Target resequencing

Genomic capture was carried out following the instructions of the SeqCap EZ Library SR User's Guide v4.2 (Roche, Inc.). Briefly, 1 μg of genomic DNA extracted from peripheral blood was fragmented into sizes of 100–500 bp using a Bioruptor UCD-250HSA instrument (Cosmo Bio, Tokyo, Japan). Then, the fragments were subjected to three enzymatic steps: end repair, A-tailing, and ligation to SeqCap Index Adapters (Roche). Once the DNA samples were indexed, they were amplified by PCR (7 cycles). After the adapter-ligated fragments from 11, 18, or 24 individuals were pooled (Supplemental Table1), they were hybridized to the custom NimbleGen SeqCap EZ Choice Library for 72 h, thoroughly washed, and amplified by PCR (14 cycles). We used sequencing reagents to generate 2 × 150 bp paired-end reads using a MiSeq instrument (Illumina).

2.5. Bioinformatics analysis of miRNA-SNPs and statistical analysis

After sequencing, the fastq data were aligned to the human reference genome build UCSC hg19 (http://genome.ucsc.edu/) using BWA (Burrows-Wheeler Aligner, http://biobwa.sourceforge.net/). Duplicated reads were assessed using Picard. Read realignment, base recalibration, and variant calling were performed with the Genome Analysis Toolkit (GATK). Data filtering and annotation were performed using VariantStudio (ver 3.0.4, Illumina). SNPs were adopted with more than 20 reads and mt/wt ratios higher than 0.2 to improve the data quality.

2.6. Statistical analysis

Fisher's exact test was used to determine the significant difference in the allele frequency of each variant. Comparisons of multiple groups were made with one-way ANOVA. When the P value was less than 0.05, the difference was considered to be statistically significant.

3. Results

3.1. Sequencing statistics

Next-generation sequencing (NGS) targeting the DNA regions encoding 1809 pre-miRNA sequences was performed. The analyses of 120 DNA samples in total (28 healthy subjects, 44 NSCLC patients, 33 CRC patients, and 15 STS patients) were performed in eight runs (Supplemental Table1). DNA libraries #1, #2–7, and #8 were analyzed using MiSeq Reagent Nano Kit v2, MiSeq Reagent Micro Kit v2, and MiSeq Reagent Kit v3, respectively, which can be used to analyze 150 Mb, 1.2 Gb, and 3.8 Gb, respectively. As shown in Table 2, the numbers of reads mapped to a reference sequence (hg19) and those to the target region encoding pre-miRNA of DNA library #1 were lower than those of DNA libraries #2–7, and those of DNA library #8 were higher than those of DNA libraries #2–7, in accordance with the abilities of the kits used. A similar difference was observed in the average coverages, 58.8 in DNA

Table 2
Parameters of NGS runs.

DNA library	#1	#2	#3	#4	#5	#6	#7	#8
Number of read mapped to hg19	276,259	558,708	493,895	462,535	484,447	481,970	589,010	1,731,750
Number of read mapped to target region	168,138	315,184	242,593	294,925	257,492	232,186	431,672	1,262,703
Average coverage (X)	58.8	109.1	84.5	105.1	90.2	81.8	151.7	440.7
Target coverage at 20*X* (%)	95.6	98.0	97.1	97.7	97.9	96.8	98.5	99.8
Q30 (%)	93.2	92.6	93.7	88.3	92.8	94.9	95.7	97.7

library #1, 81.8–151.7 in DNA library #2–7, and 440.7 in DNA library #8. The rates of target bases with at least 20*X* coverage (target coverages at 20*X*) in DNA libraries #1, #2–7, and #8 were 95.6%, 97.1–98.5%, and 99.8%, respectively. For the quality score Q30, which represents the probability that there is a 1 in 1000 chance that the base is wrongly called, the accuracy was 88.3–97.7%, exceeding the manufacturer's official guarantee of 80%. When considered together, all samples fulfilled requirements for reliable variant calling.

3.2. miRNA-SNPs in healthy Japanese subjects

By NGS analysis of 28 healthy Japanese subjects, 404 SNPs and 38 insertions/deletions (InDels) were identified. Among these findings, 19 SNPs and 6 InDels (Table 3) were not registered in the dbSNP database, indicating that they are novel mutations. On average, 144 ± 7 (127–156) SNPs and 17 ± 2 (13–25) InDels per subject were detected. The preference of the location of mutations, i.e., the region encoding the pre-mRNA sequence except that of the mature miRNA (black line in Fig. 1A), the mature miRNA sequence except for the seed sequence (green line), and the seed sequence (red line), was examined. Of the 144 ± 7 (127–156) SNPs, 103 ± 7 (90–116), 26 ± 3 (20–33), and 15 ± 3 (10–23) SNPs were located within regions encoding pre-miRNA, mature miRNA, and seed sequences, respectively (Fig. 1B). The distribution of miRNA-SNPs in each subject is shown in Fig. 1C. Of the 38 InDels we found, 27, 9, and 2 InDels were located within regions encoding pre-miRNA, mature miRNA, and seed sequences, respectively, with similar preferences for the genomic distribution of the SNPs. Even when the numbers of miRNA-SNPs were corrected to the total length of pre-miRNA, mature miRNA, and seed sequences (80,887 nucleotides, 38,765 nucleotides, and 18,578 nucleotides, respectively), the frequency of the SNPs in the mature or seed regions was lower than that in the pre-miRNA (Fig. 1D), indicating that the sequences of mature miRNAs are relatively conserved.

The allele frequencies of miRNA-SNPs found in this study were compared with those in the global population in the 1000 Genomes Project (https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/), which includes data from 1008 East Asian (Eas), 978 South Asian (Sas), 1322 African (Afr), 1006 European (Eur), and 694 American (Amr) individuals. Since the allele frequencies in the global population for 106 of 404 miRNA-SNPs were not available, the remaining 298 miRNA-SNPs were analyzed. Of 298 miRNA-SNPs, the number of miRNA-SNPs whose allele frequencies differed from those in Japanese subjects by more than 10% were 50, 122, 144, 113, and 107 in the Eas, Sas, Afr, Eur, and Amr subjects, respectively (Supplemental Figure1). Thus, the difference in the allele frequencies of miRNA-SNPs between Japanese and Eas individuals was the smallest among the populations, a finding that is reasonable because Japanese individuals are included in the Eas group. Of 298 miRNA-SNPs, 14 miRNA-SNPs showed a difference in allele frequencies of more than 10% between the Japanese and all the other populations (Table 4). For example, the frequency of SNPs on pre-miR-5191 in Japanese individuals was higher than that in other populations, whereas the frequency of SNPs on pre-miR-3183 in Japanese individuals was lower than that in other populations. For the location, the SNPs of miR-608 and miR-6763 are in the region encoding the mature miRNA sequence and seed sequence, respectively, and the remaining 12 miRNA-SNPs are in the region encoding pre-miRNA. Since the number of Japanese subjects analyzed in this study is limited, further studies with larger sample sizes are needed to confirm the ethnic differences in allele frequencies.

Table 3
miRNA SNPs and InDels which were not registered in dbSNP.

miRNA	Location[a]	Region	Variant	Type	Allele frequency
hsa-mir-7852	+74	mature	A > G	SNP	1.8
hsa-mir-7845	+45	precursor	G > A	SNP	1.8
hsa-mir-4776-1	+20	mature	A > G	SNP	1.8
hsa-mir-6822	+15	mature	A > G	SNP	1.8
hsa-mir-4636	+35	precursor	T > A	SNP	1.8
hsa-mir-2277	+30	seed	C > A	SNP	1.8
hsa-mir-2277	+31	seed	T > A	SNP	1.8
hsa-mir-1244-2	+20	precursor	A > G	SNP	1.8
hsa-mir-6720	+48	mature	C > G	SNP	1.8
hsa-mir-8055	+17	mature	G > C	SNP	1.8
hsa-mir-8055	+47	precursor	G > A	SNP	1.8
hsa-mir-101-2	+15	precursor	G > A	SNP	1.8
hsa-mir-1908	+77	precursor	C > G	SNP	1.8
hsa-mir-627	+62	mature	C > G	SNP	1.8
hsa-mir-4518	+20	precursor	T > C	SNP	1.8
hsa-mir-6782	+49	mature	A > C	SNP	1.8
hsa-mir-8069-1	+9	precursor	C > A	SNP	1.8
hsa-mir-6069	+46	precursor	C > T	SNP	1.8
hsa-mir-4763	+90	precursor	G > A	SNP	1.8
hsa-mir-4262	+29	seed	T > TC	Insertion	1.8
hsa-mir-3689b	+60	precursor	A > AAGCACGGTATCACACCTCCCAGGA	Insertion	1.8
hsa-mir-4511	+9	precursor	CT > CTT	Insertion	1.8
hsa-mir-548ae-1	+3	precursor	AGTTTTTGCCATTAAGTTGCG > A	Deletion	1.8
hsa-mir-7705	+28	precursor	TAAAAC > T	Deletion	14.3
hsa-mir-3180-4	+136	precursor	ACAGCGCGACCAGGC > A	Deletion	3.6

[a] The numbering denotes the 5′ end of the pre-miRNA as +1.

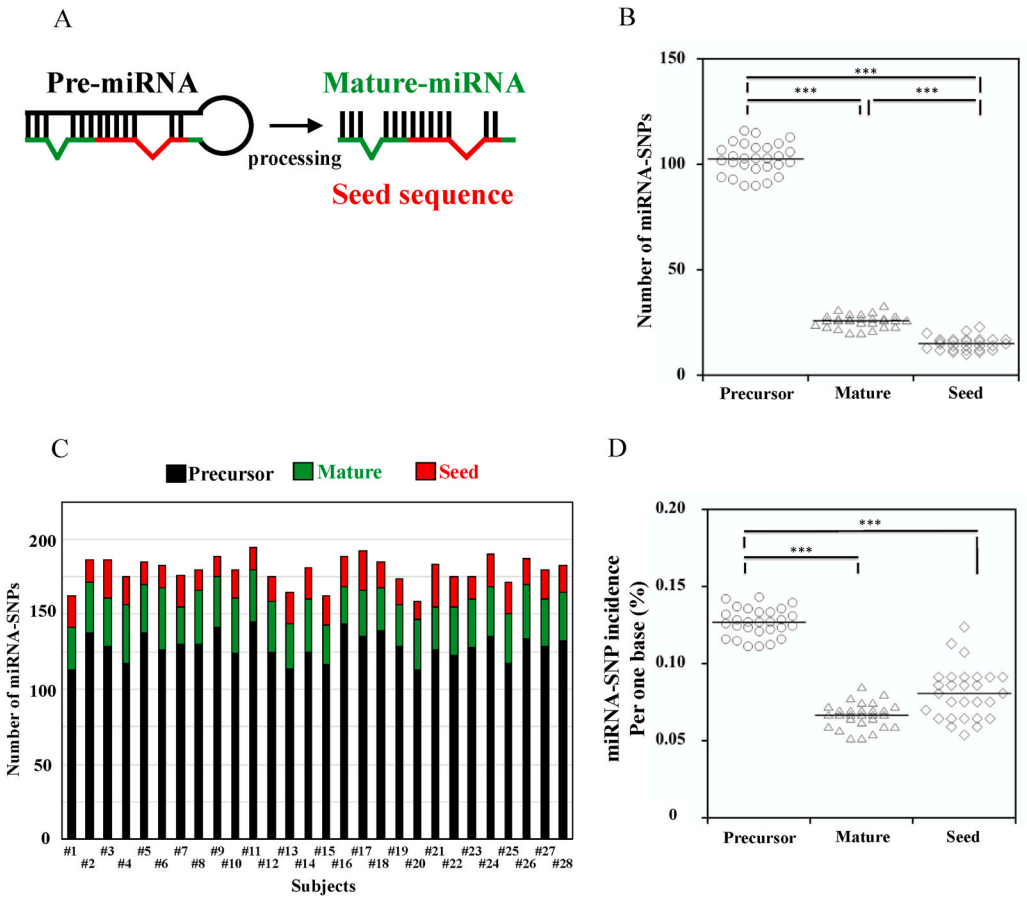


Fig. 1. Incidence of miRNA-SNPs. Biogenesis of miRNAs (A) and each region of miRNA-SNPs identified in healthy Japanese subjects (B). The incidence rate was calculated by dividing by the miRNA number (D) or total length of miRNAs (E). Each bar represents the mean (n = 28). ***$P < 0.001$ by one-way ANOVA. Seed: seed sequence, Mature: mature miRNA, Precursor: precursor miRNA.

Table 4
miRNA-SNPs whose allele frequency in 28 Japanese healthy subjects were different (>10%) from global population of allele frequency.

miRNA	dbSNP ID	SNP	Region	allele frequency					
				Japanese	Eas	Sas	Amr	Eur	Afr
hsa-mir-149	rs2292832	T > C	precursor	12.5	36.3	55.4	69.0	71.8	72.8
hsa-mir-608	rs4919510	C > G	mature	62.5	52.5	34.0	28.7	17.9	44.0
hsa-mir-1208	rs2648841	G > T	precursor	53.6	43.5	22.2	13.8	7.7	7.9
hsa-mir-1908	rs174561	T > C	precursor	41.1	54.7	12.7	56.8	30.3	2.0
hsa-mir-3180-4	rs75000738	C > A	precursor	42.9	67.7	78.4	85.3	80.6	31.6
hsa-mir-3183	rs2663345	A > G	precursor	12.5	40.7	37.4	59.5	34.2	74.8
hsa-mir-3689a	rs113454901	G > A	precursor	50.0	98.2	83.4	88.8	81.8	90.7
hsa-mir-4432	rs243080	G > A	precursor	80.4	60.9	46.7	50.6	48.0	26.3
hsa-mir-4432	rs56239160	A > G	precursor	80.4	60.4	33.1	28.0	6.2	2.5
hsa-mir-4507	rs62651104	C > T	precursor	23.2	57.0	38.5	54.9	48.7	55.9
hsa-mir-4511	rs2060455	A > G	precursor	51.8	72.3	37.3	19.9	17.1	13.1
hsa-mir-4719	rs58353328	A > G	precursor	25.0	12.4	6.7	2.0	3.6	4.8
hsa-mir-5191	rs76756293	C > T	precursor	21.4	10.1	6.9	3.9	0.5	5.4
hsa-mir-6763	rs3751304	C > T	seed	92.9	80.7	72.0	82.7	69.5	66.8

3.3. Comparison of miRNA-SNPs in NSCLC, CRC, and STS patients with those in healthy subjects

Next, we examined whether the frequencies or distribution of miRNA-SNPs are different between cancer patients and healthy individuals. The numbers of miRNA-SNPs found in 44 NSCLC patients, 33 CRC patients, and 15 STS patients were 146 ± 9 (125–167), 146 ± 9 (121–164), and 147 ± 9 (131–162), respectively (Fig. 2A), with no significant difference from those in healthy subjects, 144 ± 7 (127–156). The number of InDels in the three cancer groups were 17 ± 2 (13–23), 18 ± 2 (13–23), and 17 ± 2 (13–21) (Fig. 2B). There was no significant difference in the numbers of miRNA-SNPs or InDels between each patient group and the healthy subject group.

We sought to identify the miRNA-SNPs or InDels whose allele frequencies were significantly different between patients and healthy subjects by Fisher's exact test. As shown in Fig. 3A, we found that the allele frequencies of 14 miRNA-SNPs and 2 InDels were significantly different between NSCLC patients and healthy subjects. The frequencies of 10 miRNA-SNPs and 2 InDels were higher in NSCLC patients than in healthy subjects, and those of 4 miRNA-SNPs were lower in the NSCLC

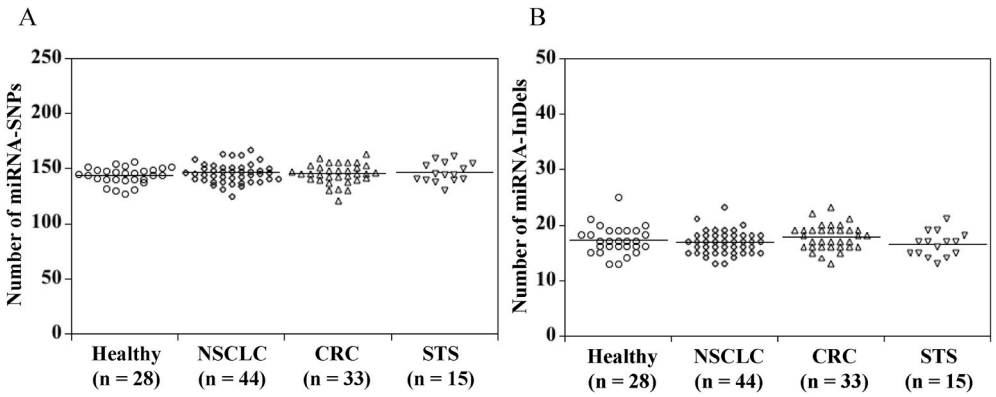
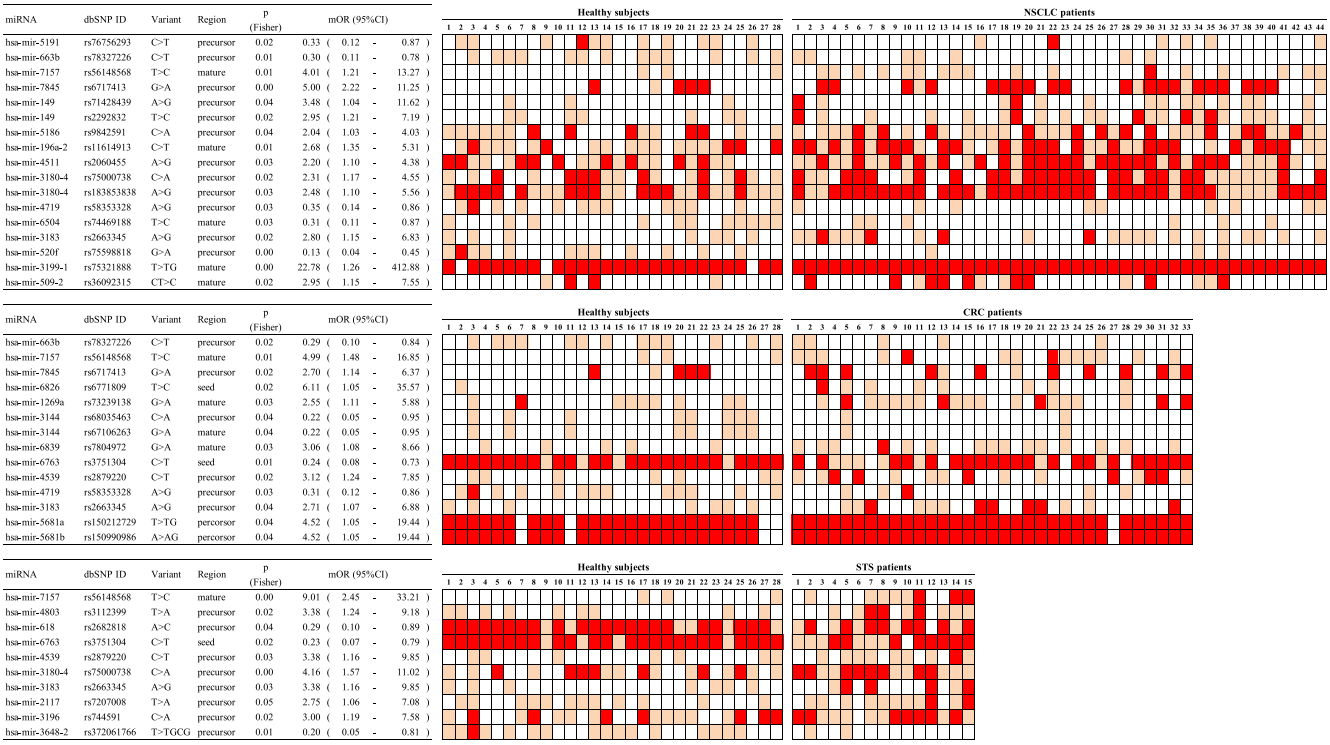


Fig. 2. Number of miRNA-SNPs (A) and InDels (B) in Japanese subjects. Each bar represents the mean. Healthy: healthy subjects. NSCLC: non-small cell lung cancer. CRC: colorectal cancer. STS: soft tissue sarcoma.



miRNA	dbSNP ID	Variant	Region	p (Fisher)	mOR (95%CI)
hsa-mir-5191	rs76756293	C>T	precursor	0.02	0.33 (0.12 - 0.87)
hsa-mir-663b	rs78327226	C>T	precursor	0.01	0.30 (0.11 - 0.78)
hsa-mir-7157	rs56148568	T>C	mature	0.01	4.01 (1.21 - 13.27)
hsa-mir-7845	rs6717413	G>A	precursor	0.00	5.00 (2.22 - 11.25)
hsa-mir-149	rs71428439	A>G	precursor	0.04	3.48 (1.04 - 11.62)
hsa-mir-149	rs2292832	T>C	precursor	0.02	2.95 (1.21 - 7.19)
hsa-mir-5186	rs9842591	C>A	precursor	0.04	2.04 (1.03 - 4.03)
hsa-mir-196a-2	rs11614913	C>T	mature	0.01	2.68 (1.35 - 5.31)
hsa-mir-4511	rs2060455	A>G	precursor	0.03	2.20 (1.10 - 4.38)
hsa-mir-3180-4	rs75000738	C>A	precursor	0.02	2.31 (1.17 - 4.55)
hsa-mir-3180-4	rs183853838	A>G	precursor	0.03	2.48 (1.10 - 5.56)
hsa-mir-4719	rs58353328	A>G	precursor	0.03	0.35 (0.14 - 0.86)
hsa-mir-6504	rs74469188	T>C	mature	0.03	0.31 (0.11 - 0.87)
hsa-mir-3183	rs2663345	A>G	precursor	0.02	2.80 (1.15 - 6.83)
hsa-mir-520f	rs75598818	G>A	precursor	0.00	0.13 (0.04 - 0.45)
hsa-mir-3199-1	rs75321888	T>TG	mature	0.00	22.78 (1.26 - 412.88)
hsa-mir-509-2	rs36092315	CT>C	mature	0.02	2.95 (1.15 - 7.55)

miRNA	dbSNP ID	Variant	Region	p (Fisher)	mOR (95%CI)
hsa-mir-663b	rs78327226	C>T	precursor	0.02	0.29 (0.10 - 0.84)
hsa-mir-7157	rs56148568	T>C	mature	0.01	4.99 (1.48 - 16.85)
hsa-mir-7845	rs6717413	G>A	precursor	0.02	2.70 (1.14 - 6.37)
hsa-mir-6826	rs6771809	T>C	seed	0.02	6.11 (1.05 - 35.57)
hsa-mir-1269a	rs73239138	G>A	mature	0.03	2.55 (1.11 - 5.88)
hsa-mir-3144	rs68035463	C>A	precursor	0.04	0.22 (0.05 - 0.95)
hsa-mir-3144	rs67106263	G>A	mature	0.04	0.22 (0.05 - 0.95)
hsa-mir-6839	rs7804972	G>A	mature	0.03	3.06 (1.08 - 8.66)
hsa-mir-6763	rs3751304	C>T	seed	0.01	0.24 (0.08 - 0.73)
hsa-mir-4539	rs2879220	C>T	precursor	0.02	3.12 (1.24 - 7.85)
hsa-mir-4719	rs58353328	A>G	precursor	0.03	0.31 (0.12 - 0.86)
hsa-mir-3183	rs2663345	A>G	precursor	0.04	2.71 (1.07 - 6.88)
hsa-mir-5681a	rs150212729	T>TG	percursor	0.04	4.52 (1.05 - 19.44)
hsa-mir-5681b	rs150990986	A>AG	percursor	0.04	4.52 (1.05 - 19.44)

miRNA	dbSNP ID	Variant	Region	p (Fisher)	mOR (95%CI)
hsa-mir-7157	rs56148568	T>C	mature	0.00	9.01 (2.45 - 33.21)
hsa-mir-4803	rs3112399	T>A	precursor	0.02	3.38 (1.24 - 9.18)
hsa-mir-618	rs2682818	A>C	precursor	0.04	0.29 (0.10 - 0.89)
hsa-mir-6763	rs3751304	C>T	seed	0.02	0.23 (0.07 - 0.79)
hsa-mir-4539	rs2879220	C>T	precursor	0.03	3.38 (1.16 - 9.85)
hsa-mir-3180-4	rs75000738	C>A	precursor	0.00	4.16 (1.57 - 11.02)
hsa-mir-3183	rs2663345	A>G	precursor	0.03	3.38 (1.16 - 9.85)
hsa-mir-2117	rs7207008	T>A	precursor	0.05	2.75 (1.06 - 7.08)
hsa-mir-3196	rs744591	C>A	precursor	0.02	3.00 (1.19 - 7.58)
hsa-mir-3648-2	rs372061766	T>TGCG	precursor	0.01	0.20 (0.05 - 0.81)

Fig. 3. miRNA-SNPs and InDels whose allele frequencies were significantly different between healthy subjects and NSCLC (A), CRC (B), or STS (C) patients. miRNA-SNPs for which allele frequency values were significantly different between the healthy subject group and patient group are listed ($P < 0.05$, Fisher's exact test). Red, orange, and white squares represent mt/mt, wt/mt, and wt/wt, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

patients. Three and 11 miRNA-SNPs were on the regions encoding mature miRNA and pre-miRNA, respectively. Two InDels were located on the regions encoding mature miRNA. As shown in Fig. 3B, the allele frequencies of 11 miRNA-SNPs and 2 InDels were significantly different between CRC patients and healthy subjects. The frequencies of 7 miRNA-SNPs and 2 InDels were higher in CRC patients than in healthy subjects, and those of 4 miRNA-SNPs were lower in the CRC patients. Two, 4, and 5 miRNA-SNPs were on the regions encoding seed sequences, mature miRNA, and pre-miRNA, respectively. Two InDels were located on the regions encoding pre-miRNA. Since hsa-mir-5681a and hsa-mir-5681b are located on chromosome 8 at positions 74,548,543–74,548,617[+] and 74,548,550–74,548,609 [-], respectively, these inDels are identical to each other. As shown in Fig. 3C, the allele frequencies of 9 miRNA-SNPs and 1 InDel were significantly different between STS patients and healthy subjects. The frequencies of 7 miRNA-SNPs were higher in STS patients than in healthy subjects, and those of 2 miRNA-SNPs and 1 InDel were lower in in the STS patients. One, 1, and 7 miRNA-SNPs were in the regions encoding seed sequences, mature miRNA, and pre-miRNA, respectively. One indel was located on the regions encoding pre-miRNA. Interestingly, among the miRNA-SNPs shown in Fig. 3A to C, the SNP on miR-7157 was common. Therefore, it may be a risk marker of solid cancer regardless of tissue type. The other miRNA-SNPs may be risk markers of NSCLC, CRC, or STS.

4. Discussion

MiRNAs, major posttranscriptional regulators, play critical roles in regulating various key biological processes, such as cell growth, proliferation, differentiation, and apoptosis [12]. Dysregulation of miRNA is relevant to the incidence or progression of diseases, including cancer

[13]. SNPs, which may affect gene expression or protein function, can be coding genes or noncoding genes. The SNPs within the pre-miRNA sequence may alter the processing of mature miRNA, and those within mature miRNA, especially in the seed sequence, may alter target recognition [12]. Earlier studies have reported the association between miRNA-SNPs and cancer incidence or the severity of adverse reactions by anticancer agents in Caucasians and Chinese individuals [14–16]. Generally, ethnic differences are observed in allele frequencies of mutations. The purposes of this study are 1) to examine miRNA-SNPs in healthy Japanese subjects by comparing allele frequencies in other populations and 2) to examine whether allele frequencies of miRNA-SNPs in cancer patients are different from those in healthy subjects, considering the possibility that miRNA-SNPs can be biomarkers of risk of cancer incidence.

First, NGS targeting the DNA regions encoding 1809 pre-miRNA sequences was performed using DNA samples from 28 healthy subjects. We found, on average, 144 miRNA-SNPs and 17 InDels per subject (Fig. 1). It was demonstrated that miRNA-SNPs on the seed sequence were relatively rare compared with those on the mature miRNA or the pre-miRNA regions. This finding is reasonable because mature miRNA sequences, especially seed sequences, are critical for the function of miRNAs, and the occurrence of mutations in genes that are essential for maintaining homeostasis is generally rare. Next, we sought to examine ethnic differences in allele frequencies of these miRNA-SNPs between Japanese and other populations in the 1000 Genomes Project. Of the 298 miRNA-SNPs examined, the numbers of miRNA-SNPs whose allele frequencies showed more than a10% difference between Japanese subjects and Eas, Sas, Afr, Eur, or Amr subjects were 50, 122, 144, 113, or 107, respectively (Supplemental Figure1). Thus, as expected, it was demonstrated that there were ethnic differences in allele frequencies of these miRNA-SNPs.

To explore the potential of miRNA-SNPs to serve as biomarkers of cancer risk in Japanese patients, allele frequencies of miRNA-SNPs and InDels in healthy subjects and cancer (NSCLC, CRC, and STS) patients were compared. We found that the allele frequencies of 12, 7, and 12 miRNA-SNPs were significantly different between NSCLC, CRC, or STS patients and healthy subjects, respectively (Fig. 3A-C). For the SNPs on genes encoding miR-196a-2 (rs11614913 T > C) or miR-149 (rs2292832 C > T) that were highly expressed in NSCLC patients, there have been several reports in other populations as follows: the allele frequency of a SNP on miR-196a-2 (rs11614913) in NSCLC patients was higher than that in healthy Korean subjects (OR, 1.42; 95% CI, 1.03 to 1.96) [17]. The allele frequencies of the SNP on pre-mir-149 (rs2292832) have been reported to be significantly higher in hepatocellular carcinoma [18] and breast cancer [19] patients than in healthy subjects. Therefore, this SNP may also be a cancer risk marker regardless of cancer type. In addition to these SNPs, it has been reported that a SNP in miR-146a (rs2910164) is associated with a reduced risk of breast, lung, and gastric cancer [20–22]. Recently, miR-SNP in miR-27a, which decreased the expression of mature miR-27a compared with the wild type and increased the expression level of RUNX-1, was reported to be involved in its regulation of chemosensitivity in bladder cancer [7]. Since miRNA-SNPs are associated with cancer risk in this study, it will also be important to investigate what kind of effects these miRNA-SNPs have.

5. Conclusion

In conclusion, we comprehensively characterized miRNA-SNPs in Japanese individuals and found ethnic differences in allele frequencies in several miRNAs. In addition, we identified miRNA-SNPs as potential biomarkers of the incidence of NSCLC, CRC, and STS. A limitation of this study is that the healthy subjects were younger than the cancer patients. Ideally, subjects whose age is matched by patient group are recruited. To evaluate the validity of miRNA-SNPs as cancer biomarkers, studies with a larger number of age-matched patients/subjects should be performed.

CRediT authorship contribution statement

Koki Katayama: Investigation, Formal analysis, Writing – original draft. **Shimon Nakashima:** Investigation, Formal analysis, Writing – review & editing. **Hiroo Ishida:** Resources. **Yutaro Kubota:** Resources. **Masataka Nakano:** Writing – review & editing. **Tatsuki Fukami:** Writing – review & editing. **Yasutsuna Sasaki:** Resources. **Ken-ichi Fujita:** Resources, Writing – review & editing. **Miki Nakajima:** Conceptualization, Project administration, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by JSPS KAKENHI (15K14999), Japan Research Foundation for Clinical Pharmacology, and Takeda Science Foundation.

References

[1] V.N. Kim, MicroRNA biogenesis: coordinated cropping and dicing, Nat. Rev. Mol. Cell Biol. 6 (2005) 376–385, https://doi.org/10.1038/nrm1644. https://www.nature.com/articles/nrm1644.
[2] Bartel MicroRNAs, Genomics, biogenesis, mechanism, and function, Cell 116 (2004) 281–297, https://doi.org/10.1016/S0092-8674(04)00045-5. https://www.sciencedirect.com/science/article/pii/S0092867404000455.
[3] L. He, G.L. Hannon, MicroRNAs: small RNAs with a big role in gene regulation, Nat. Rev. Genet. 5 (2004) 522–531, https://doi.org/10.1038/nrg1379. https://www.nature.com/articles/nrg1379.
[4] P.S. Meltzer, Cancer genomics: small RNAs with big impacts, Nature 435 (2005) 745–746, https://doi.org/10.1038/435745a. https://www.nature.com/articles/435745a.
[5] K. Srivastava, Srivastava, Comprehensive review of genetic association studies and meta-analyses on miRNA polymorphisms and cancer risk, PloS One 7 (2012), e50966, https://doi.org/10.1371/journal.pone.0050966. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0050966.
[6] S.J. Kim, J.Y. Shin, K.D. Lee, Y.K. Bae, K.W. Sung, S.J. Nam, K.H. Chun, MicroRNA let-7a suppresses breast cancer cell migration and invasion through downregulation of C-C chemokine receptor type 7, Breast Cancer Res. 14 (2012) R14, https://doi.org/10.1186/bcr3098. https://link.springer.com/article/10.1186/bcr3098.
[7] Y. Deng, H. Bai, Hu, Hrs11671784 G/A variation in miR-27a decreases chemosensitivity of bladder cancer by decreasing miR-27a and increasing the target RUNX-1 expression, Biochem. Biophys. Res. Commun. 458 (2015) 321–327, https://doi.org/10.1016/j.bbrc.2015.01.109. https://www.sciencedirect.com/science/article/pii/S0006291X15001485.
[8] J. Qi, S. Hou, Q. Zhang, D. Liao, L. Wei, J. Fang, Y. Zhou, A. Kijlstra, P. Yang, A functional variant of pre-miRNA-196a2 confers risk for Behcet's disease but not for Vogt-Koyanagi-Harada syndrome or AAU in ankylosing spondylitis, Hum. Genet. 132 (2013) 1395–1404, https://doi.org/10.1007/s00439-013-1346-8. https://link.springer.com/article/10.1007%2Fs00439-013-1346-8.
[9] C. Ciccacci, R. Morganti, D. Di Fusco, C. D'Amato, L. Cacciotti, C. Greco, S. Rufini, G. Novelli, F. Sangiuolo, G.A. Marfia, P. Borgiani, V. Spallone, Common polymorphisms in MIR146a, MIR128a and MIR27a genes contribute to neuropathy susceptibility in type 2 diabetes, Acta Diabetol. 51 (2014) 663–671, https://doi.org/10.1007/s00592-014-0582-2. https://link.springer.com/article/10.1007%2Fs00592-014-0582-2.
[10] Y. Jia, A. Zang, Y. Shang, H. Yang, Z. Song, Z. Wang, L. Ren, Y. Wei, L. Hu, H. Shi, H. Li, MicroRNA-146a rs2910164 polymorphism is associated with susceptibility to non-small cell lung cancer in the Chinese population, Med. Oncol. 31 (2014) 194–198, https://doi.org/10.1007/s12032-014-0194-2. https://link.springer.com/article/10.1007%2Fs12032-014-0194-2.
[11] F. Qian, Y. Feng, Y. Zheng, T.O. Ogundiran, O. Ojengbede, W. Zheng, W. Blot, C. B. Ambrosone, E.M. John, L. Bernstein, J.J. Hu, R.G. Ziegler, S. Nyante, E. V. Bandera, S.A. Ingles, M.F. Press, K.L. Nathanson, A. Hennis, B. Nemesure, S. Ambs, L.N. Kolonel, O.I. Olopade, C.A. Haiman, Huo, DGenetic variants in microRNA and microRNA biogenesis pathway genes and breast cancer risk among women of African ancestry, Hum. Genet. 135 (2016) 1145–1159, https://doi.org/10.1007/s00439-016-1707-1. https://link.springer.com/article/10.1007%2Fs00439-016-1707-1.
[12] G. Michlewski, J.F. Cáceres, Post-transcriptional control of miRNA biogenesis, RNA 25 (2018) 1–16, https://doi.org/10.1038/435745a. https://rnajournal.cshlp.org/content/25/1/1.long.

[13] R. Rupaimoole, F.J. Slack, MicroRNA therapeutics: towards a new era for the management of cancer and other diseases, Nat. Rev. Drug Discov. 16 (2017) 203–222, https://doi.org/10.1038/nrd.2016.246. https://www.nature.com/articles/nrd.2016.246.

[14] B. Pardini, F. Rosa, A. Naccarati, V. Vymetalkova, Y. Ye, X. Wu, C. di Gaetano, T. Buchler, J. Novotny, G. Matullo, P. Vodicka, Polymorphisms in microRNA genes as predictors of clinical outcomes in colorectal cancer patients, Carcinogenesis 36 (2015) 82–86, https://doi.org/10.1093/carcin/bgu224. https://academic.oup.com/carcin/article/36/1/82/377076.

[15] S. Peng, Z. Kuang, C. Sheng, Y. Zhang, H. Xu, Q. Cheng, Association of microRNA-196a-2 gene polymorphism with gastric cancer risk in a Chinese population, Dig. Dis. Sci. 55 (2010) 2288–2293, https://doi.org/10.1007/s10620-009-1007-x. https://link.springer.com/article/10.1007%2Fs10620-009-1007-x.

[16] C. Fang, X.P. Li, W.J. Gong, N.Y. Wu, J. Tang, J.Y. Yin, X. Li, W. Zhang, H.H. Zhou, Z.Q. Liu, Age-related common miRNA polymorphism associated with severe toxicity in lung cancer patients treated with platinum-based chemotherapy, Clin. Exp. Pharmacol. Physiol. 44 (2017) 21–29, https://doi.org/10.1111/1440-1681.12704. https://onlinelibrary.wiley.com/doi/abs/10.1111/1440-1681.12704.

[17] Y.S. Hong, H.J. Kang, J.Y. Kwak, B.L. Park, C.H. You, Y.M. Kim, H. Kim, Association between microRNA196a2 rs11614913 genotypes and the risk of non-small cell lung cancer in Korean population, J Prev Med Public Health 44 (2011) 125–130, https://doi.org/10.3961/jpmph.2011.44.3.125. https://www.jpmph.org/journal/view.php?doi=10.3961/jpmph.2011.44.3.125.

[18] H. Jia, H. Yu, Q. Liu, Single nucleotide polymorphisms of MIR-149 gene rs2292832 contributes to the risk of hepatocellular carcinoma, but not overall cancer: a meta-analysis, Minerva Med. 107 (2016) 259–269. https://www.minervamedica.it/en/journals/minerva-medica/article.php?cod=R10Y2016N04A0259.

[19] Y. Feng, F. Duan, C. Song, X. Zhao, L. Dai, S. Cui, Systematic evaluation of cancer risk associated with rs2292832 in miR-149 and rs895819 in miR-27a: a comprehensive and updated meta-analysis, Oncotarget 7 (2016) 22368–22384, https://doi.org/10.18632/oncotarget.8082. https://www.oncotarget.com/article/8082/text/.

[20] C. Bansal, K.L. Sharma, S. Misra, A.N. Srivastava, B. Mittal, U.S. Singh, Common genetic variants in pre-microRNAs and risk of breast cancer in the North Indian population, Ecancermedicalscience 8 (2014) 473–481, https://doi.org/10.3332/ecancer.2014.473. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4208924/.

[21] H.S. Jeon, Y.H. Lee, S.Y. Lee, J.A. Jang, Y.Y. Choi, S.S. Yoo, W.K. Lee, J.E. Choi, J. W. Son, Y.M. Kang, J.Y. Park, A common polymorphism in pre-microRNA-146a is associated with lung cancer risk in a Korean population, Gene 534 (2014) 66–71, https://doi.org/10.1016/j.gene.2013.10.014. https://www.sciencedirect.com/science/article/pii/S0378111913013905?via%3Dihub.

[22] J. Jiang, Z.F. Jia, D.H. Cao, Y.H. Wu, Z.W. Sun, X.Y. Cao, Association of the miR-146a rs2910164 polymorphism with gastric cancer susceptibility and prognosis, Future Oncol. 12 (2016) 2215–2226, https://doi.org/10.2217/fon-2016-0224. https://www.futuremedicine.com/doi/10.2217/fon-2016-0224?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0pubmed.