**Association study to East-Asian common SNPs in microRNAs identifies novel rheumatoid arthritis susceptibility genetic variants**

Shicheng Guo1, Yehua Jin2,3,Yanqin Bian2,3, Cen Chang2,3, Lingxia Xu2,3, Runrun Zhang2,3, Nong Nie4,5, Judith A. Smith6,7, Steven J Schrodi1,8, Dongyi He2,3

1, Department of Medical Genetics, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, USA

2, Department of Rheumatology, Shanghai Guanghua Hospital of Integrated Traditional and Western Medicine, Shanghai 200052, China.

3, Arthritis Institute of integrated Traditional and Western medicine, Shanghai Chinese Medicine Research Institute, Shanghai 200052, China.

4, Computation and Informatics in Biology and Medicine, University of Wisconsin-Madison, Madison, WI, USA

5, Department of Pediatrics, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, USA

# These authors contributed equally to this work

\*Correspondence:

Steven J Schrodi, Ph.D.

Department of Medical Genetics

School of Medicine and Public Health

University of Wisconsin-Madison, Madison

Tel: xxx-xxx-xxxx

Email: [Schrodi@Wisc.edu](mailto:Schrodi@Wisc.edu)

Dongyi He, M.D., Ph.D.

Department of Rheumatology

Shanghai Guanghua Hospital of Integrated Traditional and Western Medicine,

Shanghai, China

Email: [dongyihe@medmail.com.cn](mailto:dongyihe@medmail.com.cn)

Running title: Novel microRNA associated with rheumatoid arthritis

# Abstract

Genome-wide association studies have > 100 genetic risk or protective factors for rheumatoid arthritis. However, the reported genetic variants could only explain less than 40% heritability of rheumatoid arthritis. Majority of the heritability is still missing which require to be identified with more studies in different populations. Identification of function SNPs will explain missing heritability and reveal novel mechanism pathogenesis of rheumatoid arthritis. In this study, 225 common SNPs located in miRNA and 4 HLA SNPs which might influence the target binding or pre-miRNA stability were genotyped in 1,607 rheumatoid arthritis 1,580 matched normal individuals. We identified two novel SNPs which were significant associated with rheumatoid arthritis including rs1414273 (OR=0.84, P=8.26x10-4) and rs2620381 (OR=0.77, P=2.55x10-3). In addition, we found individuals carried 10 risk alleles showed 9.17 times more risk to be affected by RA. We also identified that rs5997893 (miR-3928) showed significant epistasis effect with HLA risk alleles including rs4947332 in HLA-DQB1 (OR=, P=) while rs2967897 (miR-5695) showed significant epistasis effect with rs7752903 in TNFAIP3 (OR=, P=).

# Background

Rheumatoid arthritis is a chronic inflammatory disorder caused by the interaction between multiple factors including genetics, epigenetics and environment. Twin studies estimated the heritability of RA is about 60%. In the past decades, genome-wide association studies have > 100 genetic risk or protective factors for rheumatoid arthritis. However, the reported genetic variants could only explain less than 40% heritability of rheumatoid arthritis. Majority of the heritability is still missing which require to be identified with more studies in different populations. Meanwhile, fine-mapping study to discover more functional SNPs to explain specific mechanisms involved in the pathogenesis of rheumatoid arthritis become the most important task in the post-GWAS era. GWAS data shown that 90% of the disease-associated variants are located in non-coding regions indicating regulatory element plays important roles in the etiology of human complex disease and RA1. Among all the regulatory elements, microRNA is one of most interesting factors and play pivotal roles in both innate and adaptive immunity as well as rheumatoid arthritis onset.

In this study, we genotyped 225 East-Asian common SNPs located in human microRNA seed regions and conducted an association and epistasis study to investigate the association between miRNA and seropositive rheumatoid arthritis in a Han Chinese population (Shanghai, China). We recruited 1,078 seropositive RA and 1,045 matched control while in the second stage, 500 RA and 500 control were included to validate the identified significant associations.

# Method

**Precision Medicine Research Cohort in Shanghai Guanghua Hospital**

We randomly selected 3,223 individuals including 1,625 RA and 1,598 normal from Guanghua hospital Precision Medicine Research Cohort (PMRC). In this study, we required all the included individuals are self-reported Han Chinese while non-Han Chinese individuals were excluded for genotyping and further analysis to avoid unexpected population structure. All cases fulfilled the 2010 European League against Rheumatism–American College of Rheumatology criteria or 1987 American College of Rheumatology revised criteria for RA. All healthy controls were required do not have personal or family history of ankylospondylitis, rheumatoid arthritis, osteoarthritis, type 1 and 2 diabetes (T1D and T2D), chronic infection and common cancers. Subjects gave their written consent prior to peripheral venipuncture. This study was reviewed and approved by the Institutional Review Board of Guanghua Hospital (No: IRB12018-K-12) and all methods were performed in accordance with the relevant guidelines and regulations.

**DNA extraction, Genotyping and Quality Control Setting**

Genomic DNA was extracted from peripheral blood samples using the LF-kit (Shanghai Lifefeng Biotechnology Co., China) according the instruction of manufacturer. SNPs were genotyped using TaqMan assays (Applied Biosystems, Foster City, CA, USA). Allelic discrimination was automated using the manufacturer’s software which has been widely described in our reported previously studies.2 Internal positive and negative control samples and following test of Hardy-Weinberg equilibrium were employed to examine and control the genotyping quality. 2.5% random sample were selected to evaluate the genotyping reproducibility. When we collected the samples we require the patients or normal should have at least three generation resident in Shanghai and nearby regions. MALDI-TOF-MS techniques for the genotyping which has been widely used in our previous studies3,4. In this study, we totally genotyped 233 SNPs. For the quality control of genotyping, 2% samples were randomly selected among different techniques for the repeat genotyping and we found the repeat accuracy was 99.98%. We applied Michigan Imputation Server (<https://imputationserver.sph.umich.edu/index.html>) to genotyping imputation and haplotype phasing with Asian samples from Genome Asian Pilot (GAsP) with GRch37/hg19 as the genomic positions and R2>0.6 as the cut-off to selected imputed SNPs of high quality and located in miRNA regions. We did not apply 1000 Genome dataset as the reference for the imputation since GAsP reference panel showed higher imputation accuracy (93%-95%) compared with 1000 Genome Asian panel (<90%). The demographic and clinical characteristics of the whole samples are presented in the **Table S2**.

**Association Analysis, Power Calculations, Epistasis Analysis and Cumulative Risk Analysis**

Associations between each genotype and rheumatoid arthritis were estimated as odds ratio (OR) and 95 % confidence interval (95% CI) using Chi-square test or Fisher test and Bayesian logistic regression (BLR) was applied in association study to adjust gender, age, BMI, smoking (Yes or No) and drinking history (Yes or No) derived from bayesm (version: 3.1.4) package. In the analysis, minor alleles or protective alleles were applied as the reference depending on the specific aims. To explore the efficacy of the study, we performed Monte Carlo simulation for the power calculations under the different models of association study including dominant model, additive model, and recessive model within the frame of Chi-square test and logistic regression test. We also applied different genetic models in the association study including dominant model (A*a* + *aa vs AA*), recessive model (aa vs AA + A*a*) and allelic model (a vs A) in which *a* represent minor alleles. The best model was selected using the Akaike information criteria (AIC). SNP-SNP interactions was analyzed by using traditional point-wise interaction analysis based on SNPassoc5, logistic test6, “fast-epistasis” in Plink7. A *P* - value less than 0.001 was considered statistically significant. Risk prediction model were fitted with cumulatively association with six most significant SNPs (rs1414273, rs4947332, rs9268839, rs9275376, rs7752903 and rs2620381) by Chi-square and logistic regression test adjusted by above mentioned covariates. All the statistical analyses were performed with the R (version: 3.6.1).

# Result

**SNP Selection, Genotyping and Population Structure**

In the stage of SNP selection, human microRNAs were downloaded from miRBase (Release 22.1). We alignment all the dbSNP (dbSNP153, updated on 08/08/2019) into human microRNA genomic regions and 40,602 SNPs were received. In order to select common SNPs (MAF>0.01), we collected the allele frequency for all the SNPs from Gnomad8, Asian100K9 and we require the allele frequency should be higher than 0.01 in both dataset and we also validated with 1000 Genome dataset10 to remove any variants whose MAF<0.01. Meanwhile, only binary alleles were included while tri-alleles SNPs were excluded in this study. Eventually, 243 SNPs were received (**Table S2**). In order to decrease the genotyping cost, we removed SNPs with high LD (R2>0.8) and only keep one of them for the genotyping assay. We also applied 4 positive association control to evaluation the phenotype quality including rs9268839 (HLA-DRB1)11,12, [rs9275376 (HLA-DQA1)](https://ard.bmj.com/content/78/6/773)13-17, [rs4947332 (tag SNP for HLA-DRB1\*04:07)](https://github.com/Shicheng-Guo/rheumatoidarthritis/blob/master/thesis/Yaman_Reena_201706_MSc_thesis.pdf) and rs7752903 (TNFAIP3)11,12 which has been reported in previous GWAS studies. In addition, a panel of five ancestry informative markers were selected to estimate the potential effects of population including rs174583 (11q12.2), rs11745587 (5q31.1), rs7515005 (1q32.3) and rs7740161 (6q23.2) which are derived from previous Han Chinese population study18 to avoid non-necessary population structure between South Han and North Han Chinese since Shanghai region is a highly mixture population from the whole parts of China (**Figure 1A**).

We received well genotyping quality with 94% SNPs showed at least 99.9% genotyping ratio and 7 SNPs were removed before the statically analysis with genotyping ratio lower than 99.0%. Missing ratio between case and control, hardy-weinberg disequilibrium in control samples was both evaluated and another 3 SNPs were removed when P<0.01 with chi-square test to excluded biased SNPs. Finally, 223 SNPs in 1,607 rheumatoid arthritis 1,580 normal individuals were received. PCA analysis based on these SNPs showed our samples well clustered with East Asian population (**Figure 1B**). Furthermore, we did not identify age, BMI, drinking history and difference between case and control (**Table S3**). However, all the variants including age, gender, BMI, drinking and smoking were adjusted as the confounding variants in the Bayesian logistic regression (BLR) based association study.

In the power analysis, given the prevalence of RA (p=0.01), case (n=1,607) and control (n=1,580) size, risk allele frequency (f=0.1) and genetic effect size (OR=1.5), we found the power for multiplicative model, additive model, dominant model were 0.856, 0.823 and 0.742 respectively when significance level reach 5x10-3. In addition, we can identify the association with P=0.05 with power of 0.80 to risk alleles frequency large then 2.5%. Therefore, our study could provide reliable genetic association estimation to majority of the alleles we have included since the allele frequency of 201 (86.3%) SNPs are large than 10% and allele frequency of 96% risk SNPs are large then 2.5% (**Table S4**).

**Association study identify novel RA associated SNPs located in miRNA-548 and miRNA-627**

Applied with Bayesian logistic regression adjusted with main covariates, we identified 6 significant SNPs located in HLA-DRB1 (rs9268839, P= 3.95x10-27and rs4947332, P=2.78x10-4), HLA-DQA1 (rs9275376, P=2.65x10-20), TNFAIP3 (rs7752903, P=2.33x10-4), miR-548ac (rs1414273, P=8.26x10-4) and miR-627(rs2620381, P=2.55x10-3) with P<2.55x10-3 and 15 SNPs with P<0.05 (**Table 1**). As expected all the SNPs located in HLA-DRB1, HLA-DQA1 and TNFAIP3 were quite significant associated with RA which is consistent with previous GWAS study. Meanwhile, we found two SNPs located in miR-548ac (P=0.01, FDR) and miR-627 (P=0.045, FDR) were significantly associated with RA even after FDR adjust.

**Epistasis analysis to identify miRNA interaction in the susceptibility of rheumatoid arthritis**

The four SNPs (rs944289T, rs965513, rs966423 and rs2439302) with significant associations with PTC risk after Bonferroni correction and FDR correction were tested for SNP-SNP interaction. No significant SNP-SNP interactions were found for four SNPs by using SNPassoc5, logistic test6, “fast-epistasis” in Plink7 and TIH in GenomeMatrix19 with four independent statistic models. While the additive effects on PTC risk of multiple risk alleles were significant (Table 3). When the four SNPs were analyzed by two-SNP pairs, the estimated odds-ratios (ORs) of risk alleles were increased when adding another risk allele to form the genotype combinations (Table 4). For example, when CC(rs944289)-GG(rs965513) genotype as the reference, the CT+TT(rs944289)-GA+AA(rs965513) genotype has the highest OR values compared with CT+TT(rs944289)-GG(rs965513) genotype or CC(rs944289)-GA+AA(rs965513) genotype. The increased OR values were caused by additive risk alleles.

**Cumulative analysis revealed increased risk effect on rheumatoid arthritis**

Since the additive effects of two-SNP exist, we estimated the combined effects of the four SNPs on PTC risk. We summed the individual accumulation of risk alleles as a variate for logistic regression analysis. As shown at table 4, the PTC risk is increasing with increased accumulative number of risk alleles after adjusted by age and gender (P = 5.929e-13, Table 4). The OR of PTC for six risk allele carriers (1.4% of Chinese population) was 23.587 compared with individuals who are non-risk homozygous (1.0% of Chinese population, with 0 risk allele). There were no individuals who are homozygous at the four SNPs (8 allele’s carrier) and only three PTC cases were 7 risk alleles carrier (0.4% of PTC cases).

# Discussion

We also introduced three East-Asian GWAS RA-associated variants including HLA-DRB1 (rs6931277), HLA-DQA1 (rs9275376) and CCR6 (rs1854853) to make sure the RA samples are well defined. We also required all the enrolled individuals were from three-generation Han Chinese family without genetic admixture with other non-Han ethnic individuals to avoid population structure difference between RA and normal individual.

**Linkage Disequilibrium (LD) Calculation and Haplotype association**

Linkage Disequilibrium was calculated both with East Asian dataset in 1000 Genome Project and with our own data to show the association of the alleles. In the association section, to show the background LD information and association status between RA and genetic variants, we applied 1000 Genome East Asian dataset in the LD calculation while in the haploview analysis section, we applied our own data to show the haplotype blocks in our own dataset. According to our analysis, these two dataset showed highly consistence in the haplotype analysis. We found rs2273626 located in seed region of miR-4707 was significantly associated with RA, OR=1.24, P=3.9x10-3 while rs2620381 located in seed region of miR-627 have a significantly associated with OR= 0.75, P=9.3x10-3. According to the miRNA-mRNA binding imputation, rs2273626 and rs2620381 will affect more than 3,041 and 3,496 mRNA-miRNA bindings respectively. For example, minor allele of rs2273626 will cause a novel binding to CLIC4 and cause [decreased IL-8 secretion](http://www.genomernai.org/ExternalLink/stableid/GR00386-A-2)20. In conclusion, we genotyped 69 common SNPs located in miRNA seed regions in a Chinese cohort and found that two miRNA showed significant association with RA. The regulation network between miR-4707 and miR-627 with CD244, CAMTA1 might be an interesting

# Declarations

**Acknowledgements**

We thank all participating subjects for their kind cooperation in this study. On 15 April 2014, the study was reviewed and approved by Guanghua Hospital Research Institute Institutional Review Board (2015-SRA-01, Title: Genetic and Epigenetic Research to Rheumatoid Arthritis). All methods were performed in accordance with the relevant guidelines and regulations recorded in IRB approved research proposal. All patients were random enrolled and fulfilled the 2010 American College of Rheumatology classification criteria for RA.

# Authors’ contributions

SG and DH contributed to the conception, design and final approval of the submitted version. YJ developed consent form, recruited individuals, collected blood samples and clinical information collection. SG, YJ contributed to statistical data analysis. SS, JS contributed to the conception, study design, manuscript review and result explanation. JS and DH performed genotyping. The final manuscript was finally completed by SG, YJ, JS, SS and DH. All authors read and approved the final manuscript.

# Competing interests

No potential conflicts of interest was disclosed for all the authors

# Funding

This work was funded by the National Natural Science Funds of China (81774114), Shanghai clinical base construction of traditional Chinese medicine (ZY3-LCPT-1-1009, ZY-LCPT-1), Shanghai intensive entity construction of integrated traditional and western medicine [rheumatoid arthritis](http://www.baidu.com/link?url=iT4ATL3sR6yg06cPHgVkbWxtLah4KbbGMOTKi-SQ9uMAjpCCVUsmrIFNHQtlWbRrIUgRFAMxUJ7_lQZh6RW0w9FcG5RyLZskoFBvt8TFc7_) (ZXBZ2012-05), Shanghai clinical intensive subject construction of traditional Chinese medicine-traditional Chinese [rheumatology](http://www.baidu.com/link?url=4Qku9Zb2wBIua7s1DyxfDyF0XZzIFL0m6nyz4LW66UiC3TyBDc31URXd5h_2afOtUNo74DSjURDzbTz6pNjZfkyaRMFKEfnfwCYRjU9oLDu&wd=&eqid=e102a46600012e760000000355827dac) (ZYXK2012012), Shanghai Municipal Planning Commission of science and Research Fund (201640192).

# Availability of data and material

Summary information are access public and provided in supplementary table for further meta-analysis. In addition, the summary statistic has been deposit in the Guthub page xxx : Raw data and materials are available upon the request to the corresponding authors on reasonable request.

URL:

dbSNP153: <https://ftp.ncbi.nlm.nih.gov/snp/latest_release/VCF/>

miRBase: <ftp://mirbase.org/pub/mirbase/CURRENT/README>

# Reference

1. Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 2012;337:1190-5.

2. Jiang XY, Zhang Q, Chen P, et al. CYP7A1 polymorphism influences the LDL cholesterol-lowering response to atorvastatin. J Clin Pharm Ther.

3. Huang L, Li Y, Guo S, et al. Different hereditary contribution of the CFH gene between polypoidal choroidal vasculopathy and age-related macular degeneration in Chinese Han people. Invest Ophthalmol Vis Sci 2014;55:2534-8.

4. Shen F, Chen J, Guo S, et al. Genetic variants in miR-196a2 and miR-499 are associated with susceptibility to esophageal squamous cell carcinoma in Chinese Han population. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine 2016;37:4777-84.

5. Landa I, Ruiz-Llorente S, Montero-Conde C, et al. The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. PLoS Genet 2009;5:e1000637.

6. Boulesteix AL, Strobl C, Weidinger S, Wichmann HE, Wagenpfeil S. Multiple testing for SNP-SNP interactions. Statistical applications in genetics and molecular biology 2007;6:Article37.

7. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics 2007;81:559-75.

8. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285-91.

9. GenomeAsia KC. The GenomeAsia 100K Project enables genetic discoveries across Asia. Nature 2019;576:106-11.

10. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature 2015;526:68-74.

11. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014;506:376-81.

12. Laufer VA, Tiwari HK, Reynolds RJ, et al. Genetic influences on susceptibility to rheumatoid arthritis in African-Americans. Hum Mol Genet 2019;28:858-74.

13. Regueiro C, Gonzalez A. Questions on 'Sequencing of the MHC region defines HLA-DQA1 as the major genetic risk for seropositive rheumatoid arthritis in Han Chinese population' by Guo et al. Ann Rheum Dis 2020.

14. Guo J, Zhang T, Cao H, et al. Sequencing of the MHC region defines HLA-DQA1 as the major genetic risk for seropositive rheumatoid arthritis in Han Chinese population. Ann Rheum Dis 2019;78:773-80.

15. Song X, Guo S, Chen Y, et al. Association between HLA-DQA1 gene copy number polymorphisms and susceptibility to rheumatoid arthritis in Chinese Han population. J Genet 2014;93:215-8.

16. Castro F, Acevedo E, Ciusani E, Angulo JA, Wollheim FA, Sandberg-Wollheim M. Tumour necrosis factor microsatellites and HLA-DRB1\*, HLA-DQA1\*, and HLA-DQB1\* alleles in Peruvian patients with rheumatoid arthritis. Ann Rheum Dis 2001;60:791-5.

17. van Kerckhove C, Luyrink L, Taylor J, et al. HLA-DQA1\*0101 haplotypes and disease outcome in early onset pauciarticular juvenile rheumatoid arthritis. J Rheumatol 1991;18:874-9.

18. Qin P, Li Z, Jin W, et al. A panel of ancestry informative markers to estimate and correct potential effects of population stratification in Han Chinese. Eur J Hum Genet 2014;22:248-53.

19. Wu X, Dong H, Luo L, et al. A novel statistic for genome-wide interaction analysis. PLoS genetics 2010;6.

20. Warner N, Burberry A, Pliakas M, McDonald C, Nunez G. A genome-wide small interfering RNA (siRNA) screen reveals nuclear factor-kappaB (NF-kappaB)-independent regulators of NOD2-induced interleukin-8 (IL-8) secretion. J Biol Chem 2014;289:28213-24.

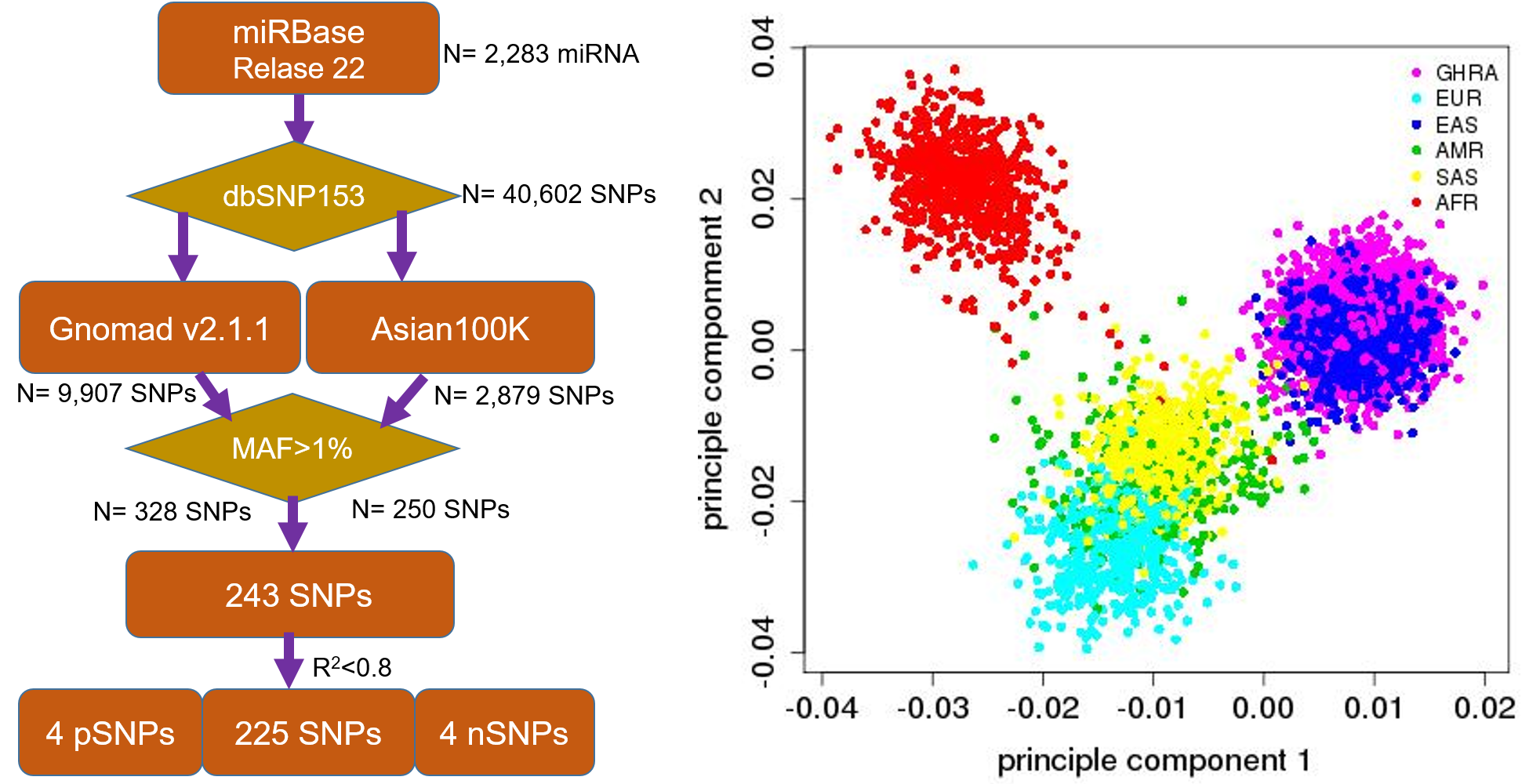


Figure 1.

Figure 2: This figure shows miR-196a2 structure and the location of rs11614913 SNP.Red regions represent mature miRNA fragments. Primary miRNA energy is 51.2 kcal/mol, while SNP-miRNA energy is 46.6 kcal/mol and ΔΔG is 4.6 kcal/mol (miRNASNP database http://bioinfo.life.hust.edu.cn/miRNASNP2/index.php).

Table 1. Summary of risk alleles, 1000 Genome frequencies and SNPs associated with rheumatoid arthritis

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CHR | SNP | A1/A2 | RA (no.) | Normal (no.) | OR | P | EAS | EUR | AFR |
| 6 | rs9268839 | A/G | 42.4%(1364) | 56.0%(1768) | 0.58(0.53-0.64) | 3.95x10-27 | 0.354 | 0.451 | 0.236 |
| 6 | rs9275376 | T/G | 25.0%(804) | 15.7%(496) | 1.79(1.58-2.03) | 2.65x10-20 | 0.192 | 0.299 | 0.285 |
| 6 | rs7752903 | G/T | 6.4%(205) | 4.3%(136) | 1.51(1.21-1.89) | 2.33x10-4 | 0.049 | 0.019 | 0.056 |
| 6 | rs4947332 | T/C | 2.2%(72) | 1.1%(34) | 2.11(1.4-3.18) | 2.78x10-4 | 0.012 | 0.029 | 0.094 |
| 1 | rs1414273 | C/T | 40.5%(1303) | 44.7%(1412) | 0.84(0.76-0.93) | 8.26x10-4 | 0.585 | 0.140 | 0.584 |
| 15 | rs2620381 | C/A | 7.9%(254) | 10.1%(318) | 0.77(0.65-0.91) | 2.55x10-3 | 0.086 | 0.004 | 0.130 |
| 14 | rs75330474 | T/C | 3.8%(121) | 5.0%(159) | 0.74(0.58-0.94) | 0.0136 | 0.065 | 0.003 | 0.064 |
| 18 | rs370878033 | G/A | 1.7%(54) | 2.6%(81) | 0.65(0.46-0.92) | 0.01435 | 0.032 | 0.001 | 0.001 |
| 7 | rs3823658 | A/G | 13.4%(432) | 11.5%(363) | 1.20(1.03-1.39) | 0.01825 | 0.138 | 0.138 | 0.006 |
| 1 | rs74085143 | A/G | 4.0%(130) | 5.3%(166) | 0.76(0.6-0.96) | 0.0219 | 0.048 | 0.022 | 0.331 |
| 6 | rs7740161 | T/A | 14.3%(459) | 16.3%(515) | 0.86(0.75-0.98) | 0.0253 | 0.102 | 0.347 | 0.614 |
| 3 | rs4687672 | A/G | 33.7%(1082) | 31.1%(983) | 1.12(1.01-1.25) | 0.02914 | 0.317 | 0.264 | 0.202 |
| 22 | rs5997893 | A/G | 46.9%(1509) | 49.5%(1565) | 0.90(0.82-1.00) | 0.03972 | 0.498 | 0.667 | 0.918 |
| 19 | rs2967897 | T/C | 16.8%(541) | 15.0%(473) | 1.15(1.00-1.31) | 0.04189 | 0.849 | 0.686 | 0.614 |
| 15 | rs76468441 | T/C | 3.2%(104) | 2.4%(76) | 1.36(1.00-1.83) | 0.0453 | 0.042 | 0.054 | 0.011 |

Table 2. meta-analysis between our study and previous Asian population based RA GWAS study

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CHR | SNP | BP | A1 | OR(meta) | P(meta) | FDR |
| 6 | rs9268839 (HLA-DRB1) | 32428772 | A | 0.52 | 4.15E-49 | 2.08E-47 |
| 6 | rs9275376 (HLA-DQB1) | 32668633 | T | 1.72 | 1.11E-47 | 1.04E-45 |
| 6 | rs7752903 (TNFAIP3) | 138227364 | T | 0.73 | 1.94E-22 | 1.21E-20 |
| 6 | rs4285314 (miR3135B) | 32717702 | A | 1.33 | 2.35E-15 | 1.10E-13 |
| 14 | rs28477407 (miR-4308) | 55344901 | T | 0.89 | 9.19E-07 | 3.44E-05 |
| 22 | rs5997893 (miR-3928) | 31556103 | A | 0.93 | 0.0001915 | 0.005968417 |
| 10 | rs45596840 (miR-4482) | 106028154 | A | 0.89 | 0.0002492 | 0.0066572 |
| 6 | rs4947332 (HLA-DRB1) | 31834197 | T | 1.41 | 0.0009501 | 0.022208588 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 3. Association statistics for rheumatoid arthritis susceptibility and genetic variants | | | | |
| Genotypes, risk allele | RA | Normal | OR(95% CI) | P-value |
| rs9268839, located at 6p21.32 | | | | |
| AA | 262(16.30%) | 494(31.27%) | 1 (reference) | 1.46x10-27 |
| GA | 840(52.27%) | 780(49.37%) | 2.03(1.69-2.44) |  |
| GG | 505(31.43%) | 306(19.37%) | 3.11(2.52-3.85) |  |
| GG+GA vs AA | 505(31.43%) | 306(19.37%) | 1.91(1.62-2.25) | 4.77x10-15 |
| GG vs GA+AA | 1345(83.70%) | 1086(68.73%) | 2.34(1.96-2.78) | 2.01x10-23 |
| G allele | 1850(57.56%) | 1392(44.05%) | 1.72(1.56-1.9) | 3.86x10-27 |
| rs9275376, located at 6p21.32 | | | | |
| GG | 887(55.20%) | 1128(71.39%) | 1 (reference) |  |
| TG | 636(39.58%) | 408(25.82%) | 1.98(1.7-2.32) |  |
| TT | 84(5.23%) | 44(2.78%) | 2.43(1.65-3.62) |  |
| TT+TG vs GG | 84(5.23%) | 44(2.78%) | 1.93(1.31-2.86) | 5.58x10-4 |
| TT vs TG+GG | 720(44.80%) | 452(28.61%) | 2.03(1.74-2.35) | 2.05x10-21 |
| T allele | 804(25.02%) | 496(15.70%) | 1.79(1.58-2.03) | 2.01x10-20 |
| rs7752903, located at 6q23.3 |  |  |  |  |
| TT | 1408(87.62%) | 1445(91.46%) | 1 (reference) |  |
| GT | 193(12.01%) | 134(8.48%) | 1.48(1.16-1.88) |  |
| GG | 6(0.37%) | 1(0.06%) | 6.16(0.75-283.06) |  |
| GG+GT vs TT | 6(0.37%) | 1(0.06%) | 5.92(0.72-272.02) | 0.1246 |
| GG vs GT+TT | 199(12.38%) | 135(8.54%) | 1.51(1.19-1.92) | 4.12x10-4 |
| G allele | 205(6.38%) | 136(4.30%) | 1.51(1.21-1.91) | 2.33x10-4 |
| rs4947332, located at 6p21.33 | | | | |
| CC | 1537(95.64%) | 1546(97.85%) | 1 (reference) |  |
| TC | 68(4.23%) | 34(2.15%) | 2.01(1.31-3.15) |  |
| TT | 2(0.12%) | 0(0.00%) | - |  |
| TT+TC vs CC | 2(0.12%) | 0(0.00%) | - | 0.4998 |
| TT vs TC+CC | 70(4.36%) | 34(2.15%) | 2.07(1.35-3.24) | 4.53x10-4 |
| T allele | 72(2.24%) | 34(1.08%) | 2.11(1.38-3.28) | 2.67x10-4 |
| rs1414273, located at 1p13.1 within miR-548ac | | | | |
| CC | 284(17.67%) | 307(19.43%) | 1 (reference) |  |
| TC | 735(45.74%) | 798(50.51%) | 1.00(0.82-1.21) |  |
| TT | 588(36.59%) | 475(30.06%) | 1.34(1.09-1.65) |  |
| TT+TC vs CC | 588(36.59%) | 475(30.06%) | 1.34(1.15-1.56) | 9.35x10-5 |
| TT vs TC+CC | 1323(82.33%) | 1273(80.57%) | 1.12(0.94-1.35) | 0.2023 |
| T allele | 1911(59.46%) | 1748(55.32%) | 1.18(1.07-1.31) | 9.04x10-4 |
| rs2620381, located at 15q15.1 within miR-627 | | | | |
| CC | 9(0.56%) | 20(1.27%) | 1 (reference) |  |
| AC | 236(14.69%) | 278(17.59%) | 1.89(0.8-4.79) |  |
| AA | 1362(84.75%) | 1282(81.14%) | 2.36(1.02-5.91) |  |
| AA+AC vs CC | 1362(84.75%) | 1282(81.14%) | 1.29(1.07-1.56) | 7.21x10-3 |
| AA vs AC+CC | 1598(99.44%) | 1560(98.73%) | 2.28(0.99-5.7) | 0.04035 |
| A allele | 2960(92.10%) | 2842(89.94%) | 1.3(1.09-1.56) | 0.00285 |

Table 3. Epistasis analysis to identify SNP-SNP interaction in rheumatoid arthritis susceptibility

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
| Epistasis | SNP Genotypes | | RA | Control | OR | P |
| rs4947332 & rs9275376 | rs4947332 | rs9275376 |  |  |  |  |
|  | CC | GG | 885 | 1119 | 1 (reference) |  |
|  | CT+TT | GG | 2 | 9 | 0.38(0.1-1.38) | 0.164 |
|  | CC | GT+TT | 652 | 427 | 1.93(1.66-2.24) | 6.59x10-18 |
|  | CT+TT | GT+TT | 68 | 25 | 3.35(2.12-5.31) | 6.70x10-8 |
| rs4947332 & rs5997893 | rs4947332 | rs5997893 |  |  |  |  |
|  | CC | GG | 330 | 392 | 1 (reference) |  |
|  | CT+TT | GG | 9 | 9 | 1.19(0.49-2.89) | 0.821 |
|  | CC | GT+TT | 1207 | 1154 | 1.24(1.05-1.47) | 0.012 |
|  | CT+TT | GT+TT | 61 | 25 | 2.83(1.75-4.58) | 1.36x10-5 |
| rs7752903 & rs2967897 | rs7752903 | rs2967897 |  |  |  |  |
|  | TT | CC | 953 | 1050 | 1 (reference) |  |
|  | TG+GG | CC | 148 | 93 | 1.75(1.33-2.29) | 5.76x10-5 |
|  | TT | CT+TT | 455 | 395 | 1.27(1.08-1.49) | 0.00370 |
|  | TG+GG | CT+TT | 51 | 42 | 1.33(0.88-2.01) | 0.20705 |

Five significant epistasis effect were selected to show the increased or decreased risk effect. rs4947332 located in C2 upstream (); rs9275376 located in HLA-DQB1 upstream (); rs5997893 located in MIR3928; rs7752903 located in downstream of TNFAIP3 and rs2967897 located in MIR5695();

Table 4. Cumulative risk effect of rheumatoid arthritis increased as more risk alleles carried by the patients

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. of risk alleles\* | RA | Normal | OR (95% CI) | P |
| 1 | 3 | 10 | 1.00(Reference) | 1.00 |
| 2 | 44 | 109 | 1.12(0.34-3.72) | 0.999 |
| 3 | 167 | 269 | 1.71(0.54-5.46) | 0.433 |
| 4 | 307 | 409 | 2.07(0.65-6.55) | 0.296 |
| 5 | 311 | 319 | 2.68(0.84-8.51) | 0.110 |
| 6 | 289 | 230 | 3.45(1.09-10.98) | 0.030 |
| 7 | 257 | 127 | 5.54(1.73-17.75) | 0.003 |
| 8 | 142 | 74 | 5.24(1.61-17.03) | 0.004 |
| 9 | 53 | 23 | 6.19(1.79-21.41) | 0.003 |
| 10 | 29 | 8 | 9.17(2.34-35.91) | 0.001 |

\*individuals carrying more than 10 risk alleles were very rare and did not counts in the table. rs1414273 (T), rs4947332 (T), rs9268839 (G), rs9275376 (T), rs7752903 (G) and rs2620381 (A) were included in the analysis. Risk alleles were showed in the parentheses followed the SNPs.

Table 2. Chi-square based genotype analysis

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CHR | SNP | A1 | A2 | TEST | AFF | UNAFF | CHISQ | DF | P |
| 6 | rs9268839 | A | G | GENO | 262/840/505 | 494/780/306 | 122 | 2 | 3.18E-27 |
| 6 | rs9275376 | T | G | GENO | 84/636/887 | 44/408/1128 | 90.9 | 2 | 1.83E-20 |
| 1 | rs1414273 | C | T | GENO | 284/735/588 | 307/798/475 | 15.27 | 2 | 0.000484 |
| 15 | rs2620381 | C | A | GENO | 9/236/1362 | 20/278/1282 | 9.797 | 2 | 0.007458 |
| 22 | rs5997893 | A | G | GENO | 339/831/437 | 401/763/416 | 8.384 | 2 | 0.01511 |
| 6 | rs7740161 | T | A | GENO | 25/409/1173 | 47/421/1112 | 8.296 | 2 | 0.0158 |
| 13 | rs1572687 | T | C | GENO | 102/719/786 | 141/674/765 | 7.769 | 2 | 0.02056 |
| 11 | rs11237828 | C | T | GENO | 190/759/658 | 188/676/716 | 7.031 | 2 | 0.02973 |
| 1 | rs12402181 | A | G | GENO | 229/699/679 | 176/712/692 | 6.951 | 2 | 0.03095 |
| 19 | rs2967897 | T | C | GENO | 35/471/1101 | 36/401/1143 | 6.191 | 2 | 0.04525 |
| 1 | rs877722 | T | A | GENO | 66/509/1032 | 40/511/1029 | 6.157 | 2 | 0.04602 |

Table 2. Chi-square based dominant model analysis

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CHR | SNP | A1 | A2 | TEST | AFF | UNAFF | CHISQ | DF | P |
| 6 | rs9275376 | T | G | DOM | 720/887 | 452/1128 | 89.89 | 1 | 2.52E-21 |
| 6 | rs9268839 | A | G | DOM | 1102/505 | 1274/306 | 61.06 | 1 | 5.55E-15 |
| 1 | rs1414273 | C | T | DOM | 1019/588 | 1105/475 | 15.27 | 1 | 9.34E-05 |
| 15 | rs2620381 | C | A | DOM | 245/1362 | 298/1282 | 7.365 | 1 | 0.006649 |
| 11 | rs11237828 | C | T | DOM | 949/658 | 864/716 | 6.205 | 1 | 0.01274 |
| 19 | rs2967897 | T | C | DOM | 506/1101 | 437/1143 | 5.607 | 1 | 0.01789 |
| 7 | rs3823658 | A | G | DOM | 406/1201 | 347/1233 | 4.815 | 1 | 0.02821 |
| 14 | rs28477407 | T | C | DOM | 1016/591 | 1055/525 | 4.409 | 1 | 0.03575 |
| 3 | rs4687672 | A | G | DOM | 901/706 | 829/751 | 4.158 | 1 | 0.04144 |

Table 3. Epistasis analysis to identify SNP-SNP interaction in rheumatoid arthritis susceptibility

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CHR1 | SNP1 | miRNA-1 | CHR2 | SNP2 | miRNA-2 | OR | STAT | P |
| 1 | rs1414273 | MIR548AC | 7 | rs117344178 | miR-595 | 0.313929 | 19.7655 | 8.76E-06 |
| 8 | rs6997249 | CCDC26 | 14 | rs2273626 | HAUS4 | 0.544402 | 19.759 | 8.79E-06 |
| 8 | rs2114358 | PVT1 | 15 | rs76468441 | MIR548AP | 0.274239 | 19.0585 | 1.27E-05 |
| 6 | rs4285314 | MIR3135B | 16 | rs16958290 | CDH13 | 1.53597 | 16.8634 | 4.02E-05 |
| 5 | rs367805 | LOC553103 | 10 | rs641071 | MIR4482 | 0.716913 | 16.0484 | 6.18E-05 |
| 7 | rs7804972 | ZNF107 | 17 | rs620301 | FBXL20 | 1.47573 | 14.5928 | 0.000134 |
| 14 | rs2296320 | MIR4706 | 22 | rs5997893 | MIR3928 | 1.93891 | 14.5927 | 0.000134 |
| 5 | rs2042253 | MIR5197 | 6 | rs9295535 | MIR5689HG | 1.34491 | 14.093 | 0.000174 |
| 10 | rs7070684 | MIR548AK | 11 | rs2155248 | MIR1304 | 1.83627 | 13.3408 | 0.00026 |
| 1 | rs73108496 | RYR2 | 11 | rs67042258 | MIR6128 | 0.558997 | 12.6479 | 0.000376 |
| 2 | rs2241347 | FASTKD2 | 5 | rs7709117 | MIR4634 | 1.35183 | 12.3538 | 0.00044 |
| 1 | rs877722 | SLC35F3 | 5 | rs62376935 | SLIT3 | 0.700077 | 12.2212 | 0.000473 |
| 8 | rs75404472 | MIR3674 | 16 | rs2925980 | CMIP | 0.634786 | 12.1725 | 0.000485 |
| 9 | rs3780662 | AK1 | 10 | rs11014002 | KIAA1217 | 0.580794 | 12.0226 | 0.000526 |
| 7 | rs4909237 | PTPRN2 | 11 | rs12801172 | EHD1 | 1.3425 | 11.831 | 0.000583 |
| 2 | rs10175383 | MIR3679 | 8 | rs10505168 | CSMD3 | 0.748224 | 11.7398 | 0.000612 |
| 9 | rs56195815 | GFI1B | 10 | rs7911488 | USMG5 | 1.53774 | 11.6583 | 0.000639 |
| 3 | rs6771809 | COPG1 | 14 | rs12894467 | MIR300 | 0.656563 | 11.5558 | 0.000676 |
| 17 | rs58390814 | EZH1 | 19 | rs56061231 | NOTCH3 | 1.9326 | 11.4028 | 0.000734 |
| 3 | rs75715827 | TP63 | 16 | rs57629257 | PDXDC2P | 0.428358 | 11.1356 | 0.000847 |
| 6 | rs68035463 | MIR3144 | 14 | rs12894467 | MIR300 | 1.50915 | 10.9572 | 0.000933 |

miRNA genomic position were obtained from miRBase 22 which is based on hg38 and then was liftover to hg19 for the further analysis. Genomic position in the table is hg19 version.

Table 1. Chi-sq based allelic analysis

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CHR | SNP | BP | A1 | F\_A | F\_U | A2 | OR | SE | P |
| 6 | rs9268839 | 32428772 | A/G | 42.4%(1,364) | 56.0%(1,768) | G | 0.58(0.53-0.64) | 0.05057 | 3.95E-27 |
| 6 | rs9275376 | 32668633 | T/G | 25.0%(804) | 15.7%(496) | G | 1.79(1.58-2.03) | 0.06364 | 2.65E-20 |
| 6 | rs7752903 | 1.38E+08 | G | 6.4%(205) | 4.3%(136) | T | 1.51(1.21-1.89) | 0.1136 | 0.000233 |
| 6 | rs4947332 | 31834197 | T | 2.2%(72) | 1.1%(34) | C | 2.11(1.4-3.18) | 0.2096 | 0.000279 |
| 1 | rs1414273 | 1.17E+08 | C | 40.5%(1303) | 44.7%(1412) | T | 0.84(0.76-0.93) | 0.05071 | 0.000827 |
| 15 | rs2620381 | 42491848 | C | 7.9%(254) | 10.1%(318) | A | 0.77(0.65-0.91) | 0.08816 | 0.002551 |
| 14 | rs75330474 | 1.02E+08 | T | 3.8%(121) | 5.0%(159) | C | 0.74(0.58-0.94) | 0.1233 | 0.0136 |
| 18 | rs370878033 | 13611175 | G | 1.7%(54) | 2.6%(81) | A | 0.65(0.46-0.92) | 0.1775 | 0.01435 |
| 7 | rs3823658 | 1.02E+08 | A | 13.4%(432) | 11.5%(363) | G | 1.2(1.03-1.39) | 0.07607 | 0.01825 |
| 1 | rs74085143 | 54519800 | A | 4.0%(130) | 5.3%(166) | G | 0.76(0.6-0.96) | 0.1199 | 0.0219 |
| 6 | rs7740161 | 1.34E+08 | T | 14.3%(459) | 16.3%(515) | A | 0.86(0.75-0.98) | 0.06972 | 0.0253 |
| 3 | rs4687672 | 52880543 | A | 33.7%(1082) | 31.1%(983) | G | 1.12(1.01-1.25) | 0.05357 | 0.02914 |
| 22 | rs5997893 | 31556103 | A | 46.9%(1509) | 49.5%(1565) | G | 0.9(0.82-1.00) | 0.05015 | 0.03972 |
| 19 | rs2967897 | 13031210 | T | 16.8%(541) | 15.0%(473) | C | 1.15(1.00-1.31) | 0.06862 | 0.04189 |
| 15 | rs76468441 | 86368922 | T | 3.2%(104) | 2.4%(76) | C | 1.36(1.00-1.83) | 0.153 | 0.0453 |

Genomic position is hg19 based