

MicroRNA variants and HLA-miRNA interactions are novel rheumatoid arthritis susceptibility factors

Shicheng Guo^{1#}, Yehua Jin^{2,3#}, Jieru Zhou^{4#}, Qi Zhu^{3,5#}, Ting Jiang³, Yanqin Bian^{3,5}, Runrun Zhang^{2,3}, Cen Chang^{2,3}, Lingxia Xu^{2,3}, Jie Shen^{3,5}, Xinchun Zheng³, Yi Shen³, Yingying Qin³, Jihong Chen³, Xiaorong Tang³, Peng Cheng³, Qin Ding³, Yuanyuan Zhang³, Jia Liu³, Qingqing Cheng³, Mengru Guo³, Zhaoyi Liu³, Weifang Qiu³, Yi Qian³, Yang Sun^{3,5}, Yu Shen^{3,5}, Hong Nie⁶, Steven J Schrod^{1*}, Dongyi He^{3,5*}

1, Department of Medical Genetics, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, USA. **2**, Shanghai University of Traditional Chinese Medicine, Shanghai, China. **3**, Department of Rheumatology, Shanghai Guanghua Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai, China. **4**, Department of Health Management, Shanghai East Hospital, Tongji University School of Medicine, Shanghai, China. **5**, Arthritis Institute of Integrated Traditional Chinese and Western Medicine, Shanghai Academy of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine. **6**, Shanghai Institute of Immunology, Shanghai Jiao Tong University School of Medicine, Shanghai, China

These authors contributed equally to this work

*Correspondence:

Steven J Schrod, Ph.D.
Department of Medical Genetics
School of Medicine and Public Health
University of Wisconsin-Madison, Madison
Tel: 608-263-3068
Email: Schrod@wisc.edu

Dongyi He, M.D., Ph.D.
Department of Rheumatology
Shanghai Guanghua Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai, 200052, China
Tel : 158-0030-0800
Email: hedongyi1967@shutcm.edu.cn

Number of Words : 2938

Abstract

Objective: Although Genome-wide association studies have identified >100 variants for rheumatoid arthritis (RA), the reported genetic variants only explain <40% of RA heritability. We conducted a systemic association study between common East-Asian miRNA SNPs with RA in a large Han Chinese cohort to explain missing heritability and identify miRNA epistatic interactions.

Methods: 4 HLA SNPs (HLA-DRB1, HLA-DRB9, HLA-DQB1 and TNFAIP3) and 225 common SNPs located in miRNA which might influence the miRNA target binding or pre-miRNA stability were genotyped in 1,607 rheumatoid arthritis and 1,580 matched normal individuals. A meta-analysis with previous GWAS studies (4,873 RA cases and 17,642 controls) was performed to discovery another novel miRNA RA-associated SNPs.

Results: 2 novel SNPs including rs1414273 (miR-548ac, OR=0.84, P=8.26x10⁻⁴) and rs2620381 (miR-627, OR=0.77, P=2.55x10⁻³) conferred significant association with RA. Individuals carried 8 risk alleles showed 15.38 (95%CI: 4.69-50.49, P<1.0x10⁻⁶) times more risk to be affected by RA. In addition, rs5997893 (miR-3928) showed significant epistasis effect with rs4947332 (HLA-DRB1, OR=4.23, P=0.04) and rs2967897 (miR-5695) with rs7752903 (TNFAIP3, OR=4.43, P=0.03). Finally, we demonstrated targets of the significant miRNAs showed enrichment in immune related genes (P=2.0x10⁻⁵) and FDA approved drug target genes (P=0.014).

Conclusions: 6 novel miRNA SNPs including rs1414273 (miR-548ac, P=8.26x10⁻⁴), rs2620381 (miR-627, P=2.55x10⁻³), rs4285314 (miR-3135b, P=1.10x10⁻¹³), rs28477407 (miR-4308, P=3.44x10⁻⁵), rs5997893 (miR-3928, P=5.9x10⁻³) and rs45596840 (miR-4482, P=6.6x10⁻³) were confirmed to be significantly associated with RA in a Chinese population. Our study suggests that miRNAs might be interesting targets to accelerate the understanding of the pathogenesis and drug development for rheumatoid arthritis.

Keywords: Rheumatoid Arthritis, miRNA, SNP, susceptibility, predispose

Introduction

Rheumatoid arthritis (RA) is a chronic inflammatory disorder caused by the interaction between multiple factors including genetics, epigenetics and the environment.¹⁻³ Twin studies estimate that an RA heritability of ~60%.^{4,5} In the past 15 years, GWAS have identified several hundred RA-associated variants. However, the reported genetic variants only explain <40% of RA heritability.⁵⁻⁸ Furthermore, studies within non-European-derived samples will aid our understanding of RA molecular pathophysiology as different variants and, potentially, different disrupted pathways may be present in those sample sets. Additionally, studying SNPs in specific functional classes of genes or motifs may reveal specific pathogenic mechanisms not previously discovered through GWAS.⁹ GWAS results indicate that 90% of disease-associated variants are located in non-coding regions, indicating regulatory elements may play important roles in complex disease etiology, including RA.¹⁰ Across regulatory elements, microRNAs are important targets to interrogate as miRNA SNPs may modify expression at numerous genes. Moreover, miRNAs play pivotal roles in both innate and adaptive immunity,^{11,12} sex-specific effects,¹³ and disease onset.^{14,15}

Recently, several preliminary RA-association studies focused on miRNAs have been conducted in Asian,¹⁶⁻¹⁹ European²⁰ and African²¹ samples. However, these studies have very limited sample sizes and only included a subset of miRNAs without genotyping all miRNA common SNPs. Additionally, the previous RA GWAS in Han Chinese was conducted in 2014²² using an array with poor coverage of miRNA SNPs. Therefore, we conducted an exhaustive study of functional miRNA SNPs in RA. miRbase annotates 1,920 primary miRNA and 2,883 mature miRNAs. From dbSNP153, there are >45,705 SNPs located in primary miRNAs and 17,570 in mature miRNA,²³ however, only 733 of these SNPs segregate alleles at moderate-high frequency (i.e. are common SNPs with MAF>1%) within primary miRNAs regions. Within specific populations, this number is expected to be substantially reduced to ~200 SNPs, providing an opportunity to use a multiplex genotyping assay in our RA samples. Hence, we conducted a systemic association study between common East-Asian

miRNA SNPs with RA in a large Han Chinese cohort to identify novel miRNA SNPs and miRNA epistatic interactions.

Methods

Precision Medicine Research Cohort in Shanghai Guanghua Hospital

Guanghua Hospital Precision Medicine Research Cohort (PMRC) is a hospital-based longitudinal cohort to investigate risk factors, genetic susceptibility, pharmacogenetics for rheumatology diseases such as RA, osteoarthritis and ankylosing spondylitis. Healthy individuals are derived from those with an annual physical exam without rheumatological disease. Currently, PMRC has enrolled >30,000 disease patients and 10,000 healthy individuals as controls. We randomly selected 3,223 individuals including 1,625 seropositive (RF+ and anti-CCP+) RA and 1,598 controls from the PMRC. Written consent was collected prior to enrollment. Non-Han Chinese individuals were excluded from the study to avoid confounding by population stratification. 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 ankylosing spondylitis, rheumatoid arthritis, osteoarthritis, type 1 and 2 diabetes, chronic infection or common cancers. We did not find significant case/control difference between gender, age, smoking, BMI after correction for multiple testing. 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. The demographic and clinical characteristics of the whole samples are presented in **Table S1**.

DNA extraction, SNP selection and Genotyping

Genomic DNA was extracted from peripheral blood using the CB-kit (CoWin Biosciences, CWBIO, China) according following manufacturer's instructions. Allelic discrimination was automated using the manufacturer's software which has been widely-described in previous studies.²⁴ Internal positive and negative control samples were used and a test of Hardy-Weinberg equilibrium was employed to assess the genotyping quality (SNPs exceeding HWE p-value cutoff were excluded from RA association testing). Cases and controls were required to have a family history of at least three generations of residency in Shanghai or neighboring regions. The SNaPshot Multiplex System (Applied Biosystem, USA) was used for the genotyping which has been widely used in our previous studies.^{25 26}

To select SNPs for subsequent association testing, human microRNAs were downloaded from miRBase methjods (Release 22.1). Mapping of SNPs within dbSNP (dbSNP153, 08/08/2019) was performed for human microRNA genomic regions, resulting in 40,602 SNPs. To select common SNPs (MAF>0.01), we collected the allele frequency for all the SNPs from Gnomad,²⁷ Asian100K²⁸ and the 1000 Genomes dataset.¹⁰ The minor allele frequency was required to be higher than 0.01 in all datasets. [t]²⁹ Only biallelic SNPs were included in the analysis, while tri-allelic SNPs were excluded from this study. Following filtering, 243 SNPs were obtained. In order to decrease the genotyping cost, one SNP from pairs of SNPs with high LD ($R^2 > 0.8$) were removed and only one of them was included in the genotyping assay. A total of 233 SNPs were genotyped. GRCH37/hg19 was used to determine genomic positions and $R^2 > 0.6$ as the cut-off to selected imputed SNPs of high quality and located in miRNA regions. The GASP reference panel showed higher imputation accuracy (93%-95%) compared with 1000 Genome Asian panel (<90%).³⁰ Four SNPs within known GWAS associations with RA were included: rs9268839 (*HLA-DRB9*),^{31 32} rs9275376 (*HLA-DQB1*),³³⁻³⁶ rs4947332 (tag-SNP for *HLA-DRB1*04:07*) and rs7752903 (*TNFAIP3*).^{31 32} In addition, a panel of five ancestry informative markers (negative control) were selected to estimate the potential effects of population stratification: rs174583 (11q12.2), rs11745587 (5q31.1), rs521188 (1p31.3) and rs7740161 (6q23.2). These SNPs were derived from a previous Han Chinese population study.³⁷ These SNPs are informative for interrogating population structure between South and North Han Chinese since the Shanghai region is a mixed population from different regions in China (**Figure 1A**).

Quality Control and Power analysis

For genotyping QC, ~2.5% samples (40 RA and 32 normal) were randomly selected for repeat genotyping. The concordance between initial and replicate genotyping was 99.98%. SNP imputation was accomplished through the Michigan Imputation Server (<https://imputationserver.sph.umich.edu/index.html>) with East Asian samples from the Genome Asian Pilot (GAsP). Missing ratio between case and control differential test and HWE in control samples were evaluated. SNPs were removed when $P < 0.01$.

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 ($OR=1.5$), the power for multiplicative model, additive model, dominant models were 0.856, 0.823 and 0.742, respectively using a significance level of $\alpha=5 \times 10^{-03}$. In addition, with our design and the above effect size, we calculate that SNPs with MAF of 0.025 can attain a statistical power of 80% with nominal $p < 0.05$. Therefore, our study could provide reliable genetic association results for the majority of the alleles we have included given that the MAF of 201 (86.3%) SNPs are $>10\%$ and MAF of 96% included SNPs exceed 2.5%.

Bioinformatics analysis of miRNA regulatory networks and enrichment analysis

To understand regulatory networks driven by miRNAs and the overlap with RA GWAS genes and immune system genes, a miRNA regulatory network analysis was performed with miRDB.³⁸ Genes predicted to be regulated by significant miRNAs were calculated among 123 previously-identified RA GWAS genes, 4,723 immune system genes collected by InnateDB database³⁹ and 373 FDA-targeted pharmacogenes. A target score threshold was set to 95 in miRDB to identify reliable gene targets for the miRNAs. Cytoscape was employed to construct and visualize the miRNA-mRNA network. We calculated a hypergeometric test of the enrichment in miRNA targets within RA GWAS significant genes, immune-related genes, and FDA approved drug targets. A random sampling technique was applied to assess the null distribution of p-values ($n=50,000$ iterations). 37,875 total genes in the genome (GENCODE v32) were assumed for the calculations. Enrichment was measured with fold-change (FC) and p-values were calculated by summing the frequency of more extreme values in the null distribution of p-values.

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

RA Association for each SNP were calculated and OR estimated (95% CI) using Chi-square test, Fisher's Exact test ($N < 5$), Cochran-Armitage Trend test and Bayesian logistic regression (BLR)⁴⁰ adjusted for gender, age, BMI, smoking and drinking history, using the bayesm (version: 3.1.4) package.⁴⁰ To explore the power of the study, a Monte Carlo simulation was performed under different models including dominant, recessive and additive models, additive models, using a Chi-square test or Bayesian logistic regression model. SNP-SNP interactions were analyzed by using traditional point-wise interaction analysis based on SNPAssoc,⁴¹ logistic test,⁴² "fast-epistasis" in Plink.⁴³ We applied $Y \sim b_0 + b_1 * A + b_2 * B + b_3 * AB + e$ to detect epistasis effect between SNP A and SNP B when b_3 not equal 0 (null hypothesis: b_3 equal 0). A $P < 0.001$ was considered significant. An RA risk prediction model was fitted with the 6 most significant SNPs (rs1414273, rs4947332, rs9268839, rs9275376, rs7752903 and rs2620381) by both Chi-square and logistic regression test to adjust by above-mentioned covariates. Q-value received after multiple test correction was conducted through a False Discover Rate (FDR) implemented in R with the function `p.adjust`.⁴⁴ SNP derived miRNA target gain or loss were imputed by miRNA-SNP database⁴⁵ with the default setting. All the statistical analyses were performed with R (version: 3.6.1).

Results

SNP Selection, Genotyping and Population Structure

Overall, the genotyping generated high quality data with 94% of the SNPs showing high genotyping quality. Seven of the SNPs were removed from being promoted to the RA association evaluation due to a genotyping ratio less than 99.0%. The missing ratio between case and control was calculated. Hardy-Weinberg equilibrium in control samples was evaluated and another 3 SNPs were removed when $P < 0.01$. Finally, 223 SNPs in 1,607 rheumatoid arthritis and 1,580 normal individuals were used in the association analysis (genotyping rate=98.88%). PCA analysis based on these SNPs showed our samples clustered with the East Asian population (**Figure 1B**). Furthermore, we did not identify significant age, BMI, drinking and smoking history

differences between cases and controls ($P > 0.01$, **Table S1**). However, as a conservative measure, association testing of all the variants included age, gender, BMI, drinking and smoking as covariates in the Bayesian logistic regression (BLR)-based tests.

Novel RA-associated SNPs located in *miRNA-548* and *miRNA-627*

Applying the Bayesian logistic regression model adjusted for covariates, we identified 6 significant SNPs located in *HLA-DRB9* (rs9268839, $P = 3.95 \times 10^{-27}$), *HLA-DRB1* (rs4947332, $P = 2.78 \times 10^{-4}$), *HLA-DQB1* (rs9275376, $P = 2.65 \times 10^{-20}$), *TNFAIP3* (rs7752903, $P = 2.33 \times 10^{-4}$), *miR-548ac* (rs1414273, $P = 8.26 \times 10^{-4}$) and *miR-627* (rs2620381, $P = 2.55 \times 10^{-3}$) and 15 additional SNPs with $P < 0.05$ (**Table 1**). As expected, all the SNPs located in *HLA-DRB1*, *HLA-DRB9*, *HLA-DQB1* and *TNFAIP3* were significantly associated with RA status, which is consistent with previous GWAS studies. We found all the population markers including rs174583 (FDR=0.88), rs11745587 (FDR=0.98), rs521188 (FDR=0.99) and rs7740161 (FDR=0.48) were not significant in the association test between RA and control. Additionally, two miRNA SNPs located in *miR-548ac* (FDR=0.01) and *miR-627* (FDR=0.045) were significantly associated with RA after FDR adjustment for multiple testing ($P < 0.05$, FDR). We also conducted mode of inheritance-based association analyses (**Table 2**) and Cochran-Armitage Trend test-based association analyses (**Table S2**). We found a dominant model of two miRNA SNPs (rs1414273 and rs2620381) showed highly significant association between the alleles and RA status. To combine previous Han Chinese association data with this study, we also implemented a meta-analysis in which 4,873 RA cases and 17,642 normal individuals from the East Asian Population (EAP) were included. The meta-analysis resulted in an additional 4 significant SNPs (**Table 3, Figure 2**) including rs4285314 (*miR-3135b*, FDR= 1.10×10^{-13}), rs28477407 (*miR-4308*, FDR= 3.44×10^{-5}), rs5997893 (*miR-3928*, FDR= 5.9×10^{-3}) and rs45596840 (*miR-4482*, FDR= 6.6×10^{-3}). Finally, we also evaluated whether rs1414273 (*miR-548ac*) and rs2620381 (*miR-627*) are significant RA susceptibility variants independent of HLA variants. We found that the association at rs1414273 (OR=1.18, $P = 1.03 \times 10^{-3}$) and rs2620381 (OR=1.31, $P = 2.55 \times 10^{-3}$) remained significant following adjustment for HLA alleles.

Epistasis analysis to identify miRNA interactions

miRNAs play multiple regulatory roles in both modification of target gene expression and larger regulatory networks. In order to identify epistatic effects between miRNAs and their role in the susceptibility of RA, we applied an epistasis analysis to reveal the interaction between above identified miRNA SNPs. We found 19 SNP-SNP epistatic interactions with $P < 7.3 \times 10^{-4}$, indicating significant interactions (**Table S3**). We found 10 SNP-SNP pairs showed significant strengthened interaction (OR>1) while nine SNP-SNP pairs showed impaired interaction (OR<1). In addition, enhanced risk epistasis was also identified across HLA alleles (rs4947332 in *HLA-DRB1* and rs9275376 in *HLA-DQB1*) between HLA and miRNAs (*HLA-DRB1* and *miR-3928*), and between *TNFAIP3* and *miR-5695* (**Table 4**). A significant positive interaction between rs4947332 (*HLA-DRB1*) and rs5997893 (*miR-3928*) with a significantly inflated OR=2.83 (95%CI: 1.75-4.58, $P = 1.36 \times 10^{-5}$, **Table 4**) for double risk allele carriers, indicating the importance of HLA and non-HLA genetic variation interaction in RA susceptibility.

Cumulative analysis revealed increased risk effect on rheumatoid arthritis

To estimate the combined effect of the 6 risk alleles (rs1414273T, rs4947332T, rs9268839G, rs9275376T, rs7752903G and rs2620381A) on RA risk, the individual accumulation of risk alleles was treated as an ordinal variable in a logistic regression analysis adjusted by BMI, age and gender. As anticipated, we found the RA risk showed a positive correlation with cumulative number of risk alleles (OR=1.4, $P = 2.0 \times 10^{-16}$, $Z = 12.54$, SE=0.027, **Table 5**). In our dataset, the largest subgroup of cases carries five risk alleles (25.6% of RA samples) while the largest subgroup of controls carries only 4 risk alleles (29.4% of control samples). The OR for RA status for carriers with eight risk alleles (2.9% of RA population) was 15.38-fold increased over individuals with only 1 risk alleles (10.76% of normal population).

Regulatory network constructed by significant miRNAs and RA GWAS associated genes

To construct regulatory networks between miRNAs and genes, we predicted target genes for the significant miRNAs identified in our study. First, we predicted candidate genes of our significant miRNAs and then mapping these targets to overall immune related genes collected by InnateDB. We found target genes of RA-associated miRNAs were significantly enriched in the immune related gene category ($P < 2.2 \times 10^{-16}$, empirical $P = 2.0 \times 10^{-5}$ and $FC = 1.43$, **Figure 3A**, **Figure S1A**). In addition, we found *miR-548ac* has 12 target genes overlapping with previous GWAS-identified RA-associated genes including *MED1*, *IL6R*, *CEP57*, *CDK6*, *RAD51B*, *RUNX1*, *RTKN2*, *RAG1*, *FADS1*, *RASGRP1*, *ETS1* and *COG6*. *miR-4308* showed seven candidate genes including *RTKN2*, *TRAF6*, *RAD51B*, *PTPN2*, *PLD4*, *TNFRSF14* and *SYNGR1*. *miR-3135b* showed 4 target genes including *GRHL2*, *CD28*, *PPIL4* and *RASGRP1*. Although a hypergeometric test showed miRNA targets significantly enriched in GWAS-identified RA candidate genes ($P = 7.66 \times 10^{-3}$, $FC = 1.59$, **Figure 3B**, **Table S4**), permutation based analysis showed a non-significant enrichment (empirical $P = 0.23$) indicating an unstable enrichment (**Figure S1B**). Finally, we also observed miRNA targets also enriched in FDA approved drug targets ($P < 2.2 \times 10^{-16}$, empirical $P = 0.014$, $FC = 1.82$) indicating miRNA as the key regulatory network hubs might be promising drug targets for autoimmune diseases (**Figure 3C**, **Figure S1C**).

Discussion

In this study, 225 East-Asian common SNPs located in human microRNA seed regions were genotyped and association and epistasis analyses were conducted to investigate the association between miRNA and seropositive RA in the Han Chinese population (Shanghai, China). To evaluate the potential association of these East-Asian-specific common miRNA SNPs with RA susceptibility, a case/control study was conducted involving 1,607 seropositive RA patients compared to 1,580 matched controls. The study identified 6 novel RA-associated miRNA SNPs (rs1414273 in *miR-548ac*, rs2620381 in *miR-627*, rs4285314 in *miR-3135b*, rs28477407 in *miR-4308*; rs5997893 in *miR-3928* and rs45596840 in *miR-4482*) and revealed the interaction between HLA alleles and miRNA SNPs, thereby advancing our understanding of RA pathogenesis.

We found rs45596840 and rs2620381 located in seed regions of *miR-4482* and *miR-627*, respectively, are significantly associated with RA. rs9268839-G (*HLA-DRB9*) was showed as a risk allele with $OR = 2.47$ in Caucasians population and we found the OR is 1.72 indicating the consistent risk effect and might be required to be validated in African population. According to the miRNA-mRNA binding imputation, rs45596840 (*miR-4482*) and rs2620381 (*miR-627*) will affect more than 5,894 (2,132 gain and 3,517 loss, see URL) and 4,845 (2,593 gain and 2,252 loss, see URL) mRNA-miRNA bindings respectively. These miRNA target change might change the immune response significantly. In addition, the common regulation network between *miR-4707* and *miR-627* with *CD244*, *CAMTA1* might be an interesting. rs1414273 is located within the *miR-548ac* stem-loop sequencing in the first intron of the *CD58* gene, which is showed a strong linkage disequilibrium with the Multiple Sclerosis (MS)-associated haplotype.⁴⁶ Hecker M et al. found that SNP rs1414273 might alter Drosha cleavage activity to provoke partial uncoupling of *CD58* gene expression and microRNA-548ac production from the shared primary transcript in immune cells and to regulate the inflammatory processes and the balance of protein folding and degradation.⁴⁷ Although *CD58-CD2* interactions recruiting lymphocyte to inflammatory sites to play a crucial role in autoimmune disease.^{48,49} Finally, in our previous research, we found DNA methylation of CD4+ cell in rheumatoid arthritis showed abnormal DNA methylation in HLA regions.⁵⁰ In this current study, we found interactions between HLA-miRNA as contributing to RA susceptibility. Overall, the interaction between regulation/epigenetics and genetic variation in RA is a promising field for future research.

There are limitations with our study. Due to the scope of the study, only a limited number of ancestry-informative SNPs were used to control for confounding by population stratification. To validate the RA phenotyping, we utilized 4 East-Asian GWAS RA-associated variants to serve as positive controls for RA samples. 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 stratification. This study did not provide functional validation of these miRNAs which is important to show biological validation of the miRNA findings and understand the mechanisms by which these miRNAs are involved in RA susceptibility. Subsequent studies should focus on functional studies of these miRNAs. Overall, we identified 6 miRNA SNPs which are significant associated with the Han Chinese RA population. Two of these SNPs are located in seed

region of their miRNAs and are expected to cause target gain/loss, while another three SNPs are located in the loop/pre-miRNA regions which may influence the stability of corresponding miRNAs.

Declarations

Acknowledgements We thank all participating subjects for their kind cooperation in this study. 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.

Contributors SG and DH contributed to the conception, design and final approval of the submitted version. YJ, YB, JZ, QZ, TJ, YB, RZ, LX, CC, JS, XZ, YS, YQ, JC, XT, PC, QD, YZ, JL, QC, MG, ZL, WQ, YQ, YS, YS, HN developed consent form, recruited individuals, collected blood samples and clinical information collection. SG, YJ contributed to statistical data analysis. JS and DH contributed to the conception, study design, manuscript review and result explanation. The final manuscript was finally completed by SG, YJ, SS and DH. All authors read and approved the final manuscript.

Funding This work was funded by the National Natural Science Funds of China (81774114), Shanghai Chinese Medicine Development Office, Shanghai Chinese and Western Medicine Clinical Pilot Project (ZY(2018–2020)-FWTX-1010), Shanghai Chinese Medicine Development Office, Shanghai Traditional Chinese Medicine Specialty Alliance Project (ZY(2018-2020)-FWTX-4017), National Administration of Traditional Chinese Medicine, Regional Chinese Medicine (Specialist) Diagnosis and Treatment Center Construction Project-Rheumatology.

Competing interests No potential conflicts of interest was disclosed for all the authors.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Patient consent for publication Not required.

Ethical approval 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).

Data availability statement Summary information are access public and provided in supplementary table for further meta-analysis. In addition, the summary statistic has been deposited in the GitHub page <https://github.com/Shicheng-Guo/miRNA-RA>. Raw data and materials are available upon the request to the corresponding authors on reasonable collaboration request.

URL:

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

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

miRNASNP: <http://bioinfo.life.hust.edu.cn/miRNASNP2/index.php>

miRDB: <http://www.mirdb.org/ontology.html>

InnateDB, 4,723 immune genes: <https://www.innatedb.com/>

rs2273626-target: http://bioinfo.life.hust.edu.cn/miRNASNP#!/snp?snp_id=rs2273626&location=Seed&one=1

rs45596840-target: http://bioinfo.life.hust.edu.cn/miRNASNP#!/snp?snp_id=rs45596840&location=Seed&one=1

rs2620381-targets: http://bioinfo.life.hust.edu.cn/miRNASNP#!/snp?snp_id=rs2620381&location=Seed&one=1

Reference

- 1 Stahl EA, Wegmann D, Trynka G, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* 2012;44:483-9.
- 2 van der Woude D, Alemayehu WG, Verduijn W, et al. Gene-environment interaction influences the reactivity of autoantibodies to citrullinated antigens in rheumatoid arthritis. *Nat Genet* 2010;42:814-6.
- 3 Cavalli G, Heard E. Advances in epigenetics link genetics to the environment and disease. *Nature* 2019;571:489-99.

- 4 Eyre S, Bowes J, Diogo D, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet* 2012;44:1336-40.
- 5 Westra HJ, Martinez-Bonet M, Onengut-Gumuscu S, et al. Fine-mapping and functional studies highlight potential causal variants for rheumatoid arthritis and type 1 diabetes. *Nat Genet* 2018;50:1366-74.
- 6 Raychaudhuri S, Sandor C, Stahl EA, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 2012;44:291-6.
- 7 Okada Y, Terao C, Ikari K, et al. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat Genet* 2012;44:511-6.
- 8 Stahl EA, Raychaudhuri S, Remmers EF, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 2010;42:508-14.
- 9 Li M, Santpere G, Imamura Kawasawa Y, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 2018;362.
- 10 Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012;337:1190-5.
- 11 Zhu X, Wu L, Mo X, et al. Identification of PBMC-expressed miRNAs for rheumatoid arthritis. *Epigenetics* 2020;15:386-97.
- 12 Lai NS, Yu HC, Tung CH, et al. The role of aberrant expression of T cell miRNAs affected by TNF-alpha in the immunopathogenesis of rheumatoid arthritis. *Arthritis Res Ther* 2017;19:261.
- 13 Khalifa O, Pers YM, Ferreira R, et al. X-Linked miRNAs Associated with Gender Differences in Rheumatoid Arthritis. *Int J Mol Sci* 2016;17:1852.
- 14 Karami J, Aslani S, Tahmasebi MN, et al. Epigenetics in rheumatoid arthritis; fibroblast-like synoviocytes as an emerging paradigm in the pathogenesis of the disease. *Immunol Cell Biol* 2020;98:171-86.
- 15 Moran-Moguel MC, Petarra-Del Rio S, Mayorquin-Galvan EE, et al. Rheumatoid Arthritis and miRNAs: A Critical Review through a Functional View. *J Immunol Res* 2018;2018:2474529.
- 16 Yang XK, Li P, Zhang C, et al. Association between IRAK1 rs3027898 and miRNA-499 rs3746444 polymorphisms and rheumatoid arthritis : A case control study and meta-analysis. *Z Rheumatol* 2017;76:622-29.
- 17 Fu L, Jin L, Yan L, et al. Comprehensive review of genetic association studies and meta-analysis on miRNA polymorphisms and rheumatoid arthritis and systemic lupus erythematosus susceptibility. *Hum Immunol* 2016;77:1-6.
- 18 Li K, Tie H, Hu N, et al. Association of two polymorphisms rs2910164 in miRNA-146a and rs3746444 in miRNA-499 with rheumatoid arthritis: a meta-analysis. *Hum Immunol* 2014;75:602-8.
- 19 Yang B, Chen J, Li Y, et al. Association of polymorphisms in pre-miRNA with inflammatory biomarkers in rheumatoid arthritis in the Chinese Han population. *Hum Immunol* 2012;73:101-6.
- 20 Hashemi M, Eskandari-Nasab E, Zakeri Z, et al. Association of pre-miRNA-146a rs2910164 and premiRNA-499 rs3746444 polymorphisms and susceptibility to rheumatoid arthritis. *Mol Med Rep* 2013;7:287-91.
- 21 Shaker OG, El Boghdady NA, El Sayed AE. Association of MiRNA-146a, MiRNA-499, IRAK1 and PADI4 Polymorphisms with Rheumatoid Arthritis in Egyptian Population. *Cell Physiol Biochem* 2018;46:2239-49.
- 22 Jiang L, Yin J, Ye L, et al. Novel risk loci for rheumatoid arthritis in Han Chinese and congruence with risk variants in Europeans. *Arthritis Rheumatol* 2014;66:1121-32.
- 23 Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res* 2019;47:D155-D62.
- 24 Jiang XY, Zhang Q, Chen P, et al. CYP7A1 polymorphism influences the LDL cholesterol-lowering response to atorvastatin. *Int J Clin Pharm* 2012;37:719-23.
- 25 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.
- 26 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 Biol* 2016;37:4777-84.
- 27 Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285-91.
- 28 GenomeAsia KC. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature* 2019;576:106-11.
- 29 Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature* 2015;526:68-74.

- 30 Belsare S, Levy-Sakin M, Mostovoy Y, et al. Evaluating the quality of the 1000 genomes project data. *BMC Genomics* 2019;20:620.
- 31 Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014;506:376-81.
- 32 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.
- 33 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
- 34 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.
- 35 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.
- 36 Castro F, Acevedo E, Ciusani E, et al. 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.
- 37 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.
- 38 Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res* 2020;48:D127-D31.
- 39 Breuer K, Foroushani AK, Laird MR, et al. InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. *Nucleic Acids Res* 2013;41:D1228-33.
- 40 Banerjee S, Zeng L, Schunkert H, et al. Bayesian multiple logistic regression for case-control GWAS. *PLoS Genet* 2018;14:e1007856.
- 41 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.
- 42 Boulesteix AL, Strobl C, Weidinger S, et al. Multiple testing for SNP-SNP interactions. *Stat Appl Genet Mol Biol* 2007;6.
- 43 Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-75.
- 44 Ghosh D. Wavelet-based Benjamini-Hochberg procedures for multiple testing under dependence. *Math Biosci Eng* 2019;17:56-72.
- 45 Gong J, Liu C, Liu W, et al. An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. *Database (Oxford)* 2015;2015.
- 46 Hecker M, Fitzner B, Blaschke J, et al. Susceptibility variants in the CD58 gene locus point to a role of microRNA-548ac in the pathogenesis of multiple sclerosis. *Mutat Res Rev Mutat Res* 2015;763:161-7.
- 47 Hecker M, Boxberger N, Illner N, et al. A genetic variant associated with multiple sclerosis inversely affects the expression of CD58 and microRNA-548ac from the same gene. *PLoS Genet* 2019;15:e1007961.
- 48 Sable R, Durek T, Taneja V, et al. Constrained Cyclic Peptides as Immunomodulatory Inhibitors of the CD2:CD58 Protein-Protein Interaction. *ACS Chem Biol* 2016;11:2366-74.
- 49 Hoffmann JC, Bayer B, Zeidler H. Characterization of a soluble form of CD58 in synovial fluid of patients with rheumatoid arthritis (RA). *Clin Exp Immunol* 1996;104:460-6.
- 50 Guo S, Zhu Q, Jiang T, et al. Genome-wide DNA methylation patterns in CD4+ T cells from Chinese Han patients with rheumatoid arthritis. *Mod Rheumatol* 2017;27:441-47.

Figure 1. Flowchart for SNP and sample selection as well as population ancestry of GHRA cohort. (A) human microRNAs were downloaded from miRBase (Release 22.1). We mapped all the dbSNP153 SNPs into human Asian 100K data with MAF>1%. 4 positive control (*HLA-DRB1*, *HLA-DRB9*, *HLA-DQB1* and *TNFAIP3*) and 4 ancestry markers (South and North Han Chinese) were added for the quality control. (B) PCA analysis to show the population ancestry of Guanghua Rheumatoid Arthritis cohort (GHRA) with 1000 Genome dataset. Symbols are as same as Super Population Code (SPC) from 1000 genome project. AFR, African; AMR, Ad Mixed American; EAS, East Asian; EUR, European; SAS, South Asian.

Figure 2. Genomic position and functional annotation to the significant SNPs in miRNAs. rs1414273 (C/T) in *miR-548ac*; rs2620381(A/C) in *miR-627-5p* which cause 2,056 target gain and 1,440 target loss. rs4285314(G/A) in *miR-3135b*; rs28477407(C/T) in *miR-4308*; rs5997893 (A/G) in *miR-3928-5p* and rs45596840 (G/A) in *miR-4482-5p* which will cause 1,677 target gain and 1,456 target loss. Target gain or loss were imputed by miRNA-SNP database⁴⁵ (see method).

Figure 3. miRNA target analysis and target distribution in GWAS significant genes, immune genes and FDA drug target genes. (A) Network connection between the top 6 miRNA predicted regulatory targets and previous GWAS identified 123 RA associated genes. (B) Network connection between the top 6 miRNA predicted regulatory targets and 4,723 InnateDB collected immune genes. (C) Network connection between the top 6 miRNA predicted regulatory targets and 672 FDA approved drug target genes. miRNA was labelled as diamond filled with green color and gene were as circle filled with yellow color.

Table 1. Summary of risk alleles, 1000 Genome frequencies and SNPs associated with rheumatoid arthritis. *FDR correction were not showed in the table while only $P < 2.55 \times 10^{-3}$ is significant in the table with FDR correction. EAS, EUR and AFR represent allele frequency of A1 in East Asian, European and African population based on 1000 Genome dataset. All the OR and P-value are referring to A1 while taking A2 as reference. Bayesian logistic regression model adjusted for covariates including gender, age and BMI, were applied as the statistical approach.

Table 2. Association statistics for rheumatoid arthritis susceptibility and genetic variants. & indicating P-values from genotype-based chi-square or fisher exact test (when $N \leq 5$) models. All the OR and P-value are calculated compared with homogeneous protective allele (reference row as the table). Chi-square or Fisher exact test were applied as the statistical approach so that ORs are comparative between different models.

Table 3. meta-analysis between our study and previous East-Asian population-based RA GWAS study.

Table 4. Epistasis analysis to identify SNP-SNP interaction in rheumatoid arthritis susceptibility. Five significant epistasis effect were selected to show the increased or decreased risk effect. rs4947332 located in C2 upstream (*HLA-DRB1*); rs9275376 located in *HLA-DQB1* upstream; rs5997893 located in *miR-3928*; rs7752903 located in downstream of *TNFAIP3* and rs2967897 located in *miR-5695*.

Table 5. Cumulative risk effect of rheumatoid arthritis increased as more risk alleles carried by the patients. *individuals carrying more than 9 risk alleles were not found in our dataset. 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.

Supplementary Matreial:

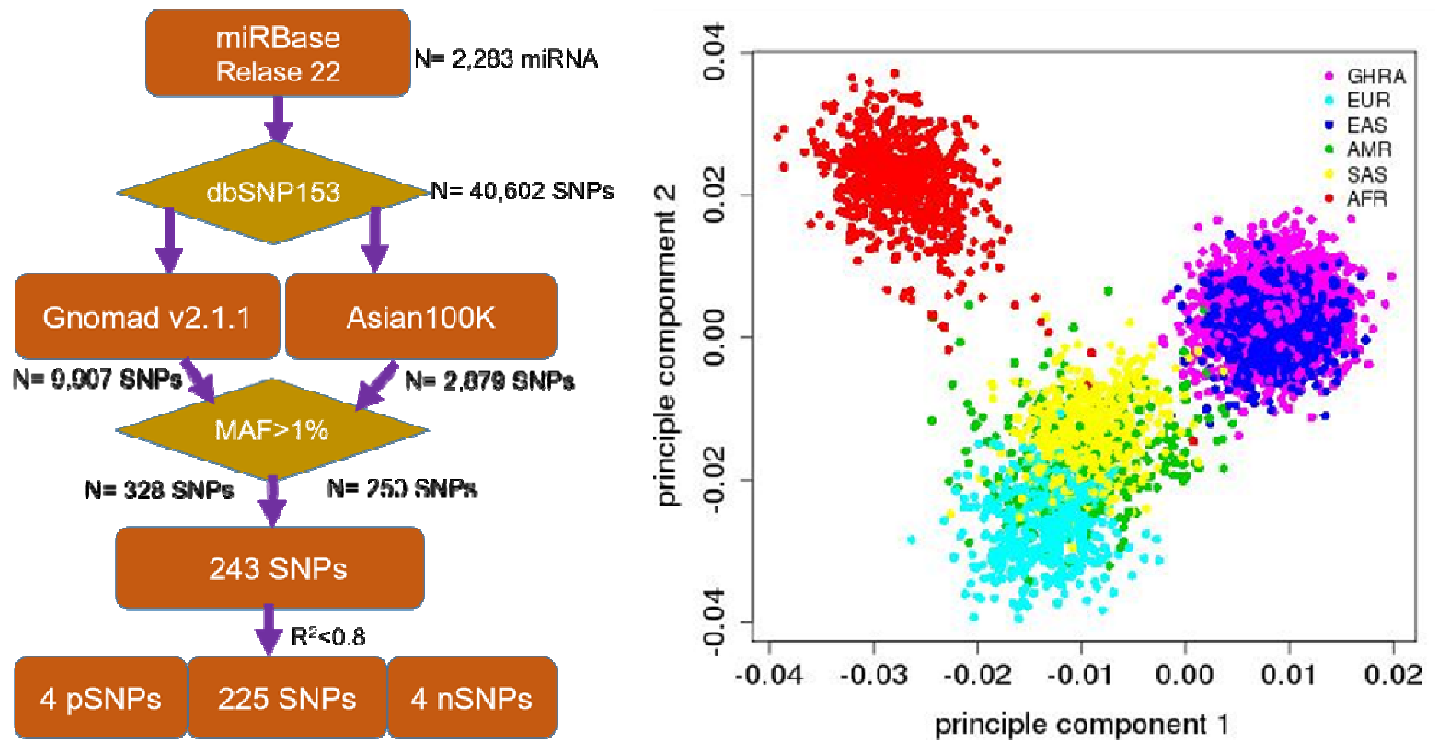
Figure S1. Re-sampling technique based enrichment analysis to miRNA regulatory targets. *A random sampling technique was applied to assess the null distribution of p-values ($n=50,000$ iterations). 37,875 total genes in the genome (GENCODE v32) were assumed for the calculations. Enrichment was measured with fold-change (FC) and p-values were calculated by summing the frequency of more extreme values in the null distribution of p-values.

Table S1. Demographic information for the GHRA cohort

Table S2. Cochran-Armitage Trend test-based association analyses.

Table S3. Epistasis analysis to identify SNP-SNP interaction in rheumatoid arthritis susceptibility

Table S4. miRDB based miRNA target prediction for the 6 significant RA-miRNA.



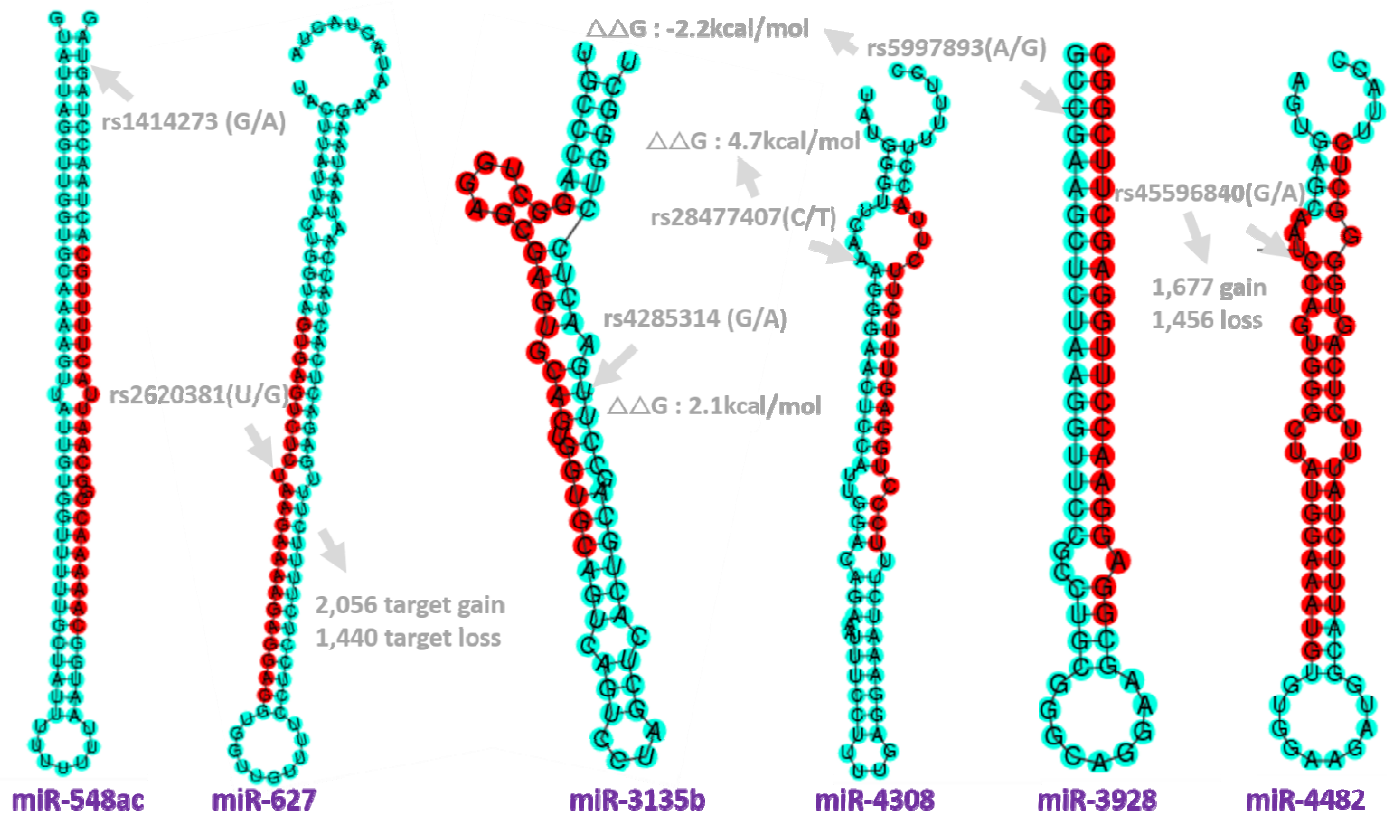


Figure 2. Genomic position and functional annotation to the significant SNPs in miRNAs

rs1414273 (C/T) in miR-548ac; rs2620381 (A/C) in *miR-627-5p* which cause 2056 target gain and 1440 target loss. rs4285314 (G/A) in miR-3135b; rs28477407 (C/T) in miR-4308; rs5997893 (A/G) in miR-3928-5p and rs45596840 (G/A) in miR-4482-5p which will cause 1677 target gain and 1456 target loss. Target gain or loss were imputed by miRNA-SNP database⁴⁵ (see method).

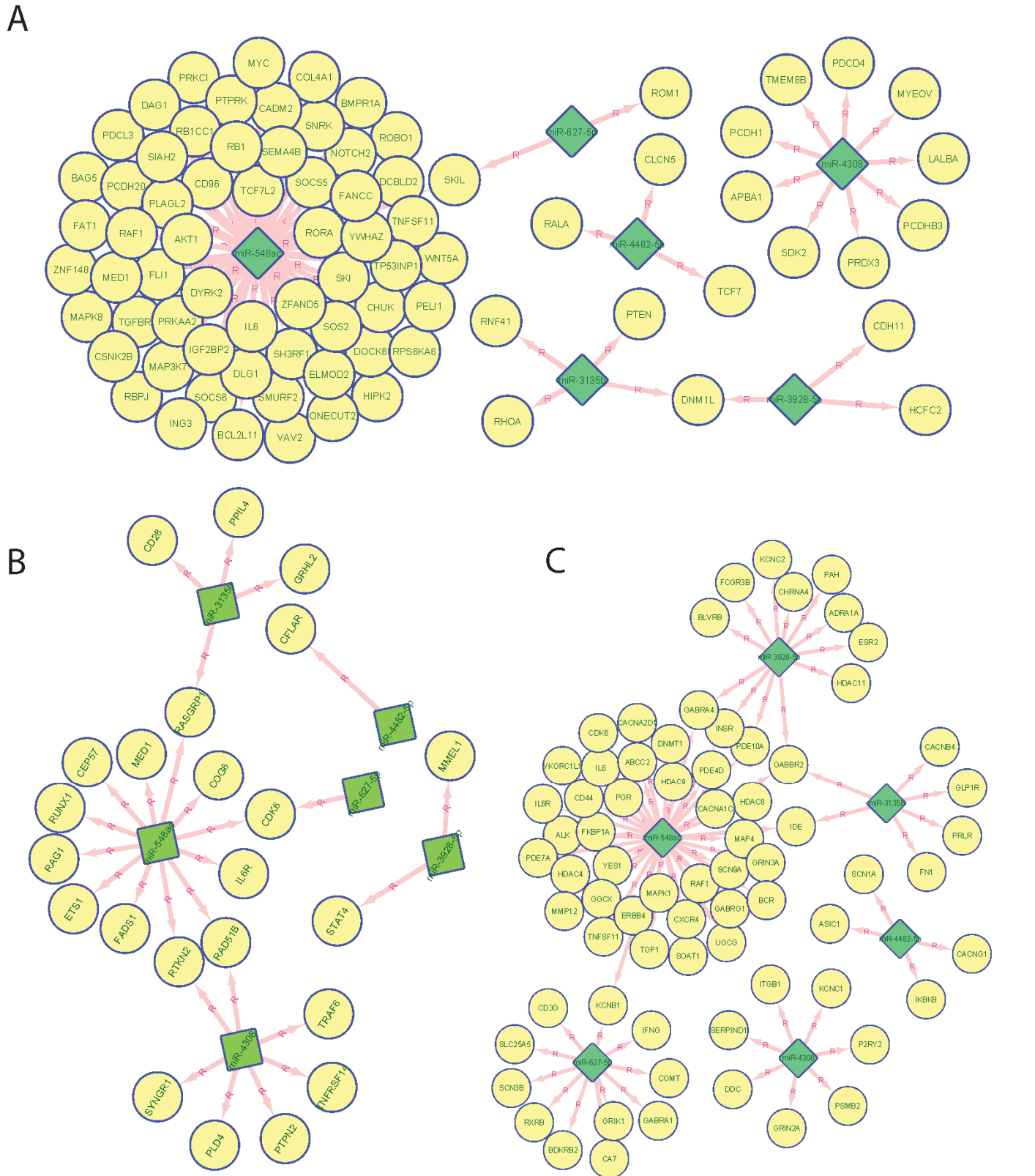


Table 1. Summary of risk alleles, 1000 Genome frequencies and SNPs associated with rheumatoid arthritis

CHR	SNP	A1/A2	RA (no.)	Normal (no.)	OR	P	EAS	EUR	AFR
6p21.32	rs9268839	A/G	42.4%(1364)	56.0%(1768)	0.58(0.53-0.64)	3.95x10 ^{-27*}	0.354	0.451	0.236
6p21.32	rs9275376	T/G	25.0%(804)	15.7%(496)	1.79(1.58-2.03)	2.65x10 ^{-20*}	0.192	0.299	0.285
6q23.3	rs7752903	G/T	6.4%(205)	4.3%(136)	1.51(1.21-1.89)	2.33x10 ^{-4*}	0.049	0.019	0.056
6p21.33	rs4947332	T/C	2.2%(72)	1.1%(34)	2.11(1.40-3.18)	2.78x10 ^{-4*}	0.012	0.029	0.094
1p13.1	rs1414273	C/T	40.5%(1303)	44.7%(1412)	0.84(0.76-0.93)	8.26x10 ^{-4*}	0.585	0.140	0.584
15q15.1	rs2620381	C/A	7.9%(254)	10.1%(318)	0.77(0.65-0.91)	2.55x10 ^{-3*}	0.086	0.004	0.130
14q32.31	rs75330474	T/C	3.8%(121)	5.0%(159)	0.74(0.58-0.94)	0.01360	0.065	0.003	0.064
18p11.21	rs370878033	G/A	1.7%(54)	2.6%(81)	0.65(0.46-0.92)	0.01435	0.032	0.001	0.001
7q22.1	rs3823658	A/G	13.4%(432)	11.5%(363)	1.20(1.03-1.39)	0.01825	0.138	0.138	0.006
1p32.3	rs74085143	A/G	4.0%(130)	5.3%(166)	0.76(0.60-0.96)	0.02190	0.048	0.022	0.331
6q23.2	rs7740161	T/A	14.3%(459)	16.3%(515)	0.86(0.75-0.98)	0.02530	0.102	0.347	0.614
3p21.1	rs4687672	A/G	33.7%(1082)	31.1%(983)	1.12(1.01-1.25)	0.02914	0.317	0.264	0.202
22q12.2	rs5997893	A/G	46.9%(1509)	49.5%(1565)	0.90(0.82-1.00)	0.03972	0.498	0.667	0.918
19p13.2	rs2967897	T/C	16.8%(541)	15.0%(473)	1.15(1.00-1.31)	0.04189	0.849	0.686	0.614
15q25.3	rs76468441	T/C	3.2%(104)	2.4%(76)	1.36(1.00-1.83)	0.04530	0.042	0.054	0.011

*FDR correction were not showed in the table while only $P < 2.55 \times 10^{-3}$ is significant in the table with FDR correction. EAS, EUR and AFR represent allele frequency of A1 in East Asian, European and African population based on 1000 Genome dataset. All the OR and P-value are referring to A1 while taking A2 as reference. Bayesian logistic regression model adjusted for covariates including gender, age and BMI, were applied as the statistical approach.

Table 2. Association statistics for rheumatoid arthritis susceptibility and genetic variants

Genotypes, risk allele	RA	Normal	OR (95% CI)	P-value
rs9268839, located at 6p21.32, <i>HLA-DRB9</i>				
AA	262(16.30%)	494(31.27%)	1.00 (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 vs GA+AA	505(31.43%)	306(19.37%)	1.91(1.62-2.25)	4.77x10 ⁻¹⁵
GG+GA vs 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, <i>HLA-DQB1</i>				
GG	887(55.20%)	1128(71.39%)	1.00 (reference)	1.26x10 ⁻²⁰ &
TG	636(39.58%)	408(25.82%)	1.98(1.70-2.32)	
TT	84(5.23%)	44(2.78%)	2.43(1.65-3.62)	
TT vs TG+GG	84(5.23%)	44(2.78%)	1.93(1.31-2.86)	5.58x10 ⁻⁴
TT+TG vs 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, <i>TNFAIP3</i>				
TT	1408(87.62%)	1445(91.46%)	1.00(reference)	4.19x10 ⁻⁴⁸ &
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 vs GT+TT	6(0.37%)	1(0.06%)	5.92(0.72-272.02)	0.12460
GG+GT vs 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, Tag-SNPs of <i>HLA-DRB1*04:07</i>				
CC	1537(95.64%)	1546(97.85%)	1.00 (reference)	5.51x10 ⁻⁴⁸ &
TC	68(4.23%)	34(2.15%)	2.01(1.31-3.15)	
TT	2(0.12%)	0(0.00%)	-	
TT vs TC+CC	2(0.12%)	0(0.00%)	-	0.49980
TT+TC vs 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 <i>miR-548ac</i>				
CC	284(17.67%)	307(19.43%)	1.00 (reference)	4.84x10 ⁻⁴⁸ &
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 vs TC+CC	588(36.59%)	475(30.06%)	1.34(1.15-1.56)	9.35x10 ⁻⁵
TT+TC vs CC	1323(82.33%)	1273(80.57%)	1.12(0.94-1.35)	0.20230
T allele	1911(59.46%)	1748(55.32%)	1.18(1.07-1.31)	9.04x10 ⁻⁴
rs2620381, located at 15q15.1 within <i>miR-627</i>				
CC	9(0.56%)	20(1.27%)	1.00 (reference)	7.27x10 ⁻³⁸ &
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 vs AC+CC	1362(84.75%)	1282(81.14%)	1.29(1.07-1.56)	7.21x10 ⁻³
AA+AC vs CC	1598(99.44%)	1560(98.73%)	2.28(0.99-5.7)	0.04035
A allele	2960(92.10%)	2842(89.94%)	1.30(1.09-1.56)	0.00285

& indicating P-values from genotype-based chi-square or fisher exact test (when N<=5) models. All the OR and P-value are calculated compared with homogeneous protective allele (reference row as the table). Chi-square or Fisher exact test were applied as the statistical approach so that ORs are comparative between different models.

Table 3. meta-analysis between our study and previous East-Asian population-based RA GWAS study

CHR	SNP	BP	A1	OR ¹	P ¹	FDR
6	rs9268839 (<i>HLA-DRB9</i>)	32428772	A	0.52	4.15x10 ⁻⁴⁹	2.08 x10 ⁻⁴⁷
6	rs9275376 (<i>HLA-DQB1</i>)	32668633	T	1.72	1.11x10 ⁻⁴⁷	1.04x10 ⁻⁴⁵
6	rs7752903 (<i>TNFAIP3</i>)	138227364	T	0.73	1.94x10 ⁻²²	1.21x10 ⁻²⁰
6	rs4285314 (<i>miR-3135b</i>)	32717702	A	1.33	2.35x10 ⁻¹⁵	1.10x10 ⁻¹³
14	rs28477407 (<i>miR-4308</i>)	55344901	T	0.89	9.19x10 ⁻⁰⁷	3.44x10 ⁻⁵
22	rs5997893 (<i>miR-3928</i>)	31556103	A	0.93	1.92x10 ⁻⁰⁴	5.97x10 ⁻³
10	rs45596840 (<i>miR-4482</i>)	106028154	A	0.89	2.49x10 ⁻⁰⁴	6.65x10 ⁻³
6	rs4947332 (<i>HLA-DRB1</i>)	31834197	T	1.41	9.50x10 ⁻⁰⁴	2.22x10 ⁻²

Table 4. 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	1.00 (reference)	-
	CT+TT	GG	2	9	0.38(0.10-1.38)	0.164
	CC	GT+TT	652	427	1.93(1.66-2.24)	6.59x10 ⁻¹⁸
	CT+TT	68	25	3.35(2.12-5.31)	6.70x10 ⁻⁸	
rs4947332& rs5997893	rs4947332	rs5997893				
	CC	GG	330	392	1.00 (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	61	25	2.83(1.75-4.58)	1.36x10 ⁻⁵	
rs7752903& rs2967897	rs7752903	rs2967897				
	TT	CC	953	1050	1.00 (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	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 (*HLA-DRB1*); rs9275376 located in *HLA-DQB1* upstream; rs5997893 located in miR-3928; rs7752903 located in downstream of *TNFAIP3* and rs2967897 located in miR-5695.

Table 5. 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	4	17	1.00 (Reference)	-
2	65	133	1.77(0.63-4.99)	0.348209
3	206	340	2.19(0.80-5.97)	0.129698
4	388	466	3.00(1.10-8.15)	0.031822
5	411	359	4.12(1.51-11.21)	0.002819
6	344	185	6.68(2.44-18.27)	6.30x10 ⁻⁰⁵
7	138	67	7.36(2.62-20.66)	4.26x10 ⁻⁰⁵
8	46	10	15.38(4.69-50.49)	1.00x10 ⁻⁰⁶
9	5	3	5.40(1.08-26.93)	0.049413

*individuals carrying 0 or more than 9 risk alleles was not found in our dataset. 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.