Comparison of the Genetic Structure of Rheumatoid Arthritis Between European and Asian Population

Chun'e Li
Xi'an Jiaotong University https://orcid.org/0000-0001-9497-9026
Xiaomeng Chu
Xi'an Jiaotong University
Shiqiang Cheng
Xi'an Jiaotong University
Yan Wen
Xi'an Jiaotong University
Chuyu Pan
Xi'an Jiaotong University
Huijie Zhang
Xi'an Jiaotong University
Yujing Chen
Xi'an Jiaotong University
Jingxi Zhang
Xi'an Jiaotong University
Zhen Zhang
Xi'an Jiaotong University
Xuena Yang
Xi'an Jiaotong University
Peilin Meng
Xi'an Jiaotong University
Yumeng Jia
Xi'an Jiaotong University
Feng Zhang (fzhxjtu@mail.xjtu.edu.cn)
Xi'an Jiaotong University

Research article

Keywords: Genetic Structure, Rheumatoid Arthritis, PTPN2 (rs284728), PADI2 (rs223592), STAT4 (rs12612769)

Posted Date: July 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-642514/v1

Comparison of the Genetic Structure of Rheumatoid Arthritis Between European and Asian Population

Chun'e Li*, Xiaomeng Chu, Shiqiang Cheng, Yan Wen, Chuyu Pan, Huijie Zhang, Yujing Chen, Jingxi Zhang, Zhen Zhang, Xuena Yang, Peilin Meng, Yumeng Jia#, Feng Zhang#

Key Laboratory of Trace Elements and Endemic Diseases of National Health and Family Planning Commission, School of Public Health, Health Science Center, Xi'an Jiaotong University, Xi'an, China

#Corresponding author:

Yumeng Jia

Key Laboratory of Trace Elements and Endemic Diseases, National Health Commission of the People's Republic of China. School of Public Health, Health Science Center, Xi'an Jiaotong University

Email: jiayumemg@mail.xjtu.edu.cn

Xi'an, P. R. China 71006

Feng Zhang

Key Laboratory of Trace Elements and Endemic Diseases, National Health Commission of the People's Republic of China. School of Public Health, Health Science Center, Xi'an Jiaotong University

Email: fzhxjtu@mail.xjtu.edu.cn

Xi’an, P. R. China 71006

Abstract

Background: Rheumatoid arthritis (RA) is one of the chronic inflammatory diseases that primarily influences the joints, and its prevalence is 0.5-1.0%. Previous studies have shown that there are differences in the genetic structure of RA between European and Asian populations, and most of the studies have been conducted using meta-analysis. This study analyzed the genetic structure of rheumatoid arthritis in European and Asian populations using a new method.

Methods: The Genome-Wide Association Study (GWAS) summary statistics of RA from Europe (N=8383) and Asia (N=19190) were derived from an article published in *Nature*. First, the GWAS data was divided into 1368 blocks, in which SNPs were approximately independent of linkage disequilibrium (LD). Second, we calculated the LD matrix of SNP in each block by using PLINK 1.9. Then, PESCA analysis was performed to detect population-specific/shared risk genes of RA. Finally, Metascape platform was used to perform gene set enrichment analysis.

Results: In European population, we found multiple genes which were associated with RA, including HLA-DPA1, HLA-DPB1 (rs2856822, PTP=1.000), MICA (rs2844518, PTP=1.000). In Asian population, C6orf10 (rs3129915, PTP=1.000), PTPN2 (rs2847288, PTP=0.995), were significant related to RA. The population-shared genes included PADI2 (rs2235920, PTP=1.000), STAT4 (rs12612769, PTP=1.000). Furthermore, gene sets enrichment analysis reported population-specific/shared pathway terms, such as interferon-gamma-mediated signaling pathway ($P=2.884\times10^{-9}$), negative regulation of innate immune response ($P=1.841\times10^{-7}$), protein-arginine deiminase activity ($P=7.047\times10^{-8}$).

Conclusions: The results of our study indicate differences in risk loci between Asian and European populations, which provided clues for exploring the population-specific/shared genetics and

pathogenesis of RA.

Keywords: Genetic Structure, Rheumatoid Arthritis, PTPN2 (rs284728), PADI2 (rs223592), STAT4 (rs12612769)

Introduction

Rheumatoid arthritis (RA) is one of the chronic inflammatory diseases that primarily influences the joints, leading to chronic inflammation in synovium, which finally causes the destruction of joints, deformity, and disability. The prevalence of RA is 0.5-1.0% (1).

Although the etiology of RA is unknown, significant progress has been made in identifying risk factors. The genetics, environmental factors and autoimmunity are demonstrated to play an important role in the pathogenesis of RA. For example, smoking is one of the most important environmental factors affecting RA susceptibility, and the alcohol intake, BMI, dietary habits, and even coffee consumption (2-5). However, twin studies and family studies from European countries estimated 53-65% of the risk for development of RA is attributable to genetic factors, the numbers for Asian populations were similar and comparable(6-8). In the past years, Genome-wide Association Studies (GWAS) studies have been widely used to identify the susceptibility risk loci of RA, for example, RAD51B, BACH2 et al. have been found to be related to RA (9).

It is widely believed that many common risk variants for most of the complex diseases are shared among different populations. However, the causal variants and disease-related SNPs varies significantly among ethnic groups. This variation might attribute to genetic or environmental factors. Trans-populations analysis suggested that the genetic architecture of RA is similar across ethnic groups, but the incidence of RA is not completely consistent between Asian and European populations, which suggested only modest overlap in RA risk related SNPs between different populations (6, 10). This genetic structural heterogeneity has been shown to limit some post-GWAS analyses, such as the transferability of polygenic risk scores (PRSs) in different populations (11-13). Therefore, it is necessary to identify the shared and population-specific components of genetic

structure, which can improve cross-population analysis and study of disease etiology.

Previous GWAS studies have reported some common/specific genetic risk loci of RA in different ethnicities. However, most of the previous trans-population studies used meta-analysis which has some limitations, heterogeneity caused by different genetic backgrounds impact on the interpretation of meta-analysis results. Besides, although previous studies have included a large number of samples which from different genetic backgrounds, the European sample was still a large proportion, which limited the generalization of the results. PESCA (population-specific/shared causal variants) (14) is a method which can estimate the genome-wide proportions of SNPs with nonzero effects in a single population or in two different populations from GWAS summary statistics and estimates of Linkage Disequilibrium (LD). Comparing with meta-analysis, PESCA could improve the transferability of GWAS findings between different populations by correcting the population-specific LD and allele frequencies (14), which has been used in several complex traits, such as schizophrenia.

In our study, we used PESCA to analyze the RA individuals of Asian and European ancestry, in order to identify population-specific/shared SNPs, and all the GWAS summary statistics were from a published study (15). Our study can help to identify the potential risk loci of RA, and provide some new clues to the pathophysiology of RA.

Method

GWAS Summary data of European and Asian RA

GWAS summary data for European and Asian were derived from articles published in *Nature* (15) and were available from the JENGER (Japanese ENcyclopedia of GEnetic Associations by Riken)

(http://jenger.riken.jp/en/result). European data included 2843 cases and 5540 controls, and Asian data included 3636 cases and 15,554 controls. All RA cases met the 1987 criteria of the American College of Rheumatology for RA diagnosis. GWAS genotyping and quality control standards included sample call rate, SNP call rate, minor allele frequency (MAF), and Hardy-Weinberg equilibrium (HWE) P value. Briefly, monomorphic or singleton SNPs or SNPs with deviation of HWE ($P<1\times10^{-7}$) have been excluded from each of the reference panels, SNPs with MAF<0.005 in RA cases or controls, or with low imputation score (Rsq<0.5 for genome-wide array and<0.7 for Immunochip) from each study have been filtered.

Data Processing

We removed redundant SNPs based on the 1000 Genomes Phase 3 reference panel (16, 17) following the criterion of PESCA, then calculated Z-score of all SNPs in each population according to Z=Effect/StdErr, Huwenbo Shi et al.(18) used LDetect and the 1000 Genomes Phase 3 reference panel to create 1,368 LD blocks (with an average length of 2 Mb), which are approximately independent in EAS and EUR individuals. According to the LD blocks, we split the GWAS results of two populations into 1368 blocks, in which SNPs were approximately independent of LD. Finally, we calculated LD matrix of SNPs in each block by using the PLINK 1.9 (19), and then used LD Score Regression (LDSC) software to calculate the heritability of RA in two populations (20).

Identifying Population-Specific/Shared risk genes for RA

PESCA takes the causal relationships i of SNPs between the two populations as a binary vector C_i, and assumes that C_i follows a multivariate Bernoulli (MVB) distribution (21):

$$C_i \sim MVB(f_{00}, f_{01}, f_{10}, f_{11})$$

Where $(f_{00}, f_{01}, f_{10}, f_{11})$ represents the parameter vector of MVB distribution. Using Expectation-

maximization (EM) and Markov Chain Monte Carlo (MCMC) to estimate the genome-wide proportions of population-specific/shared risk SNPs. Then, genome-wide proportions are used as the prior probability to estimate the posterior probability of each SNP as a risk factor in a single population or two populations.

In this study, there are 557,717 SNPs for a given genotype (RA) in European and Asian populations, respectively, the effects of the i^{th} risk SNP in each population, β_{1i} and β_{2i} can be computed by the following equation:

$$\beta_{1C_1}|c_1 \sim MVN(0, \frac{h_{g1}^2}{|c_1|})I_{C1}$$

$$\beta_{1C_2}|c_2 \sim MVN(0, \frac{h_{g2}^2}{|c_2|})I_{C_2}$$

Where $|C_1| = \sum_{i-1}^{M} C_{i1}$ and $|C_2| = \sum_{i-1}^{M} C_{i2}$ represent the total numbers of SNPs in each population, h_{g1}^2 and h_{g2}^2 are the total SNP-heritability in each population. The published research contained more detailed information(14). The SNPs with high posterior probability (PTP>0.8) were reported.

GO Enrichment Analysis

The gene enrichment analysis was performed on Metascape platform(22). GO Biological Processes, GO Cellular Components were used in the analysis for each of the given gene list. All genes in the genome participated in the enrichment analysis. P-values are calculated according to the accumulative hypergeometric distribution, as well as q-values are computed using the Banjamini-Hochberg(23) procedure to account for multiple tests.

Ethic statement

All participants provided written informed consent for participation in the study as approved by the ethical committees of each of the institutional review boards, and more ethical details were

given in the published studies(15).

Result

European-specific genes and pathways for RA

In the European population, multiple candidate genes were identified with a nonzero effect on RA, such as HLA-DPA1, HLA-DPB1 (rs2856822, PTP=1.000), MICA (rs2844518, PTP=1.000), TRIM10 (rs2517653, PTP=1.000), HLA-B (rs35795116, PTP=1.000). Table 1 summarized the top 20 European-specific SNPs.

In the gene enrichment analysis of European population-specific genes, we observed several specific pathways. Most of these pathways were related to receptor signaling pathways and membrane, such as GO:0060333~*interferon-gamma-mediated signaling pathway* ($P=2.884 \times 10^{-9}$) and GO:0071556~*integral component of lumenal side of endoplasmic reticulum membrane* ($P=7.303 \times 10^{-8}$).

Asian-specific genes and pathways for RA

In the Asian population, our research found a group of candidate genes with a nonzero effect on RA, including C6orf10 (rs3129915, PTP=1.000), LEMD2 (rs6904716, PTP=1.000), PTPN2 (rs2847288, PTP=0.995), FOXJ3 (rs7526424, PTP=0.847). Table 2 summarized the top 20 Asian-specific SNPs.

In the gene enrichment analysis result of Asian population-specific genes, we found multiple specific pathways which related to immune regulation, including GO:0045824~*negative regulation of innate immune response* ($P=1.841 \times 10^{-7}$) and GO:0050777~*negative regulation of immune response* ($P=6.637 \times 10^{-6}$).

European-Asian shared genes and pathways for RA

In the European and Asian populations, we detected multiple significant shared genes with a nonzero effect on RA, including PADI2 (rs2235920, PTP=1.000), ARID5B (rs10740058, PTP=1.000), PADI2 (rs2235924, PTP=1.000), STAT4 (rs12612769, PTP=1.000). Table 3 summarized the European-Asian populations share risk genes. In the results of gene enrichment analysis, we observed significant enriched pathways which associated with enzyme and cell by using population-shared genes, these pathways including GO:0004668~*protein-arginine deiminase activity* (PTP=7.047×10^{-8}), GO:0018101 ~*protein citrullination* (PTP=7.047×10^{-8}), GO:0046631~*alpha-beta T cell activation* (PTP=2.239×10^{-5}).

Discussion

In this study, we explored the risk loci of RA across populations by using PESCA based on the published GWAS data. A total of 51 risk gene loci in the European population and 21 risk gene loci in the Asian population, as well as the 52 shared risk loci between European and Asian were identified. The risk genes of Asian and European populations were clustered together within chromosome 6, while Asian-European shared genes are scattered on multiple chromosomes.

Our results support that there was a common genetic risk for RA in Asians and Europeans population, and these conclusions have been published in previous studies. Laura B et al. have reported that most of the risk genes of RA among European ancestry showed the similar ORs in the population of African Americans with RA (24), and Jan Freudenberg et al. suggested that the most of the specific risk loci of RA to Asian or European populations were also shared in the both populations(10). However, although there was a certain overlap in the RA risk loci between the two populations, several complex traits have been reported to have strong cross-ethnic genetic

association (18) . Consistent with our study, we found a significant difference in the genetic structure of RA between the two populations. This difference was not only caused by genetic susceptibility, the environmental factors such as culture, migration histories, geodemographic events, but also caused by the variantion of genetic structure (25, 26).

Some of population-shared risk genes and GO pathway terms have been reported to be associated with RA. Citrullination, which plays an important role in RA, was catalyzed by a group of peptidylarginine deiminases (PADs). Previous studies have shown that the PADI4 was a susceptibility factor for RA (27, 28), but their research results showed that PADI2 was also significantly related to RA, and may even be involved in the pathogenesis of RA(29). Signal transducer and activator of transcription 4 (STAT4) is a member of the STAT family and locates in the cytoplasm. STAT4 can activate the pathogenesis of a variety of human autoimmune diseases and inflammatory diseases, through the Janus kinase (JAK)-STAT signaling pathway (30). A few researches have shown that STAT4 polymorphism was related to RA susceptibility, STAT4 was even associated with systemic lupus erythematosus in the Mexican population (31, 32). In addition, Chromosome 5 open reading frame 30 (C5orf30) has been repeatedly reported to be associated with the risk of many autoimmune diseases, including RA. C5orf30 is a negative regulator of joint and tissue damage in the patients with RA (33). Population-shared GO pathway terms included *protein arginine deiminase activity* and *protein citrullination* have been indicated to be related with RA (34).

For the result of European population-specific risk genes, we also found several genes which are associated with RA. The HLA gene family, found on the major histocompatibility complex (MHC) of human chromosome 6, is divided into HLA Class I and Class II genes, the HLA-DRB1

allele constitutes the strongest single genetic association of RA and may account for at least 30% of the total genetic component of RA(35, 36). In addition to HLA-DRB1, other members of the HLA gene family were associated with the susceptibility of RA, including the HLA-DPB1 and HLA-DPA1 which have been found in our study. For example, Lu Jiang et al. have reported that the high-frequency expression of DPB1*0401 and *0601 was significantly related to the susceptibility of RA, and that may be a risk factor for the occurrence of RA. On the contrary, the low-frequency expression of DPB1*0101, *0402 and *0501 may be negatively correlated with RA, which may be a protective factor for the occurrence of RA (37). MHC class I polypeptide-related sequence A (MICA) is a stress-induced protein, Achour et al. have suggested that the MICA-250 polymorphism was involved in the genetic susceptibility and severity of RA in south Tunisian population(38). Holger Kirsten et al. have proved that the association between MICA-250 (rs1051794) and RA was independent on the HLA-DRB1 risk allele, and MICA was a susceptibility gene of RA(39). In the peptidylarginine deiminases (PADs), PADI4 was the most reported to be associated with RA, multiple reports have shown that PADI4 was not only related to RA susceptibility in Asian populations, but also in European populations (28, 40).

In addition, in the GO pathway terms of European population, *MHC protein complex, T cell receptor signaling pathway* were related to the RA to some extent, it was mentioned that the MHC class I polypeptide-related sequence A (MICA) was a stress-induced protein involved in activation of NK and T cells through interaction with NKG2D receptor in the research of Milena Iwaszko et al. (38). Several Asian population-specific genes were reported in the previous researches. JU YEON BAN et al. have reported that the fork head box J3 (FOXJ3) was related to the development of RA by using the SNP Stats, Haploview and Helixtree programs (41). Chikashi et al. have shown

that the rs6904716 in LEMD2 of the HLA locus showed a borderline association with ACPA-negative RA based on the meta-analysis in a Japanese population (42). Furthermore, the Asian population-specific GO pathway terms were related to the regulation of immunity, such as *negative regulation of innate immune response*, and *regulation of leukocyte mediated immunity*. As we all know, the immune and inflammatory systems are controlled by a variety of cytokines, including interleukins and interferons. The improper activation of immune and inflammatory systems were also considered to be responsible for the generation of autoimmunity, so autoimmune arthritis such as RA, systemic lupus erythematosus (SLE), multiple sclerosis (MS) are all related to the regulation of the innate immune system (43, 44).

The total prevalence of RA in adults worldwide was about 0.5-1%(1). However, there were an obvious difference in the prevalence between different populations, the prevalence of European descent (0.3-1.1%) was higher than the prevalence of Asian descent (0.1-0.5%) (45, 46). In addition, it was reported that the prevalence rate is even higher in the Native American population (approximately 5-7%)(47). RA is an autoimmune and complex disease, the reasons for this population heterogeneity may include environmental, genetic, and autoimmune factors according to reports(45, 48). The HLA gene family with the highest polymorphism in the human genome is located on chromosome 6. There is a significant difference in the incidence and prevalence of RA among different populations or ethnic groups, and this difference could partly due to genetic variation in the HLA region, and variation in the prevalence of “shared epitope” in different populations (46, 49).

Furthermore, environmental differences are also part of the reason for this significant difference. Epidemiological studies have shown that differences in dietary structure will affect the

occurrence and development of RA to a certain extent. For example, a Mediterranean diet can reduce the risk of RA and prevent the disease from getting worse (46). The effect of smoking on RA has been demonstrated many times, Yuta Kochi et al. have shown that the variant in smoking prevalence could explain the difference in the contribution of PADI4 to RA susceptibility among various ethnic groups (45). In addition, the diverse allele frequencies of susceptible variants were also the main factors that cause this distinct (45).

In our study, GWAS summary statistics and LD estimation have been used to estimate the population-specific/shared risk genes of RA by the PESCA analysis. It is difficult to detect random SNPs shared between the two populations in meta-analysis, because of the difference in LD or allele frequency. Moreover, the conditions of meta-analysis were relatively strict, and the result interpretation bias will produce misleading information (50). Compared with traditional meta-analysis methods, PESCA analysis can help us to improve performance by using the LD model of specific populations. This way clearly corrected the population-specific LD and allele frequencies by locating the shared components of the genetic structure. Besides, a large amount of sample data has been included in this study, both of the PESCA analysis and sample data help to improve the accuracy and portability of the results. However, the "causal SNPs" which discovered by PESCA refer to non-zero effects on traits through LD, or non-zero effects on unmeasurable variant markers. These non-zero effects may be influenced by the environmental factors or gene-environment interactions, so the interpretation for our results should be cautious.

Conclusions

In summary, we observed a large number of SNPs with a PTP> 0.8 in European and Asian

populations, and most of them have been confirmed to have a clear association with RA, including the HLA gene family that accounts for about 30% for the RA genetic component. Our further research showed that there was certain overlap of non-MHC loci among two populations, suggesting the genetic correlation of RA existed in Asian and European populations was greater than our expectations. Therefore, this study provided clues for exploring the population-specific/ shared genetics and pathogenesis of RA.

Abbreviations

RA: Rheumatoid arthritis

GWAS: The Genome-Wide Association Study

LD: linkage disequilibrium

PRSs: polygenic risk scores

MAF: minor allele frequency

HWE: Hardy-Weinberg equilibrium

LDSC: LD Score Regression

MHC: major histocompatibility complex

Acknowledgement

This research has been conducted using the Biobank Japan Project. Chun'e Li contributes to analysis of data and draft the paper; Feng Zhang contributes to the design of the work and the acquisition of Biobank Japan Project and agree all aspects of the work in ensuring the work to be appropriately investigated and resolved.

Ethical Approval and Consent to participate

All participants provided written informed consent for participation in the study as approved by the ethical committees of each of the institutional review boards

Funding

No specific funding was received from any bodies in the public, commercial or not-for-profit sectors to carry out the work described in this article.

Availability of supporting data:

Data from JENGER (Japanese ENcyclopedia of GEnetic Associations by Riken) (http://jenger.riken.jp/en/result), data use has been licensed.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Authors' Contributions

Miss Li had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Chun'e Li and Feng Zhang.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Chun'e Li.

Critical revision of the manuscript for important intellectual content: Yumeng Jia Shiqiang Cheng, Yan Wen, Chuyu Pan, Huijie Zhang, Yujing Chen, Jingxi Zhang, Zhen Zhang, Xuena Yang, Peilin Meng

Statistical analysis: Chun'e Li and Xiaomeng Chu.

Supervision: Feng Zhang, Yumeng Jia

References

1. Myasoedova E, Crowson CS, Kremers HM, Therneau TM, Gabriel SE. Is the Incidence of Rheumatoid Arthritis Rising? Results From Olmsted County, Minnesota, 1955-2007. Arthritis Rheum-Us. 2010;62(6):1576-82.

2. Sugiyama D, Nishimura K, Tamaki K, Tsuji G, Nakazawa T, Morinobu A, et al. Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. Ann Rheum Dis. 2010;69(1):70-81.

3. Bae SC, Lee YH. Alcohol intake and risk of rheumatoid arthritis: a Mendelian randomization study. Z Rheumatol. 2019;78(8):791-6.

4. Bae SC, Lee YH. Causal association between body mass index and risk of rheumatoid arthritis: A Mendelian randomization study. Eur J Clin Invest. 2019;49(4).

5. Sparks JA, Barbhaiya M, Tedeschi SK, Leatherwood CL, Tabung FK, Speyer CB, et al. Inflammatory dietary pattern and risk of developing rheumatoid arthritis in women. Clin Rheumatol. 2019;38(1):243-50.

6. Terao C, Ikari K, Nakayamada S, Takahashi Y, Yamada R, Ohmura K, et al. A twin study of rheumatoid arthritis in the Japanese population. Mod Rheumatol. 2016;26(5):685-9.

7. Alamanos Y, Voulgari PV, Drosos AA. Incidence and prevalence of rheumatoid arthritis, based on the 1987 American College of Rheumatology criteria: a systematic review. Semin Arthritis Rheum. 2006;36(3):182-8.

8. Silman AJ, MacGregor AJ, Thomson W, Holligan S, Carthy D, Farhan A, et al. Twin concordance rates for rheumatoid arthritis: results from a nationwide study. Br J

Rheumatol. 1993;32(10):903-7.

9. McAllister K, Yarwood A, Bowes J, Orozco G, Viatte S, Diogo D, et al. Identification of BACH2 and RAD51B as Rheumatoid Arthritis Susceptibility Loci in a Meta-Analysis of Genome-Wide Data. Arthritis Rheum-Us. 2013;65(12):3058-62.

10. Freudenberg J, Lee HS, Han BG, Shin HD, Kang YM, Sung YK, et al. Genome-Wide Association Study of Rheumatoid Arthritis in Koreans. Arthritis Rheum-Us. 2011;63(4):884-93.

11. Kerminen S, Martin AR, Koskela J, Ruotsalainen SE, Havulinna AS, Surakka I, et al. Geographic Variation and Bias in the Polygenic Scores of Complex Diseases and Traits in Finland. Am J Hum Genet. 2019;104(6):1169-81.

12. Arya R, Escalante A, Farook VS, Restrepo JF, Battafarano DF, Almeida M, et al. Data on genetic associations of carotid atherosclerosis markers in Mexican American and European American rheumatoid arthritis subjects. Data Brief. 2018;17:820-9.

13. Vilhjalmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. Am J Hum Genet. 2015;97(4):576-92.

14. Shi HWB, Burch KS, Johnson R, Freund MK, Kichaev G, Mancuso N, et al. Localizing Components of Shared Transethnic Genetic Architecture of Complex Traits from GWAS Summary Data. Am J Hum Genet. 2020;106(6):805-17.

15. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014;506(7488):376-+.

16. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et

al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-65.

17. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.

18. Shi H, Burch KS, Johnson R, Freund MK, Kichaev G, Mancuso N, et al. Localizing Components of Shared Transethnic Genetic Architecture of Complex Traits from GWAS Summary Data. Am J Hum Genet. 2020;106(6):805-17.

19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.

20. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015;47(11):1228-+.

21. Dai B, Ding SL, Wahba G. Multivariate Bernoulli distribution. Bernoulli. 2013;19(4):1465-83.

22. Zhou YY, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10.

23. Hochberg Y, Benjamini Y. More Powerful Procedures for Multiple Significance Testing. Stat Med. 1990;9(7):811-8.

24. Hughes LB, Reynolds RJ, Brown EE, Kelley JM, Thomson B, Conn DL, et al. Most Common Single-Nucleotide Polymorphisms Associated With Rheumatoid Arthritis in

Persons of European Ancestry Confer Risk of Rheumatoid Arthritis in African Americans. Arthritis Rheum-Us. 2010;62(12):3547-53.

25. Campbell MC, Tishkoff SA. The Evolution of Human Genetic and Phenotypic Variation in Africa. Curr Biol. 2010;20(4):R166-R73.

26. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am J Hum Genet. 2017;100(4):635-49.

27. Goh LL, Yong MY, See WQ, Chee EYW, Lim PQ, Koh ET, et al. NLRP1, PTPN22 and PADI4 gene polymorphisms and rheumatoid arthritis in ACPA-positive Singaporean Chinese. Rheumatol Int. 2017;37(8):1295-302.

28. Lu C, Xu K, Guo H, Peng K, Yang Z, Hao YQ, et al. The relationship of PADI4_94 polymorphisms with the morbidity of rheumatoid arthritis in Caucasian and Asian populations: a meta-analysis and system review. Clin Rheumatol. 2018;37(2):289-96.

29. Chang XT, Xia YF, Pan JH, Meng QS, Zhao Y, Yan XF. PADI2 Is Significantly Associated with Rheumatoid Arthritis. Plos One. 2013;8(12).

30. Yang C, Mai HM, Peng JX, Zhou B, Hou JL, Jiang DK. STAT4: an immunoregulator contributing to diverse human diseases. Int J Biol Sci. 2020;16(9):1575-85.

31. El-Lebedy D, Raslan H, Ibrahim A, Ashmawy I, Abd El-Aziz S, Mohammed AM. Association of STAT4 rs7574865 and PTPN22 rs2476601 polymorphisms with rheumatoid arthritis and non-systemically reacting antibodies in Egyptian patients. Clin Rheumatol. 2017;36(9):1981-7.

32. Lee YH, Woo JH, Choi SJ, Ji JD, Song GG. Association between the rs7574865

polymorphism of STAT4 and rheumatoid arthritis: a meta-analysis. Rheumatol Int. 2010;30(5):661-6.

33. Muthana M, Hawtree S, Wilshaw A, Linehan E, Roberts H, Khetan S, et al. C5orf30 is a negative regulator of tissue damage in rheumatoid arthritis. P Natl Acad Sci USA. 2015;112(37):11618-23.

34. Mondal S, Thompson PR. Protein Arginine Deiminases (PADs): Biochemistry and Chemical Biology of Protein Citrullination. Accounts Chem Res. 2019;52(3):818-32.

35. Dedmon LE. The genetics of rheumatoid arthritis. Rheumatology. 2020;59(10):2661-70.

36. van der Helm-van Mil AH, Huizinga TW, Schreuder GM, Breedveld FC, de Vries RR, Toes RE. An independent role of protective HLA class II alleles in rheumatoid arthritis severity and susceptibility. Arthritis Rheum. 2005;52(9):2637-44.

37. Jiang L, Jiang DD, Han Y, Shi X, Ren CL. Association of HLA-DPB1 polymorphisms with rheumatoid arthritis: A systemic review and meta-analysis. Int J Surg. 2018;52:98-104.

38. Achour Y, Kammoun A, Ben Hamad M, Mahfoudh N, Chaabane S, Marzouk S, et al. Association study of MICA gene polymorphisms with rheumatoid arthritis susceptibility in south Tunisian population. Int J Immunogenet. 2014;41(6):486-92.

39. Kirsten H, Petit-Teixeira E, Scholz M, Hasenclever D, Hantmann H, Heider D, et al. Association of MICA with rheumatoid arthritis independent of known HLA-DRB1 risk alleles in a family-based and a case control study. Arthritis Res Ther. 2009;11(3).

40. Zhang SB, Meng TT, Li SD. Meta-analysis on the association between PADI4

polymorphisms and rheumatoid arthritis in a Chinese population. Int J Clin Exp Med. 2019;12(5):5440-6.

41. Ban JY, Park HJ, Kim SK, Kim JW, Lee YA, Choi IA, et al. Association of forkhead box J3 (FOXJ3) polymorphisms with rheumatoid arthritis. Mol Med Rep. 2013;8(4):1235-41.

42. Terao C, Ohmura K, Kochi Y, Ikari K, Okada Y, Shimizu M, et al. Anti-citrullinated peptide/protein antibody (ACPA)-negative RA shares a large proportion of susceptibility loci with ACPA-positive RA: a meta-analysis of genome-wide association study in a Japanese population. Arthritis Res Ther. 2015;17.

43. Saferding V, Bluml S. Innate immunity as the trigger of systemic autoimmune diseases. J Autoimmun. 2020;110.

44. Yoshimura A, Nishinakamura H, Matsumura Y, Hanada T. Negative regulation of cytokine signaling and immune responses by SOCS proteins. Arthritis Res Ther. 2005;7(3):100-10.

45. Kochi Y, Suzuki A, Yamada R, Yamamoto K. Ethnogenetic heterogeneity of rheumatoid arthritis-implications for pathogenesis. Nat Rev Rheumatol. 2010;6(5):290-5.

46. Alamanos Y, Drosos AA. Epidemiology of adult rheumatoid arthritis. Autoimmun Rev. 2005;4(3):130-6.

47. Ferucci ED, Templin DW, Lanier AP. Rheumatoid arthritis in American Indians and Alaska Natives: A review of the literature. Semin Arthritis Rheu. 2005;34(4):662-7.

48. Karami J, Aslani S, Jamshidi A, Garshasbi M, Mahmoudi M. Genetic implications in the pathogenesis of rheumatoid arthritis; an updated review. Gene. 2019;702:8-16.

49. Yamamoto K, Okada Y, Suzuki A, Kochi Y. Genetics of rheumatoid arthritis in Asia-present and future. Nat Rev Rheumatol. 2015;11(6):375-9.

50. Walker E, Hernandez AV, Kattan MW. Meta-analysis: Its strengths and limitations. Clev Clin J Med. 2008;75(6):431-9.

Supplementary Files

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

- Table1.pdf
- Table2.pdf
- Table3.pdf