Detecting novel micro RNAs in rheumatoid arthritis with gene-based association testing

A. Lenert[1], D.W. Fardo[2]

[1]Department of Internal Medicine, Division of Rheumatology, University of Kentucky, Kentucky Clinic J507, Lexington, USA; [2]Department of Biostatistics, College of Public Health, University of Kentucky, Lexington, USA.

Abstract

Objective

To identify novel risk genes by gene-based association analysis in rheumatoid arthritis (RA).

Methods

We performed gene-based association testing with GATES (Gene-based Association Test using Extended Simes procedure) to augment the power of genomewide-association study (GWAS) results from the largest meta-GWAS by Okada et al. in 14,361 RA cases and 43,923 controls of European ancestry using 8,694,488 SNPs.

Results

We identified 115 genes significantly associated with RA by gene-based association testing corresponding to 43 RA risk loci; 23 risk loci contained a single top risk gene, while 20 risk loci contained two or more risk genes. We replicated 39 of the genomewide significant risk loci identified by Okada et al. in Europeans with RA; we found identical top genes for 26 loci. Our gene-based testing identified 6 new top gene hits for each of the following 6 RA risk loci: RPP14 *(for* DNASE1L3-ABHD6-PXK*),* PXT1 *(for* ETV7*),* MIR5708 *(for* TPD52*),* DDX6 *(for* CXCR5*),* SUOX *(for* CDK2*), and* PCAT29 *(for* LOC145837*). We also identified a potential novel RA risk locus (11q23.3, start position 118528941 bp) which contains the following 3 genes:* TREH-PHLDB1-MIR6716*; the locus was not identified previously but may be a proxy for* CXCR5*.*

Conclusion

Through novel comprehensive gene-based association testing in >50,000 Europeans with RA using ~8 million SNPs, we confirmed prior RA risk loci and identified novel risk genes including non-coding regulatory miRNAs, providing further insight into the complex genetics of RA.

Key words

rheumatoid arthritis, gene-based association testing, genomewide association testing, risk genes, micro RNA

Aleksander Lenert, MD FRCPC
David W. Fardo, PhD

Please address correspondence to:
Dr Aleksander Lenert,
Department of Internal Medicine,
Division of Rheumatology,
University of Kentucky,
Kentucky Clinic J507,
740 South Limestone St,
Lexington, KY 40536, USA.
E-mail: ale259@uky.edu

Received on September 28, 2016; accepted in revised form on December 14, 2016.

Competing interests: none declared.

Introduction

Rheumatoid arthritis (RA) is a chronic inflammatory arthritis affecting up to 2% of individuals worldwide (1). Underlying complex genetics and immune system dysfunction with loss of self-tolerance and auto-antibody production against joints and other tissues lead to RA (2). The genetic susceptibility to RA has been established via traditional genetics studies, with the shared epitope (SE), *HLA-DRB1*, considered to be the most important susceptibility factor for RA (3). However the HLA locus does not explain the entire phenotypic variance, *i.e.* heritability, observed in RA (4). The genome-wide association study (GWAS) approach using single nucleotide polymorphisms (SNP) as the units of analysis has shed light into the polygenic inheritance of RA beyond the HLA region (2, 5-7). Despite the successful identification of 100 non-HLA risk loci in RA, several challenges of GWAS findings remain: 1) the detected variants explain only a small proportion of heritability; 2) the majority of identified SNPs are not causal but in linkage with variants with true effects; and 3) the contribution of non-genome-wide significant variants is ignored due to stringent statistical thresholds (8, 9).

Gene-based analysis can improve on prior GWAS results and translate knowledge from single SNP hits to significant genes in RA. Gene-based association tests (GBAT) evaluate the association between a gene and a trait of interest by means of aggregation. The advantages of using the gene as the basic unit of analysis have been previously highlighted (10). Genes are functional units of the human genome across different populations and unlike SNPs are not influenced by allele frequencies, linkage disequilibrium (LD) structure and heterogeneity across populations (10). Correcting for multiple-testing is lower given that one accounts for approximately 25,000 genes in the human genome rather than millions of SNPs compared in a typical GWAS (10). GBAT methods include the combined multivariate and collapsing method, kernel-based adaptive cluster, versatile gene-based association study and sequence kernel association test (11, 12). The main disadvantages of these regression-based methods are the requirement for the individual raw phenotype and genotype data, as well as computational demands of permutation with genomewide data for successful analysis (10).

The GATES (gene-based association test with extended Simes procedure) approach has shown promise for rapid gene-based analysis (10). GATES assesses the gene-level statistical association based on millions of SNPs, by combining the *p*-values of SNPs within a gene without requiring raw individual genotypes, to produce gene-based *p*-values for the trait of interest (10). The GATES approach has been successfully employed for post-GWAS analysis in systemic lupus erythematosus and osteoporosis (13, 14).

Despite tremendous successes of GWAS in RA, novel approaches for post-GWAS data analysis are needed in order to identify candidate risk genes for functional exploration and drug targeting. Our hypothesis is that regulatory genetic elements, specifically non-coding miRNAs, are associated with seropositive RA patients compared with controls. To achieve our aim, we performed gene-based association testing in RA patients of European ancestry using over 8 million SNP variants from the largest trans-ethnic meta-GWAS (15).

Materials and methods

Data source and subjects

Genomewide SNP data was obtained from the results of the publically available meta-GWAS in RA (15). Study design, characteristics of subjects, genotyping, quality control and SNP-based association analysis were described in the original publication (15). The authors of the meta-GWAS obtained written informed consent from all of the participants, and the study was approved by the relevant ethics committees of the institutions involved in compliance with the Declaration of Helsinki (15). In brief, this meta-GWAS was performed in >100,000 subjects of European and Asian ancestries by evaluating ~10 million SNPs, and revealed 42 novel RA risk loci. For our study, the dataset consisted of 14,361 RA cases and 43,923 controls with 8,694,488

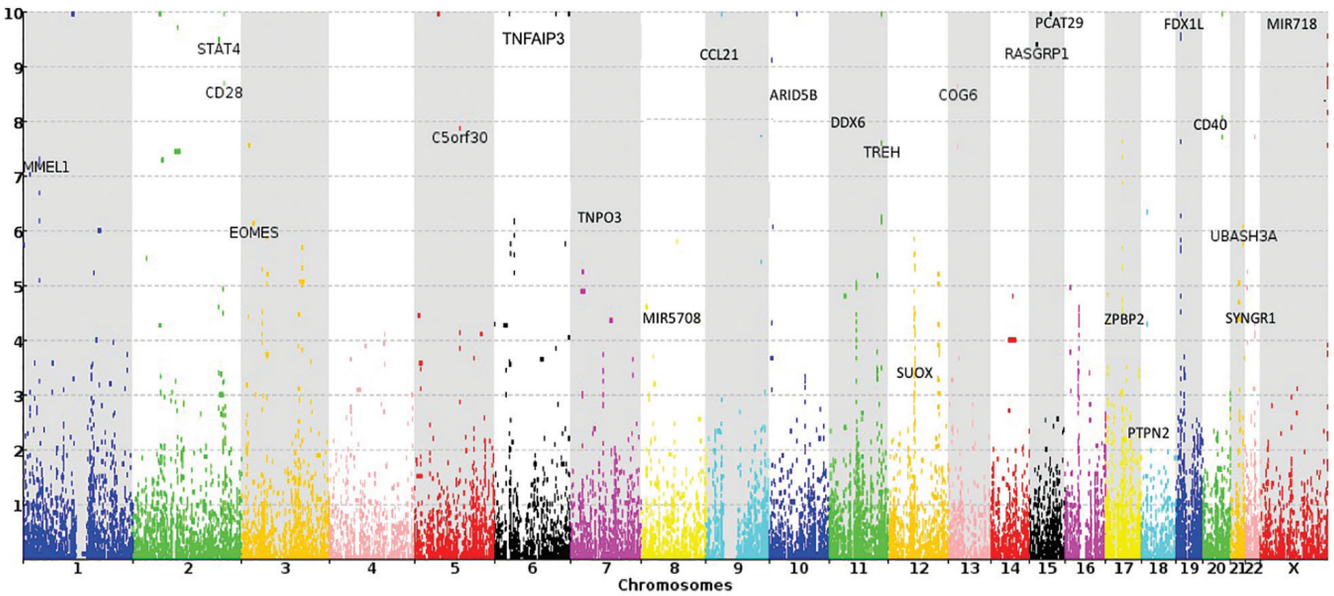

Fig. 1. Manhattan plot of significant genes from gene-based analysis in Europeans with rheumatoid arthritis.

genotype and imputed SNPs excluding the extended MHC region [chromosome 6, 25-33 megabases (Mb)] from stage I of the GWAS in Europeans from 18 international cohorts of patients with seropositive RA (positive for rheumatoid factor or ACPA or both). All RA cases fulfilled the 1987 American College of Rheumatology classification criteria or were diagnosed by a rheumatologist (16).

Gene-based association testing

Gene-based association analysis was performed using the GATES method (Gene-based Association Test using Extended Simes procedure) implemented in KGG4, a systematic biological Knowledge-based mining system for Genome-wide Genetic studies version 4 (8, 10). This powerful method does not require the raw genotype and phenotype data as inputs. GATES assumes that the test of association for the phenotype of interest and each SNP has been carried out with resulting SNP-based p-values and pair-wise correlation coefficients for all SNPs. It can rapidly combine the *p*-values of SNPs within a gene to obtain an overall *p*-value for association of the entire gene with the phenotype of interest (10). The null hypothesis of this gene-based test is that no SNP within the gene is associated with the phenotype of interest, while the alternative hypothesis is that at least one SNP in the gene is associated with the phenotype of interest. For gene-based association analysis, the data file for cases (RA) and controls contained the following 4 input variables: SNP ID (with rs number), chromosome, position (based on hg19 or NCBI build 37) and SNP *p*-value. Genes were defined as ± 5kb from the 5' and 3' untranslated-regions (UTRs), respectively. We accounted for LD between SNPs by using information from 1000 Genomes Project Phase I for Europeans (17). Genomic control was calculated by median of Chi-square statistic. Bonferroni correction was used for multiple testing (18). LocusZoom was used for visual interpretation of significant genes and risk loci (19).

We viewed significant genes hits in the UCSC Genome Browser including their physical location and surrounding genes (20). Genes were assigned to risk loci based on their locations. We viewed significant risk loci in LocusZoom to identify the most significant regional SNP. We identified SNPs in high LD (r^2 >0.7) with SNP Annotation and Proxy Search (SNAP) (21). We compared our gene-based association results with published RA risk loci (15, 22).

Results

Gene-based association analysis with GATES was performed with a total of 8,694,488 SNPs (excluding the extended HLA region on chr.6 ~25-33 Mb). SNPs were assigned to genes (defined as ±5 Kb from 5' and 3' UTRs) to yield gene-based *p*-values with significance threshold of 1.96×10^{-6} ($p<0.05$, Bonferroni corrected). A total of 25,539 genes were analysed. Out of all SNPs, 51.1% SNPs were located within genes and 48.9% were located outside genes. We identified 115 genes significantly associated with RA by gene-based association testing, with selected top genes highlighted in the Manhattan plot (Fig. 1). We identified several miRNAs significantly associated with RA patients compared with controls, including *MIR5708*, *MIR6716*, *MIR4305*, *MIR718* and *MIR3202*. The majority of our top genes were located on chromosomes 6, 1, 2 and 19 (22.6%, 20%, 7.8% and 7.8%, respectively).

From our gene-based analysis, the most significant 115 genes were assigned to 43 RA risk loci (Tables I and II). Out of these, 23 risk loci contained a single top risk gene, while 20 risk loci contained two or more risk genes (Tables I and II). Hence our analysis identified 43 RA risk loci and replicated 39 of the genomewide significant risk loci identified in Europeans with RA. In the Stage 1 meta-GWAS in Europeans by Okada *et al*., 44 RA risk loci (36 known and 8 novel) were identified at a genomewide significance ($p<5 \times 10^{-8}$) and an additional 31 RA risk loci had a nominal

Table I. Top gene hits (chromosomes 1 to 6) in rheumatoid arthritis from the gene-based association analysis.

Chr	RA risk locus	Top gene	Gene *p*-value (corrected)	Genes in locus	Additional Genes in Locus (ordered by gene-based *p*-value)
1	*MMEL1*	*FAM213B*	4.06×10^{-4}	6	*LOC100996583, MMEL1, LOC115110, TNFRSF14, TTC34*
	PADI4	*PADI4*	2.41×10^{-3}	1	-
	MTF1-INPP5B	*C1orf122*	1.22×10^{-3}	4	*YRDC, MTF1, MANEAL*
	PTPN22	*RSBN1*	4.52×10^{-144}	11	*PHTF1, PTPN22, AP4B1-AS1, MAGI3, BCL2L15, DCLRE1B, AP4B1, HIPK1-AS1, HIPK1, OLFML3*
	LOC100506023	*LOC100506023*	2.55×10^{-2}	1	-
2	*REL*	*LINC01185*	2.43×10^{-8}	3	*REL, PUS10*
	SPRED2	*SPRED2*	1.31×10^{-3}	1	-
	AFF3	*LINC01104*	5.08×10^{-6}	2	*AFF3*
	STAT4	*STAT4*	8.42×10^{-6}	1	-
	CD28	*CD28*	5.33×10^{-5}	1	-
	CTLA4	*CTLA4*	3.47×10^{-14}	1	-
3	*PLCL2*	*PLCL2*	6.94×10^{-4}	1	-
	EOMES	*EOMES*	1.92×10^{-2}	1	-
	DNASE1L3-ABHD6-PXK	*RPP14*	3.44×10^{-2}	1	-
5	*ANKRD55*	*ANKRD55*	3.60×10^{-16}	1	-
	C5orf30	*C5orf30*	3.50×10^{-4}	1	-
6	*HLA-DRB1*	*ZBTB9**	2.65×10^{-87}	17*	*PHF1, KIFC1, BAK1, CUTA, SYNGAP1, ITPR3, IP6K3, LINC00336, MIR5004, UQCC2, MIR3934, LEMD2, MLN, LINC01016, LOC101929188, GGNBP1*
	ETV7	*PXT1*	4.47×10^{-2}	1	-
	NFKBIE	*NFKBIE*	1.67×10^{-2}	3	*TMEM151B, SLC35B2*
	TNFAIP3	*TNFAIP3*	3.19×10^{-11}	3	*LOC100130476, LOC102723649*
	TAGAP	*TAGAP*	4.57×10^{-2}	1	-
	CCR6	*CCR6*	8.60×10^{-9}	1	-

*These genes are located in the HLA region which is driving overall significance due to high linkage disequilibrium. RA risk locus as defined by Okada (15). Chr: chromosome; RA: rheumatoid arthritis.

level of significance ($5 \times 10^{-6} > p > 5 \times 10^{-8}$) (15). Further, the most significant gene was identical to Okada's top gene for 26 out of 43 loci.

The remaining 17 out of 43 RA risk loci contained a different top gene compared to Okada. From these 17 RA risk loci with discordant top genes, 10 loci contained multiple genes per locus (mean 5.6 genes/locus) and included Okada's top gene at each locus. However, for these loci, gene-based testing identified a different top gene based on a lower gene-based *p*-value.

Out of the 7 remaining risk loci, 6 contained a single top gene which was different from Okada's top gene for these loci. Our gene-based testing identified 6 new top gene hits for each of the following 6 Okada risk loci: *RPP14* (for *DNASE1L3-ABHD6-PXK*), *PXT1* (for *ETV7*), *MIR5708* (for *TPD52*), *DDX6* (for *CXCR5*), *SUOX* (for *CDK2*), and *PCAT29* (for *LOC145837*). The remaining locus identified by gene-based testing represents a potential novel RA risk locus located on chromosome 11 (start position 118528941 bp) which contains the following 3 genes: *TREH-PHLDB1-MIR6716*. This locus was not identified in Stage 1 GWAS in Europeans (15) but may be a proxy for the *CXCR5* finding.

Discussion

Our gene-based testing revealed 115 significant gene hits using one of the richest resources for genetics analysis in RA (15). The genes with the most significant gene-based *p*-values were: *PTPN22* (4.52×10^{-144}), *HLA-DRB1* (2.65×10^{-87}), *ANKRD55* (3.60×10^{-16}), *CTLA4* (3.47×10^{-14}), *TNFAIP3* (3.19×10^{-11}), *CCR6* (8.60×10^{-9}), *CXCR5* (4.06×10^{-9}), *CCL21* (6.66×10^{-8}) and *REL* (2.43×10^{-8}). These genes have confirmed associations in Europeans with RA (7, 23-27).

The most significant RA risk genes were located predominantly on chromosomes 6 (22.6%), 1 (20%), 2 (7.8%) and 19 (7.8%). This is consistent with prior findings which have established *HLA-DRB1* on chromosome 6 as the most significant genetic contributor to RA susceptibility (4, 7). Given the complex polygenetic inheritance of RA, non-HLA genes are hypothesized to contribute additional susceptibility to RA (7). The most significant non-HLA gene contributors include *PTPN22* and *PADI4* (on chromosome 1), *CTLA4* and *STAT4* (on chromosome 2) and *TYK2* (on chromosome 19). *PTPN22* has a regulatory role in B and T cell function and is considered as the archetypal non-HLA autoimmunity risk gene given its association with multiple autoimmune diseases, including RA, systemic lupus erythematosus, juvenile idiopathic arthritis, type 1 diabetes mellitus and Hashimoto's thyroiditis (28). In RA, specific *PTPN22* SNPs have been associated with increased risk of disease, positive auto-antibody status with ACPA production and prediction of disease onset (28). Additionally, therapeutic targeting of *CTLA4* with Abatacept, and *STAT4-TYK2* with Tofacitinib has been successfully employed in clinical practice in RA patients failing non-bio-

Table II. Top gene hits (chromosomes 7 to X) in rheumatoid arthritis from the gene-based association analysis.

Chr	RA risk locus	Top gene	Gene *p*-value (corrected)	Genes in locus	Additional Genes in Locus (ordered by gene-based *p*-value)
7	*IRF5*	*TNPO3*	6.09×10^{-6}	3	*IRF5, TPI1P2*
8	*TPD52*	*MIR5708*	3.98×10^{-2}	1	-
9	*CCL19-CCL21*	*CCL21*	6.66×10^{-8}	2	*FAM205A*
	TRAF1-C5	*PHF19*	4.47×10^{-3}	2	*TRAF1*
10	*IL2RA*	*IL2RA*	1.93×10^{-5}	1	-
	GATA3	*GATA3*	7.53×10^{-5}	2	*AS1*
	ARID5B	*ARID5B*	8.04×10^{-7}	1	-
11	*CXCR5*	*DDX6*	4.07×10^{-9}	1	-
	-	*TREH*	6.53×10^{-4}	3	*PHLDB1, MIR6716*
12	*CDK2*	*SUOX*	3.60×10^{-2}	1	-
13	*COG6*	*COG6*	5.71×10^{-5}	2	*MIR4305*
15	*RASGRP1*	*RASGRP1*	9.95×10^{-6}	1	-
	LOC145837	*PCAT29*	3.04×10^{-11}	1	-
17	*IKZF3-CSF3*	*ZPBP2*	6.23×10^{-4}	4	*IKZF3, GSDMB, GRB7*
18	*PTPN2*	*PTPN2*	1.17×10^{-2}	1	-
19	*TYK2*	*FDX1L*	1.67×10^{-10}	7	*TYK2, RAVER1, ICAM1, ICAM4, ICAM5, ICAM3, ILF3, AP1M2*
	ILF3	*ILF3*	6.10×10^{-4}	2	*AP1M2*
20	*CD40*	*CD40*	3.32×10^{-8}	3	*NCOA5, SLC12A5*
21	*UBASH3A*	*UBASH3A*	2.17×10^{-2}	2	*TMPRSS3*
22	*SYNGR1*	*SYNGR1*	4.87×10^{-4}	1	-
X	*IRAK1*	*MIR718*	5.31×10^{-7}	11	*IRAK1, MECP2, MIR3202-2, MIR3202-1, TMEM187, NAA10, RENBP, HCFC1, ARHGAP4, HCFC1-AS1*

RA risk locus as defined by Okada (15). Chr: chromosome; RA: rheumatoid arthritis.

logic DMARDs and anti-TNF therapy, thus establishing a link between genetic risk association, pathways and drug targeting (15).

Our top 115 gene-based genes were assigned to 43 RA risk loci with 23 loci containing a single top risk gene and 20 loci contained two or more risk genes (Tables I and II). We confirmed 39 out of 44 genomewide risk loci identified in the meta-GWAS of Europeans with RA (15). In 26 RA loci, we identified the same top gene by gene-based testing as in the Okada's GWAS. For example, this was observed for *PADI4*, *STAT4*, *CTLA4*, *CCR6*, *IL2RA* and *GATA3* (Tables I and II). In 17 RA loci, we identified a different top gene compared with the meta-GWAS results (Tables I and II). Ten RA risk loci contained multiple genes in the region with a mean of 5.6 genes/locus, and included the top gene from the meta-GWAS. For example, for the RA risk locus *PTPN22*, our results were significant for 11 gene hits in this region, with the most significant genes in order *RSBN1*, *PHTF1* and *PTPN22*. Gene-dense regions contain SNPs which may be shared by adjacent genes, given variable gene length definitions, can contribute to the gene-based *p*-value of the highest risk gene in the region (*i.e. PTPN22*) as well as surrounding genes (*i.e. RSBN1* and *PHTF1*). Additionally, for 6 RA risk loci, we identified a different top gene compared to the meta-GWAS (Tables I and II). For example, for the novel RA risk locus *ETV7*, our gene-based testing identified *PXT1* as the top hit gene. It is hypothesised that these newly identified genes could represent additional RA risk genes, since these three RA risk loci (*ETV7*, *TPD52* and *LOC145837*) are novel and require further validation in independent cohorts. Our gene-based testing also identified a potential novel RA risk locus located on chromosome 11 (start position 118528941 bp) which contains the following 3 genes: *TREH-PHLDB1-MIR6716*. Despite robust association results from large-scale GWAS in RA, none of the identified risk SNPs and genes have been shown to be causal in RA. Hence, it is imperative to consider additional coding and non-coding regulatory genetic elements for further functional studies to link genetics with aberrant biological mechanisms in RA.

To this extent, our study revealed gene-based associations of non-coding genetic elements, specifically miRNAs, in RA patients. We identified several miRNAs significantly associated with RA patients compared with controls, including *MIR5708*, *MIR6716*, *MIR4305*, *MIR718* and *MIR3202*. MiRNAs are short non-coding RNAs, with an average length of 22 nucleotides, involved in the regulation of gene expression by binding to messenger RNA (mRNA) and affecting its stability and translation (29, 30). It is estimated that miRNAs constitute 1–2% of the human genome and potentially regulate 30% of protein-coding genes, mainly as negative regulators of expression of their target genes (31). MiRNAs show

very high stability due to complexing with proteins or their packaging within microvesicles, and thus are thought to be promising candidate biomarkers in autoimmune diseases such as systemic lupus erythematosus and RA (29, 32, 33). Additionally, miRNAs are essential for the development, differentiation and function of immune cells, hence influencing immunopathological processes in autoimmune diseases (34, 35). In RA, the most extensively studied miRNAs include *mir-16* which targets the 3' UTR of TNF-α and regulates its signalling; *mir-21* which targets *STAT3* and may be involved in regulating the regulatory T cell (Treg) and Th17 pathways; *mir-22* which targets the 3' UTR of *CYR61* inhibiting its expression, thus leading to increased IL-6 expression and synovial tissue hyperplasia; *mir-146a*/*mir-146b* which inhibit T-helper cell responses and suppress Tregs; and *mir-155* which is essential for homeostasis of Tregs (29). Despite the discovery of differentially expressed miRNAs in RA patients compared with healthy subjects and osteoarthritis, our understanding of the exact role of miRNA in RA remains unknown.

MIR5708, located on chromosome 8 at RA risk locus *TPD52*, was significantly associated with RA in our study. This is a novel association with RA which has not been previously reported; however, *mir-5708* has shown promising results in cancer studies. Recently *mir-5708* was identified as a novel putative miRNA associated with metastatic prostate cancer (36, 37). Watahiki *et al*. used next generation sequencing in a transplantable metastatic compared with a non-metastatic prostate cancer xenograft line to successfully identify 36 novel differentially expressed miRNAs. They identified differentially expressed miRNAs, including *mir-5708*, which could serve as biomarkers and future drug targets for metastatic prostate cancer (37). *MIR5708* targets *YWHAE*, a member of the 14-3-3 family of proteins which are novel joint-derived pro-inflammatory mediators implicated in RA pathogenesis and useful for earlier diagnosis in the clinic (38, 39).

MIR6716, located on chromosome 11 (start position 118528941 bp), was significantly associated with RA in our study which has not been reported previously. Li *et al*. systematically investigated miRNAs in healthy human epididymis and identified 18 novel miRNA genes which will serve as a database for further functional studies in the male reproductive system (40). This miRNA has not been studied in autoimmune disease. *MIR6716* targets *TCF7L2*, a transcription factor involved in the Wnt signalling pathway (39). Hence, *MIR6716* may play a role in modulating this key signalling pathway involved in fibroblast-like synoviocyte activation, bone resorption and joint destruction in RA (41). Similarly, *MIR4305*, located on chromosome 13 at RA risk locus *COG6*, was associated with RA and represents a novel association with RA. There are currently no published reports of its association with a specific disease phenotype. Goff *et al*. performed an exploratory study to characterize miRNAs including *mir-4305* in human embryonic stem cells (42).

MIR718 and *MIR3202*, located on chromosome X at RA risk locus *IRAK1*, were both associated with RA in our gene-based study. There are no reported associations of neither *mir-718* nor *mir-3202* with autoimmune disease; however, promising results in cancer are emerging for the role of these miRNAs. Schrauder *et al*. preformed microarray-based miRNA profiling in serum of early breast cancer patients compared with healthy controls; they found a total of 59 differentially expressed miRNAs and validated 2 miRNAs, *mir-202* and *mir-718*, in an independent cohort as potential biomarkers for detection of early breast cancer (43). Xue *et al*. identified *mir-718* as a key inhibitor of PTEN expression in an animal model of Kaposi's sarcoma in HIV, leading to activation of the AKT/mTOR signalling pathway in tumourigenesis, thus demonstrating its importance as a promising therapeutic target (44). The mTOR pathway is also targeted by immunosuppressive agents used for treatment of autoimmune diseases including RA (15). *Mir-718* has also been shown to directly target and suppress VEGF, which is an important driver of angiogenesis in tumours such as ovarian cancer (45). Stark *et al*. undertook a comprehensive analysis of miRNAs in normal pigmented and melanoma cells to detect novel miRNAs for earlier detection of distant metastatic disease (46). As suggested by the findings of the role of *mir-718*s in cancer pathogenesis, this non-coding regulatory gene has a promising role in RA given its regulatory role in mTOR signalling and VEGF-induced angiogenesis which are key drivers of break in immune tolerance, pro-inflammatory cascade and synovial hypertrophy in RA.

Our study focused on seropositive RA patients of European ancestry and hence its results may not be applicable to non-Europeans consistent with established findings of ancestry-specific GWAS associations in RA (47). Despite rigorous quality control and conservative correction for multiple statistical testing, some of our gene-based associations may represent false positive findings, and will need confirmation in other independent cohorts and across different ancestries.

Conclusion

Through comprehensive and novel gene-based association testing with >50,000 individuals and ~8 million SNPs, we confirmed prior risk loci and identified novel miRNAs associations with seropositive RA of European ancestry. In addition to protein-coding genes, novel regulatory elements such as miRNAs show promise for understanding onset of disease, break in immune tolerance and upregulation of pro-inflammatory pathways at play in RA. MiRNAs are non-coding genetic elements with a promising role as future biomarkers and novel drug targets in RA.

Acknowledgements

We would like to thank Okada *et al*. for the publically available data used in our study. This work was supported in part by an NIH grant K25AG043546 (DWF).

References

1. SPECTOR TD: Rheumatoid arthritis. *Rheum Dis Clin North Am* 1990; 16: 513-37.
2. PERRICONE C, CECCARELLI F, VALESINI G: An overview on the genetic of rheumatoid arthritis: a never-ending story. *Autoimmun Rev* 2011; 10: 599-608.

3. MACGREGOR AJ, SNIEDER H, RIGBY AS *et al*.: Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 2000; 43: 30-7.
4. GREGERSEN PK, SILVER J, WINCHESTER RJ: The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 1987; 30: 1205-13.
5. VIATTE S, PLANT D, RAYCHAUDHURI S: Genetics and epigenetics of rheumatoid arthritis. *Nat Rev Rheumatol* 2013; 9: 141-53.
6. RAYCHAUDHURI S: Recent advances in the genetics of rheumatoid arthritis. *Curr Opin Rheumatol* 2010; 22: 109-18.
7. 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.
8. LI MX, SHAM PC, CHERNY SS, SONG YQ: A knowledge-based weighting framework to boost the power of genome-wide association studies. *PLoS One* 2010; 5: e14480.
9. YARWOOD A, HUIZINGA TW, WORTHINGTON J: The genetics of rheumatoid arthritis: risk and protection in different stages of the evolution of RA. *Rheumatology* (Oxford) 2016; 55: 199-209.
10. LI MX, GUI HS, KWAN JS, SHAM PC: GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am J Hum Genet* 2011; 88: 283-93.
11. LIU JZ, MCRAE AF, NYHOLT DR *et al*.: A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* 2010; 87: 139-45.
12. LEE S, ABECASIS GR, BOEHNKE M, LIN X: Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 2014; 95: 5-23.
13. LEI SF, DENG FY: Identification of susceptibility genes for systemic lupus erythematosus with a genome-wide gene-based association study. *Scand J Rheumatol* 2014; 43: 426-8.
14. MO XB, LU X, ZHANG YH, ZHANG ZL, DENG FY, LEI SF: Gene-based association analysis identified novel genes associated with bone mineral density. *PLoS One* 2015; 10: e0121811.
15. OKADA Y, WU D, TRYNKA G *et al*.: Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014; 506: 376-81.
16. ARNETT FC, EDWORTHY SM, BLOCH DA *et al*.: The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988; 31: 315-24.
17. ABECASIS GR, AUTON A, BROOKS LD *et al*.: An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; 491: 56-65.
18. TARONE RE: A modified Bonferroni method for discrete data. *Biometrics* 1990; 46: 515-22.
19. PRUIM RJ, WELCH RP, SANNA S *et al*.: LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010; 26: 2336-7.
20. KENT WJ, SUGNET CW, FUREY TS *et al*.: The human genome browser at UCSC. *Genome Res* 2002; 12: 996-1006.
21. JOHNSON AD, HANDSAKER RE, PULIT SL, NIZZARI MM, O'DONNELL CJ, DE BAKKER PI: SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008; 24: 2938-9.
22. WELTER D, MACARTHUR J, MORALES J *et al*.: The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014; 42: D1001-6.
23. WTCCC T: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; 447: 661-78.
24. PLENGE RM, COTSAPAS C, DAVIES L *et al*.: Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 2007; 39: 1477-82.
25. KOCHI Y, OKADA Y, SUZUKI A *et al*.: A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. *Nat Genet* 2010; 42: 515-9.
26. ZHERNAKOVA A, STAHL EA, TRYNKA G *et al*.: Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet* 2011; 7: e1002004.
27. GREGERSEN PK, AMOS CI, LEE AT *et al*.: REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat Genet* 2009; 41: 820-3.
28. STANFORD SM, BOTTINI N: PTPN22: the archetypal non-HLA autoimmunity gene. *Nat Rev Rheumatol* 2014; 10: 602-11.
29. CHUROV AV, OLEINIK EK, KNIP M: MicroRNAs in rheumatoid arthritis: altered expression and diagnostic potential. A*utoimmun Rev* 2015; 14: 1029-37.
30. VICENTE R, NOEL D, PERS YM, APPARAILLY F, JORGENSEN C: Deregulation and therapeutic potential of microRNAs in arthritic diseases. *Nat Rev Rheumatol* 2016; 12: 211-20.
31. FURER V, GREENBERG JD, ATTUR M, ABRAMSON SB, PILLINGER MH: The role of microRNA in rheumatoid arthritis and other autoimmune diseases. *Clin Immunol* 2010; 136: 1-15.
32. GHODKE-PURANIK Y, NIEWOLD TB: Immunogenetics of systemic lupus erythematosus: A comprehensive review. *J Autoimmun* 2015; 64: 125-36.
33. STYPINSKA B, PARADOWSKA-GORYCKA A: Cytokines and microRNAs as candidate biomarkers for systemic lupus erythematosus. *Int J Mol Sci* 2015; 16: 24194-218.
34. CHEN JQ, PAPP G, SZODORAY P, ZEHER M: The role of microRNAs in the pathogenesis of autoimmune diseases. *Autoimmun Rev* 2016; 15: 1171-80.
35. HUSAKOVA M: MicroRNAs in the key events of systemic lupus erythematosus pathogenesis. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 2016; 160: 327-42.
36. GRIFFITHS-JONES S, GROCOCK RJ, VAN DONGEN S, BATEMAN A, ENRIGHT AJ: miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 2006; 34: D140-4.
37. WATAHIKI A, WANG Y, MORRIS J *et al*.: MicroRNAs associated with metastatic prostate cancer. *PLoS One* 2011; 6: e24950.
38. MAKSYMOWYCH WP, BOIRE G, VAN SCHAARDENBURG D *et al*.: 14-3-3eta autoantibodies: diagnostic use in early rheumatoid arthritis. *J Rheumatol* 2015; 42: 1587-94.
39. WONG N, WANG X: miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res* 2015; 43: D146-52.
40. LI Y, WANG HY, WAN FC *et al*.: Deep sequencing analysis of small non-coding RNAs reveals the diversity of microRNAs and piRNAs in the human epididymis. *Gene* 2012; 497: 330-5.
41. MIAO CG, YANG YY, HE X *et al*.: Wnt signaling pathway in rheumatoid arthritis, with special emphasis on the different roles in synovial inflammation and bone remodeling. *Cell Signal* 2013; 25: 2069-78.
42. GOFF LA, DAVILA J, SWERDEL MR *et al*.: Ago2 immunoprecipitation identifies predicted microRNAs in human embryonic stem cells and neural precursors. *PLoS One* 2009; 4: e7192.
43. SCHRAUDER MG, STRICK R, SCHULZ-WENDTLAND R *et al*.: Circulating microRNAs as potential blood-based markers for early stage breast cancer detection. *PLoS One* 2012; 7: e29770.
44. XUE M, YAO S, HU M *et al*.: HIV-1 Nef and KSHV oncogene K1 synergistically promote angiogenesis by inducing cellular miR-718 to regulate the PTEN/AKT/mTOR signaling pathway. *Nucleic Acids Res* 2014; 42: 9862-79.
45. LENG R, ZHA L, TANG L: MiR-718 represses VEGF and inhibits ovarian cancer cell progression. *FEBS Lett* 2014; 588: 2078-86.
46. STARK MS, TYAGI S, NANCARROW DJ *et al*.: Characterization of the Melanoma miRNAome by Deep Sequencing. *PLoS One* 2010; 5: e9685.
47. YAMAMOTO K, OKADA Y, SUZUKI A, KOCHI Y: Genetics of rheumatoid arthritis in Asia - present and future. *Nat Rev Rheumatol* 2015; 11: 375-9.