



■ RESEARCH

Genome-wide association studies and gene expression profiles of rheumatoid arthritis

AN ANALYSIS

**X. Xiao,
J. Hao,
Y. Wen,
W. Wang,
X. Guo,
F. Zhang**

*Xi'an Jiaotong
University, Shaanxi,
China*

■ X. Xiao, Graduate Student, Key Laboratory of Trace Elements and Endemic Diseases of National Health and Family Planning Commission,

■ J. Hao, Graduate Student, Key Laboratory of Trace Elements and Endemic Diseases of National Health and Family Planning Commission,

■ Y. Wen, Graduate Student, Key Laboratory of Trace Elements and Endemic Diseases of National Health and Family Planning Commission,

■ W. Wang, Graduate Student, Key Laboratory of Trace Elements and Endemic Diseases of National Health and Family Planning Commission,

■ X. Guo, BA, Professor, Key Laboratory of Trace Elements and Endemic Diseases of National Health and Family Planning Commission,

■ F. Zhang, PhD, Associate Professor, 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, Yanta West Road 76, Xi'an, Shaanxi, China.

Correspondence should be sent to F. Zhang; email: fzhxjtu@mail.xjtu.edu.cn

doi:10.1302/2046-3758.57.2000502

Bone Joint Res 2016;5:314–319.

Received: 16 July 2015;

Accepted: 7 June 2016

Objectives

The molecular mechanism of rheumatoid arthritis (RA) remains elusive. We conducted a protein-protein interaction network-based integrative analysis of genome-wide association studies (GWAS) and gene expression profiles of RA.

Methods

We first performed a dense search of RA-associated gene modules by integrating a large GWAS meta-analysis dataset (containing 5539 RA patients and 20 169 healthy controls), protein interaction network and gene expression profiles of RA synovium and peripheral blood mononuclear cells (PBMCs). Gene ontology (GO) enrichment analysis was conducted by DAVID. The protein association networks of gene modules were generated by STRING.

Results

For RA synovium, the top-ranked gene module is HLA-A, containing TAP2, HLA-A, HLA-C, TAPBP and LILRB1 genes. For RA PBMCs, the top-ranked gene module is GRB7, consisting of HLA-DRB5, HLA-DRA, GRB7, CD63 and KIT genes. Functional enrichment analysis identified three significant GO terms for RA synovium, including antigen processing and presentation of peptide antigen via major histocompatibility complex class I (false discovery rate (FDR) = 4.86×10^{-4}), antigen processing and presentation of peptide antigen (FDR = 2.33×10^{-3}) and eukaryotic translation initiation factor 4F complex (FDR = 2.52×10^{-2}).

Conclusion

This study reported several RA-associated gene modules and their functional association networks.

Cite this article: *Bone Joint Res* 2016;5:314–319

Keywords: Rheumatoid arthritis; Genome-wide association studies; Microarray; Gene modules; Protein interaction network

Article focus

■ The molecular mechanism of rheumatoid arthritis (RA) remains elusive. Integrating the information of genome-wide association studies (GWAS), microarray and protein interaction networks can provide a novel insight into the mechanism of RA.

Key messages

- dmGWAS analysis identified several RA-associated gene modules, mainly including HLA-A and GRB7.
- Enrichment analysis detected three significant gene ontology terms for RA functionally involved in antigen processing and presentation.

Strengths and limitations

■ Our study results may help to reveal the pathogenesis of RA. Further studies are needed to confirm our findings.

Introduction

Rheumatoid arthritis (RA) is one of the most common systemic autoimmune chronic inflammatory disorders. It is characterised by progressive joint destruction and multi-system inflammation, causing joint pain, swelling, stiffness and deformities. RA affects more than 1% of the population worldwide,¹ and results in a reduction of lifespan of RA patients of about ten years.² Due to its high morbidity and mortality, RA

has a significant societal impact in terms of medical cost and lost productivity.

Benefiting from the rapid development of high-throughput omics technologies, researchers are able to monitor hundreds of thousands of genes and proteins involved in the development of RA. For instance, in the last three years more than ten genome-wide association studies (GWAS) have been conducted and have reported a set of RA-associated genes.³ But due to the analysing strategy of focusing on the most significant loci, the susceptibility genes identified by GWAS are mostly limited and explain a small proportion of the genetic risk of RA. At the RNA level, extensive gene expression profile studies have been performed and have identified some important differentially expressed genes implicated in the development of RA.^{4,5} However, the rich pathogenetic information contained in the abundant omics data was not fully used, limiting our efforts to clarify the elusive molecular mechanism and develop effective treatments for RA.

As the pathogenesis of RA is complicated, the traditional single-omic study approach is usually insufficient for fully unraveling the functional principles and dynamics of the global biological process of complex diseases. Multi-omics integrative studies have gained more and more attention in pathogenetic studies.⁶⁻¹⁰ These studies are expected to provide more pathogenetic information and could potentially reduce the impact of data error and statistical bias on study results. Recently, the protein-protein interaction (PPI) network was introduced into omics data analysis, which could also enhance the identification of disease genes and help us to understand their relationships.^{6,10,11}

In this study, we conducted a PPI network-based integrative analysis of GWAS and gene expression profiles of RA. Briefly, we first performed a dense search of RA-associated gene modules by integrative analysis of a large GWAS meta-analysis dataset (containing 5539 RA patients and 20 169 healthy controls) and gene expression profiles of RA synovium and peripheral blood mononuclear cells (PBMC). The biological function of the identified gene modules was then investigated by gene ontology (GO) enrichment analysis and protein-association network analysis. Our study identified several RA-associated gene modules and their functional association networks in the pathogenesis of RA. We also illustrated the performance of multi-omics data integration and mining for RA pathogenetic studies.

Materials and Methods

GWAS datasets. We used a large GWAS meta-analysis dataset involving 5539 autoantibody-positive RA patients and 20 169 healthy controls of European ancestry.¹² This large sample consisted of: 483 RA cases and 1449 healthy controls (genotyped by Affymetrix 6.0; Affymetrix, Santa Clara, California) from the Brigham Rheumatoid Arthritis Sequential Study; 589 RA cases and 1472 healthy controls (genotyped by Illumina 370K; Illumina, San Diego,

California) from the CANADA RA study; 1173 RA cases and 1089 healthy controls (genotyped by Illumina 317K; Illumina) from the Epidemiological Investigation of Rheumatoid Arthritis; 1769 RA cases and 5551 healthy controls (genotyped by Illumina 317K and Illumina 370K; Illumina) from the North American Rheumatoid Arthritis Consortium I and III; 1525 RA cases and 10 608 healthy controls (genotyped by Affymetrix 500K; Affymetrix) from the Wellcome Trust Case Control Consortium (WTCCC). RA was diagnosed according to the 1987 American College of Rheumatology criteria for diagnosis of RA or by board-certified experts of rheumatology.

All study samples passed a genetic relatedness check based on pairwise identity by state testing.¹² Genetically related subjects were removed. The single nucleotide polymorphisms (SNPs) with call rates < 95.00% (< 99% for WTCCC GWAS data), minor allele frequencies < 0.01 or Hardy-Weinberg Equilibrium testing p -values < 10^{-6} (< 10^{-5} for WTCCC GWAS data) were excluded. EIGENSTRAT analysis was performed and did not detect significant impact of population stratification on the GWAS results (genomic control inflation factor $\lambda = 1.04$).¹³ Ungenotyped SNPs across the genome were imputed by IMPUTE software (MathWorks, Natick, Massachusetts) against the reference panel of HapMap Phase 2 European Central European University founders.¹⁴ After quality control, a total of 2 554 714 SNPs with high quality genotype data were used. Meta-analysis was performed by summing inverse variance-weighted β and z cores. The final genomic control corrected meta-analysis results were reported. A detailed description of study subjects, experimental procedures and analysis approaches can be found in the original study.¹²

Gene expression profile of RA synovium. The gene expression profile data of RA synovium were downloaded from the Gene Expression Omnibus (GEO, access number: GSE1919) database.¹⁵ Briefly, synovial tissue samples were collected from eight RA patients undergoing synovectomy or arthroplasty, and from 15 normal subjects who died from fatal accidents, and were matched for age and gender. All donors were Caucasians living in Berlin, Germany. Total ribonucleic acid (RNA) was isolated from the synovium using RNeasy spin columns (Qiagen, Hilden, Germany). The isolated total RNA was labelled and hybridised to Affymetrix Human Genome U95A Array (containing 12 600 oligonucleotide probes) following standard protocol. Global normalisation of raw signal intensities was performed by GeneChip software (MAS 5.0, Affymetrix). Data analysis was conducted using DMT 3.0 software (Affymetrix). A detailed description of study subjects, experimental procedures and analysis approaches can be found in the previous study.⁴

Gene expression profile of RA PBMC. The gene expression profile of PBMC of RA was downloaded from the GEO database (access number: GSE15573). Briefly, peripheral blood specimens were collected from 18 RA patients and 15 healthy controls, matched for age and gender.

Table I. The top three gene modules identified for rheumatoid arthritis synovium

Module name	Gene	GWAS meta-analysis p-value
HLA-A	TAP2	2.46E ⁻¹²⁶
	HLA-C	1.57E ⁻⁴⁰
	HLA-A	1.37E ⁻²⁴
	TAPBP	7.63E ⁻¹¹
	LILRB1	2.00E ⁻⁰²
PTPN2	DAXX	9.19E ⁻²⁰
	PTPN2	4.33E ⁻⁰⁵
	EIF4E	2.12E ⁻⁰³
	EIF4G2	3.42E ⁻⁰³
	VCL	1.00E ⁻⁰²
	STAT3	9.00E ⁻⁰²
	EIF4B	2.00E ⁻⁰²
STAT3	PML	1.60E ⁻⁰¹
	VCL	1.00E ⁻⁰²
	STAT3	9.00E ⁻⁰²
	DAXX	9.19E ⁻²⁰
	PML	1.60E ⁻⁰¹
	EIF4B	2.00E ⁻⁰²
	EIF4E	2.12E ⁻⁰³
EIF4G2	3.42E ⁻⁰³	

GWAS, genome-wide association study

All were Caucasians living in Paris, France. Total RNA was isolated from PBMC using the PAXgene RNA isolation kit (PreAnalytix, Hombrechtikon, Switzerland). The isolated total RNA was transcribed into cRNA and labelled biotin-16-UTP. Illumina human-6 v2.0 expression beadchip (Illumina, San Diego, California) (containing 12 600 oligonucleotide probes) was applied for microarray hybridisation following standard protocol. Hybridisation signals were scanned by Illumina BeadArray 500GX Reader and analysed by Illumina BeadScan software 2.3.0.13 (Illumina). BeadStudio software (version 1.5.0.34) was used for preliminary data analysis and quality control. The generated microarray data were expressed as log₂ ratios of the fluorescence intensities of RA samples divided by the fluorescence intensities of control samples. Normalisation was conducted using the "normalise quantiles" function in the BeadStudio software. A detailed description of study subjects, experimental procedures and analysis approaches can be found in the original study.⁵

Dense gene module search. The integrative analysis of RA GWAS and gene expression profiles data was performed by dmGWAS 3.0 (Bioinformatics, Houston, Texas).^{16,17} dmGWAS implemented a PPI network-based Dense Module Search (DMS) algorithm.^{16,17} The detailed description of the dmGWAS algorithm can be found in previous studies^{16,17} The brief analysis procedures of dmGWAS can be summarised as follows: the node weights of PPI network are derived from GWAS results. The edge weights of PPI network are derived from gene expression profiling results. A greedy algorithm was applied to the PPI network for searching dense modules enriched in the gene sets showing association evidence and abnormal expression for target diseases. dmGWAS has been successfully applied to the causal gene studies of multiple complex

Table II. The top three gene modules identified for rheumatoid arthritis peripheral blood mononuclear cells

Module name	Gene	GWAS meta-analysis p-value
GRB7	HLA-DRB5	3.51E ⁻²³⁵
	HLA-DRA	1.00E ⁻²⁸⁵
	KIT	2.65E ⁻⁰³
	GRB7	7.70E ⁻⁰¹
	CD63	7.00E ⁻⁰²
MAP4K5	HLA-DQA2	4.08E ⁻²⁸⁵
	HLA-DQB1	4.08E ⁻²⁸⁵
	PIK3R1	1.72E ⁻⁰³
	CRKL	1.30E ⁻⁰¹
	CD4	6.00E ⁻⁰²
	CD3E	5.00E ⁻⁰²
	MAP4K5	7.00E ⁻⁰²
ZFP36L1	VAV3	1.00E ⁻⁰²
	IL1R1	5.05E ⁻⁰³
	PTPN22	7.36E ⁻²⁷
	EGFR	1.94E ⁻⁰⁴
	NRG1	2.61E ⁻⁰³
	ITGAS	3.31E ⁻⁰³
	ZFP36L1	4.45E ⁻⁰³
	PTPRE	1.00E ⁻⁰²
	GRB2	2.00E ⁻⁰²
	MAPK14	3.00E ⁻⁰²

GWAS, genome-wide association study

diseases.^{6,10,18} The PPI network used by Goh et al¹⁹ was applied here.⁶ This PPI network consisted of 10 174 proteins and 61 070 protein-protein interactions.^{20,21} In this study, dmGWAS was running under default parameters recommended by the developers. During the processing of RA GWAS data, a SNP was assigned to a gene if it was located within the gene or up to 20 kb immediately upstream or downstream of the gene. We assigned each gene a gene-level p-value, which was the smallest p-value of the SNPs assigned to the gene. The candidate genes residing in the top three identified gene modules were subjected to further function evaluation.

Protein-association network. To understand the functional association network of the identified gene modules, the genes within the top three gene modules were input into STRING (version 9.1)²² for protein-protein interaction analysis. STRING is a widely used protein-association network database^{23,24} which can help to assemble and evaluate the relationships of disease-related proteins.

Functional enrichment analysis. The biological functional significance of identified gene modules was evaluated by functional enrichment analysis. The genes within the top three gene modules were input into the DAVID tool²⁵ for GO enrichment analysis. An enrichment p-value and a false discovery rate (FDR) were calculated by DAVID for each GO term. Significant GO terms were identified as FDR < 0.05.

Results

For RA synovium, 1149 gene modules were detected by dmGWAS analysis. Table I summarises the top three gene modules consisting of 16 genes. The top-ranked gene module is named HLA-A, which contains TAP2, HLA-A,

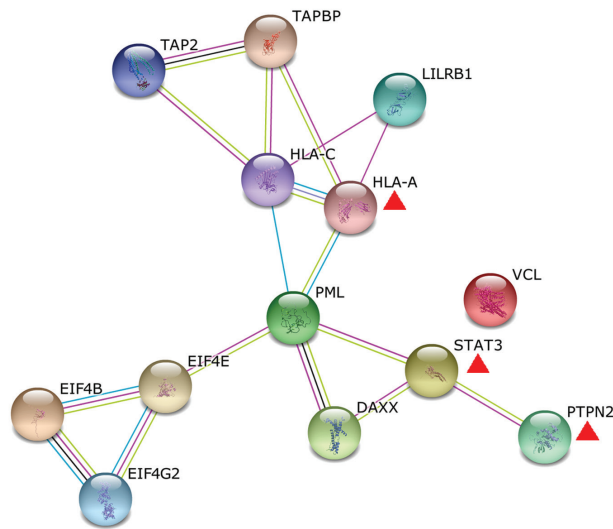


Fig. 1a

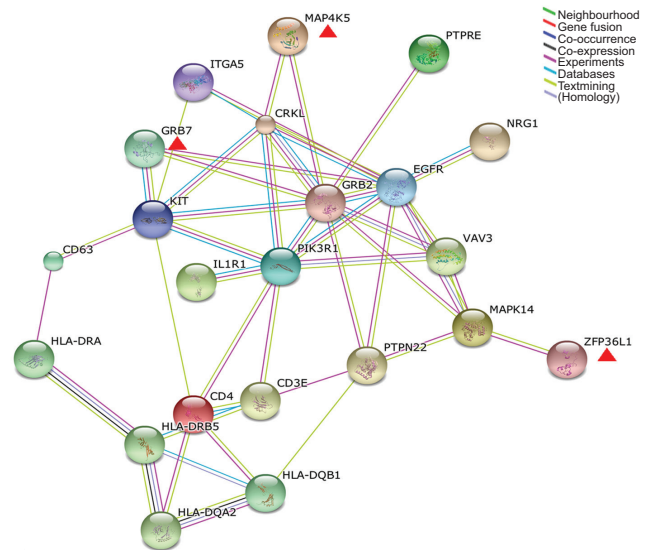


Fig. 1b

Images showing protein-association networks of gene modules identified for rheumatoid arthritis synovium (a) and peripheral blood mononuclear cells (b). The lines with different colours represent different types of association evidence. The red triangles denote the key genes of the top three significant gene modules.

Table III. Gene ontology enrichment analysis results of identified gene modules

	Gene ontologyname	Gene ontology number	Enrichment score*	p-value	False discovery rate
Synovium	Antigen processing and presentation of peptide antigen via MHC class I	0002474	244.85	3.60E-07	4.86E-04
	Antigen processing and presentation of peptide antigen	0048002	148.66	1.73E-06	2.33E-03
	Eukaryotic translation initiation factor 4F complex	0016281	355.06	2.42E-05	2.52E-02
Peripheral blood	Transmembrane receptor protein tyrosine kinase signalling pathway	GO:0007169	21.96	2.97E-08	4.34E-05
	MHC class II protein complex	GO:0042613	97.95	6.99E-06	7.87E-03
	Plasma membrane part	GO:0044459	3.87	1.74E-05	1.95E-02
	Protein kinase cascade	GO:0007243	11.63	1.54E-05	2.25E-02
	Antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	GO:0002504	74.53	1.71E-05	2.50E-02
	Positive regulation of kinase activity	GO:0033674	15.97	2.26E-05	3.31E-02
	Positive regulation of transferase activity	GO:0051347	15.37	2.72E-05	3.98E-02

*Enrichment score calculated by the DAVID tool
MHC, major histocompatibility complex

HLA-C, TAPBP and LILRB1 genes. Within the HLA-A module, TAP2 achieved the most significant p-value (2.46×10^{-126}) in the GWAS meta-analysis of RA.¹² The gene expression study also found that TAP2 was upregulated in RA synovium (ratio = 2.10 SD 1.66).⁴ HLA-C (GWAS p-value = 1.57×10^{-40} , ratio = 19.96 SD 19.96) is another notable gene in the HLA-A modules. The protein-association network analysis result of the top three gene modules of RA synovium is shown in Figure 1a.

In total, 1675 gene modules were detected for RA PBMC. Table II summarises the top three gene modules, which contain 26 genes. The top-ranked gene module is named GRB7, consisting of CD63, HLA-DRA, KIT, GRB7 and HLA-DRB5 genes. Within the GRB7 module, HLA-DRB5 achieved the most significant association signal (GWAS p-value = 3.51×10^{-235}),¹² and was also upregulated (ratio = 4.53 SD 4.83) in RA PBMC.⁵ Figure 1b

presents the protein-association network analysis result of the top three gene modules of PBMC.

The functional enrichment analysis results of identified gene modules are presented in Table III. For RA synovium, we identified three significant GO terms, including antigen processing and presentation of peptide antigen via major histocompatibility complex (MHC) class I (FDR = 4.86×10^{-4}), antigen processing and presentation of peptide antigen (FDR = 2.33×10^{-3}) and eukaryotic translation initiation factor 4F complex (FDR = 2.52×10^{-2}). For RA PBMC, seven significant GO terms were identified, including transmembrane receptor protein tyrosine kinase signalling pathway (FDR = 4.34×10^{-5}), MHC class II protein complex (FDR = 7.87×10^{-3}), plasma membrane (FDR = 1.95×10^{-2}), protein kinase cascade (FDR = 2.25×10^{-2}), antigen processing and presentation of peptide or polysaccharide antigen via MHC class II

(FDR = 2.50×10^{-2}), positive regulation of kinase activity (FDR = 3.31×10^{-2}) and positive regulation of transferase activity (FDR = 3.98×10^{-2}). Interestingly, we found that antigen processing and presentation-related GO terms were enriched in the top three gene modules of both RA synovium and PBMC.

Discussion

In this study, we performed a PPI network-based integrative analysis of GWAS and gene expression profiles of RA. We identified several RA-associated genes, gene modules and their association networks. Further functional enrichment analysis of identified gene modules highlighted the importance of antigen processing and presentation in the pathogenesis of RA.

It is well documented that genetic factors contribute greatly to the risk of RA.²⁶ The identified gene modules contain several reported candidate genes of RA, such as PTPN22,²⁷ CD4,²⁸ HLA-C,²⁹ MAPK14³⁰ and TAP2.²⁹ Integrating the GWAS and gene expression profile data confirmed the relevance of several reported susceptibility genes to RA. However, the molecular mechanism of multiple RA relevant genes interactively leading to RA remains elusive. The identified gene modules and their functional association networks may help to answer this question. For instance, the protein-association network analysis of RA-associated gene modules suggests that PTPN22, CD4 and MAPK14 indirectly interacted with each other to contribute to the development of RA (Fig. 1b). General functional interactions were also observed among HLA, TAP2 and LILRB1 genes (Fig. 1a). Furthermore, we provide functional association networks of RA-associated gene modules, which may help to clarify the complex pathogenesis of RA. Further studies are required to confirm our findings.

Another finding of this study is the antigen processing and presentation-related GO terms enriched in the identified gene modules of both RA synovium and PBMCs. Antigen processing and presentation is an immunological process, in which antigen-presenting cells prepare antigen and present generated peptide–MHC class II complexes to CD4 + T cells for activation of the immune system. The mechanism of antigen processing and presentation implicated in the imbalance of the immune system of RA remains elusive. Our study results confirm the importance of antigen processing and presentation in the pathogenesis of RA. We also provide some potential key gene modules involved in the imbalance of the immune system of RA.

This study also illustrates the application of multi-omics integrative analysis in the pathogenetic studies of complex diseases. Multi-omics integrative studies are expected to provide more pathogenetic information. Through the mutual complement and validation of various omics data, multi-omics integrative studies have the potential to reduce the impact of data error and statistical

bias on study results. Compared with single-omic studies, multi-omics integrative studies have significant advantages for revealing the functional principles and dynamics of the global pathogenetic process of complex diseases. However, the availability of multi-omics integrative study approaches and tools is limited and they warrant further study.

We performed a PPI network-based integrative analysis of GWAS and gene expression profiles of RA. We identified several RA-associated gene modules and their functional association networks in the development of RA. This study provides novel clues for understanding the pathogenesis of RA. We also demonstrate the performance of multi-omics integrative analysis for mechanism studies of complex diseases.

References

1. Lee DM, Weinblatt ME. Rheumatoid arthritis. *Lancet* 2001;358:903-911.
2. Pincus T, Brooks RH, Callahan LF. Prediction of long-term mortality in patients with rheumatoid arthritis according to simple questionnaire and joint count measures. *Ann Intern Med* 1994;120:26-34.
3. Yamamoto K, Okada Y, Suzuki A, Kochi Y. Genetics of rheumatoid arthritis in Asia—present and future. *Nat Rev Rheumatol* 2015;11:375-379.
4. Ungethuem U, Haeupl T, Witt H, et al. Molecular signatures and new candidates to target the pathogenesis of rheumatoid arthritis. *Physiol Genomics* 2010;42A:267-282.
5. Teixeira VH, Olaso R, Martin-Magniette ML, et al. Transcriptome analysis describing new immunity and defense genes in peripheral blood mononuclear cells of rheumatoid arthritis patients. *PLoS One* 2009;4:e6803.
6. He H, Zhang L, Li J, et al. Integrative analysis of GWASs, human protein interaction, and gene expression identified gene modules associated with BMDs. *J Clin Endocrinol Metab* 2014;99:E2392-E2399.
7. Chen C, Yang B, Zeng Z, et al. Genetic dissection of blood lipid traits by integrating genome-wide association study and gene expression profiling in a porcine model. *BMC Genomics* 2013;14:848.
8. Yoo S, Takikawa S, Geraghty P, et al. Integrative analysis of DNA methylation and gene expression data identifies EPAS1 as a key regulator of COPD. *PLoS Genet* 2015;11:e1004898.
9. International Multiple Sclerosis Genetics Consortium. Network-based multiple sclerosis pathway analysis with GWAS data from 15,000 cases and 30,000 controls. *Am J Hum Genet* 2013;92:854-865.
10. Han S, Yang BZ, Kranzler HR, et al. Integrating GWASs and human protein interaction networks identifies a gene subnetwork underlying alcohol dependence. *Am J Hum Genet* 2013;93:1027-1034.
11. Rossin EJ, Lage K, Raychaudhuri S, et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet* 2011;7:e1001273.
12. Stahl EA, Raychaudhuri S, Remmers EF, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 2010;42:508-514.
13. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904-909.
14. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;39:906-913.
15. No authors listed. Gene Expression Omnibus. <http://www.ncbi.nlm.nih.gov/gds> (date last accessed 01 July 2016).
16. Jia P, Zheng S, Long J, Zheng W, Zhao Z. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. *Bioinformatics* 2011;27:95-102.
17. Wang Q, Yu H, Zhao Z, Jia P. EW_dmGWAS: edge-weighted dense module search for genome-wide association studies and gene expression profiles. *Bioinformatics* 2015;31:2591-2594.
18. Jia P, Wang L, Fanous AH, et al. Network-assisted investigation of combined causal signals from genome-wide association studies in schizophrenia. *PLoS Comput Biol* 2012;8:e1002587.

19. **Goh KI, Cusick ME, Valle D, et al.** The human disease network. *Proc Natl Acad Sci U S A* 2007;104:8685-8690.
20. **Rual JF, Venkatesan K, Hao T, et al.** Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 2005;437:1173-1178.
21. **Stelzl U, Worm U, Lalowski M, et al.** A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 2005;122:957-968.
22. **No authors listed.** STRING v9.1. <http://string-db.org/> (date last accessed 04 July 2016).
23. **Choudhary C, Kumar C, Gnad F, et al.** Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 2009;325:834-840.
24. **Kravchenko-Balasha N, Levitzki A, Goldstein A, et al.** On a fundamental structure of gene networks in living cells. *Proc Natl Acad Sci U S A* 2012;109:4702-4707.
25. **No authors listed.** DAVID. <http://david.abcc.ncifcrf.gov/> (date last accessed 04 July 2016).
26. **Scott DL, Wolfe F, Huizinga TW.** Rheumatoid arthritis. *Lancet* 2010;376:1094-1108.
27. **Costenbader KH, Chang SC, De Vivo I, Plenge R, Karlson EW.** Genetic polymorphisms in PTPN22, PADI-4, and CTLA-4 and risk for rheumatoid arthritis in two longitudinal cohort studies: evidence of gene-environment interactions with heavy cigarette smoking. *Arthritis Res Ther* 2008;10:R52.
28. **Lo SF, Wan L, Lin HC, Huang CM, Tsai FJ.** Association of CD4 enhancer gene polymorphisms with rheumatoid arthritis and systemic lupus erythematosus in Taiwan. *J Rheumatol* 2008;35:2113-2118.
29. **Lee HS, Lee AT, Criswell LA, et al.** Several regions in the major histocompatibility complex confer risk for anti-CCP-antibody positive rheumatoid arthritis, independent of the DRB1 locus. *Mol Med* 2008;14:293-300.
30. **Coulthard LR, Taylor JC, Eyre S, et al.** Genetic variants within the MAP kinase signalling network and anti-TNF treatment response in rheumatoid arthritis patients. *Ann Rheum Dis* 2011;70:98-103.

Funding Statement

- Funding has been received from the National Natural Scientific Fund of China (81472925).

Author Contributions

- X. Xiao: Data analysis, Writing the paper.
- J. Hao: Data collection.
- Y. Wen: Data collection and analysis.
- W. Wang: Data analysis.
- X. Guo: Data collection.
- F. Zhang: Study design, Data analysis.

ICMJE conflict of interest

- None declared.

© 2016 Guo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution licence (CC-BY-NC), which permits unrestricted use, distribution, and reproduction in any medium, but not for commercial gain, provided the original author and source are credited.