American Medical Journal 3 (1): 56-66, 2012
ISSN 1949-0070

High-Order Interactions in Rheumatoid Arthritis Detected by Bayesian Method using Genome-Wide Association Studies Data

[1,3]Jing Zhang, [2]Zheyang Wu, [1]Chao Gao and [4,5]Michael Q. Zhang
[1]Department of Statistics, Yale University,
24 Hillhouse Ave., New Haven, CT 06511, USA
[2]Department of Mathematical Sciences,
WPI, 100 Institute Road, Worcester, MA 01609, USA
[3]Program of Computational Biology and Bioinformatics Yale University,
300 George Street, New Haven, CT 06051, USA
[4]MCB and Center for Systems Biology, The University of Texas at Dallas,
800 West Campbell Road, RL11, Richardson, Dallas, 75080, USA
[5]NTLIST, Tsinghua University, Beijing, 100084, China

Abstract: Problem statement: In order to reveal the missing genetic component of Rheumatoid Arthritis (RA) susceptibility, we carried out a genome-wide high-order epistatic interaction study for RA. **Approach:** A powerful Bayesian strategy was applied to analyze the data of Genome-Wide Association Studies (GWAS) from the Welcome Trust Case Control Consortium (WTCCC), where 319 high-order interactions were found across the whole genome and many of which were validated by the GWAS data from the North American Rheumatoid Arthritis Consortium (NARAC). **Results:** This is the first study intensively searching for high-order epistatic interactions genome-widely for RA. **Conclusion:** Our results suggest that high-order interactions might explain a big proportion of missing genetic component of RA. In the meanwhile, synapse, calcium ion binding and membrane part likely have interactive associations with RA. This finding implies that not only autoimmune system but also nervous system can play an important role in RA.

Key words: Welcome Trust Case Control Consortium (WTCCC), North American Rheumatoid Arthritis Consortium (NARAC), the results suggest, autoimmune system, genetic component

INTRODUCTION

Rheumatoid Arthritis (RA) is the most common chronic inflammatory autoimmune disease that leads to progressive joint destruction with the prevalence up to 1% in adult populations (Gabriel, 2001; Hu *et al*., 2011). Twin studies estimated that genetic factors contribute to 60% of the susceptibility of RA (MacGregor *et al*., 2000). Multiple loci associated with RA susceptibility have been identified by genome-wide linkage and association studies (Cornelis *et al*., 1998; WTCCC, 2007; Plenge *et al*., 2007; Stahl *et al*., 2010). However, current findings only account for a small portion of the genetic component to the RA susceptibility (Stahl *et al*., 2010; Raychaudhuri *et al*., 2008; Imboden, 2009), while the most recent GWAS findings are ethnicity-specific (Terao *et al*., 2011; Freudenberg *et al*., 2011; Julia *et al*., 2008). In order to find the missing heritability of RA, it is critical to study epistatic interactions in which genetic variations may show weak marginal penetrance, but may interact with each other in complex ways (Manolio *et al*., 2009; Wu and Zhao, 2009). The complicated interacting structures likely exist in the pathogenesis of RA due to the sophisticated regulatory mechanisms encoded in the human genome.

This study provides the first genome-wide high-order interaction analysis for RA using Bayesian epistasis association mapping (BEAM and BEAM2) methods (Zhang and Liu, 2007; Zhang *et al*., 2011). Specifically, BEAM uses Markov chain Monte Carlo (MCMC) to 'interrogate' each marker conditional on the current status of other markers iteratively and has been proved to provide a higher statistical power than many commonly used interaction-mapping methods in GWAS (Culverhouse *et al*., 2004; Cook *et al*., 2004; Ritchie *et al*., 2001; Nelson *et al*., 2001). Furthermore, to capture the block-wise structure of human genome,

Corresponding Author: Jing Zhang, Department of Statistics, Yale University, 24 Hillhouse Ave., New Haven, CT 06511, USA

Zhang *et al*. (2011) extended BEAM model to BEAM2, the latter incorporates LD blocks into the original Bayesian partition model. BEAM2 is able to simultaneously infer haplotype-blocks and select SNPs within blocks that are associated with the disease, either individually or through epistatic interactions with other SNPs across the genome. BEAM2 has shown great success in studying type-1 diabetes (Zhang *et al*., 2011). We applied BEAM and BEAM2 to the RA data from WTCCC (2007) and found fruitful interesting interacting structures for RA, many of which were validated by NARAC data (Plenge *et al*., 2007). In particular, our results show that not only autoimmune system but also nerve system plays an important role in genetic mechanisms of RA.

MATERIALS AND METHODS

Analysis strategy: Figure 1 shows a flow chart for our analysis strategy. We first applied BEAM2 (Zhang *et al*., 2011) to analyze the WTCCC RA data on each chromosome. BEAM2 is a sophisticated method that takes care of LD block structure while searching for epistasis. In the second step, we took the advantage of BEAM (Zhang and Liu, 2007) as a more efficient tool to search for genome-wide high-order interactions. In particualr, we pooled all BEAM2-selected candidate SNPs with posterior associtions probabilities > 0.5 from 22 autosomal chromosomes and ran the Bayesian variable partition model to search for high-order interactions across all the 22 chromosomes. Finally, these identified high-order interactions were validated by NARAC data (Plenge *et al*., 2007).

Data description: The RA data set from the WTCCC (2007) contains 1999 RA patients, 1504 controls from the 1958 Birth Cohort (58C) and 1500 additional controls from National Blood Service (NBS). We removed all SNPs from the sample if they have genotype scores less than 0.9 in more than 20 individuals within each group of RA, 58C, or NBS. In addition, we removed all non-polymorphic SNPs, SNPs violating Hardy-Weinberg Equilibrium at a Bonferroni adjusted 0.05 level and SNPs with bad clustering quality according to the WTCCC summary report (WTCCC, 2007). After SNP filtration the dataset contains 301,653 high quality SNPs.

From the 90 unique SNPs involved in the 319 interactions identified above, we retrieved 18 SNPs from the RA GWAS data from NARAC (Plenge *et al*., 2007) that have good genotype data quality: Hardy-Weinberg Equilibrium P-values > 0.001, minor allele frequencies > 0.01 and the SNP- and subject-missing rates < 10%. The final data contain 2,002 subjects (862 cases and 1,140 controls). The missing genotypes were eliminated at each test on site.

RESULTS

Search for chromosome-wise high-order interactions: Using BEAM2 to analyze each autosomal chromosome individually, we obtained the chromosome-wise posterior probabilities of SNPs associated with RA, which are shown in Fig. 2. It is clear that MHC region on Chromosome 6 strongly associates with RA: there are 72 SNPs located in MHC region with posterior probabilities > 0.5. At the same time, we also identified many associated SNPs outside MHC region: 85 SNPs across autosomal chromosomes except chr16 and chr21. Table 1 shows that our results are highly consistent with the original paper (WTCCC, 2007) on detecting single SNP effects.

Detection of inter-chromosomal high-order interactions: We applied BEAM on all the SNPs that have posterior probabilities of association greater than 0.5 (by BEAM2) to search for inter-chromosomal high-order interactions among these SNPs. After 2000 runs of BEAM to explore as many local modes as possible, we obtained 319 interactions. The supplementary files completetableX.txt and SNPinfoX.txt provide the diplotype P-values (Fisher's exact test) and the annotations of the involved SNPs, respectively. Table 2 summarizes some representatives of the significant interactions that have P-values less than 4.03e-11 (at family-wise significant level 0.05 after Bonferroni adjustment for about 1.24e9 possible diplotypes of these 319 interactions) and have diplotype frequencies larger than 0.05 in either controls or cases for stable results. The rows were sorted by decreasing genetic disease relative risk (DRR). Table 3 lists the SNP annotations for these selected interactions in Table 2.

Table 1 Comparison with Previous WTCCC Analysis (WTCCC, 2007)

Strongest loci		
Previously replicated loci	SNP or region	Posterior probability
RA	rs6679677	0.9992
RA	MHC	Many >0.5
Moderate loci		
RA	rs6684865	0.4602
RA	rs11162922	0.333
RA	rs3816587	0.8894
RA	rs6920220	0.7692
RA	rs11761231	0.9468
RA	rs2104286	0.36
RA	rs9550642	0.7124
RA	rs2837960	does not pass quality filter
RA	rs743777	0.8724

Table 2: Significant interactions obtained from the WTCCC data. The columns are: Inter: interaction index; Diplo.: diplotype, where number 0, 1 or 2 represents the copy number of Allele 2 (in Table 3) at the corresponding SNP; Contr.1: diplotype frequency in 58C controls; Contr.2: diplotype frequency in NBS controls. Case: diplotype frequency in cases; P-value.12: Fisher's exact test P-values between two control groups; P-value: Fisher's exact test P-values between pooled controls and cases; DRR: disease relative risk

Inter	Diplo.	Contr.1	Contr.2	Cases	P-value.12	P-value	DRR
303	002022002022020	0.04	0.04	0.16	0.92	1.96E-50	2.02
121	02222002022020	0.04	0.04	0.16	0.92	3.73E-50	2.02
11	00202220002022020	0.04	0.04	0.16	0.92	7.16E-50	2.01
182	00202222002022020	0.04	0.04	0.16	0.92	7.16E-50	2.01
228	2220020	0.08	0.06	0.20	0.03	5.71E-39	1.76
153	22002000	0.09	0.07	0.20	0.09	2.14E-37	1.73
1	1002	0.06	0.07	0.12	0.46	2.34E-12	1.45
290	1022	0.05	0.06	0.11	0.39	2.52E-11	1.45
113	1000	0.06	0.07	0.12	0.42	7.80E-12	1.44
75	1000	0.07	0.08	0.14	0.58	1.32E-12	1.44
109	100200	0.06	0.06	0.12	0.50	3.22E-11	1.44
114	1000	0.07	0.07	0.13	0.52	2.39E-11	1.42
109	000200	0.61	0.63	0.52	0.43	1.07E-11	0.79
176	220002022	0.41	0.42	0.32	0.48	8.34E-12	0.78
113	0000	0.62	0.63	0.52	0.50	2.86E-13	0.78
1	0002	0.65	0.67	0.56	0.26	1.15E-13	0.77
77	000000	0.66	0.67	0.55	0.76	7.76E-15	0.76
75	0000	0.72	0.73	0.62	0.71	3.23E-15	0.75
29	0011	0.39	0.36	0.27	0.19	4.64E-14	0.75
41	202202	0.27	0.28	0.19	0.74	7.69E-12	0.74
174	22000202022	0.26	0.25	0.17	0.93	1.20E-11	0.74
114	0000	0.71	0.71	0.59	0.78	1.28E-18	0.73
290	0022	0.59	0.58	0.44	0.97	6.52E-25	0.70
274	20011	0.15	0.16	0.09	0.76	7.89E-12	0.67
60	00112000	0.14	0.14	0.08	0.96	1.54E-12	0.64
155	20001102	0.14	0.14	0.07	0.83	2.10E-13	0.62
180	2002000	0.14	0.15	0.07	0.61	1.58E-15	0.60
285	2220020	0.16	0.16	0.08	1.00	1.18E-19	0.57
16	0020020	0.17	0.17	0.08	0.77	2.69E-20	0.57
302	2002002	0.16	0.16	0.07	0.96	8.10E-20	0.56
35	002220020	0.15	0.15	0.07	0.88	1.02E-18	0.56
233	002002002	0.14	0.15	0.06	0.64	1.64E-19	0.54

Table 3: SNP annotations for the representative high-order interactions listed in Table 2

Interaction	SNPs	Allele1	Allele2	Chromosome	Location
228	rs10490886	C	T	chr3	62684434
	rs17067111	A	G	chr3	62697011
	rs3134926	C	G	chr6	32308125
	rs6936204	A	G	chr6	32325070
	rs12524063	A	T	chr6	32405288
	rs4959093	C	T	chr6	32421075
	rs6907322	A	G	chr6	32432923
290	rs6679677	A	C	chr1	114015850
	rs3811019	C	G	chr1	114183625
	rs4713376	A	C	chr6	30881293
	rs12195469	A	T	chr6	30897587
302	rs2736172	C	T	chr6	31698877
	rs707974	C	T	chr6	31737478
	rs2075800	A	G	chr6	31885925
	rs2072633	A	G	chr6	32027557
	rs17421624	C	T	chr6	32174155
	rs2292365	A	T	chr9	91112710
	rs10991868	A	G	chr9	91129311
233	rs4312689	G	T	chr3	45503637
	rs2128361	C	T	chr3	45508228
	rs2736172	C	T	chr6	31698877
	rs707974	C	T	chr6	31737478
	rs2075800	A	G	chr6	31885925
	rs2072633	A	G	chr6	32027557
	rs17421624	C	T	chr6	32174155
	rs2292365	A	T	chr9	91112710
	rs10991868	A	G	chr9	91129311

Table 4: Representative interactions validated using NARAC data. The columns are: Inter: interaction index; Diplo.: diplotype, where number 0, 1 or 2 represents the copy number of Allele 2 (in Table 5); Contr.: diplotype frequency in controls; Case: diplotype frequency in cases; P-value: Fisher's exact test P-values; DRR: disease relative risk; Contr.1: diplotype frequency in 58C controls; Contr.2: diplotype frequency in NBS controls.

		NARAC				WTCCC				
Inter	Diplo.	Contr.	Cases	P-value	DRR	Contr.1	Contr.2	Cases	P-value	DRR
26	200	0.07	0.120	1.32E-05	1.41	0.10	0.09	0.15	4.58E-09	1.33
31	200	0.14	0.010	2.78E-32	0.09	0.15	0.16	0.06	5.69E-29	0.46
	010	0.10	0.240	1.66E-16	1.64	0.10	0.08	0.18	1.18E-20	1.52
	000	0.02	0.100	4.87E-14	1.86	0.03	0.03	0.07	3.09E-09	1.51
	210	0.05	0.002	3.97E-13	0.07	0.05	0.06	0.02	1.37E-11	0.44
	020	0.09	0.160	9.75E-08	1.45	0.08	0.07	0.12	1.12E-08	1.36
52	000	0.26	0.400	1.13E-11	1.43	0.27	0.27	0.36	1.78E-11	1.28
	010	0.24	0.140	1.05E-08	0.66	0.22	0.24	0.17	2.74E-06	0.81
96	102	0.09	0.030	1.26E-07	0.48	0.10	0.10	0.05	1.21E-12	0.57
180	020	0.15	0.050	9.88E-15	0.42	0.16	0.17	0.09	6.41E-16	0.61
	200	0.08	0.170	2.50E-09	1.51	0.07	0.08	0.13	4.02E-11	1.41
	100	0.12	0.200	2.87E-06	1.35	0.13	0.11	0.16	8.78E-05	1.21
236	2000	0.12	0.010	9.25E-29	0.09	0.14	0.14	0.05	2.91E-25	0.46
	0110	0.04	0.150	1.88E-17	1.83	0.05	0.04	0.11	6.57E-17	1.60
	0020	0.09	0.160	3.69E-07	1.43	0.08	0.07	0.12	1.12E-08	1.36
246	1200	0.09	0.005	5.07E-20	0.09	0.09	0.08	0.03	3.83E-19	0.40
	1011	0.02	0.080	1.60E-10	1.83	0.03	0.02	0.05	1.77E-05	1.40
255	010	0.03	0.120	1.29E-16	1.88	0.05	0.04	0.09	6.98E-12	1.51
297	0010	0.01	0.060	1.26E-08	1.82	0.03	0.02	0.06	3.10E-06	1.44
303	000	0.09	0.340	2.86E-44	2.10	0.11	0.09	0.27	1.61E-52	1.81
307	000	0.26	0.400	1.13E-11	1.43	0.27	0.27	0.36	1.78E-11	1.28
	010	0.24	0.140	1.05E-08	0.66	0.22	0.24	0.17	2.74E-06	0.81

Table 5: SNP annotations for the representative interactions listed in Table 4

Interaction	SNPs	Allele1	Allele2	Chromosome	Location
26	rs4538338	A	C	chr3	141436333
	rs6907322	A	G	chr6	32432923
	rs3135363	C	T	chr6	32497626
52	rs6907322	A	G	chr6	32432923
	rs3135363	C	T	chr6	32497626
	rs634435	A	G	chr9	3997714
180	rs2075800	A	G	chr6	31885925
	rs2072633	A	G	chr6	32027557
	rs634435	A	G	chr9	3997714
236	rs6457617	C	T	chr6	32771829
	rs3916765	A	G	chr6	32793528
	rs9461799	C	T	chr6	32797507
	rs634435	A	G	chr9	3997714
303	rs6907322	A	G	chr6	32432923
	rs3135363	C	T	chr6	32497626
	rs6457617	C	T	chr6	32771829

The full versions of Table 2 and 3 are given in supplementary files bestEpistasesFromWTCCC.txt and bestEpistasesFromWTCCC-SNPs.txt, respectively. From Table 2 and 3, we can see that most inter-chromosomal high-order interactions are among chr1, chr3, chr6 and chr9. A lot of them are more than 4-way interactions with very high Disease Relative Risk (DRR). Many protective diplotypes with DRR < 0.8 can also be found in Table 2. Diplotype frequencies in two control populations, 58C and NBS, are highly consistent, suggesting that these frequency estimations are stable.

Validation of interactions using NARAC data: Using NARAC data (Plenge *et al*., 2007) we sought to validate the high-order interactions identified by the WTCCC data. From the 90 unique SNPs involved in the 319 interactions, we retrieved 18 SNPs from NARAC data. For each supplementary table completetableX.txt obtained from the WTCCC data, whenever at least two NARAC SNPs are available for this interaction, we generated a supplementary table NARAC-completetableX.txt to list the diplotype frequencies and the P-values in the similar format. The supplementary file bestEpistasesFromWTCCC-NARAC.txt summarizes the significant NARAC-validated high-order interactions involving at least three SNPs. These interactions have P-values < 2.94e-5 for a family-wise significance level of 0.05 after the Bonferroni adjustment based on 1701 possible diplotypes of these interactions.

Table 6: Best GO terms for associated RA genes.

Best GOs	Accession	Genes(chr)	Count (24)	Total (33972)	P-Value (Correct-Method: Benjamini)
GO:0045202	Synapse (exact: synaptic junction)	CADPS (chr3) CLSTN2(chr3) ERC2(chr3) PCDH15(chr10)	4	254	0.0115
GO:0005509	Calcium ion binding (related: calcium ion storage activity)	LRP1B(chr2) CADPS(chr3) CLSTN2(chr3) CACNB2(chr10) PCDH15(chr10) NELL1(chr11)	6	1458	0.0705
GO:0044425	membrane part	KCNH7(chr2) LRP1B(chr2) CLSTN2(chr3) CADPS (chr3) ERC2(chr3) NDST4(chr4) LHFPL3(chr7) CSMD3(chr8) PTPRD(chr9) CACNB2(chr10) PCDH15(chr10) SORCS3(chr10) GRM5(chr11)	13	7726	0.0705

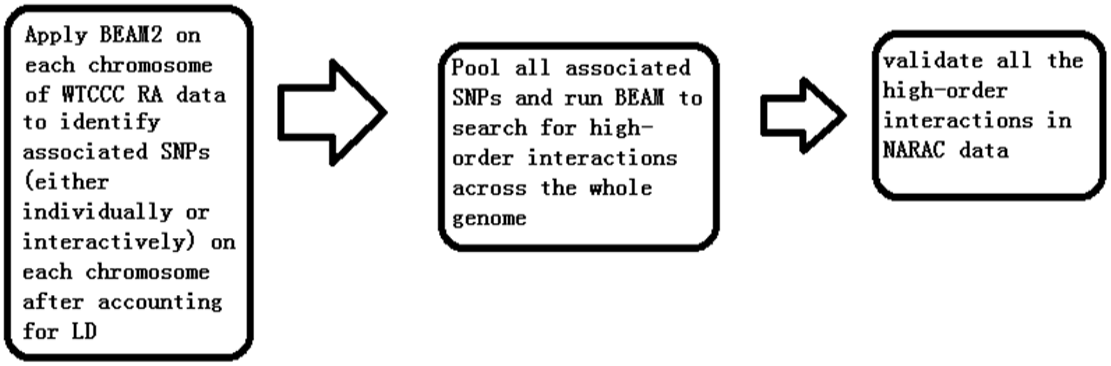
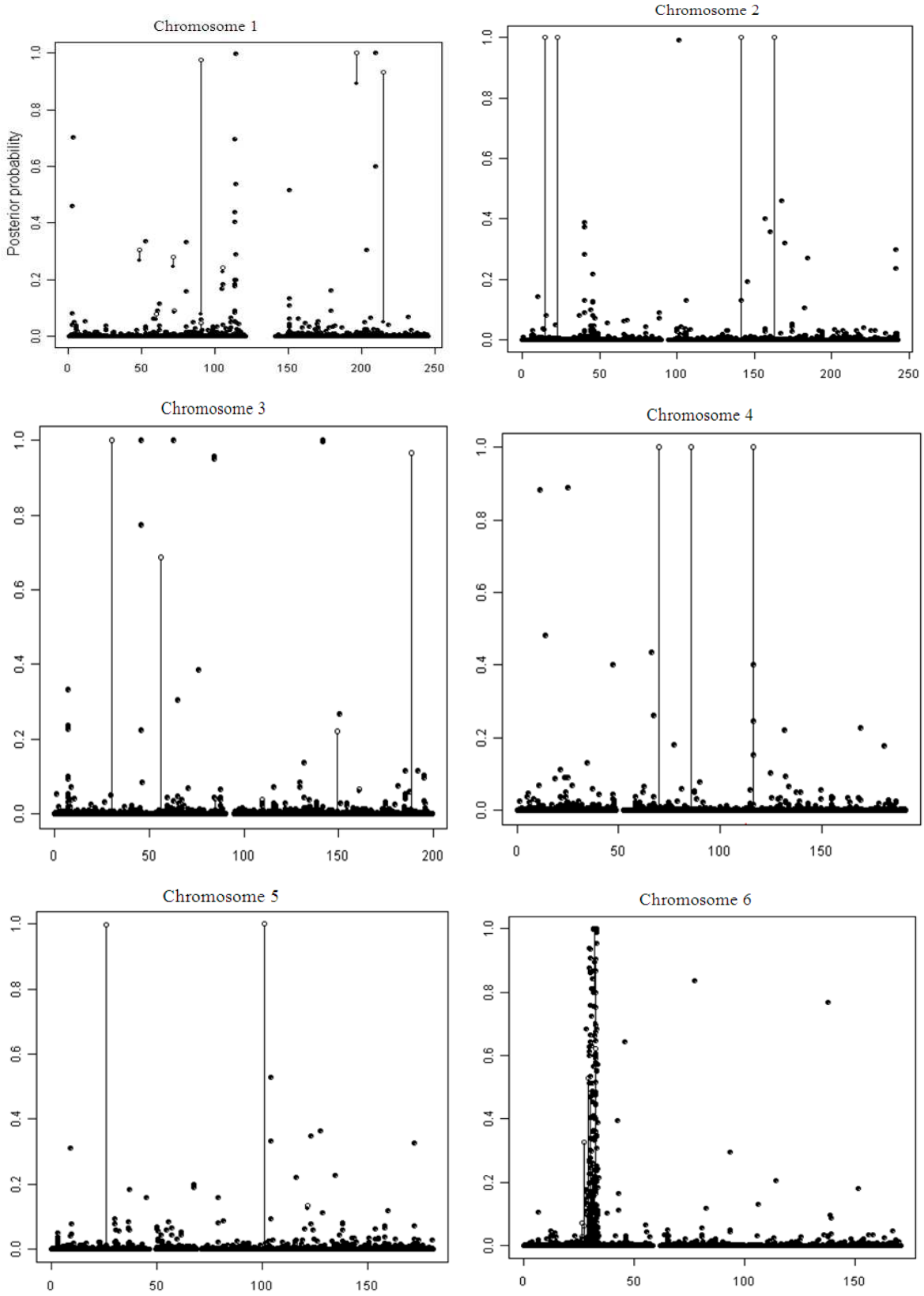
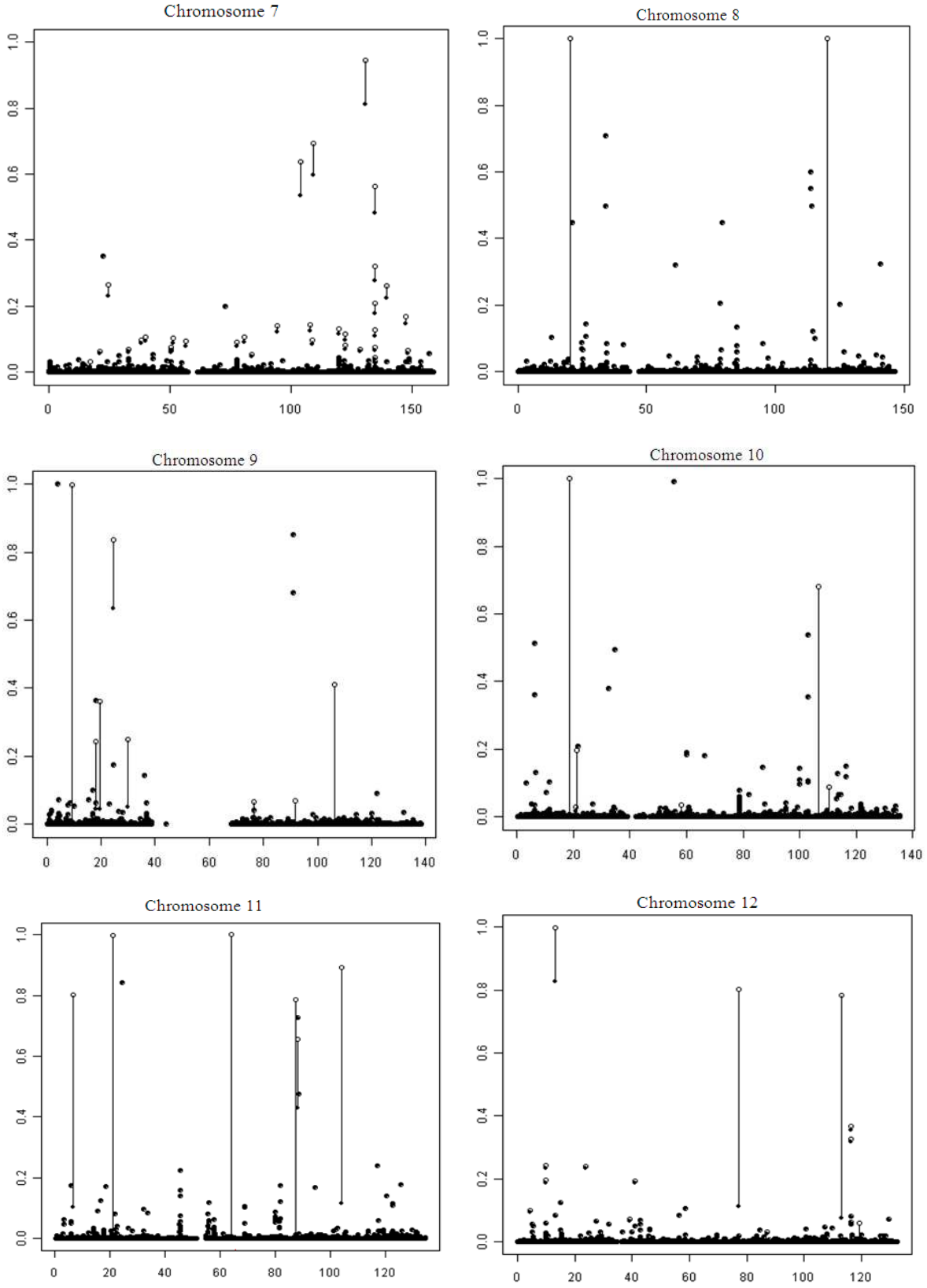


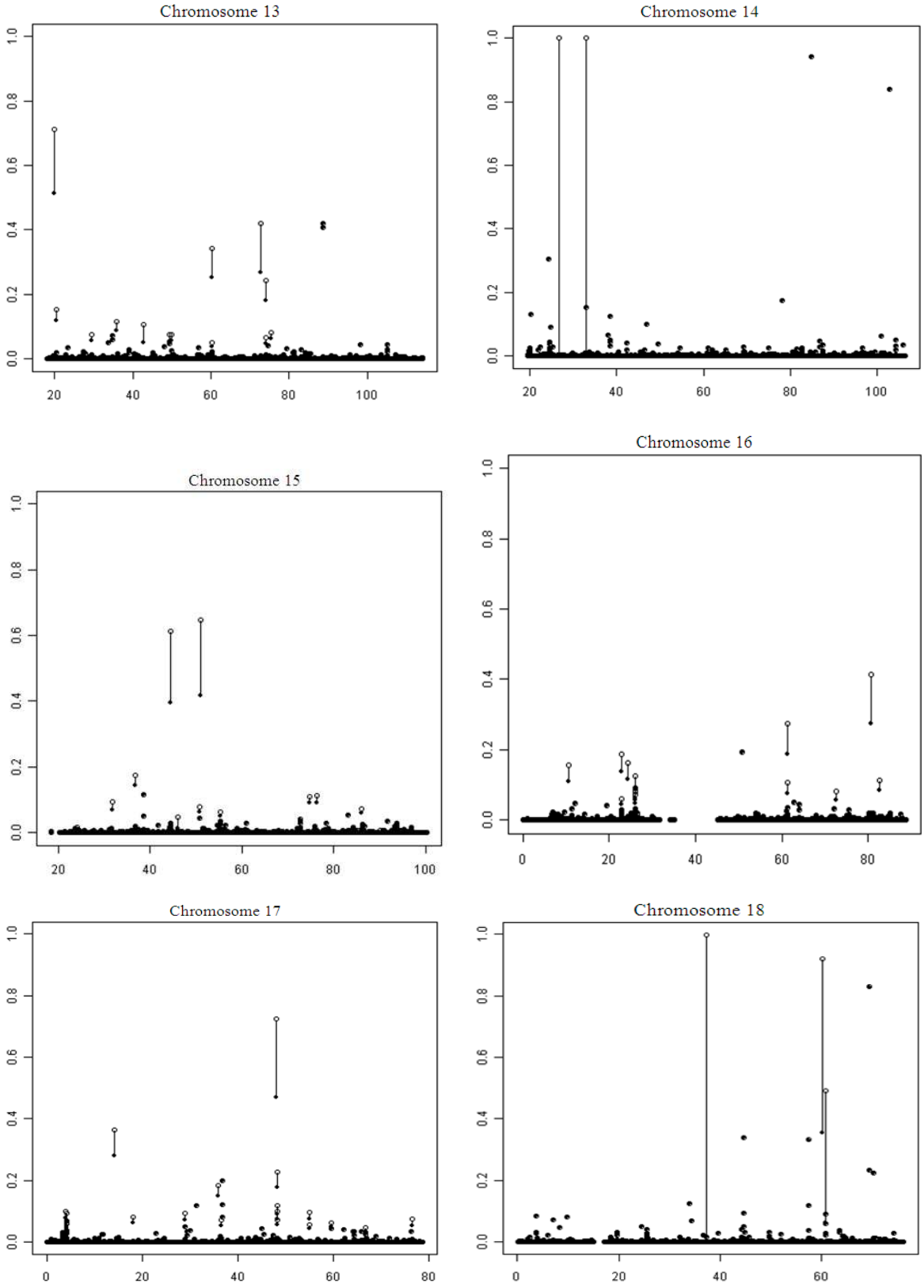
Fig. 1: The flow chart of our Bayesian analysis strategy

At the same time, these interactions also have moderately small P-values < 1e-4 for WTCCC data, as well as stable diplotype frequencies > 0.05 in either cases or controls for both NARAC and WTCCC data. The SNP annotations corresponding to these interactions are given in the supplementary file bestEpistasesFromWTCCC-NARAC-SNP.txt. Table 4 shows the selected high-order NARAC-SNP interactions that have DRR > 2 or are located on more than one chromosome (i.e., inter-chromosomal interactions). Table 5 shows the SNP annotations for selected interactions in Table 4. The validated inter-chromosomal interactions are on chromosomes 6 and 9 and chromosomes 6 and 3.

Synapse, calcium ion binding and membrane part associated with RA: MHC region on chromosome 6 is well-known for RA (Kozyryev and Zhang, 2012; Zhang *et al*., 2012). But very few other loci were detected without accounting for epistasis interaction. In our study, 31 associated SNPs were detected in 24 genes with posterior probabilities of association larger than 0.5 across 12 chromosomes (excluding chromosome 6) in WTCCC data. In order to test overrepresentation of biological pathways in these RA-associated genes, we use GOstat (Beissbarth and Speed, 2004) (http://gostat.wehi.edu.au/) to search enriched GO terms in these 24 RA-associated genes. Table 6 shows the top three GO terms: Synapse, calcium ion binding and membrane part (with Benjamini-corrected false discovery rates 0.0115, 0.0705 and 0.0705 respectively), the correspondingly associated genes and their chromosomes.


Chromosome 1
Chromosome 2
Chromosome 3
Chromosome 4
Chromosome 5
Chromosome 6
Posterior probability


Chromosome 7
Chromosome 8
Chromosome 9
Chromosome 10
Chromosome 11
Chromosome 12


Chromosome 13
Chromosome 14
Chromosome 15
Chromosome 16
Chromosome 17
Chromosome 18

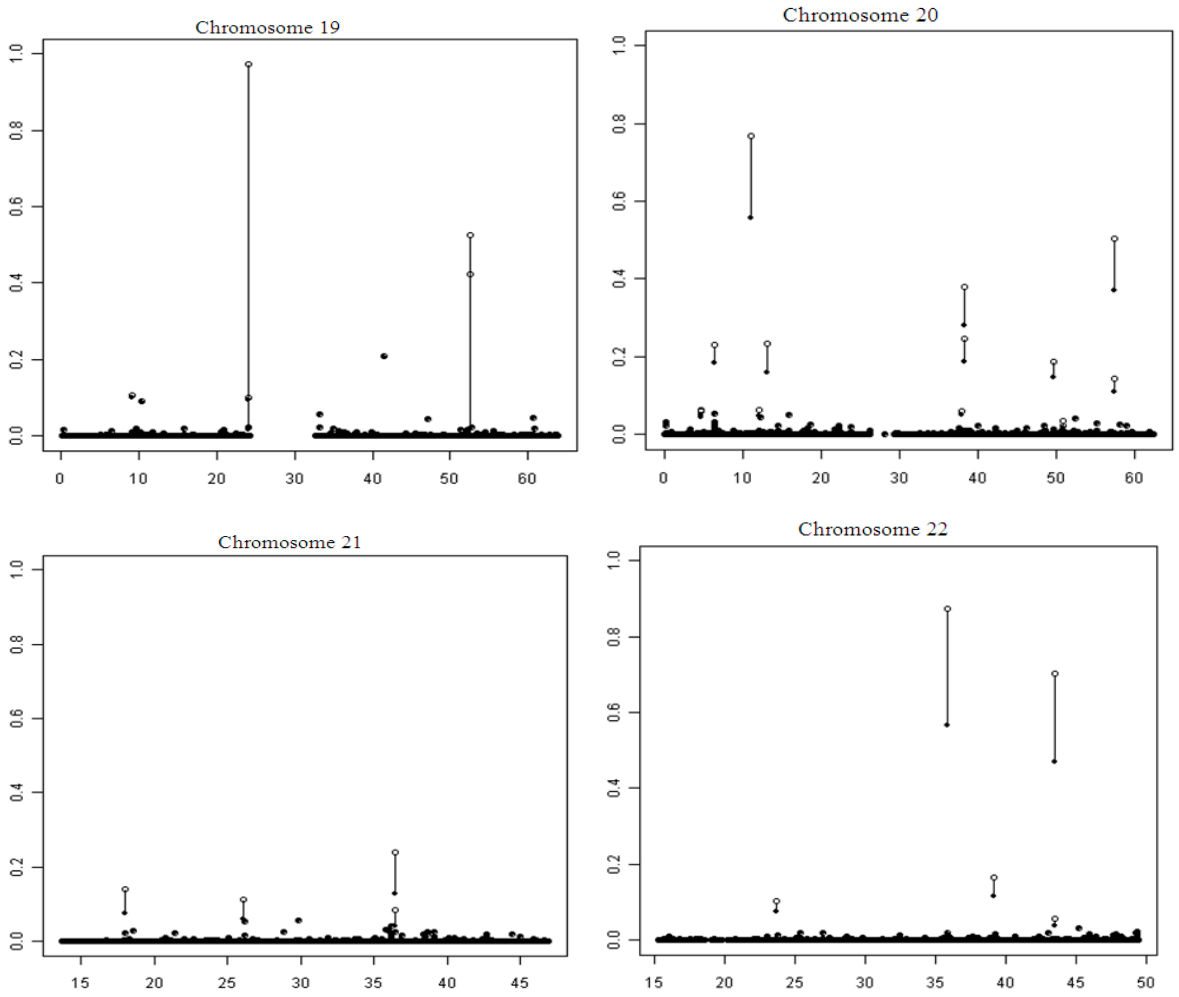


Fig. 2: Chromosome-wise posterior probabilities of SNPs associated with RA on each chromosome. Dot indicates the marginal posterior probability of association per SNP, circle indicates the total posterior probability of association per SNP (i.e., the marginal plus the joint association probabilities). We connected the dot and circle for each SNP for better illustration. X-axis indicates the chromosomal position (Mb), y-axis shows the posterior probability

DISCUSSION

Rheumatoid arthritis is an inflammatory disease, primarily of the joints, with autoimmune features and a complex genetic component. To our limited knowledge this is the first time that synapse, calcium ion binding and membrane part are reported to be interactively associated with RA in whole-genome association studies. Rheumatoid arthritis is characterized by a chronic inflammation of the synovial joints (Feldmann *et al*., 1996) and factors like Fibroblast-Like Synoviocyte (FLS) and T cells are actively involved in joint deconstruction and synapse is the contact point of these factors (Tran *et al*., 2007). Also it is well-known that calcium ions and membrane are important parts in synaptic neurotransmission and peripheral and central nervous system is very important for joint protection (O'Connor and Vilensky, 2003). It is already known that neurogenic factors play very important roles in the etiopathogenesis of osteoarthritis (O'Connor and Vilensky, 2003). Our results suggest that synaptic neurotransmission could be as important for RA.

CONCLUSION

Our results suggest that high-order interactions might explain a big proportion of missing genetic component of RA. In the meanwhile, synapse, calcium

ion binding and membrane part likely have interactive associations with RA. This finding implies that not only autoimmune system but also nerve system could play an important role in RA.

ACKNOWLEDGEMENT

J Zhang was supported by start-up funding and Sesseel Award from Yale University. The computation was done with the help from the Yale University Biomedical High Performance Computing Center, which was supported by the NIH grant RR19895. This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113. All the chromosomal positions are in NCBI build 35 coordinates. M. Zhang would like to thank the STARS fund of Texas and NIH grant HG001696.

REFERENCES

Beissbarth, T. and T.P. Speed, 2004. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics, 20: 1464-1465. DOI: 10.1093/bioinformatics/bth088

Cook, N.R., R.Y. Zee and P.M. Ridker, 2004. Tree and spline based association analysis of gene-gene interaction models for ischemic stroke. Stat. Med. 23: 1439-1453. PMID: 15116352

Cornelis, F., S. Faure, M. Martinez, J.F. Prud'homme and P. Fritz *et al*., 1998. New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. Proc. Nat. Acad. Sci., 95: 10746-10750.

Culverhouse, R., T. Klein and W. Shannon, 2004. Detecting epistatic interactions contributing to quantitative traits. Genet. Epidemiol., 27: 141-152. PMID: 15305330

Feldmann, M., F.M. Brennan and R.N. Maini, 1996. Role of cytokines in rheumatoid arthritis. Annu. Rev. Immunol., 14: 397-440. PMID: 8717520

Freudenberg, J., H.S. Lee, B.G. Han, H.D. Shin and Y.M. Kang *et al*., 2011. Genome-wide association study of rheumatoid arthritis in Koreans: Population-specific loci as well as overlap with European susceptibility loci. Arthritis Rheum., 63: 884-893. DOI: 10.1002/art.30235

Gabriel, S.E., 2001. The epidemiology of rheumatoid arthritis. Rheum. Dis. Clin. North Am., 27: 269-281. PMID: 11396092

Hu, H.J., E.H. Jin, S.H. Yim, S.Y. Yang and S.H. Jung *et al*., 2011. Common variants at the promoter region of the APOM confer a risk of rheumatoid arthritis. Exp. Mol. Med., 43: 613-621. PMID: 21844665

Imboden, J.B., 2009. The immunopathogenesis of rheumatoid arthritis. Annu. Rev. Pathol., 4: 417-434. PMID: 18954286

Julia, A., J. Ballina, J.D. Canete, A. Balsa and J. Tornero-Molina *et al*., 2008. Genome-wide association study of rheumatoid arthritis in the Spanish population: KLF12 as a risk locus for rheumatoid arthritis susceptibility. Arthritis Rheum., 58: 2275-2286. PMID: 18668548

Kozyryev, I. and J. Zhang, 2012. Bayesian determination of disease associated differences in haplotype blocks. Am. J. Bioinform., 1: 20-29 DOI: 10.3844/ajbsp.2012.20.29

MacGregor, A.J., H. Snieder, A.S. Rigby, M. Koskenvuo and J. Kaprio *et al*., 2000. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum., 43: 30-37. PMID: 10643697

Manolio, T.A., F.S. Collins, N.J. Cox, D.B. Goldstein and L.A. Hindorff *et al*., 2009. Finding the missing heritability of complex diseases. Nature, 461: 747-753. PMID: 19812666

Nelson, M.R., S.L.R. Kardia, R.E. Ferrell and C.F. Sing, 2001. A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation. Genome Res., 11: 458-470. DOI: 10.1101/gr.172901

O'Connor, B.L. and J.A. Vilensky, 2003. Peripheral and central nervous system mechanisms of joint protection. Am. J. Orthop. (Belle Mead NJ)., 32: 330-336. PMID: 12892277

Plenge, R.M., M. Seielstad, L. Padyukov, A.T. Lee and E.F. Remmers, 2007. TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. N. Engl. J. Med., 357: 1199-1209. PMID: 17804836

Raychaudhuri, S., E.F. Remmers, A.T. Lee, R. Hackett and C. Guiducci *et al*., 2008. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. Nat. Genet., 40: 1216-1223. PMID: 18794853

Ritchie, M.D., L.W. Hahn, N. Roodi, L.R. Bailey and W.D. Dupont *et al*., 2001. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am. J. Hum. Genet., 69: 138-147. PMID: 11404819

Stahl, E.A., S. Raychaudhuri, E.F. Remmers, G. Xie and S. Eyre *et al*., 2010. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat. Genet., 42: 508-514. PMID: 20453842

Terao, C., R. Yamada, K. Ohmura, M. Takahashi and T. Kawaguchi *et al*., 2011. The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. Hum. Mol. Genet., 20: 2680-2680. DOI: 10.1093/hmg/ddr161

Tran, C.N., S.K. Lundy, P.T. White, J.L. Endres and C.D. Motyl *et al*., 2007. Molecular interactions between T cells and fibroblast-like synoviocytes: Role of membrane tumor necrosis factor-alpha on cytokine-activated T cells. Am. J. Pathol., 171: 1588-1598. PMID: 17823284

WTCCC, 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 447: 661-678. PMID: 17554300

Wu, Z. and H. Zhao, 2009. Statistical power of model selection strategies for genome-wide association studies. PLoS Genet., 5: 849-911. PMID: 19649321

Zhang, Y. and J.S. Liu, 2007. Bayesian inference of epistatic interactions in case-control studies. Nat. Genet., 39: 1167-1173. DOI: 10.1038/ng2110

Zhang, Y., J. Zhang and J.S. Liu, 2011. Block-based Bayesian epistasis association mapping with application to WTCCC type 1 diabetes data. Ann. Applied Stat., 5: 2052-2077. DOI: 10.1214/11-AOAS469SUPP

Zhang, J., Q. Zhang, D. Lewis and M. Zhang, 2012. A bayesian method for disentangling dependent structure of epistatic interaction. Am. J. Biostat., 2: 1-10. DOI: 10.3844/amjbsp.2011.1.10