Allelic based gene-gene interactions in rheumatoid arthritis

Jeesun Jung*1, Joon Jin Song2 and Deukwoo Kwon3

Addresses: 1Department of Medical and Molecular Genetics, Indiana University School of Medicine, 410 West 10th Street, HTIS 5000, Indianapolis, Indiana 46202, USA, 2Department of Mathematical Sciences, University of Arkansas, 301 SCEN, Fayetteville, Arkansas 72701, USA and 3Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Boulevard, EPS Room 7056, Rockville, Maryland, 20852, USA

E-mail: Jeesun Jung* - jeejung@iupui.edu; Joon Jin Song - jjsong@uark.edu; Deukwoo Kwon - kwonde@mail.nih.gov

*Corresponding author

from Genetic Analysis Workshop 16
St Louis, MO, USA 17-20 September 2009

Published: 15 December 2009

BMC Proceedings 2009, 3(Suppl 7):S76 doi: 10.1186/1753-6561-3-S7-S76

This article is available from: http://www.biomedcentral.com/1753-6561/3/S7/S76

© 2009 Jung et al; licensee BioMed Central Ltd.
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The detection of gene-gene interaction is an important approach to understand the etiology of rheumatoid arthritis (RA). The goal of this study is to identify gene-gene interaction of SNPs at the allelic level contributing to RA using real data sets (Problem 1) of North American Rheumatoid Arthritis Consortium (NARAC) provided by Genetic Analysis Workshop 16 (GAW16). We applied our novel method that can detect the interaction by a definition of nonrandom association of alleles that occurs when the contribution to RA of a particular allele inherited in one gene depends on a particular allele inherited at other unlinked genes. Starting with 639 single-nucleotide polymorphisms (SNPs) from 26 candidate genes, we identified ten two-way interacting genes and one case of three-way interacting genes. SNP rs2476601 on PTPN22 interacts with rs2306772 on SLC22A4, which interacts with rs881372 on TRAF1 and rs2900180 on C5, respectively. SNP rs2900180 on C5 interacts with rs2242720 on RUNX1, which interacts with rs881375 on TRAF1. Furthermore, rs2476601 on PTPN22 also interacts with three SNPs (rs2905325, rs1476482, and rs2106549) in linkage disequilibrium (LD) on IL6. The other three SNPs (rs2961280, rs2961283, and rs2905308) in LD on IL6 interact with two SNPs (rs477515 and rs2516049) on HLA-DRB1. SNPs rs660895 and rs532098 on HLA-DRB1 interact with rs2834779 and four SNPs in LD on RUNX1. Three-way interacting genes of rs10229203 on IL6, rs4816502 on RUNX1, and rs10818500 on C5 were also detected.

Background

Rheumatoid arthritis (RA) is a complex, chronic inflammatory disease whose etiology remains unknown. It has been known that RA is a result of the complicated networks of multiple genes along with the environmental factors such as smoking. It is more common in females. Through a combined linkage and association study [1], the HLA gene cluster on 6p21 has been shown to have the most likely predisposing loci for RA. In addition to HLA, numerous genetic variants influence the pathology of RA.
Unfortunately, detection of gene-gene interaction has been challenging due to an issue of high dimensionality from multi-locus combinations that require a large sample size.

In this study, we applied a novel approach to detect gene-gene interaction influencing RA using the case-control subjects provided by the North American Rheumatoid Arthritis Consortium (NARAC). In contrast to the previously available method searching for the interaction at the genotype level, our approach focuses on the detection of interaction at the allelic level with a novel definition: the allele-based gene-gene interaction occurs when a particular allele in one gene and a particular allele at another unlinked gene are dependent on the contribution to RA (Figure 1) [1]. Based on the 639 SNPs from 26 candidate genes related to RA pathology, we performed a score test based on logistic regression and a $F$-test based on the Cochran-Armitage regression model developed for the detection of allelic based gene-gene interaction [2,3].

**Methods**

**Characteristics of data**

As a regular quality control procedure, population stratification analysis using all 531,688 single-nucleotide polymorphisms (SNPs) was performed by EIGENSTRAT as we included the subjects of the four populations from HapMap database (Yoruba, CEPH, Japanese, and Han Chinese). The result showed that the case and control subjects are confirmed as European Americans. Additionally, we tested sex inconsistency between X chromosome and the clinical report and removed seven subjects whose data were discrepant. After the removal, 866 cases and 1189 controls were used for the further analysis. The 26 candidate genes were selected based on the following reasons: 1) previous reported results (HLA-DRB1, PTPN22, and TNF) [4,5]; 2) genes related to macrophage migration inhibitory factor and linked to the production of inflammatory cytokines (MIF, IL6, IL1B, IL3, ILA, and IL13) [8,9]; 3) genes playing an immunologically important role in down-regulating the immune response (CTLA4, RUNX1, STAT4, and SLC22A4) [8]; 4) genes used to test for interaction by Mei et al. and Ding et al. [6,7]; 5) genes relevant to inflammatory disease [8]. We also removed SNPs deviating from Hardy-Weinberg equilibrium ($p$-value < $10^{-5}$) and having a minor allele frequency of less than 0.01. The names, locations, and the number of SNPs being tested for the 26 genes are provided in Table 1.

**Statistical model**

The underlying principle of our method is to identify the association of allelic combination between two unlinked markers with a disease trait so that subjects are assigned an allelic score given their observed genotype information. The score is a conditional probability of obtaining the particular allelic combination given the observed genotypes at the two loci of each subject. For example, a subject with AA (at marker $M_1$) and Bb (at marker $M_2$) genotype has $1/2$ in the AB combination and $1/2$ in the Ab combination, $X_{AB} = P(AB|M_1 = AA, M_2 = Bb) = 1/2$, $X_{Ab} = P(AB|M_1 = AA, M_2 = Bb) = 1/2$. Table 2 shows the allelic scores of a subject whose genotype is given [2].

**Score statistic by logistic regression model**

Denote $y_i = 1$ if $i^{th}$ subject is affected by RA and $y_i = 0$ otherwise. In the non-parametric maximum likelihood

| Gene Symbol | Locus | No. of SNPs |
|-------------|-------|-------------|
| TNFRSF1B    | 1p36.22 | 20          |
| PADI4       | 1p36.13 | 16          |
| PTPN22      | 1p13.3  | 7           |
| FCGR3A      | 1q23.3  | 1           |
| IL1B        | 2q14    | 14          |
| ITGB6       | 2q32.1  | 16          |
| STAT4       | 2q32.3  | 35          |
| CTLA4       | 2q33    | 16          |
| IL3         | 5q31.1  | 3           |
| IL13        | 5q31.1  | 3           |
| IL4         | 5q31.1  | 4           |
| SLC22A4     | 5q31.1  | 14          |
| HAVCR1      | 5q33.3  | 13          |
| NFKBIL1     | 6p21.3  | 7           |
| HLA-DRB1    | 6p21.3  | 6           |
| LFA         | 6p21.3  | 4           |
| TNF         | 6p21.3  | 1           |
| MAP3K7IP2   | 6q25.1  | 52          |
| IL6         | 7p21    | 96          |
| TRAF1       | 9q33    | 3           |
| C5          | 9q33    | 27          |
| DLG5        | 10q22.3 | 26          |
| MS4A1       | 11q12.2 | 11          |
| CARD15      | 16q12.1 | 10          |
| RUNX1       | 21q22.12| 216         |
| MIF         | 22q11.23| 18          |
| Total       |         | 639         |

Table 1: List of genes selected for analysis

![Cell combinations of two unlinked markers; $M_1$ has A and a alleles, and $M_2$ has B and b alleles.](image)
solution that allows an arbitrary covariate distribution, fitting a standard prospective logistic regression in case-control sampling design is equivalent to fitting a retrospective logistic regression except that an intercept in case-control sampling needs the information of sampling fraction of cases and controls [10,11]. Therefore, the prospective logistic regression model is used due to the equivalence in parameter estimates of interaction effect. The likelihood function of the standard logistic regression is

\[
U(\beta_{AB}, \beta_{AB}, \beta_{ab}) = \prod_{i=1}^{n} \left[ P(Y_i=1|X_{i,AB}, X_{i,Ab}, X_{i,ab}) ]^{Y_i} \times \left[ 1 - P(Y_i=1|X_{i,AB}, X_{i,Ab}, X_{i,ab}) \right]^{1-Y_i} \right].
\]

where \( P(Y_i=1|X_{i,AB}, X_{i,Ab}, X_{i,ab}) = \frac{\exp(\alpha + \sum_{k=2}^{K} \beta_{k} X_{i,k})}{1 + \exp(\alpha + \sum_{k=2}^{K} \beta_{k} X_{i,k})} \).

\( X_{i,AB} = P(AB|M_1, M_2) \) is the allelic score of A allele and B allele from \( M_1 \) and \( M_2 \) genotypes of \( i \)th subject and \( \beta_k \) is interaction effect of \( k \)th \( \in \{ AB, Ab, aB \} \) allelic combination. The overall proportion of \( y \) is \( R/N \), where \( R \) is the number of case subjects and \( N \) is the number of total subjects. Under the assumption of no covariates, let \( U' = (U_{AB}, U_{AB}, U_{AB})' \) be the score function, which is a derivative of the log-likelihood function with respect to \( \beta = (\beta_{AB}, \beta_{Ab}, \beta_{ab}) \) respectively. Under the null hypothesis of no interaction effect \( H_0: \beta_{AB} = \beta_{Ab} = \beta_{ab} = 0 \) the efficient score test statistic under the null hypothesis is

\[
S = U'^{T}V^{-1}U \sim \chi^2_{df=3},
\]

where \( U' = \{u_{AB}, u_{Ab}, u_{ab}\}' = (\sum_{x_{i,AB}}r_{i} - \tau) \sum_{x_{i,Ab}}r_{i} - \tau) \sum_{x_{i,ab}}(r_{i} - \tau) \) and \( V^{-1} \) is the submatrix of \( I' \{\alpha, \beta_{AB}, \beta_{Ab}, \beta_{ab}\} \), which is the observed Fisher information matrix corresponding to \( U' = (U_{AB}, U_{Ab}, U_{ab})' \). Detailed derivation and theoretical justification were published by Jung and Zhao [3].

**Extension of Cochran-Armitage trend regression**

With the same allelic scores in Table 2 at the two markers, we can model a linear trend of proportion of cases over total number of subjects at each allelic combination, \( p_{k,j} = r_{k,j}/n_{k,j} \) where \( n_{k,j} = r_{k,j}/s_{k,j} \) for \( k \in \{ AB, Ab, aB, ab \} \), \( j \in \{ 0, 0.1, 0.2, 0.3 \} \) for two markers. \( r_{k,j} \) and \( s_{k,j} \) are the number of affected subjects and unaffected subjects having \( j \) score at \( k \) allelic combination, respectively. It has been shown that regressing \( p_{k,j} \) on \( Z_{AB,j}, Z_{Ab,j}, Z_{ab,j} \) is equivalent to regressing \( y_i \) on \( Z_{AB,j}, Z_{Ab,j}, Z_{ab,j} \) [12]. As an extension of Cochran-Armitage trend method, the interaction effect of two markers on RA trait can be modeled as

\[
y_i = \alpha + Z_{AB,1}\beta_{AB} + Z_{Ab,1}\beta_{Ab} + Z_{ab,1}\beta_{ab} + \epsilon_i.
\]

Under the assumption of no covariates, the theoretical regression coefficients are functions of linkage disequilibrium (LD) between a marker and a disease locus as follows:

\[
\beta = \frac{\alpha}{\beta_{AB}} = K^{-1} \left[ \begin{array}{c}
R \frac{P_{PA}}{P_{PA} + P_{PA}} + (f_{D,D} - f_{D,d} + f_{d,d}) \frac{P_{DA,DA}}{P_{DA,DA} + P_{DA,DA}} \\
\frac{P_{PA}}{P_{PA} + P_{PA}} + (f_{D,D} - f_{D,d} + f_{d,d}) \frac{P_{DA,DA}}{P_{DA,DA} + P_{DA,DA}} \\
\frac{P_{PA}}{P_{PA} + P_{PA}} + (f_{D,D} - f_{D,d} + f_{d,d}) \frac{P_{DA,DA}}{P_{DA,DA} + P_{DA,DA}} \\
\frac{P_{PA}}{P_{PA} + P_{PA}} + (f_{D,D} - f_{D,d} + f_{d,d}) \frac{P_{DA,DA}}{P_{DA,DA} + P_{DA,DA}}
\end{array} \right],
\]

where

\[
f_{D,D} - f_{D,d}, f_{D,d} - f_{d,d}, f_{d,d} - f_{d,d}
\]

are the average effect of the gene substitution at each disease loci and \( f_{D,D}, f_{D,d} \) is the magnitude of interaction effect.

The global test statistic for interaction effect over all allelic combinations under the null hypothesis \( H_0: \beta_{AB} = \beta_{Ab} = \beta_{ab} = 0 \) is

\[
F = \frac{(H\hat{\beta})'(H(X'X)^{-1}H'\hat{\beta})}{(Y'(YX')(YX')^{-1}Y)'3},
\]

which follows \( F(3, N-4) \) distribution with \( \lambda = 0 \). The analytical properties of two methods were derived by Jung and Zhao [3] and simulation studies showed that the score test and \( F \) test by Cochran-Armitage trend are asymptotically equivalent.

**Simulation study of power and type I error rates and comparison of genotype-based method**

A simulation study was performed to study power and type I error rates at the 1% significance level [3]. Six two-way interaction models were simulated using the simulation of software SNaP [3]. Most of the models were designed based on the combination of dominant and recessive inheritance at the genotypic level at each marker. These models are 1)
Under the assumption of normal errors, an artificial nesting approach called the $J$ test [14] was utilized as follows:

$$y_i = (1 - \alpha) \cdot f(X, \beta) + \alpha \cdot g(Z, \delta) + \epsilon_i.$$ (5)

Because $f$ is linear in $\beta$, the comparison requires that one estimates $\delta$ and then fits a linear regression and test for $\alpha = 0$ using the ordinary $t$-statistic $[13, 14]$. On the other hand, we can compare Akaike information criterion (AIC) and Bayesian information criterion (BIC) for a two-way model with that of a three-way interaction model.

### Analysis procedure

The procedure to search for the best interaction model consists of multiple steps based on the proposed methods.

#### Step 1

When performing a two-way interaction analysis [Model (1) and (3)] of two SNPs, each is selected from each gene and the global test for an interaction is performed. Note that two SNPs in the same gene are removed from the interaction analysis. We then compared the interaction model (3) with a main effect model (6) in order to search for the pure interacting SNPs that are not confounded with the main effects, and selected the best two-way interaction models which met three criteria: 1) the $p$-value less than $2.5 \times 10^{-7}$ from interaction test (the total 203,841 combination; adjusted for Bonferroni correction), 2) the $p$-value of the test for comparison between the interaction model and the main effect model less than 0.01, and 3) the testing SNPs should have both the smallest AIC and the smallest BIC. The following models were considered:

- No genetic effect model: $H_0: y_i = \mu + \epsilon_i$
- Main effect model: $H_{10}: y_i = \mu + X_{i, A} \beta_A + X_{i, B} \beta_B + \epsilon_i$
- Two-way model: $H_{11}: y_i = f(X, \beta) + \epsilon_i = \mu + X_{i, A} \beta_A + X_{i, B} \beta_B + X_{i, A, B} \beta_{AB} + \epsilon_i$

### Non-nested model comparison using Cochran-Armitage regression method

Technically, the proposed two-way interaction and three-way interaction model are not nested models, so an artificial nesting approach was employed to select the best model as follows:

- Two-way: $H_0: y_i = f(X, \beta) + \epsilon_i = \mu + X_{i, A} \beta_A + X_{i, B} \beta_B + X_{i, A, B} \beta_{AB} + \epsilon_i$
- Three-way: $H_0: y_i = g(Z, \delta) + \epsilon_i = \mu + \sum_{k=1}^{2^l-1} Z_{i, k} \beta_k + \epsilon_i$, $k \in \{A, B, AB\}$

### Table 3: Type I error rates and power over six two-way interaction models

| Model                | Allele-based method | Genotype-based method |
|----------------------|---------------------|-----------------------|
|                      | Score | $F$-test | Score | $F$-test | Score | Power | MDR |
| Dom $\cup$ Rec       | 1.1   | 1       | 17.3  | 16.65 | 9.85  | 6.3  |
| Modified             | 0.78  | 0.84    | 23.95 | 23.45 | 13.75 | 9.3  |
| Dom $\cap$ Dom       | 0.8   | 0.84    | 46.75 | 46.15 | 29.3  | 24.5 |
| Rec $\cup$ Rec       | 1.22  | 1.26    | 58.2  | 57.7  | 38.1  | 31.9 |
| Threshold            | 1     | 1.04    | 92    | 91.75 | 80.45 | 73.45|
| Dom $\cap$ Rec       | 0.92  | 0.94    | 96.45 | 96.3  | 88.95 | 82.6 |
Step 2  
Based on the pair-wise SNPs selected by Step 1, we conducted three-way interaction model analysis as we added one SNP at a time from one of the remaining genes, which is the scheme of the forward selection procedure. With the same procedure of a two-way model selection and an additional comparison of the three-way model with two-way interaction model, the best three-way interaction models were selected by the same criteria described in Step 1.

Step 3  
We continued these steps until no further high-dimensional interaction model was identified.

Results  
Table 4 lists ten pairs of two-way interacting genes and the characteristics of genes identified. Because we analyzed all SNPs in LD with a gene, there are multiple SNPs in a gene interacting with a SNP of the other gene. SNP rs2476601 on *PTPN22* interacts with rs2306772 on *SLC22A4*, which interacts with rs881372 on *TRAF1* and rs2900180 on *C5*, respectively. SNP rs2900180 interacts with rs2242720 on *RUNX1*, which interacts with rs881375 on *TRAF1*. SNPs rs881375 and rs2900180 are in LD ($R^2 = 0.89$). Furthermore, rs2476601 on *PTPN22* interacts with three SNPs (rs2905325, rs1476482, and rs2106549) on *IL6*. Three SNPs that are not in LD interact with two SNPs (rs477515 and rs2516049) on *TRAF1*.

### Table 4: Results of two-way (two genes) interaction and the characteristics of genes

| Gene 1 symbol (location) | Gene 2 symbol (location) | SNP1a from gene 1 | SNP 2a from gene 2 | Function of SNP 1 | Function of SNP 2 | Score | F-test | Main vs. interaction |
|--------------------------|--------------------------|-------------------|-------------------|-------------------|-------------------|-------|-------|---------------------|
| *PADI4* (1p36.13)       | *TRAF1* (9q33)           | rs6586516,        | rs3761847         | 5' UTR            | 5' UTR            | 1.66 × 10^-8 | 1.43 × 10^-8 | 0.005 |
| *PTPN22* (1p13.3)       | *SLC22A4* (5q31.1)       | rs2477142,        | rs2306772         | Coding intron     | Coding intron     | 1.79 × 10^-12 | 1.25 × 10^-12 | 0.0054 |
| *PTPN2* (1p13.3)        | *IL6* (7p21)             | rs2476601,        | rs2905325,        | Coding intron     | 5' UTR            | 3.80 × 10^-13 | 2.53 × 10^-13 | 0.0003 |

*The SNPs listed under each gene are in LD (within $R^2 > 0.8$).

The smallest $p$-value of each combination is reported.

rs4947324 on *NFKBIL1* and three SNPs (rs477515, rs2516049, rs532098) are not in LD.

rs4816502 and rs2242720 on *RUNX1* are not in LD, and the function of rs4816502 is intron, that of rs2242720 is 5' UTR, respectively.
Figure 2
Graphical view of the interaction of two-way and three-way interaction (red).

HLA-DRB1. SNP rs660895 on the same gene interacts with rs2834779 on the 5’ UTR region on RUNX1, and rs660895 and rs532098 interact with four SNPs on RUNX1. Additionally, rs4947324 on NFKB1 is not in LD with three SNPs on HLA-DRB1, but it interacts with them. Two SNPs (rs6586516 and rs2477142) on PADI4 interact with rs3761847 on TRAF1, which interacts with five SNPs in LD on RUNX1. Furthermore, we detected three-way interacting genes which are rs10229203 on IL6, rs4816502 on RUNX1 and rs10818500 on C5. Figure 2 summarized the pathway of ten pairs genes and one group of three interacting genes (indicated in red).

Discussion
In this study, the allele-based gene-gene interaction analyses were applied to case-control data sets of RA. Based on the analysis with 639 SNPs from 26 candidate genes that were previously detected through linkage study or fine mapping, we identified ten two-way interacting genes with multiple SNPs in LD from a gene and one three-way interaction. We have not identified any four-way interaction effects. However, the 26 candidate genes selected in this study may not represent all candidate genes for RA and we observed that Illumina 550k chip may not have a good gene-wide coverage for SNPs because no SNPs of SUMO4 and VEGFA in the platform are available.

A more standard interaction model using a logistic regression consisting of two main effect terms (X = 0, 1, 2 according to the number of alleles) and a multiplicative term of the main effect (additive × additive) was applied to the same data set. There is no interacting SNPs by an even more lenient criteria (p-value<10^{-5}).

Three criteria to justify the significant interaction models were used. For the interaction models, Bonferroni correction was used for multiple testing, and for the comparison of the interaction model with a main-effect model (significance level of 0.01), the smaller AIC and BIC were utilized. The final selected interacting SNPs satisfied all of the criteria, which may be conservative and may cause false-negative error. There still remains the issue of multiple comparisons in the high-dimensional interactions and the complexity of the procedure to screen the interaction effects.

Conclusion
As shown in the results, the proposed allele-based approach allows us to identify multiple interactions that may not have been identified as risk factors for RA. PTPN22, SLC22A4, HLA-DRB1, IL6, PADI4, TRAF1, NFKB1, C5, and RUNX1 may play interactive roles for RA, especially PTPN22 and SLC22A4, which are related to the reaction of antigen for RA. Therefore, our method taking into account the nonrandom association of all allelic combinations may help detect novel genetic variants and interpret biological pathways.

List of abbreviations used
AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; GAW16: Genetic Analysis Workshop 16; LD: Linkage disequilibrium; MDR: Multifactor dimensionality reduction; NARAC: North American Rheumatoid Arthritis Consortium; RA: Rheumatoid arthritis; SNP: Single-nucleotide polymorphism.

Competing interests
The authors declare that have no competing interests.

Authors’ contributions
JJ developed statistical models, performed the analysis, and wrote the manuscript. IJS checked SAS/IML codes that JJ has written. DK contributed on the interpretation of the analysis.

Acknowledgements
The Genetic Analysis Workshops are supported by NIH grant R01 GM031575 from the National Institute of General Medical Sciences.

This article has been published as part of BMC Proceedings Volume 3 Supplement 7, 2009: Genetic Analysis Workshop 16. The full contents of the supplement are available online at http://www.biomedcentral.com/1753-6561/3/S7.

References
1. Dieudo P and Gornelis F: Genetic basis of rheumatoid arthritis. Joint Bone Spine 2005, 72:520–526.
2. Jung J, Sun B, Kwon D, Koller DL and Foroud TM: Allelic based gene-gene interaction association with quantitative traits. Genet Epidemiol 2009, 33:323–343.
3. Jung J and Zhao Y: Allelic based gene-gene interaction association in case-control study. Hum Hered in press.
4. Kallberg H, Padyukov L, Plenge RM, Ronnelid J, Gregersen PK, Helm-van Mil van der AH, Toes RE, Huizinga TW, Klarenkog L, Alfredsson L and Epidemiological Investigation of Rheumatoid Arthritis study group: Gene-gene and gene-environment interactions involving HLA-DRB1,
PTPN22, and smoking in two subsets of rheumatoid arthritis. Am J Hum Genet 2007, 80:867–875.

5. Mu H, Chen JJ, Jiang Y, King MC, Thomson G and Criswell LA: Tumor necrosis factor a microsatellite polymorphism is associated with rheumatoid arthritis severity through an interaction with the HLA-DRB1 shared epitope. Arthritis Rheum 1999, 42:438–442.

6. Mei L, Li X, Yang K, Cui J, Fang B, Guo X and Rotter JI: Evaluating gene × gene and gene × smoking interaction in rheumatoid arthritis using candidate genes in GAW 15. BMC Proc 2007, 1(suppl 1):S17.

7. Ding Y, Cong L, Ionita-Laza I, Lo SH and Zheng T: Constructing gene association networks for rheumatoid arthritis using the backward genotype-trait association (BGTA) algorithm. BMC Proc 2007, 1(suppl 1):S13.

8. Phelan JD, Thompson SD and Glass DN: Susceptibility to JRA/JIA: complementing general autoimmune and arthritis traits. Genes Immun 2006, 7:1–10.

9. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, Liew A, Khalili H, Chandrasekaran A, Davies LR, Li W, Tan AK, Bonnard C, Ong RT, Thalamuthu A, Pettersson S, Liu C, Tian C, Chen WY, Carulli JP, Beckman EM, Altshuler D, Alfredsson L, Criswell LA, Amos CI, Seldin MF, Kastner DL, Klarskov L and Gregersen PK: TRAF1-C5 as a risk locus for rheumatoid arthritis—a genomewide study. N Engl J Med 2007, 357:1199–1208.

10. Anderson JA: Separate sample logistic discrimination. Biometrika 1972, 59:19–35.

11. Prentice RL and Pyke R: Logistic disease incidence models and case-control studies. Biometrika 1979, 66:403–411.

12. Armitage P: Tests for linear trends in proportions and frequencies. Biometrics 1955, 11:375–386.

13. McAleer M: Exact tests of a model against non-nested alternatives. Biometrika 1983, 70:285–288.

14. Warton M, Johnson W and Bedrick EJ: Non-nested linear model selection revisited. Commun Statist 2001, 30:1–20.