Acknowledgements We thank Rachel Morra, AP-HP/UPMC liver center, Hôpital Pitié-Salpêtrière, Paris, France, and la Banque de Cellules - Cochin APHP with all its staff for their logistic help and for providing patient lymphoblastoid cell lines. We are grateful to David-Alexandre Tregouët for helpful discussions about the haplotype analysis, and to the Centre National de Génotypage (CNG) for its help in genotyping. We also thank all members of the laboratory of Human Genetics of Infectious Diseases for fruitful discussions.

Funding This work was supported by grants from Institut National de la Santé et de la Recherche Médicale (4CH08G), Japanese Society for Promotion of Science (JSPS)/Inserm cooperation agreement, and Agence Nationale de Recherche sur le SIDA et les hépatites virales (ANRS HC EP13). LA was supported, in part, by a grant from Assistance Publique-Hôpitaux de Paris.

Competing interests None.

Ethics approval This study was conducted with the approval of the institutional review board (CPP: Comité de Protection des Personnes) of Ile de France - Paris - Saint Antoine, on 5 March 2002.

Contributors All authors participated in the collection, the management, and the interpretation of clinical data, and in the writing and the final approval of the manuscript. BN, ML, CB, TP, FM, SP and LA designed the study. SP, BR, EP, SH and LA were involved in statistical analysis. RLM and FM were involved in the sequencing of IFNGR2.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Lauer GM,** Walker BD. Hepatitis C virus infection. *N Engl J Med* 2001;**345**:41–52.
2. **Shepard CW,** Finelli L, Alter MJ. Global epidemiology of hepatitis C virus infection. *Lancet Infect Dis* 2005;**5**:558–67.
3. **Dustin LB,** Rice CM. Flying under the radar: the immunobiology of hepatitis C. *Annu Rev Immunol* 2007;**25**:71–99.
4. **Marcellin P,** Asselah T, Boyer N. Fibrosis and disease progression in hepatitis C. *Hepatology* 2002;**36**:S47–56.
5. **Missiha SB,** Ostrowski M, Heathcote EJ. Disease progression in chronic hepatitis C: modifiable and nonmodifiable factors. *Gastroenterology* 2008;**134**:1699–714.
6. **Poynard T,** Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. *Lancet* 1997;**349**:825–32.
7. **Poynard T,** Ratziu V, Charlotte F, *et al*. Rates and risk factors of liver fibrosis progression in patients with chronic hepatitis c. *J Hepatol* 2001;**34**:730–9.
8. **Ortiz V,** Berenguer M, Rayon JM, *et al*. Contribution of obesity to hepatitis C-related fibrosis progression. *Am J Gastroenterol* 2002;**97**:2408–14.
9. **Pol S,** Fontaine H, Carnot F, *et al*. Predictive factors for development of cirrhosis in parenterally acquired chronic hepatitis C: a comparison between immunocompetent and immunocompromised patients. *J Hepatol* 1998;**29**:12–19.
10. **Osterreicher CH,** Stickel F, Brenner DA. Genomics of liver fibrosis and cirrhosis. *Semin Liver Dis* 2007;**27**:28–43.
11. **Bataller R,** North KE, Brenner DA. Genetic polymorphisms and the progression of liver fibrosis: a critical appraisal. *Hepatology* 2003;**37**:493–503.
12. **Huang H,** Shiffman ML, Cheung RC, *et al*. Identification of two gene variants associated with risk of advanced fibrosis in patients with chronic hepatitis C. *Gastroenterology* 2006;**130**:1679–87.
13. **Huang H,** Shiffman ML, Friedman S, *et al*. A 7 gene signature identifies the risk of developing cirrhosis in patients with chronic hepatitis C. *Hepatology* 2007;**46**:297–306.
14. **Bataller R,** Brenner DA. Liver fibrosis. *J Clin Invest* 2005;**115**:209–18.
15. **Friedman SL.** Molecular regulation of hepatic fibrosis, an integrated cellular response to tissue injury. *J Biol Chem* 2000;**275**:2247–50.
16. **Lotersztajn S,** Julien B, Teixeira-Clerc F, *et al*. Hepatic fibrosis: molecular mechanisms and drug targets. *Annu Rev Pharmacol Toxicol* 2004;**45**:605–28.
17. **Bedossa P,** Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996;**24**:289–93.
18. **Fan JB,** Oliphant A, Shen R, *et al*. Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol* 2003;**68**:69–78.
19. **Dizier MH,** Bouzigon E, Guilloud-Bataille M, *et al*. Genome screen in the French EGEA study: detection of linked regions shared or not shared by allergic rhinitis and asthma. *Genes Immun* 2005;**6**:95–102.
20. **Takahashi M,** Matsuda F, Margetic N, *et al*. Automated identification of single nucleotide polymorphisms from sequencing data. *J Bioinform Comput Biol* 2003;**1**:253–65.
21. **Barrett JC,** Fry B, Maller J, *et al*. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;**21**:263–5.
22. **Tregouet DA,** Garelle V. A new JAVA interface implementation of THESIAS: testing haplotype effects in association studies. *Bioinformatics* 2007;**23**:1038–9.
23. **Hnatyszyn HJ.** Chronic hepatitis C and genotyping: the clinical significance of determining HCV genotypes. *Antivir Ther* 2005;**10**:1–11.
24. **Bedossa P,** Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology* 2003;**38**:1449–57.
25. **Falleti E,** Fabris C, Toniutto P, *et al*. Genetic polymorphisms of inflammatory cytokines and liver fibrosis progression due to recurrent hepatitis C. *J Interferon Cytokine Res* 2007;**27**:239–46.
26. **Dessein AJ,** Hillaire D, Elwali NE, *et al*. Severe hepatic fibrosis in Schistosoma mansoni infection is controlled by a major locus that is closely linked to the interferon-gamma receptor gene. *Am J Hum Genet* 1999;**65**:709–21.
27. **Blanton RE,** Salam EA, Ehsan A, *et al*. Schistosomal hepatic fibrosis and the interferon gamma receptor: a linkage analysis using single-nucleotide polymorphic markers. *Eur J Hum Genet* 2005;**13**:660–8.
28. **Shi Z,** Wakil AE, Rockey DC. Strain-specific differences in mouse hepatic wound healing are mediated by divergent T helper cytokine responses. *Proc Natl Acad Sci U S A* 1997;**94**:10663–8.
29. **Mallat A,** Preaux AM, Blazejewski S, *et al*. Interferon alfa and gamma inhibit proliferation and collagen synthesis of human Ito cells in culture. *Hepatology* 1995;**21**:1003–10.
30. **Bajwa EK,** Ayas NT, Schulzer M, *et al*. Interferon-gamma1b therapy in idiopathic pulmonary fibrosis: a metaanalysis. *Chest* 2005;**128**:203–6.
31. **Henri S,** Chevillard C, Mergani A, *et al*. Cytokine regulation of periportal fibrosis in humans infected with Schistosoma mansoni: IFN-gamma is associated with protection against fibrosis and TNF-alpha with aggravation of disease. *J Immunol* 2002;**169**:929–36.
32. **Arnaud V,** Li J, Wang Y, *et al*. Regulatory role of interleukin-10 and interferon-gamma in severe hepatic central and peripheral fibrosis in humans infected with Schistosoma japonicum. *J Infect Dis* 2008;**198**:418–26.
33. **Napoli J,** Bishop GA, McGuinness PH, *et al*. Progressive liver injury in chronic hepatitis C infection correlates with increased intrahepatic expression of Th1-associated cytokines. *Hepatology* 1996;**24**:759–65.
34. **Fieschi C,** Dupuis S, Picard C, *et al*. High levels of interferon gamma in the plasma of children with complete interferon gamma receptor deficiency. *Pediatrics* 2001;**107**:E48.
35. **Bach EA,** Szabo SJ, Dighe AS, *et al*. Ligand-induced autoregulation of IFN-gamma receptor beta chain expression in T helper cell subsets. *Science* 1995;**270**:1215–18.
36. **Bernabei P,** Coccia EM, Rigamonti L, *et al*. Interferon-gamma receptor 2 expression as the deciding factor in human T, B, and myeloid cell proliferation or death. *J Leukoc Biol* 2001;**70**:950–60.
37. **Poynard T,** Muntenau M, Morra R, *et al*. Methodological aspects of the interpretation of non-invasive biomarkers of liver fibrosis: a 2008 update. *Gastroenterol Clin Biol* 2008;**32**:8–21.
38. **Guidotti LG,** Chisari FV. Immunobiology and pathogenesis of viral hepatitis. *Annu Rev Pathol* 2006;**1**:23–61.
39. **Thimme R,** Oldach D, Chang KM, *et al*. Determinants of viral clearance and persistence during acute hepatitis C virus infection. *J Exp Med* 2001;**194**:1395–406.

Genetic Epidemiology 34:543–551 (2010)

Estimation of *P*-value of MAX Test with Double Triangle Diagram for 2 × 3 SNP Case-Control Tables

Katsura Hirosawa, Takahisa Kawaguchi, Fumihiko Matsuda, and Ryo Yamada*

Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan

Single nucleotide polymorphisms (SNPs) are the most popular markers in genetic epidemiology. Multiple tests have been applied to evaluate genetic effect of SNPs, such as Pearson's test with two degrees of freedom, three tests with one degree of freedom (χ^2 tests for dominant and recessive modes and Cockran-Armitage trend test for additive mode) as well as MAX3 test and MAX test, which are combination of four tests mentioned earlier. Because MAX test is a combination of Pearson's test of two degrees of freedom and two tests of one degree of freedom, the probability density function (pdf) of MAX statistics does not match pdf of χ^2 distribution of either one or two degrees of freedom. In order to calculate *P*-value of MAX test, we introduced a new diagram, Double Triangle Diagram, which was an extension of de Finetti diagram in population genetics which characterized all of the tests for 2 × 3 tables. In the diagram the contour lines of MAX statistics were consisted of elliptic curves and two tangent lines to the ellipses in the space. We normalized the ellipses into regular circles and expressed *P*-value of MAX test in an integral form. Although a part of the integral was not analytically solvable, it was calculable with arbitrary accuracy by dividing the area under pdf into finite rectangles. We confirmed that *P*-values from our method took uniform distribution from 0 to 1 in three example marginal count sets and concluded that our method was appropriate to give *P*-value of MAX test for 2 × 3 tables. *Genet. Epidemiol.* 34:543–551, 2010.

Key words: SNP; MAX test; association study; trend test

*Correspondence to: Ryo Yamada, Yoshida-konoecho, Sakyo-ku, Kyoto 606-8501, Japan. E-mail: ryamada@genome.med.kyoto-u.ac.jp
Received 9 December 2009; Revised 16 March 2010; Accepted 29 March 2010
Published online 17 August 2010 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/gepi.20510

INTRODUCTION

Genetic epidemiology has been one of the most active research fields in genetics. Since the human genome project published the reference sequence of human beings, genome-wide case-control association studies (GWAS) have been carried out on a large scale with remarkable results. In GWAS, single nucleotide polymorphisms (SNP) have been used as principal genetic markers. In individuals, SNPs have two alleles, major (M) and minor (m), and three genotypes, MM, Mm and mm. Therefore, case-control studies in GWAS consist of 2 × 3 contingency tables for the two groups (case and control) and three genotypes. The technology of molecular genetics has been progressing very rapidly and SNPs are no longer the only genetic markers to be tested in GWAS studies [Balding, 2006]. However, the importance of 2 × 3 tables has not become obsolete, because any genetic factor in DNA can be evaluated with 2 × 3 tables in case-control studies.

For 2 × 3 tables, Pearson's test of two degrees of freedom can be applied. When three categories are in order, the Cockran-Armitage trend test (CAT) of one degree of freedom is the best choice. In many cases in genetics, it is reasonable to consider that the risk of the heterozygous type (Mm) is between the risks of two homozygous types (MM and mm). Therefore, CAT has been frequently used for analyzing the additive effect, which considers the middle category as the average of the other two categories. However, dominant and recessive effects are also well known in genetics and these effects are tested frequently with 2 × 2 tables in which the risk of the heterozygous type is considered the same as the risk of two homozygous types [Balding, 2006; Cochran, 1954]. Sometimes the MAX3 test is used, which consists of three tests (CAT and dominant and recessive tests) and adopts the maximum of the three as its statistical value. In fact, MAX3 test was used with successful identification of disease-associated markers in a genome-wide association study (GWAS) [Sladek et al., 2007]. Alternatively, the MAX test or the optimal mode trend test (OMTT test) can be used [Campbell, 2005]. The MAX method or the OMTT method tests all modes between dominant and recessive including the additive mode. The MAX and OMTT methods are conceptually the same and the OMTT offered exact calculation of *P*-value of the test. In fact, all the abovementioned tests are trend tests with different types of scores [Yamada and Okada, 2009; Zheng et al., 2009]. Because the MAX test best represents the genetic hypothesis in many situations and has the highest power among these tests under the hypothesis, it seems to be the best test for 2 × 3 table tests for SNP genetic studies. However, the *P*-value of the MAX test is not analytically calculable, which is a drawback of the test. Although we previously proposed a method to calculate the exact *P*-value [Yamada and Okada, 2009], it requires a high computational load.

A method to approximation of *P*-value of the MAX test was proposed by Li et al. with good performance [Li et al., 2008]. In this paper, we introduce a diagram to display a 2×3 table test in which the contour lines of the MAX test are drawn as a combination of an ellipse and its tangent lines, and we propose a method to estimate the *P*-value of the MAX test using the diagram.

DOUBLE TRIANGLE DIAGRAM: A GEOMETRIC LAYOUT OF 2×3 TABLES IN TWO-DIMENSIONAL SPACE

DOUBLE TRIANGLE DIAGRAM AS AN EXTENSION OF THE DE FINETTI DIAGRAM

A de Finetti diagram is a ternary plot to graph the genotype frequencies of populations, where there are two alleles and the population is diploid. The diagram locates the conditions of genotype frequencies in an equilateral triangle. It is based on Viviani's theorem that at any point within the triangle, given the three lines from that point that are perpendicular to the sides of the triangle, the sum of the lengths of the lines is a fixed value, regardless of the position of the point [Cannings and Edwards, 1968].

Because the marginal counts of a 2×3 table are given, the sum of three genotypes is fixed for both cases and controls. Therefore, three genotype counts of each group can be plotted as a point in an equilateral triangle. In our double triangle diagram, two triangles for cases and controls are drawn and 2×3 tables are plotted as described below.

Let $\tau = \{\{n_{11}, n_{12}, n_{13}\}, \{n_{21}, n_{22}, n_{23}\}\}$ denote the observed table and its cell counts and $m = \{n_{1.}, n_{2.}, n_{.1}, n_{.2}, n_{.3}, n_{..}\}$ denote its marginal counts. Let $t_e = \{\{e_{11}, e_{12}, e_{13}\}, \{e_{21}, e_{22}, e_{23}\}\}$ denote the expected table and its counts under the null assumption and $\tau - t_e = \{\{d_{11}, d_{12}, d_{13}\}, \{d_{21}, d_{22}, d_{23}\}\}$ denote the difference between τ and t_e. Then, we have the following tables.

	AA	*Aa*	*aa*	*Sum*
Case	n_{11}	n_{12}	n_{13}	$n_{1.}$
Control	n_{21}	n_{22}	n_{23}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{..}$

$n_{ij} \geq 0$

The difference between τ and t_e is as follows:

	AA	*Aa*	*aa*	Sum
Case	e_{11}	e_{12}	e_{13}	$n_{1.}$
Control	e_{21}	e_{22}	e_{23}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{..}$

$$e_{ij} = n_{i.} n_{.j} / n_{..}.$$

	AA	*Aa*	*aa*
Case	d_{11}	d_{12}	$d_{13} = -(d_{11} + d_{12})$
Control	$d_{21} = -d_{11}$	$d_{22} = -d_{12}$	$d_{23} = -d_{13} = d_{11} + d_{12}$

$$d_{ij} = n_{ij} - e_{ij} \geq -e_{ij}$$

$$\sum_{i=1}^{2} d_{ij} = 0, \sum_{j=1}^{3} d_{ij} = 0.$$

We introduce following coordinates:

$$(x, y) = \left(d_{12}, \frac{1}{\sqrt{3}} (d_{11} - d_{13}) \right). \quad (1)$$

These co-ordinates are based on the idea that geometric arrangement of multiple categories as below. Two categories are expressed as two vectors in opposite directions to each other. They are in one-dimensional space. Three categories are expressed as three vectors in two-dimensional space and their three tips are the vertices of a regular triangle. Four categories make a regular tetrahedron. In general, k categories make $k-1$ simplex in $k-1$ dimensional space, which has k vertices. Equation (1) is one way to give Cartesian coordinates to vertices of regular triangle (2 simplex for 3 categories). With these co-ordinates,

	AA	*Aa*	*aa*
Case	$d_{11} = -\frac{1}{2}x + \frac{\sqrt{3}}{2}y$	$d_{12} = x$	$d_{13} = -\frac{1}{2}x - \frac{\sqrt{3}}{2}y$
Control	$d_{21} = -d_{11} = \frac{1}{2}x - \frac{\sqrt{3}}{2}y$	$d_{22} = -d_{12} = -x$	$d_{23} = -d_{13} = \frac{1}{2}x + \frac{\sqrt{3}}{2}y.$

Because $n_{ij} \geq 0$ and because $d_{1j} + d_{2j} = 0$ and $d_{1j} \geq -e_{1j}$ and $d_{2j} \geq -e_{2j}$, therefore $d_{1j} \geq -e_{1j}$ and $-d_{2j} = d_{1j} \leq -e_{2j}$,

$$-e_{1j} \leq d_{1j} \leq e_{2j},$$

which can be re-written as,

$$\frac{1}{\sqrt{3}}x - \frac{2}{\sqrt{3}}e_{11} \leq y \leq \frac{1}{\sqrt{3}}x + \frac{2}{\sqrt{3}}e_{21}$$

$$-e_{12} \leq x \leq e_{22}$$

$$-\frac{1}{\sqrt{3}}x - \frac{2}{\sqrt{3}}e_{23} \leq y \leq -\frac{1}{\sqrt{3}}x + \frac{2}{\sqrt{3}}e_{13}$$

These three equations demarcate the field with three sets of parallel lines which make two equilateral triangles.

Figure 1A is the diagram of the following table, in which the case-to-control ratio is 1.5 and the total allele frequency is 0.2 and the samples are in Hardy-Weinberg equilibrium. The size of triangles is proportional to the sample size of groups and the larger and the smaller triangles represent controls and cases, respectively.

	AA	*Aa*	*aa*	*Sum*
Case	$n_{11} = 150$	$n_{12} = 520$	$n_{13} = 330$	$n_{1.} = 1{,}000$
Control	$n_{21} = 250$	$n_{22} = 680$	$n_{23} = 570$	$n_{2.} = 1{,}500$
Total	$n_{.1} = 400$	$n_{.2} = 1{,}200$	$n_{.3} = 900$	$n_{..} = 2{,}500$

DISTRIBUTION OF TEST STATISTICS IN DOUBLE-TRIANGLE DIAGRAM

CONTOUR LINES OF TEST STATISTICS

For 2×3 contingency tables in SNP case-control association studies, the above-mentioned multiple tests, Pearson's genotype test of two degrees of freedom, the three tests of one degree of freedom for additive, dominant and recessive mode, the MAX test [Zheng et al., 2009]

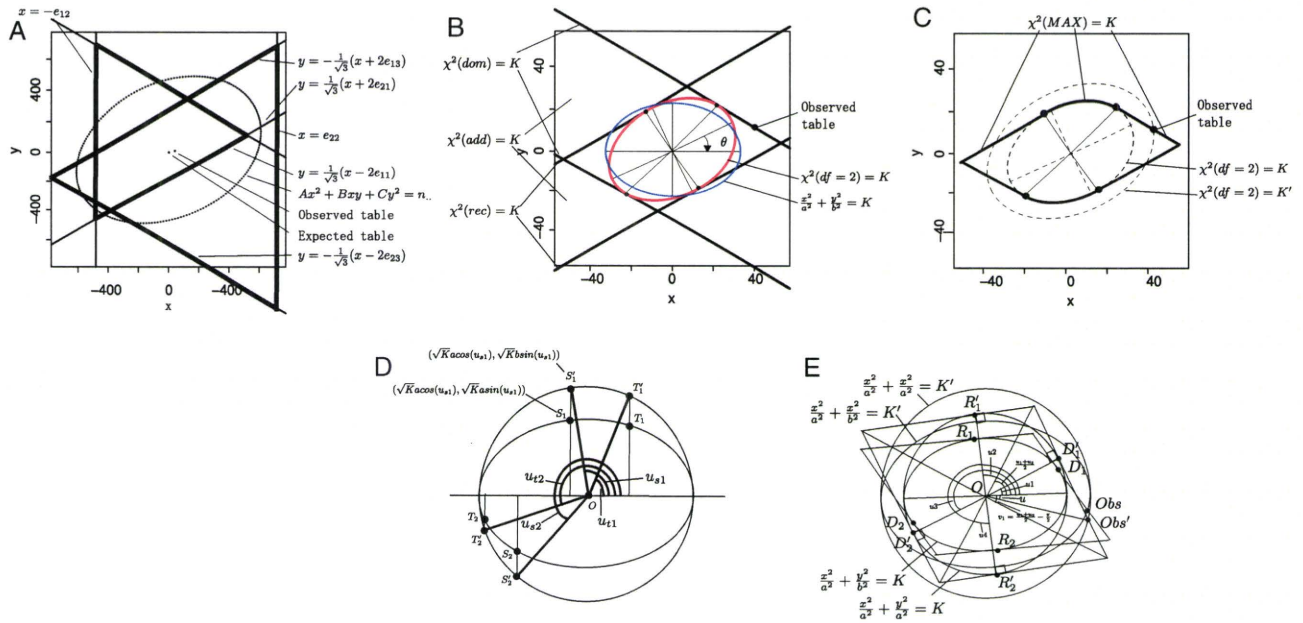


Fig. 1. Double triangle diagrams and contour lines of statistical tests. (A) The space of the tables sharing the marginals. The origin (0,0) represents the expected table and the neighboring point is the table given in the text. Six lines demarcate the space where tables exist, which share the marginals with the observed table. The overlap of two triangles is a pentagon. The dashed ellipse is the contour curve of χ^2(df2) = n. (B) The central area of (A) is enlarged. The solid ellipse of χ^2(df2) = K with its major and minor axis is shown. The solid ellipse is rotated by θ to the dashed ellipse, which is in the normalized form, $(x^2/a^2)+(y^2/b^2) = K$. Three lines, χ^2(dom) = K, χ^2(add) = K and χ^2(rec) = K are tangent to χ^2(df2) = K. The gradients of the lines are $-(1/\sqrt{3})$, 0 and $1/\sqrt{3}$, respectively. The observed table is indicated by a dot on χ^2(dom) = K and outside of the ellipse, χ^2(df = 2) = K. (C) The contour line of χ^2(MAX) = K consists of χ^2(df = K), χ^2(dom) = K and χ^2(rec) = K. In the observed table, χ^2(df2) = K' is more than K. The contour line of the ellipse, χ^2(df2) = K', is indicated by a larger dashed ellipse. Because χ^2(dom) and χ^2(df2) of the observed table are K and K', respectively, the dot of the observed table is the intersection of χ^2(dom) = K and χ^2(df2) = K'. Description on K' appears in the section "Geometric evaluation of χ^2(MAX). Therefore ignore K' when this figure was indicated earlier in the main text. (D) The ellipse $(x^2/a^2)+(y^2/b^2) = K$ is enlarged by $\frac{a}{b}$ in the y-axis direction and it becomes a circle, $(x^2/a^2)+(y^2/b^2) = K$. The coordinates of S_i, S_i', T_i and T_i' are parameterized with a, b, K, K' and u_i. The ratio of the area of the sector in the ellipse and the corresponding sector in the circle is b/a. In the figure, $u_{s1} - u_{t1} = u_{s2} - u_{t2}$. Therefore, the areas of the two sectors, S_1OT_1 and S_2T_2, are the same. (E) The ellipses, $(x^2/a^2)+(y^2/b^2) = K$ and $(x^2/a^2)+(y^2/b^2) = K'$, and the corresponding circles, enlarged by a/b in the y-axis direction, are drawn in the left and right sides, respectively.. Four tangent points on the ellipse $(x^2/a^2)+(y^2/b^2) = K$, D_1, D_2, R_1 and R_2 and their corresponding points on the circle, D_1', D_2', R_1' and R_2', are plotted. The Ds and Rs are on the lines of χ^2(dom) = K and χ^2(rec) = K, respectively. Four tangent lines to the ellipse form a parallelogram. The enlargement in the y-axis moves the lines tangent to the ellipse to the lines tangent to the corresponding circle. The tangent lines to the circle form a rhombus and the radii to D's and R's, and the tangent lines are perpendicular. Two diagonals quadrisect the rhombus. Obs and Obs′ are the point of the observed table and its corresponding point in the enlarged circle, respectively. Because the observed table's χ^2(MAX) is K and its χ^2(df = 2) is K', Obs is the intersection of the straight part of the contour line of χ^2(MAX) = K and the ellipse $(x^2/a^2)+(y^2/b^2) = K'$. Lines Obs′D_1 and OD_1 are perpendicular and angle Obs′$D_1 = u_1 - u$. The length is $D_1O = \sqrt{K}a$ and the distance is $Obs'O = \sqrt{K'}a$. Therefore, $\sqrt{K'}a/\sqrt{K}a = \cos(u_1 - u)$.

[or the OMTT test; Yamada and Okada, 2009], can be applied. As reported, these tests are all trend tests with different types of scores, as expressed below.

$$\chi^2(\text{df2}) = \max(Y^2(\{0, r, 1\}); -\infty \le r \le \infty)$$
$$\chi^2(\text{dom}) = Y^2(\{0, 0, 1\})$$
$$\chi^2(\text{rec}) = Y^2(\{0, 1, 1\})$$
$$\chi^2(\text{add}) = Y^2(\{0, 0.5, 1\})$$
$$\chi^2(\text{MAX}) = \max(Y^2(\{0, r, 1\}); 0 \le r \le 1)$$

where

$$Y^2(\{0, r, 1\}) = \frac{n_{..}^2}{n_{1.}n_{2.}} \frac{(d_{11}(-W)+d_{12}(r-W)+d_{13}(1-W))^2}{n_{.1}(-W)^2+(n_{.2}(r-W)^2+n_{.3}(1-W))^2}.$$
$$W = \frac{n_{.2} \times r + n_{.3}}{n_{..}}$$

Excluding χ^2(MAX), they are expressed with x and y as shown below.

$$\chi^2(\text{df2}) = \frac{n_{..}^2}{n_{1.}n_{2.}} \left(\frac{1}{4}\left(\frac{1}{n_{.1}} + \frac{1}{n_{.3}}\right) + \frac{1}{n_{.2}}\right) x^2 + \frac{\sqrt{3}}{2}\left(\frac{1}{n_{.3}} - \frac{1}{n_{.1}}\right)xy + \frac{3}{4}\left(\frac{1}{n_{.1}} + \frac{1}{n_{.3}}\right)y^2)$$
$$\chi^2(\text{dom}) = \frac{n_{..}^2}{n_{1.}n_{2.}} \left(\frac{1}{n_{.1}+n_{.2}} + \frac{1}{n_{.3}}\right)\left(-\frac{1}{2}x - \frac{\sqrt{3}}{2}y\right)^2 \quad (2)$$
$$\chi^2(\text{rec}) = \frac{n_{..}^2}{n_{1.}n_{2.}} \left(\frac{1}{n_{.1}} + \frac{1}{n_{.2}+n_{.3}}\right)\left(-\frac{1}{2}x + \frac{\sqrt{3}}{2}y\right)^2$$
$$\chi^2(\text{add}) = \frac{n_{..}^2}{n_{1.}n_{2.}} \frac{n_{.1}n_{.2}n_{.3}}{\frac{1}{4}\left(\frac{1}{n_{.1}} + \frac{1}{n_{.3}}\right) + \frac{1}{n_{.2}}} \left(\frac{\sqrt{3}}{2}y\right)^2$$

Let K denote χ^2(MAX) of the observed table. Figure 1B indicates the contour lines of χ^2(df2) = K, χ^2(dom) = K, χ^2(rec) = K and χ^2(add) = K. Figure 1C indicates the contour lines of χ^2(MAX) = K. The contour line of χ^2(df2) is an ellipse and the contour lines of χ^2(dom), χ^2(rec) and χ^2(add) are pairs of parallel lines. The lines of χ^2(dom) and χ^2(rec) are parallel to the lines of the triangles. The lines of χ^2(add) are horizontal. The contour lines of χ^2(MAX) consist of the elliptic curve of χ^2(df2) and the straight lines of χ^2(dom) and χ^2(rec).

The contour lines of χ^2(dom) = K, χ^2(rec) = K and χ^2(add) = K are tangent to the ellipse, which can be shown by the simple transformation of equations (not shown).

ELLIPSE NORMALIZATION

In general, the ellipse can be normalized by rotation.

The ellipse of χ^2(df2) = K is normalized to $(x^2/a^2)+(y^2/b^2) = K$, $(a \geq b)$ by rotating θ in the clockwise direction as shown below (Fig. 1B).

$$\begin{aligned}
&\frac{x^2}{a^2}+\frac{y^2}{b^2}=K\\
&a=\sqrt{\frac{2}{A+C-\sqrt{B^2+(A-C)^2}}}\\
&b=\sqrt{\frac{2}{A+C+\sqrt{B^2+(A-C)^2}}}\\
&\theta=\frac{1}{2}sin^{-1}\frac{B}{A-C}\\
&A=\frac{n^2_{..}}{n_{1.}n_{2.}}\left(\frac{1}{4}\left(\frac{1}{n_{.1}}+\frac{1}{n_{.3}}\right)+\frac{1}{n_{.2}}\right)\\
&B=\frac{n^2_{..}}{n_{1.}n_{2.}}\frac{\sqrt{3}}{2}\left(\frac{1}{n_{.3}}-\frac{1}{n_{.1}}\right)\\
&C=\frac{n^2_{..}}{n_{1.}n_{2.}}\frac{3}{4}\left(\frac{1}{n_{.1}}+\frac{1}{n_{.3}}\right)
\end{aligned} \tag{3}$$

When the ellipse is normalized, the coordinates of the points on the ellipse are given as $(\sqrt{K}a \times \cos(u), \sqrt{K}b \times \sin(u))$.

We enlarge the figure by a/b in the y-axis direction, and the ellipse becomes a regular circle. The coordinates of the points change from $(\sqrt{K}a \times \cos(u), \sqrt{K}b \times \sin(u))$ to $(\sqrt{K}a \times \cos(u), \sqrt{K}a \times \sin(u))$ (Fig. 1D).

The tangent lines to the ellipse change their gradients but remain tangent to the circle (Fig. 1E).

GEOMETRIC EVALUATION OF X^2(MAX)

The contour line of χ^2(MAX) = K consists of the tangent lines of the dominant and recessive models and the elliptic curve (see Fig. 1E). The four tangent points, R_1, R_2, D_1 and D_2 are on the ellipse and their locations are given as $(\sqrt{K}a\cos(u_i), \sqrt{K}b\sin(u_i))$. The observed table is on the larger ellipse and its location is given as $(\sqrt{K'}a\cos(u), \sqrt{K'}b\sin(u))$. The observed table is indicated as Obs at the intersection of the ellipse and the tangent line from one of the four tangent points, D_1. After enlargement in the y-axis direction, Obs and D_1 are moved to Obs′ and D_1, respectively. Because the radius and tangent line of a regular circle are perpendicular, the line Obs′−D_1 is perpendicular to the radius to D_1. Therefore,

$$\frac{\sqrt{K}a}{\sqrt{K'}a}=\cos(u_i-u),$$

and,

$$K'(u|u_i)=\frac{K}{(\cos(u_i-u))^2}.$$

We will use $K'(u)$ instead of $K'(u \mid u_i)$ for simplicity.

The χ^2(df2) values of points on χ^2(MAX) = K, $K'(u)$ are

$$K'(u)=\begin{cases}K & (u_1 \leq u \leq u_2, u_3 \leq u \leq u_4)\\ \min\left(\dfrac{K}{(\cos(u_i-u))^2}\right) & \text{otherwise}\end{cases} \tag{4}$$

GEOMETRIC CALCULATION OF *P*-VALUE OF MAX TEST

The area of the ellipse, $(x^2/a^2)+(y^2/b^2) = K$, is $A = \pi Kab$ (see Fig. 1D).

Assume two sectors of the ellipse, S_1OT_1 and S_2OT_2 $(S_i = (\sqrt{K}a\cos(u_{si}), \sqrt{K}b\sin(u_{si}))$ and $T_i = (\sqrt{K}a\cos(u_{ti}), \sqrt{K}b\sin(u_{ti})))$. When $u_{s1}-u_{t1} = u_{s2}-u_{t2}$, the area of the sectors, $((u_{s1}-u_{t1})/2\pi)A$ and $((u_{s2}-u_{t2})/2)A$, are equal.

Let $\Pr(x,y)$ denote the probability density function (pdf) of χ^2(df = 2) in two-dimensional space. The P-value of χ^2(df = 2) = K, $P_{df2}(K) = e^{-\frac{K}{2}}$ is given as,

$$P_{df2}(K)=\int_{\frac{x^2}{a^2}+\frac{y^2}{b^2}\geq K}\Pr(x,y)\,dxdy. \tag{5}$$

Because $\Pr(x_i,y_i) = \Pr(x_j,y_j)$, when

$$\frac{x_i^2}{a^2}+\frac{y_i^2}{b^2}=\frac{x_j^2}{a^2}+\frac{y_j^2}{b^2}$$

the integral of $\Pr(x,y)$ in the sector S_1OT_1 is,

$$\begin{aligned}&\int_{\frac{x^2}{a^2}+\frac{y^2}{b^2}\geq K, u_{t1}\leq u\leq u_{s1}}\Pr(x,y)\frac{u_{s1}-u_{t1}}{2\pi}\,dxdy\\ &=P_{df2}(K)\frac{u_{s1}-u_{t1}}{2\pi}.\end{aligned} \tag{6}$$

This can be applied to du as,

$$\int_{\frac{x^2}{a^2}+\frac{y^2}{b^2}\geq K(u_t),u_t\leq u\leq u_t+du}\Pr(x,y)\frac{du}{2\pi}dxdy=P_{df2}(K(u_t))\frac{du}{2\pi}. \tag{7}$$

For the points on the contour line χ^2(MAX) = K, χ^2(df = 2) is not constant because it is a function of u, $K'(u)$ (Equation (4)). Figure 2A,B shows $K'(u)$ and $P_{df2}(K(u))=e^{-\frac{K(u)}{2}}$.

The P-value of the MAX test when χ^2(MAX) = K, $P_{MAX}(K)$ is geometrically defined as,

$$P_{MAX}(K)=\int_0^{2\pi}P(K'(u))\frac{du}{2\pi}. \tag{8}$$

Because the four tangent lines form a rhombus in the enlarged coordinates where the ellipse is the regular circle,

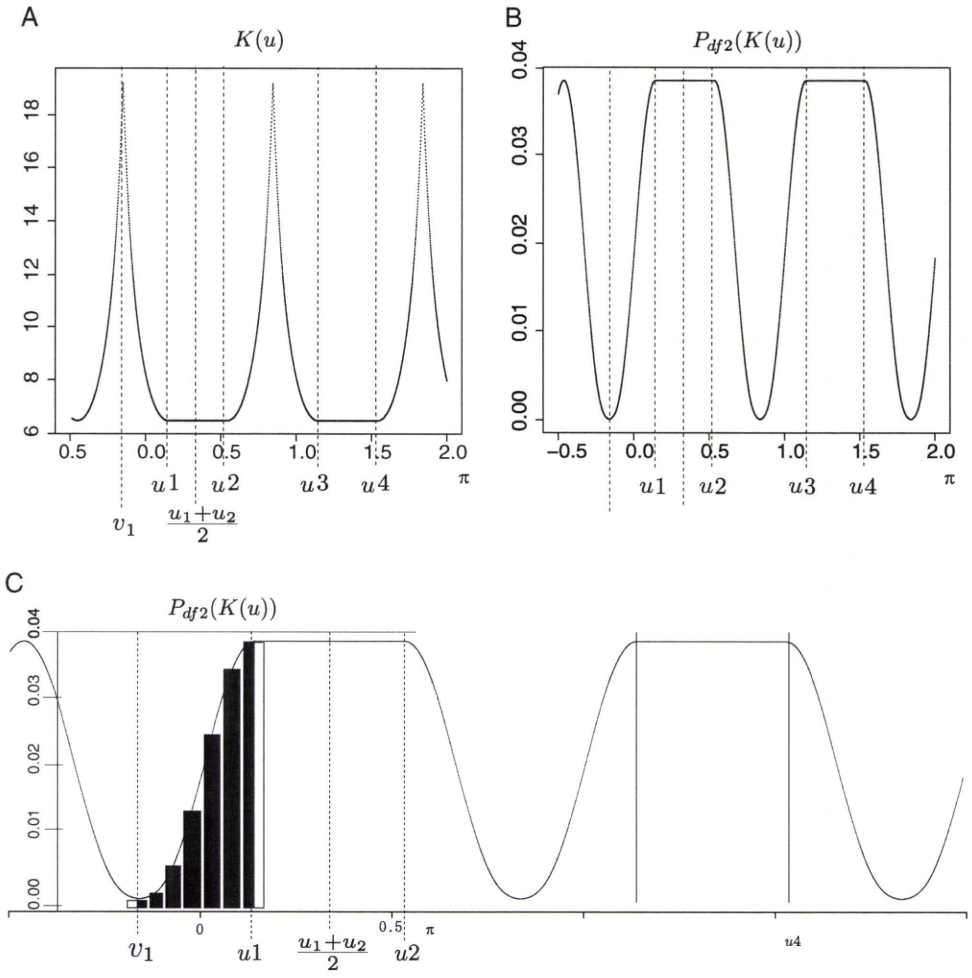


Fig. 2. Plots of χ^2(df = 2) and $P_{df2}(K)$ on the contour line of χ^2(MAX) = K to the circular angles. (A) $K'(u)$, χ^2(df2) values of points on the contour line χ^2(MAX) = K = 6.51 are plotted along the axis of $\frac{u}{\pi}$. The line is cyclic. Between u_1 and u_2 and between u_3 and u_4, the line is flat at 6.51; otherwise, $K'(u)$ is a curve larger than the value of the flat segments. (B) $P_{df2}(K(u))$ were plotted on the same horizontal axis. The flat segments correspond to the flat segments in (A). Otherwise, the line is a curve smaller than the value of the flat segments. (C) A quarter from $v_1 = \frac{u_1+u_2}{2} - \frac{\pi}{2}$ to $\frac{u_1+u_2}{2}$ of (B) is enlarged. The black bars represent the area of equation (9) for estimating $P_{MAX}(K)$. The height of the bars is $P_{MAX}(K(u)$ of the midpoint. The width of the right- and left-most black bars are half of the others.

and because a rhombus consists of four congruent regular triangles, a quarter of the rhombus can be considered. The first quarter of the rhombus of $\chi^2(\mathrm{MAX}) = K$ is parameterized with u from $v_1 = ((u_1+u_2)/2) - \pi/2$ to $v_1+(\pi/2) = (u_1+u_2)/2$ (Fig. 1E). Therefore,

$$P_{\mathrm{MAX}}(K) = 4\int_{v_1}^{v_1+\frac{\pi}{2}} P_{\mathrm{df2}}(K'(u))\frac{\mathrm{d}u}{2\pi}$$

$$= 4\int_{v_1}^{u_1} P_{\mathrm{df2}}\left(\frac{K}{(\cos(u_1-u))^2}\right)\frac{\mathrm{d}u}{2\pi} + \int_{u_1}^{v_1+4\frac{\pi}{2}} P_{\mathrm{df2}}(K)\frac{\mathrm{d}u}{2\pi}.$$

The first term of the right side is

$$4\int_{v_1}^{u_1} P_{\mathrm{df2}}\left(\frac{K}{(\cos(u_1-u))^2}\right)\frac{\mathrm{d}u}{2\pi} = \frac{2}{\pi}\int_{v_1}^{u_1} \mathrm{e}^{-\frac{K}{2(\cos(u_1-u))^2}}\mathrm{d}u.$$

This integral cannot be analytically solved. The second term of the right side is

$$4\int_{u_1}^{v_1+\frac{\pi}{2}} P_{\mathrm{df2}}(K)\frac{\mathrm{d}u}{2\pi} = \frac{2}{\pi}P(K)\left(v_1+\frac{\pi}{2}-u_1\right) = \frac{1}{\pi}P(K)(u_2-u_1).$$

Therefore,

$$P(\chi^2(\mathrm{MAX}) = K) = \frac{1}{\pi}\left(2\int_{v_1}^{u_1} \mathrm{e}^{-\frac{K}{2(\cos(u_1-u))^2}}\frac{\mathrm{d}u}{2\pi} + \mathrm{e}^{-\frac{K}{2}}(u_2-u_1)\right).$$

Although the first term cannot be analytically solved, it can be calculated by summing the thin rectangles, as expressed below (Fig. 2C):

$$\frac{u_1-t_1}{2\pi N}\left(P_{\mathrm{df2}}(K'(u_1-t_1)) + P_{\mathrm{df2}}(K'(u_1)) \times \frac{1}{2} + \sum_{i=1}^{N-1} P_{\mathrm{df2}}\left(K'\left(u_1-\left(1-\frac{i}{N}\right)t\right)\right)\right). \quad (9)$$

The precision can be adjusted by changing N. In the following calculation, the first N started with 2 and it was repeatedly doubled until the difference of the estimated $P_{MAX}(K)$ by each update of N was less than $P_{df2}(K) \times 10^{-3}$. For example, when $P_{df2}(K) \times 10^{-4}$, calculation is continued until the difference between iterations χbecomes less than $10^{-4-3} = 10^{-7}$.

VALIDATION OF THE METHOD TO CALCULATE $P(\chi^2(MAX) = K)$

Previous reports have confirmed that χ^2(MAX) ranks the observed tables appropriately under the condition that r is in the hypothesized range [Yamada and Okada, 2009; Zheng et al., 2009]. Because P-values need to be observed in uniform distribution from 0 to 1 when tests are repeated under the null hypothesis, the appropriateness of the method to estimate $P_{MAX}(K)$ proposed above is confirmed by observing that the distribution of $P(\chi^2(MAX))$ is uniform for multiple tables sampled under the null hypothesis. As Equation (4) indicates, χ^2(df2) is a function of u_i, which determines the location of the tangent points. Therefore, the eccentricity of the ellipse, $\sqrt{1-(b/a)^2}$, and the angle θ affect the estimation of $P(\chi^2(MAX)) = K$. Both b/a and θ are parameterized only by n_j based on Equation (3) (details are not shown). So, we selected three marginal count sets as examples $(n_{\cdot 1}, n_{\cdot 2}, n_{\cdot 3}) = (3333, 3333, 3334), (100, 9000, 900)$ and (9000,100,900), with $(n_{1\cdot}, n_{2\cdot}) = (5000, 5000)$, for evaluation of the method to estimate $P_{MAX}(K)$. Figure 3A shows the double triangle diagrams of the three examples. First, (3333,3333,3334) has similar values for the three categories and its diagram gives an ellipse that is almost a regular circle and the fraction where $\chi^2(MAX) = \chi(df = 2)$ is approximately one-third. In the second example, (100,9000,900) has a very large value of $n_{\cdot 2}$ and its diagram gives an ellipse that is long in the vertical axis and the fraction where $\chi^2(MAX) = \chi(df = 2)$ is very small. In the third example, (9000,100,900) has a very large value of $n_{\cdot 1}$ instead and its diagram gives an ellipse that is long in the horizontal axis and the fraction where $\chi^2(MAX) = \chi(df = 2)$ is very large. A total of 1,000 tables that had the marginal counts were randomly sampled for each example and P-P plots of $P_{MAX}(K)$ and $P_{df2}(K)$ were drawn on a linear scale and logarithmic scale (Fig. 3A). The P-P plots for P_{MAX} and P_{df2} indicate that they are uniformly distributed. There is no difference among the examples, which indicates that the difference of eccentricity or rotation of the ellipses does not affect the distributions. Figure 3B displays the relationship among P_{df2}, P_{MAX}, P_{dom}, P_{rec} and P_{add} for the 1,000 table samples of (3333,3333,3334). The correlation between P_{df2} and P_{MAX} is the strongest. Although three tests of one degree of freedom show considerable correlation with P_{MAX}, a fraction of the tables show a substantial difference.

COMPARISON OF THE POWER OF NEW METHOD WITH PEARSON'S TEST

We compared power of our method with Pearson's test and asymptotic estimation of P-value of MAX by Zang

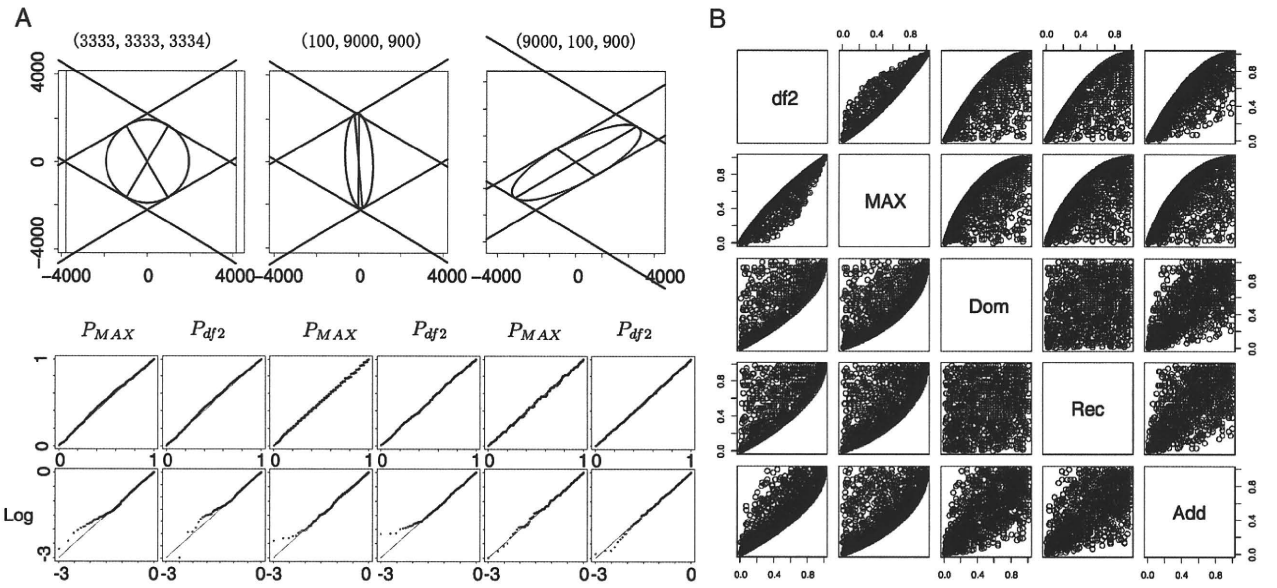


Fig. 3. Distribution of P_{MAX} for three sets of marginal counts. (A) A double triangle diagram and four P-P plots below it are drawn for three marginal counts, (3333,3333,3334), (100,9000,900) and (9000,100,900), as indicated. The two darker lines in each double triangle diagram indicate the lines connecting tangent points. They separate the sections where the contour line of $\chi^2(MAX) = K$ corresponds to the ellipse, the elliptic sections (the upper and the lower sections) and the sections where the contour line of $\chi^2(MAX) = K$ corresponds to the tangent lines (the right and left sections). The angles where the contour line of $\chi^2(MAX) = K$ is a part of ellipse is middle, narrow and wide for three examples, respectively. For each marginal count set, 1000 tables were randomly simulated under the null hypothesis. (B)Four P-P plots of P_{MAX} and P_{df2} were drawn below each diagram. The left-side plots are P_{MAX} and the right-side plots are P_{df2}. The upper plots are on a linear scale and the lower plots are on a logarithmic scale.

et al. in "Rassoc" package in CRAN (http://cran.r-project.org/), for six genetic models [Zang et al., 2010]. For each model, Hardy-Weinberg equilibrium was assumed in a population and allele frequency of risk allele in the population and prevalence of phenotype were set at 0.3 and 0.01, respectively. Genotypic relative risks for each model were given as (1.5,1.5,1), (1.5,1.25,1), (1.5,1,1) for dominant, additive and recessive models, respectively. A model between dominant and additive (half dominant model) and a model between additive and recessive (half recessive model) were also defined with (1.5,1.375,1) and (1.5,1.125,1). Relative risk of the sixth model (heterozygote-specific model) was given as (1,1.5,1). Five hundreds of cases and five hundreds of controls were randomly sampled from the population and 1,000 2×3 tables were created for each model and tested with three tests. The result was shown in Table I. Powers of proposed method and MAX3 with asymptotic P-value were almost identical for all models. Power of Pearson's test was less powerful than the other two for all models.

APPLICATION TO REAL GENOTYPE DATA

In response to Decision Letter, we applied our proposing method to two types of real GWAS study data. The first data were 17 SNPs that were reported with statistical significance in three papers and that were used by Li et al. to evaluate their method to approximate P-value of MAX test [Li et al., 2008]. The second data were 10,000 SNPs among WTCCC study for rheumatoid arthritis [The Wellcome Trust Case Control Consortium, 2007]. The ten thousands SNPs were selected from the top of the list of markers in the order of chromosomal location. For the first 17 SNPs, we applied our new method (P_{MAX}) and asymptotic P-value estimation of MAX3 test [($P_{MAX3asy}$ Zang et al., 2010 ("Rassoc" package in CRAN (http://cran.r-project.org/)], and the exact P of the MAX or the OMTT [Yamada and Okada, 2009] (P_{OMITex}) and they were shown in Table II with P-value based on Rhombus formula $P_{rhombus}$. In the report by Li et al., they compared $P_{rhombus}$ with empirical P-values of bootstraps and permutations. In this report, we adopted P_{OMITex}, instead. As shown in Table II, all methods gave similar values for all SNPs.

For 10,000 SNPs in WTCCC study, we compared P_{MAX}, $P_{MAX3asy}$ and P_{OMITex}. Figure 4A, B shows their co-plots in regular and logarithmic scales, respectively. P_{MAX} showed stronger correlation with P_{OMITex} than $P_{MAX3asy}$. Figure 4C–F shows relation of difference between P_{MAX} and P_{OMITex} or $P_{MAX3asy}$ and P_{OMITex} with allele frequency or the minimum value of 2×3 table cells. The differences between P_{MAX} and P_{OMITex} or $P_{MAX3asy}$ and P_{OMITex} were both larger when allele frequency of minor allele was smaller and the minimum value of table was smaller. When the minimum value of table was very small, any asymptotic method deviates from the exact method and this was why some of P_{MAX} were relatively more deviated from P_{OMITex}. However, excluding these exceptions, P_{MAX} tended to give closer value to P_{OMITex} than $P_{MAX3asy}$ than P_{MAXasy} regardless of the minimum of table cells. It

TABLE I. Power comparison of proposed method and Pearson's test and MAX3

	Recessive			Half recessive			Additive			Half dominant			Dominat			Heterozygout-specric		
Cut off	Proposed	Pearson's	MAX3	Proposed	Pearson's	MAX3	Proposed	Pearson's	MAX3	Proposed	Pearson's	MAX3	Proposed	Pearson's	MAX3	Proposed	Pearson's	MAX3
0.01	198	157	204	224	186	217	332	288	330	460	406	468	651	599	660	445	392	461
0.001	54	37	54	63	48	55	130	102	125	200	170	203	342	322	378	196	166	211
0.0001	14	11	15	14	12	14	41	31	45	72	60	74	189	164	191	65	55	78
0.00001	5	4	4	4	1	4	11	9	11	26	21	30	82	60	93	21	17	22
0.000001	1	0	1	0	0	0	3	3	3	8	6	9	22	18	25	8	6	8
0.0000001	0	0	0	0	0	0	1	0	1	0	0	2	9	8	11	2	2	2

Number of tables with P less than cut-off values in 1,000 simulations.

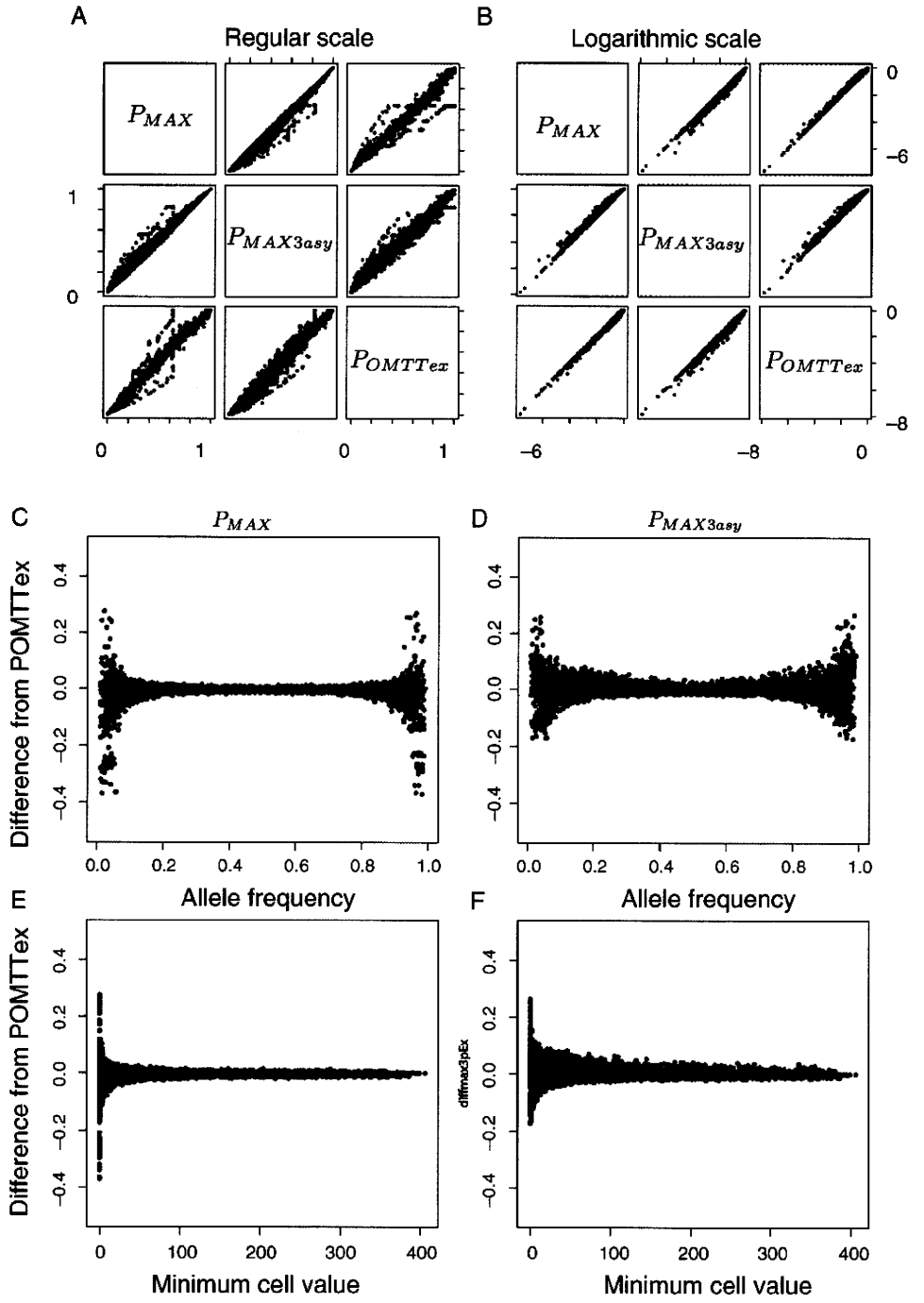


Fig. 4. P_{MAX}, $P_{MAX3asy}$ and P_{OMITex} of 10,000 SNP tables in a GWAS were compared. (A) and (B) are co-plots among P_{MAX}, $P_{MAX3asy}$ and P_{OMITex} in regular and logarithmic scale. All three were well correlated and P_{MAX} and P_{OMITex} were better. (C) and (D) Differences of P_{MAX} and $P_{MAX3asy}$ from P_{OMITex} were plotted along allele frequency. (E) and (F) Differences of P_{MAX} and $P_{MAX3asy}$ from P_{OMITex} were plotted along the minimum value of 2×3 table cells.

seemed reasonable that P_{MAX} performed relatively better with tables of low minor allele frequency or of small minimal value of cells, because the ellipses of those tables tend to have high eccentricity, to which our method was designed to correct.

DISCUSSION

In this paper, the de Finetti diagram, a diagram for genotype frequency of diallelic markers in population genetics, was applied to 2×3 contingency tables for case-control

TABLE II. *P*-values of identified SNPs in GWASs of diabetes, breast, and prostate cancers

SNPID	A11	A12	A22	B11	B12	B22	PMAX	PMAXasy	PRhombus	POMTTex
8 confirmed SNPs associated with Type 2 diabetes										
rs7903146	197	348	149	335	254	65	6.26E-19	0.00E+00	1.58E-18	3.73E-19
rs13266634	54	229	411	53	293	307	2.40E-05	1.89E-05	1.84E-05	2.27E-05
rs1111875	77	302	315	119	308	227	8.44E-06	7.04E-06	6.78E-06	7.74E-06
rs7923837	66	300	328	116	296	242	1.52E-06	2.36E-06	2.28E-06	1.43E-06
rs7480010	301	327	66	363	246	45	2.56E-05	2.24E-05	2.18E-05	2.46E-05
rs3740878	25	273	386	65	249	353	2.43E-05	1.89E-05	1.84E-05	1.94E-05
rs11037909	25	274	387	65	251	353	2.44E-05	1.89E-05	1.85E-05	2.00E-05
rs1113132	25	271	390	63	251	355	5.30E-05	4.22E-05	4.12E-05	4.11E-05
6 reported SNPs associated with breast cancer										
rs10510126	955	180	10	854	272	14	2.21E-06	1.41E-06	1.42E-06	1.63E-06
rs12505080	608	477	50	628	408	99	1.06E-04	8.46E-05	8.27E-05	1.02E-04
rs17157903	777	316	18	862	220	26	8.41E-05	6.17E-05	6.20E-05	7.56E-05
rs1219648	352	543	250	433	538	170	4.08E-06	4.99E-06	4.80E-06	3.80E-06
rs7696175	353	605	187	396	496	249	2.30E-03	2.07E-03	1.98E-03	2.37E-03
rs2420946	357	546	242	440	537	165	4.83E-06	5.34E-06	5.14E-06	4.73E-06
3 reported SNPs associated with prostate cancer										
rs1447295	25	283	864	10	218	929	1.17E-04	1.09E-04	1.10E-04	1.00E-04
rs6983267	351	598	223	277	579	301	1.95E-05	2.16E-05	2.06E-05	1.93E-05
rs7837688	861	283	27	939	206	11	9.10E-06	6.66E-06	6.67E-06	9.06E-06

association tests with SNPs, and a novel diagram, called the double triangle diagram, was proposed. Using the new diagram, test statistics of 2×3 tables were geometrically described. Given that the tests for 2×3 tables, Pearson's test, dominant test, recessive test, additive test and MAX test are all trend tests with different scores, the results were interpreted within this context. The occurrence probability distribution in the diagram space was elliptic, and the eccentricity and rotation were functions of the marginal counts. Once the distribution of the probability density was expressed in algebraic geometry, it was easy to transform the ellipse into a regular circle. In the normalized figure, integration of the probability was easy even when the integral could not be solved analytically. Subsequently, the integration of probability of the MAX test was implemented and validated by a simulation in which the geometrically estimated *P*-values of the MAX test were in a uniform distribution. Although our method performed well in terms of type I error and power and gave close value to the exact MAX test, there were limitations. In recent GWASs, tests of SNP data have to consider covariates in many cases and actually the rhombus method proposed by Li et al. [2008] was designed to be able to adjust for the covariates. However, our approach was not applicable to such conditions. Besides the handling of covariates, we evaluated our geometric approach only for 2×3 tables and no further extensions to other statistical tests of higher dimensions and more complexed data structure were investigated.

The web tool to estimate the *P*-values of the MAX test is available from the author's web site along with the R source code (http://www.genome.med.kyoto-u.ac.jp/wiki_tokyo/index.php/EllipseMAXP).

REFERENCES

Balding DJ. 2006. A tutorial on statistical methods for population association studies. Nat Rev Genet 7:781–791.

Campbell M. 2005. $\chi 2$ test for linear trend—what's that? Midwifery 32:127–130.

Cannings C, Edwards AWF. 1968. Natural selection and the de Finetti diagram. Ann Hum Gen 31:421–428.

Cochran WG. 1954. Some methods for strengthening the common $\chi 2$ tests. Biometrics 10:417–451.

Li Q, Zheng G, Li Z, Yu K. 2008. Efficient approximation of P-value of the maximum of correlated tests, with applications to genome-wide association studies. Ann Hum Genet 72:397–406.

Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P. 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445:881–885.

The Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678.

Yamada R, Okada Y. 2009. An optimal dose-effect mode trend test for SNP genotype tables. Genet Epidemiol 33:114–127.

Zang Y, Wing F K, Zheng G. 2010. Simple algorithms to calculate asymptotic null distributions of robust tests in case-control genetic association studies in R. J Stat 33:1–24.

Zheng G, Joo J, Yand Y. 2009. Pearson's test, trend test, and MAX are all trend tests with different types of scores. Ann Hum Genet 73:133–140.

RHEUMATOLOGY

Rheumatology 2010;49:2298–2304
doi:10.1093/rheumatology/keq273
Advance Access publication 9 September 2010

Original article

Anti-citrullinated peptide antibody-negative RA is a genetically distinct subset: a definitive study using only bone-erosive ACPA-negative rheumatoid arthritis

Koichiro Ohmura[1], Chikashi Terao[1,2], Etsuko Maruya[3], Masaki Katayama[1], Kenichiro Matoba[4], Kota Shimada[5], Akira Murasawa[6], Shigeru Honjo[7], Kiyoshi Takasugi[4], Shigeto Tohma[5], Keitaro Matsuo[8], Kazuo Tajima[8], Naoichiro Yukawa[1], Daisuke Kawabata[1], Takaki Nojima[1], Takao Fujii[1], Ryo Yamada[2], Hiroo Saji[3], Fumihiko Matsuda[2] and Tsuneyo Mimori[1]

Abstract

Objectives. ACPA is a highly specific marker for RA. It was recently reported that ACPA can be used to classify RA into two disease subsets, ACPA-positive and ACPA-negative RA. ACPA-positive RA was found to be associated with the HLA-DR shared epitope (SE), but ACPA negative was not. However, the suspicion remained that this result was caused by the ACPA-negative RA subset containing patients with non-RA diseases. We examined whether this is the case even when possible non-RA ACPA-negative RA patients were excluded by selecting only patients with bone erosion.

Methods. We genotyped HLA-DRB1 alleles for 574 ACPA-positive RA, 185 ACPA-negative RA (including 97 erosive RA) and 1508 healthy donors. We also tested whether HLA-DR SE is associated with RF-negative or ANA-negative RA.

Results. ACPA-negative RA with apparent bone erosion was not associated with SE, supporting the idea that ACPA-negative RA is genetically distinct from ACPA-positive RA. We also tested whether these subsets are based on autoantibody-producing activity. In accordance with the ACPA-negative RA subset, the RF-negative RA subset showed a clearly distinct pattern of association with SE from the RF-positive RA. In contrast, ANA-negative as well as ANA-positive RA was similarly associated with SE, suggesting that the subsets distinguished by ACPA are not based simply on differences in autoantibody production.

Conclusions. ACPA-negative erosive RA is genetically distinct from ACPA-positive RA.

Key words: Rheumatoid arthritis, Anti-citrullinated peptide antibody, HLA, Shared epitope, Subset, Genetics, Association study.

[1]Department of Rheumatology and Clinical Immunology, [2]Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, [3]HLA Laboratory, Kawabata-Marutamachi, Sakyo-ku, Kyoto, [4]Department of Rheumatology, Dohgo Spa Hospital, Dohgo Himezuka, Matsuyama, [5]Department of Rheumatology, Sagamihara National Hospital, National Hospital Organization, Sakuradai, Sagamihara, [6]Department of Rheumatology, Niigata Rheumatic Center, Shibata, Niigata, [7]Department of Rheumatology, Saiseikai Takaoka Hospital, Takaoka, Toyama and [8]Division of Epidemiology and Prevention, Aichi Cancer Center Hospital and Research Institute, Chikusa-ku, Nagoya, Japan.

Submitted 30 March 2010; revised version accepted 19 July 2010.

Correspondence to: Koichiro Ohmura, Department of Rheumatology and Clinical Immunology, Graduate School of Medicine, Kyoto University, 54 Shogoin-Kawahara-cho Sakyo-ku, Kyoto 606-8507, Japan. E-mail: ohmurako@kuhp.kyoto-u.ac.jp

Introduction

RA is an inflammatory arthritic disorder that is characterized by inflammatory cell infiltration, synovial cell proliferation and destruction of cartilage and subcartilageous bones, which can lead to joint deformity. However, the clinical course of RA varies from patient to patient, as do autoantibody profiles such as RF and ACPA. Such heterogeneity may be derived from genetic and environmental factors. In the early stage of arthritis, the diagnosis of RA is often difficult and such patients can be classified as

SCIENCE

undifferentiated arthritis (UA). According to Thabet *et al.* [1], about half of the UA patients remit spontaneously, while ~30% develop RA. At baseline, 28.6% of UA patients have bone erosions and it is a good prognostic value for the development of RA.

ACPA is an autoantibody that recognizes peptides or proteins whose arginine residues are changed to citrulline by post-translational modification. The target protein is not a single protein, but filaggrin [2], vimentin [3], fibrin [4], α-enolase [5] and so on. ACPA is a useful diagnostic marker for RA because of its very high specificity (>95%) and reasonably high sensitivity (65–88%) [6–8]. It has also been proposed that ACPA is a useful marker for predicting destructive RA [9, 10].

Genetic predisposition to RA has been investigated intensively. HLA is a major determinant of RA susceptibility and *HLA-DRB1 *0101, *0102, *0401, *0404, *0405, *0408, *0410, *1001, *1303* and **1402* are reported to be associated with RA development. There is a common amino acid sequence among such HLA-DR molecules at the 70th–74th residues of the HLA-DRβ1 chain, which is called a shared epitope (SE) [11]. The association of carrying this SE and developing RA has been repeatedly reported for different ethnic groups. However, recently, a Dutch group reported that the association of SE was only exhibited with ACPA-positive RA and no association was seen with the ACPA-negative RA patients [12]. They also showed that the influence of SE on joint damage was abrogated when stratified by ACPA. In addition to *HLA-DRB1* (SE), other RA susceptibility genes such as *PTPN22, CTLA4, TRAF1/C5* and *STAT4* were also investigated for association by stratifying RA with ACPA [13–16]. In almost all cases, such susceptibility genes were found to be associated with ACPA-positive RA but not with ACPA-negative RA. Although genetic differences are clear between ACPA-positive and ACPA-negative RA, there still remains the possibility that such differences might be caused by the contamination of non-RA diseases such as seronegative SpA and PMR in the ACPA-negative RA subset. In this article, we re-evaluated the association analysis by selecting only patients with bone-eroding arthritis for the ACPA-negative population.

Materials and methods

Patients and healthy control subjects

A total of 1411 patients who were diagnosed with RA in five hospitals (Kyoto University Hospital, Dohgo Spa Hospital, Sagamihara National Hospital, Niigata Rheumatic Center and Saiseikai Takaoka Hospital) were enrolled in this study. All patients were Japanese and fulfilled the ACR (formerly ARA) 1987 revised criteria for the classification of RA. RA patients overlapped with other collagen vascular diseases were excluded. SS was not excluded because the prevalence of SS in our cohort was quite low (<2%) compared with the reported prevalence of 10–24%, probably due to incomplete clinical information. The ethics committee of each hospital approved the study and genomic DNA was extracted from peripheral blood of patients and healthy individuals after written informed consent was obtained. Out of 1411 RA patients, 1182 (83.8%) were ACPA positive and 229 (16.2%) were ACPA negative. Five hundred and seventy-four ACPA-positive and 185 ACPA-negative RA patients were selected and genotyped for *HLA-DRB1*. Out of the 185 ACPA-negative RA patients, radiographic data were available in 160 patients, of whom 97 patients had typical bone erosions. Such patients are denoted as ACPA-negative erosive RA patients in this article. DNA samples from 1508 healthy control subjects were collected at Aichi Cancer Center Hospital and from the DNA banks for healthy Japanese volunteers of the Pharma SNP Consortium [17] after written informed consent was obtained.

Genotyping and autoantibody detection

HLA-DRB1 genotyping was carried out with a high-throughput, high-resolution genotyping method (WAKFlow WAKUNAGA) by combining PCR and sequence-specific oligonucleotide probe protocols with the Luminex 100 xMAP flow cytometry dual-laser system to quantify fluorescently labelled oligonucleotides attached to colour-coded microbeads. The following *HLA-DRB1* alleles were classified as SE positive: *DRB1*0101, *0102, *0401, *0404, *0405, *0408, *0413, *0416, *1001, *1303* and **1402*.

ACPA in sera or plasma was detected using a second-generation anti-CCP antibody (Ab) ELISA kit (MESACUP CCP; Medical & Biological Laboratories Co. Ltd, Nagoya, Japan) in accordance with the manufacturer's instructions. A cut-off value of 4.5 U/ml was used for anti-CCP Ab positivity. RF was quantified by latex immunoturbidimetry and the cut-off values of each detection kit in each hospital were employed. ANA was semi-quantified by IIF for most samples, but some were measured by ELISA (MESACUP ANA; Medical & Biological Laboratories Co. Ltd). The cut-off values of each hospital were employed.

Statistical analysis

Chi-squared test, Student's *t*-test, Jonckheere–Terpstra trend test and the 95% CI of odds ratio (OR) were used to assess the statistical significance and magnitude of association for categorical outcomes.

Results

ACPA-positive RA is distinct from ACPA-negative RA on the basis of SE association

One hundred and eighty-five ACPA-negative patients and 574 ACPA-positive patients, as well as 1508 healthy individuals, were genotyped for *HLA DRB1*. SE was determined as described in the 'Materials and methods' section. ACPA was not tested for healthy individuals because its positivity among healthy people was reported to be only ~1% [6, 18]. As shown in Tables 1 and 2, SE was the clear risk factor for ACPA-positive RA development ($P = 8.7 \times 10^{-32}$ and 5.3×10^{-28} for double- and

Table 1 Association of SE with ACPA-positive or ACPA-negative RA

SE status	Control ($n = 1508$)	ACPA-positive RA ($n = 574$)		ACPA-negative RA ($n = 185$)		ACPA-negative erosive RA ($n = 97$)	
	n (%)	*n* (%)	OR (95% CI)	*n* (%)	OR (95% CI)	*n* (%)	OR (95% CI)
SE (+/+)	74 (5)	93 (16)	6.6 (4.7, 9.3)	6 (3)	0.7 (0.3, 1.7)	3 (3)	0.7 (0.2, 2.3)
SE (+/−)	492 (33)	302 (53)	3.2 (2.6, 4.0)	71 (38)	1.3 (0.9, 1.7)	39 (40)	1.4 (0.9, 2.1)
SE (−/−)	942 (62)	179 (31)	1.0	108 (58)	1.0	55 (57)	1.0

SE (+/+): double-SE carrier; SE (+/−): single-SE carrier; SE (−/−): no SE carrier.

Table 2 *P*-values for association of SE between each group

Groups for comparison	*P*-value		
	SE (+/+) *vs* SE (/)	SE (+/) *vs* SE (/)	SE (+) *vs* SE ( )
Control *vs* ACPA-positive RA	8.7×10^{-32}	5.3×10^{-28}	1.8×10^{-37}
Control *vs* ACPA-negative RA	0.43	0.16	0.28
Control *vs* ACPA-negative erosive RA	0.54	0.16	0.26
ACPA-positive RA *vs* ACPA-negative RA	2.9×10^{-9}	1.0×10^{-7}	3.3×10^{-11}
ACPA-positive RA *vs* ACPA-negative erosive RA	1.0×10^{-5}	1.2×10^{-4}	1.1×10^{-6}
ACPA-negative RA *vs* ACPA-negative erosive RA	0.98	0.77	0.79

P-values were calculated by chi-squared test.

single-SE carriers, respectively), but not for ACPA-negative RA development ($P = 0.43$ and 0.16 for double- and single-SE carriers, respectively). There was also a dose effect of SE number for ACPA-positive RA (ORs were 6.6 and 3.2 for double- and single-SE carriers, respectively), but not for ACPA-negative RA (ORs were 0.71 and 1.3 for double- and single-SE carriers, respectively). When combining the double- and single-SE carriers, *P*-values for ACPA-positive RA *vs* control, ACPA-negative RA *vs* control and ACPA-positive RA *vs* ACPA-negative RA were 1.8×10^{-37}, 0.28 and 3.3×10^{-11}, respectively. These results are similar to those obtained for Caucasian [12] and Japanese subjects [19].

SE was not associated with ACPA-negative RA even when selecting only bone-destructive RA patients

As reported previously [12], no association was observed between SE and ACPA-negative RA. However, some of the patients who were diagnosed with ACPA-negative RA might be non-RA patients, such as those with seronegative SpA, PMR, palindromic rheumatism, OA and other collagen vascular diseases. Indeed, during a survey of the medical records, we found three patients in the ACPA-negative RA subset who had been diagnosed with MCTD, SLE or PMR and were subsequently recorded as presenting with RA. Although we cannot tell which diagnosis is correct, such cases led us to the idea that it is important to exclude possible non-RA patients in ACPA-negative RA subset in order to reveal whether SE is really not associated with ACPA-negative RA. We first excluded the patients who had suffered from RA for <3 years in order to exclude patients with potentially false-negative results for ACPA, on the basis of the fact that the sensitivity of ACPA is lower in the early stage of RA than in the established stage of RA (disease duration ⩾3 years) [7]. Then, we excluded possible non-RA patients who do not have bone erosions by X-ray. Ninety-seven ACPA-negative RA patients showed typical bone erosions and were denoted as ACPA-negative erosive RA patients. As shown in Table 3, the baseline characteristics of ACPA-negative erosive RA patients are similar to those of ACPA-positive RA patients. However, an association of SE with ACPA-negative erosive RA was not observed (Tables 1 and 2). The *P*-value for ACPA-negative erosive RA against the control was 0.26; in contrast, that for ACPA-positive RA against the control was 1.8×10^{-37}. This result clearly shows that ACPA-negative erosive RA is a distinct subset from ACPA-positive RA ($P = 1.1 \times 10^{-6}$), and HLA-DRs containing SE are not causative alleles for developing ACPA-negative RA.

RF, but not ANA, positivity classified RA in terms of SE association

Since it was previously reported that SE was associated only with RF-positive RA [20, 21], we also tested this with our cohort. RF data were available for 843 RA patients and 85.6% were positive for RF. As shown in Table 4, SE was significantly associated with RF-positive RA ($P = 1.0 \times 10^{-44}$, OR 3.7), while the association was much weaker with RF-negative RA ($P = 2.2 \times 10^{-4}$, OR 2.0), showing similar results to Caucasian subjects.

We hypothesized that SE may be related to autoantibody production in general, because not only ACPA and RF but also anti-calpastatin antibodies are reported to be

associated with SE [22]. Therefore, we further examined the association between SE and ANA positivity in RA. ANA data were available for 491 RA patients: 385 (78.4%) patients were ANA positive (Table 5). In contrast with ACPA and RF results, SE was equally associated with both ANA-positive and ANA-negative RA ($P = 3.1 \times 10^{-29}$, OR 3.8 and $P = 6.4 \times 10^{-9}$, OR 3.2, respectively), indicating that ANA does not classify RA in terms of SE. Even when the cut-off value of ANA was set higher, the result was similar (data not shown).

Table 3 Baseline characteristics of ACPA-positive RA and ACPA-negative erosive RA

Characteristics	ACPA-positive RA ($n = 574$)	ACPA-negative erosive ($n = 97$)	P-value*
Age, mean (S.D.), years	63.0 (12.8)	62.1 (12.6)	0.83
Sex: women, %	81.6	86.0	0.29
Disease duration, mean (S.D.) years	18.3 (11.9)	18.0 (13.9)	0.95
Stage, *n* (%)			
1	52 (9.1)	0 (0)	
2	100 (21.4)	26 (27)	
3	69 (14.8)	25 (26)	
4	246 (52.7)	46 (47)	
Class, mean (S.D.)	1.82 (0.69)	2.07 (0.65)	

As not all of the X-ray films for ACPA-positive RA patients were available, the total number of patients and the sum of patients for stage classification do not match. *Student's *t*-test was used for statistical analysis. The *P*-values for stage and class classification are not shown because non-erosive patients were intentionally excluded from the ACPA-negative subset.

SE (especially *DRB1*0405*) is associated with ACPA titre but not RF nor ANA titre

We also investigated whether SE is related to autoantibody titres. ACPA, RF and ANA titres were measured only for the sera from the Kyoto University cohort. The sera with an ACPA titre >100 IU/ml were further diluted to obtain a correct titre. Among those for whom HLA data were available, 252, 248 and 173 RA patients were positive for ACPA, RF and ANA, respectively. Only samples positive for each autoantibody were selected and the association of each autoantibody titre with SE number was tested by Jonckheere–Terpstra trend test. As shown in Fig. 1A–C, the number of SEs is associated with ACPA titre, but not with RF or ANA titre. When we focused on the *DRB1*0405* allele (the most popular SE allele in Japanese subjects), the association of ACPA titre and *DRB1*0405* allele number was statistically significant ($P = 0.000127$) as shown in Fig. 2.

Discussion

Here, we have demonstrated that *HLA-DRB1* SE is associated with ACPA-positive RA, but not with ACPA-negative RA in Japanese subjects. No association of SE with ACPA-negative RA was observed even when eliminating possible non-RA patients from the ACPA-negative RA group. We further demonstrated that ANA did not classify RA into two subsets in terms of SE association, in contrast with RF and ACPA.

The fact that ACPA-positive and ACPA-negative RA are genetically distinct subsets was first reported by a Dutch group studying Caucasian subjects [12], followed by a group studying Japanese subjects [19]. However, the number of patients enrolled in the Japanese

Table 4 Association of SE with RF-positive or RF-negative RA

SE status	Control ($n = 1508$) *n* (%)	RF-positive RA ($n = 722$)		RF-negative RA ($n = 121$)		P-value*
		n (%)	OR (95% CI)	*n* (%)	OR (95% CI)	
SE (+/+)	74 (5)	113 (16)	6.5 (4.7, 9.0)	11 (9)	2.5 (1.3, 5.1)	0.0061
SE (+/−)	492 (33)	387 (54)	3.3 (2.7, 4.1)	55 (45)	1.9 (1.3, 2.8)	0.0072
SE (−/−)	942 (62)	222 (31)	1.0	55 (45)	1.0	

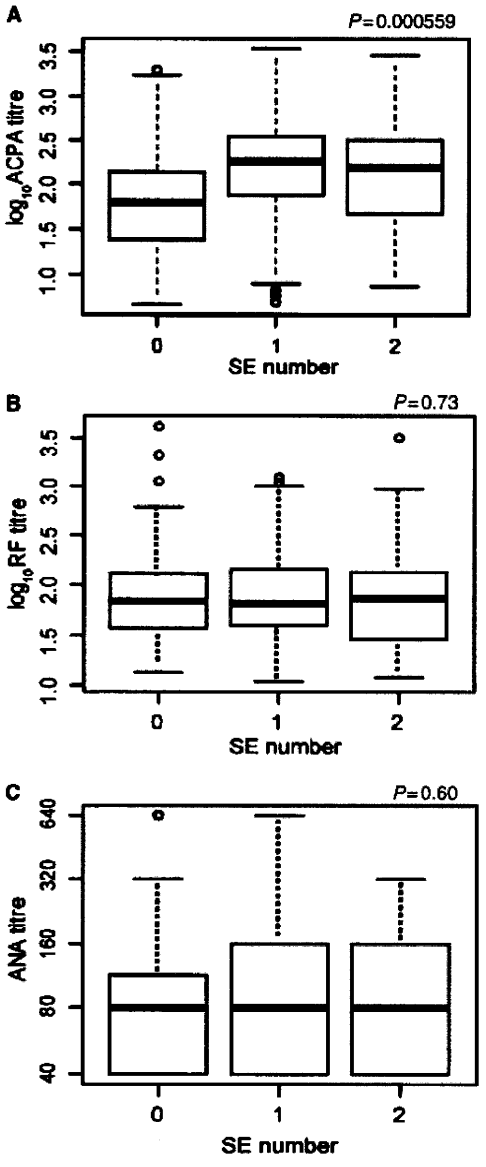
*P-value for RF-positive *vs* RF-negative RA by chi-squared test.

Table. 5 Association of SE with ANA-positive or ANA-negative RA

SE status	Control ($n = 1508$) *n* (%)	ANA-positive RA ($n = 385$)		ANA-negative RA ($n = 106$)		P-value*
		n (%)	OR (95% CI)	*n* (%)	OR (95% CI)	
SE (+/+)	74 (5)	53 (14)	5.7 (3.8, 8.5)	20 (19)	7.1 (3.9, 12.8)	0.51
SE (+/−)	492 (33)	214 (56)	3.5 (2.7, 4.5)	50 (47)	2.7 (1.7, 4.1)	0.28
SE (−/−)	942 (62)	118 (31)	1.0	36 (34)	1.0	

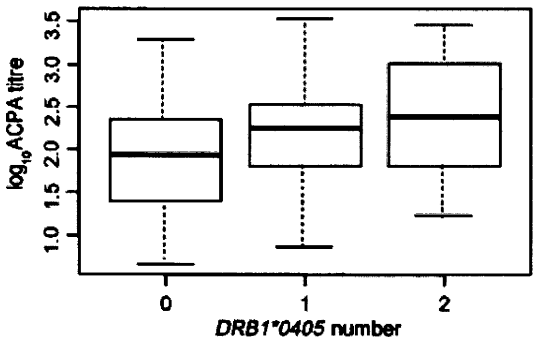
*P-value for ANA-positive *vs* ANA-negative RA by chi-square test.

FIG. 1 Association of number of SE alleles and titre of ACPA, RF or ANA. ACPA-positive (**A**), RF-positive (**B**) or ANA-positive (**C**) RA patients were selected from the Kyoto University cohort, and the serum ACPA titre (**A**), RF titre (**B**) or ANA titre (**C**) was plotted stratified by the number of SE alleles present. The *P*-values were calculated by Jonckheere–Terpstra trend test.



study was only 110 RA patients (82 ACPA-positive and 28 ACPA-negative) and the *P*-values were 0.017 and 0.033 for double-SE and single-SE carriers, respectively. Furthermore, both the Dutch and the Japanese groups enrolled only early RA patients. Therefore, as we discuss later, their cohorts might have contained non-RA patients, especially in the ACPA-negative group. Since our ACPA-negative RA cohort consisted only of patients with established RA (disease duration >3 years) and the *P*-value

FIG. 2 Association of *HLA-DRB1*0405* allele number and ACPA titre. Only ACPA-positive RA samples were selected from the Kyoto University cohort, and ACPA titres and the number of *HLA-DRB1*0405* alleles (the most popular SE allele in Japanese subjects) in each sample are box plotted. The *P*-value by Jonckheere–Terpstra trend test for this association is 0.000127.



reached 3.3×10^{-11}, our study may be the first that has clearly shown that ACPA-positive and ACPA-negative RA subsets are distinct based on SE association using an established RA cohort.

One of the major issues that we aimed to clarify was the suspicion that the lack of an association of SE with ACPA-negative RA was due to ACPA-negative RA groups including non-RA patients. Since the specificity of ACR (formerly ARA) 1987 revised criteria for the classification of RA has been reported to be 89% [23] and it is probable that many non-RA patients will fall into the ACPA-negative group, it is clear that the ACPA-negative RA patient group contains some non-RA patients, which affects the calculated association. From our survey of medical records, 77 out of 174 ACPA-negative patients for whom records were available did not show any bone erosion by X-ray. These patients might not have RA, although we believe that many of these patients do have RA because the group should include RA patients in remission as well as some with slightly active RA without the exhibition of clear changes detectable by radiography. Since all of the patients in our ACPA-negative erosive RA cohort have bone erosion as determined by X-ray, the number of non-RA patients should be minimal. As shown in Table 3, 73% of ACPA-negative erosive RA patients are classified in Steinbrocker's Stage III or IV with joint deformity. Often ACPA-negative RA is described as a less severe arthritic subset, but our erosive cohort consists of patients with RA of a severity similar to that of ACPA-positive RA. Nonetheless, it is interesting that ACPA-negative RA is genetically distinct from ACPA-positive RA.

The next question we addressed was whether such subsets may be formed generally by autoantibody-producing ability. Since it has already been reported that SE was not associated with RF-negative RA [20, 21] or anti-calpastatin-negative RA [22], it appears that SE is related to autoantibody

production in general. However, ANA did not classify RA into two subsets on the basis of the association with SE. Therefore, SE is related to at least ACPA, RF and anti-calpastatin production, but not ANA, suggesting that HLA-DR molecules with SE consensus amino acid sequence present rather specific autoantigens. The dosage effect of the *DRB1*0405* allele for ACPA titre (Fig. 2), but not RF titre, (data not shown) supports this.

Genetic polymorphisms of *PTPN22, CTLA4, TRAF1/C5* and *STAT4* are also reported to be associated with only ACPA-positive RA but not with ACPA-negative RA [13–16]. There is a circumstantial evidence that smoking may promote citrullination of protein/peptides [24] and the affinity of citrullinated vimentin peptide for SE-containing HLA-DR molecules, *HLA-DRB1*0101, *0401* and **0404*, is higher than that of non-citrullinated vimentin peptide [25]. From these findings, one may assume that SE and other genetic polymorphisms, together with smoking, promote the production of ACPA, resulting in joint inflammation [26]. Although there are no direct evidences that ACPAs cause arthritis, aggravation of experimental arthritis by transferring anti-citrullinated fibrinogen mAbs was demonstrated [27], suggesting an arthritis-promoting activity of ACPA. So, we assume that antigen-presenting cells expressing HLA with SE may preferentially present citrullinated peptides to Th2 cells, which may support ACPA-producing B lymphocytes to differentiate into plasma cells. In contrast, there are no plausible explanations for the pathogenesis of ACPA-negative RA. Unknown autoantibodies under a different genetic background might cause arthritis in ACPA-negative RA, or antibody-independent mechanism might be a major pathogenesis in ACPA-negative RA. *HLA-DRB1*03* and **0901* were reported to be weakly associated with ACPA-negative RA patients in Caucasian [28, 29] and Japanese [19] groups, respectively, and only a few genetic determinants of ACPA-negative RA among non-HLA genes have been reported [30, 31]. So far, no genome-wide association study for ACPA-negative RA has been reported, and genetic and environmental factors of ACPA-negative RA development is to be elucidated.

Rheumatology key messages

- ACPA-negative RA, even of bone-erosive type, is distinct subset from ACPA-positive RA.
- HLA-DRB1 SE is not associated with ACPA-negative RA.
- SE is associated with ACPA titre, but not with RF or ANA titres.

Acknowledgements

We would like to thank Ms Miki Kokubo for extraction and preparation of DNA. We would also like to thank Mr Taishi Shigeki for his excellent work in making clinical database software in Dohgo Spa Hospital. We would also like to thank all the doctors and co-medical people who worked together to collect patients' samples.

Funding: This work was supported by Grants-in-aid from the Ministry of Health, Labor and Welfare of Japan and from the Ministry of Education, Culture, Sports, Science and Technology of Japan as well as by research grants from the Japan Rheumatism Foundation, the Waksman Foundation and the Mitsubishi Pharma Research Foundation. Funding to pay the Open Access publication charges for this article was provided by the Japan Rheumatism Foundation.

Disclosure statement: The authors have declared no conflicts of interest.

References

1 Thabet MM, Huizinga TW, van der Heijde DM, van der Helm-van Mil AH. The prognostic value of baseline erosions in undifferentiated arthritis. Arthritis Res Ther 2009;11:R155.

2 Schellekens GA, de Jong BA, van den Hoogen FH, van de Putte LB, van Venrooij WJ. Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. J Clin Invest 1998;101:273–81.

3 Vossenaar ER, Despres N, Lapointe E *et al*. Rheumatoid arthritis specific anti-Sa antibodies target citrullinated vimentin. Arthritis Res Ther 2004;6:R142–50.

4 Masson-Bessiere C, Sebbag M, Girbal-Neuhauser E *et al*. The major synovial targets of the rheumatoid arthritis-specific antifilaggrin autoantibodies are deiminated forms of the alpha- and beta-chains of fibrin. J Immunol 2001;166:4177–84.

5 Lundberg K, Kinloch A, Fisher BA *et al*. Antibodies to citrullinated alpha-enolase peptide 1 are specific for rheumatoid arthritis and cross-react with bacterial enolase. Arthritis Rheum 2008;58:3009–19.

6 van Venrooij WJ, Hazes JM, Visser H. Anticitrullinated protein/peptide antibody and its role in the diagnosis and prognosis of early rheumatoid arthritis. Neth J Med 2002; 60:383–8.

7 Dubucquoi S, Solau-Gervais E, Lefranc D *et al*. Evaluation of anti-citrullinated filaggrin antibodies as hallmarks for the diagnosis of rheumatic diseases. Ann Rheum Dis 2004;63: 415–9.

8 Suzuki K, Sawada T, Murakami A *et al*. High diagnostic performance of ELISA detection of antibodies to citrullinated antigens in rheumatoid arthritis. Scand J Rheumatol 2003;32:197–204.

9 Kroot EJ, de Jong BA, van Leeuwen MA *et al*. The prognostic value of anti-cyclic citrullinated peptide antibody in patients with recent-onset rheumatoid arthritis. Arthritis Rheum 2000;43:1831–5.

10 Meyer O, Labarre C, Dougados M *et al*. Anticitrullinated protein/peptide antibody assays in early rheumatoid arthritis for predicting five year radiographic damage. Ann Rheum Dis 2003;62:120–6.

11 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.

12 Huizinga TW, Amos CI, van der Helm-van Mil AH *et al.* Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins. Arthritis Rheum 2005; 52:3433–8.

13 Plenge RM, Padyukov L, Remmers EF *et al.* Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. Am J Hum Genet 2005;77:1044–60.

14 Barton A, Thomson W, Ke X *et al.* Re-evaluation of putative rheumatoid arthritis susceptibility genes in the post-genome wide association study era and hypothesis of a key pathway underlying susceptibility. Hum Mol Genet 2008;17:2274–9.

15 Lee HS, Remmers EF, Le JM, Kastner DL, Bae SC, Gregersen PK. Association of STAT4 with rheumatoid arthritis in the Korean population. Mol Med 2007;13: 455–60.

16 Kurreeman FA, Padyukov L, Marques RB *et al.* A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. PLoS Med 2007;4: e278.

17 Kamatani N, Sekine A, Kitamoto T *et al.* Large-scale single-nucleotide polymorphism (SNP) and haplotype analyses, using dense SNP Maps, of 199 drug-related genes in 752 subjects: the analysis of the association between uncommon SNPs within haplotype blocks and the haplotypes constructed with haplotype-tagging SNPs. Am J Hum Genet 2004;75:190–203.

18 Rantapaa-Dahlqvist S, de Jong BA, Berglin E *et al.* Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. Arthritis Rheum 2003;48:2741–9.

19 Furuya T, Hakoda M, Ichikawa N *et al.* Differential association of HLA-DRB1 alleles in Japanese patients with early rheumatoid arthritis in relationship to autoantibodies to cyclic citrullinated peptide. Clin Exp Rheumatol 2007; 25:219–24.

20 Dobloug JH, Forre O, Kass E, Thorsby E. HLA antigens and rheumatoid arthritis. Association between HLA-DRw4 positivity and IgM rheumatoid factor production. Arthritis Rheum 1980;23:309–13.

21 Olsen NJ, Callahan LF, Brooks RH *et al.* Associations of HLA-DR4 with rheumatoid factor and radiographic severity in rheumatoid arthritis. Am J Med 1988;84: 257–64.

22 Auger I, Roudier C, Guis S, Balandraud N, Roudier J. HLA-DRB1*0404 is strongly associated with anticalpastatin antibodies in rheumatoid arthritis. Ann Rheum Dis 2007;66:1588–93.

23 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.

24 Klareskog L, Stolt P, Lundberg K *et al.* A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. Arthritis Rheum 2006;54:38–46.

25 Hill JA, Southwood S, Sette A, Jevnikar AM, Bell DA, Cairns E. Cutting edge: the conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. J Immunol 2003;171:538–41.

26 Kallberg H, Padyukov L, Plenge RM *et al.* Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. Am J Hum Genet 2007;80:867–75.

27 Kuhn KA, Kulik L, Tomooka B *et al.* Antibodies against citrullinated proteins enhance tissue injury in experimental autoimmune arthritis. J Clin Invest 2006;116:961–73.

28 Verpoort KN, van Gaalen FA, van der Helm-van Mil AH *et al.* Association of HLA-DR3 with anti-cyclic citrullinated peptide antibody-negative rheumatoid arthritis. Arthritis Rheum 2005;52:3058–62.

29 Irigoyen P, Lee AT, Wener MH *et al.* Regulation of anti-cyclic citrullinated peptide antibodies in rheumatoid arthritis: contrasting effects of HLA-DR3 and the shared epitope alleles. Arthritis Rheum 2005;52:3813–8.

30 Sigurdsson S, Padyukov L, Kurreeman FA *et al.* Association of a haplotype in the promoter region of the interferon regulatory factor 5 gene with rheumatoid arthritis. Arthritis Rheum 2007;56:2202–10.

31 Daha NA, Kurreeman FA, Marques RB *et al.* Confirmation of STAT4, IL2/IL21, and CTLA4 polymorphisms in rheumatoid arthritis. Arthritis Rheum 2009;60:1255–60.

OPEN ACCESS Freely available online

PLoS one

The Progression of Liver Fibrosis Is Related with Overexpression of the miR-199 and 200 Families

Yoshiki Murakami[1]*, Hidenori Toyoda[2], Masami Tanaka[3], Masahiko Kuroda[3], Yoshinori Harada[4], Fumihiko Matsuda[1], Atsushi Tajima[5¤], Nobuyoshi Kosaka[6], Takahiro Ochiya[6], Kunitada Shimotohno[7]

1 Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, **2** Department of Gastroenterology, Ogaki Municipal Hospital, Ogaki, Japan, **3** Department of Molecular Pathology, Tokyo Medical University, Tokyo, Japan, **4** Department of Pathology and Cell Regulation, Kyoto Prefectural University of Medicine, Kyoto, Japan, **5** Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Japan, **6** Division of Molecular and Cellular Medicine, National Cancer Center Research Institute, Tokyo, Japan, **7** Research Institute, Chiba Institute of Technology, Narashino, Japan

Abstract

Background: Chronic hepatitis C (CH) can develop into liver cirrhosis (LC) and hepatocellular carcinoma (HCC). Liver fibrosis and HCC development are strongly correlated, but there is no effective treatment against fibrosis because the critical mechanism of progression of liver fibrosis is not fully understood. microRNAs (miRNAs) are now essential to the molecular mechanisms of several biological processes. In order to clarify how the aberrant expression of miRNAs participates in development of the liver fibrosis, we analyzed the liver fibrosis in mouse liver fibrosis model and human clinical samples.

Methodology: In a CCL_4-induced mouse liver fibrosis model, we compared the miRNA expression profile from CCL_4 and olive oil administrated liver specimens on 4, 6, and 8 weeks. We also measured expression profiles of human miRNAs in the liver biopsy specimens from 105 CH type C patients without a history of anti-viral therapy.

Principle Findings: Eleven mouse miRNAs were significantly elevated in progressed liver fibrosis relative to control. By using a large amount of human material in CH analysis, we determined the miRNA expression pattern according to the grade of liver fibrosis. We detected several human miRNAs whose expression levels were correlated with the degree of progression of liver fibrosis. In both the mouse and human studies, the expression levels of miR-199a, 199a*, 200a, and 200b were positively and significantly correlated to the progressed liver fibrosis. The expression level of fibrosis related genes in hepatic stellate cells (HSC), were significantly increased by overexpression of these miRNAs.

Conclusion: Four miRNAs are tightly related to the grade of liver fibrosis in both human and mouse was shown. This information may uncover the critical mechanism of progression of liver fibrosis. miRNA expression profiling has potential for diagnostic and therapeutic applications.

Citation: Murakami Y, Toyoda H, Tanaka M, Kuroda M, Harada Y, et al. (2011) The Progression of Liver Fibrosis Is Related with Overexpression of the miR-199 and 200 Families. PLoS ONE 6(1): e16081. doi:10.1371/journal.pone.0016081

Editor: Chad Creighton, Baylor College of Medicine, United States of America

Received September 15, 2010; **Accepted** December 6, 2010; **Published** January 24, 2011

Funding: This work was supported by the Japanese Ministry of Health, Labour and Welfare (Y.M, and K.S). This work was also supported by the 'Strategic Research-Based Support' Project for private universities; with matching funds from the Ministry of Education, Culture, Sports, Science and Technology (M.K). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ymurakami@genome.med.kyoto-u.ac.jp

¤ Current address: Department of Human Genetics and Public Health, Institute of Health Biosciences, The University of Tokushima Graduate School, Tokushima, Japan

Introduction

Chronic viral hepatitis is a major risk factor for hepatocellular carcinoma (HCC) [1]. Worldwide 120–170 million persons are currently chronically Hepatitis C Virus (HCV) infected [2]. Due to repetitive and continuous inflammation, these patients are at increased risk of developing cirrhosis, subsequent liver decompensation and/or hepatocellular carcinoma. However, the current standard of care; pegylated interferon and rivabirin combination therapy is unsatisfied in the patients with high titre of HCVRNA and genotype 1b. Activated human liver stellate cells (HSC) with chronic viral infection, can play a pivotal role in the progression of liver fibrosis [3]. Activated HSC produce a number of profibrotic cytokines and growth factors that perpetuate the fibrotic process through paracrine and autocrine effects.

MicroRNAs (miRNAs) are endogenous small non-coding RNAs that control gene expression by degrading target mRNA or suppressing their translation [4]. There are currently 940 identifiable human miRNAs (The miRBase Sequence Database - Release ver. 15.0). miRNAs can recognize hundreds of target genes with incomplete complementary; over one third of human genes appear to be conserved miRNA targets [5][6]. miRNA is associated several pathophysiologic events as well as fundamental cellular processes such as cell proliferation and differentiation. Aberrant expression of miRNA can be associated with the liver diseases [7][8][9][10]. Recently reported miRNAs can regulate the activation of HSCs and thereby regulate liver fibrosis. miR-29b, a negative regulator for the type I collagen and SP1, is a key regulator of liver fibrosis [11]. miR-27a and 27b allowed culture-activated rat HSCs to switch to a more quiescent HSC phenotype,

with restored cytoplasmic lipid droplets and decreased cell proliferation [12].

In this study, we aimed to reveal the association between miRNA expression patterns and the progression of liver fibrosis by using a chronic liver inflammation model in mouse. We also sought to identify the miRNA expression profile in chronic hepatitis (CH) C patients according to the degree of liver fibrosis, and to clarify how miRNAs contribute to the progression of liver fibrosis. We observed a characteristic miRNA expression profile common to both human liver biopsy specimens and mouse CCL_4 specimens, comprising the key miRNAs which are associated with the liver fibrosis. This information is expected to uncover the mechanism of liver fibrosis and to provide a clearer biomarker for diagnosis of liver fibrosis as well as to aid in the development of more effective and safer therapeutic strategies for liver fibrosis.

Results

The expression level of several mouse miRNAs was increased by introducing mouse liver fibrosis

In order to identify changes in the miRNA expression profile between advanced liver fibrosis and non-fibrotic liver, we intraperitoneally administered CCL_4 in olive oil or olive oil alone twice a week for 4 weeks and then once a week for the next 4 weeks. Mice were sacrificed at 4, 6, or 8 weeks and then the degree of mouse liver fibrosis was determined by microscopy (Figure S1). miRNA expression analysis was performed from the liver tissue collected at the same time. Histological examination revealed that the degree of liver fibrosis progressed in mice that received CCL_4 relative to mice receiving olive oil alone (Figure 1A). Microarray analysis revealed that in CCL_4 mice, the expression level of 11 miRNAs was consistently higher than that in control mice (Figure 1B).

miRNA expression profile in each human liver fibrosis grade

We then established human miRNAs expression profile by using 105 fresh-frozen human chronic hepatitis (CH) C liver tissues without a history of anti-viral therapy, classified according to the grade of the liver fibrosis (F0, F1, F2, and F3 referred to METAVIR fibrosis stages)(Figure 2, Table S2). Fibrosis grade F0 was considered to be the negative control because these samples were derived from patients with no finding of liver fibrosis. In zebrafish, most highly tissue-specific miRNAs are expressed during embryonic development; approximately 30% of all miRNAs are expressed at a given time point in a given tissue [13]. In mammals, the 20–30% miRNA call rate has recently been validated [14]. Such analysis revealed that the diversity of miRNA expression level among specimens was small. Therefore, we focused on miRNAs with a fold change in mean expression level greater than 1.5 ($p<0.05$) in the two arbitrary groups of liver fibrosis.

Expression of several miRNAs was dramatically different among grades of fibrosis. In the mice study 11 miRNAs were related to the progression of liver fibrosis (mmu-let-7e, miR-125-5p, 199a-5p, 199b, 199b*, 200a, 200b, 31, 34a, 497, and 802). In the human study 10 miRNAs were extracted, and the change in their expression level varied significantly between F0 and F3 (F0<F3: hsa-miR-146b, 199a, 199a*, 200a, 200b, 34a, and 34b, F0>F3: hsa-miR-212, 23b, and 422b). The expression level of 6 miRNAs was significantly different between F0 and F2 (F0<F2: hsa-miR-146b, 200a, 34a, and 34b, F0>F2: hsa-miR-122 and 23b). 5 extracted miRNAs had an expression level that was significantly different between F1 and F2 (F1<F2: hsa-miR-146b, F1>F2: hsa-miR-122, 197, 574, and 768-5p). The expression level of 9 miRNAs changed significantly between F1 and F3 (F1<F3: hsa-miR-146b, 150, 199a, 199a*, 200a, and 200b, F1>F3: hsa-miR-378, 422b, and 768-5p). The miRNAs related to liver fibrosis were extracted using two criteria: similar expression pattern in both the human and the mice specimens and shared sequence between human and mouse. We compared the sequences of mouse miRNAs as described on the Agilent Mouse MiRNA array Version 1.0 (miRbase Version 10.1) and human miRNAs as described on the Agilent Human MiRNA array Version 1.5 (miRbase Version 9.1). The sequences of mmu-miR-199a-5p, mmu-miR-199b, mmu-miR-199b, mmu-miR-200a, and mmu-miR-200b in mouse miRNA corresponded to the sequences of hsa-miR-199a, hsa-miR-199a*, hsa-miR-199a, hsa-miR-200a, and hsa-miR-200b in human miRNA, respectively (Table S3).

Validation of the microarray result by real-time qPCR

The 4 human miRNAs (miR-199a, miR-199a*, miR-200a, and miR-200b) with the largest difference in fold change between the F1 and F3 groups were chosen to validate the microarray results using stem-loop based real-time qPCR. The result of real-time qPCR supported the result of that microarray analysis. The expression level of these 4 miRNAs was significantly different between F0 and F3 and spearman correlation analysis also showed that the expressions of these miRNAs were strongly and positively correlated with fibrosis grade ($n = 105$, $r = 0.498$(miR-199a), 0.607(miR-199a*), 0.639(miR-200a), 0.618(miR-200b), p-values<0.0001) (Figure 3).

Over expression of miR-199a, 199a*, 200a, and 200b was associated with the progression of liver fibrosis

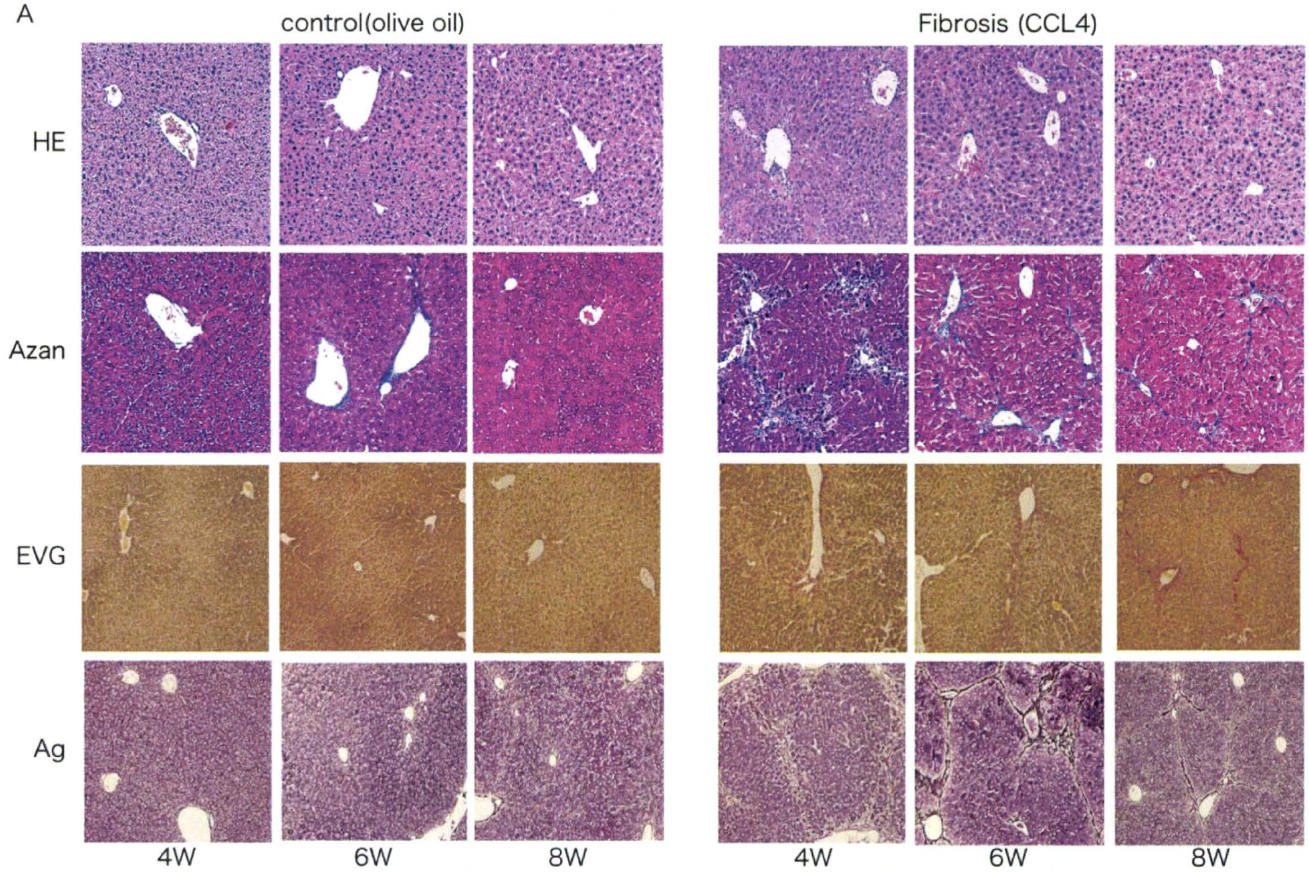
In order to reveal the function of miR-199a, miR-199a*, miR-200a, and miR-200b, we investigated the involvement of these miRNAs in the modulation of fibrosis-related gene in LX-2 cells. The endogenous expression level of these 4 miRNAs in LX2 and normal liver was low according to the microarray study (Figure S2). Transforming growth factor (TGF)β is one of the critical factors for the activation of HSC during chronic inflammation [15] and TGFβ strongly induced expression of three fibrosis-related genes include a matrix degrading complex comprised of α1 procollagen, matrix remodeling complex, comprised of metalloproteinases-13 (MMP-13), tissue inhibitors of metalloproteinases-1 (TIMP-1) in LX-2 cells (Figure 4A). Furthermore, overexpression of miR-199a, miR-199a*, miR-200a and miR-200b in LX-2 cells resulted significant induction of above fibrosis-related genes compared with control miRNA (Figure 4B). Finally we validated the involvement of TGFβ in the modulation of these miRNAs. In LX-2 cells treated with TGFβ, the expression levels of miR-199a and miR-199a* were significantly higher than in untreated cells; the expression levels of miR-200a and miR-200b were significantly lower than in untreated cells. Thus, our in vitro analysis suggested a possible involvement of miR-199a, 199a*, 200a, and 200b in the progression of liver fibrosis.

Discussion

Our comprehensive analysis showed that the aberrant expression of miRNAs was associated with the progression of liver fibrosis. We identified that 4 highly expressed miRNAs (miR-199a, miR-199a*, miR-200a, and miR-200b) that were significantly associated with the progression of liver fibrosis both human and mouse. Coordination of aberrant expression of these miRNAs may contribute to the progression of liver fibrosis.

Prior studies have discussed the expression pattern of miRNA found in liver fibrosis samples between previous and present study. In this report and prior mouse studies and the expression pattern of

A



B

miRNA	Fold change CCL4 mouse/control		
	4week	6week	8week
mmu-let-7e	2.92	2.09	2.64
mmu-miR-125a-5p	4.29	2.35	2.67
mmu-miR-199a-5p	2.71	2.12	2.38
mmu-miR-199b	3.25	2.55	2.49
mmu-miR-199b*	5.13	4.47	3.94
mmu-miR-200a	2.23	1.63	2.02
mmu-miR-200b	2.38	2.06	2.20
mmu-miR-31	3.4	2.35	3.19
mmu-miR-34a	3.89	3.22	3.5
mmu-miR-497	2.25	2.00	2.55
mmu-miR-802	2.02	2.87	2.74

Figure 1. The change of liver fibrosis in mouse model. A. Representative H&E-stained, Azan-stained, Ag-stained, and EVG-stained histological sections of liver from mice receiving olive oil alone or CCL_4 in olive oil. Magnification is ×10. B. The expression level of mmu-miRNA in mouse liver with olive oil or CCL_4 at 4W, 6W, and 8W respectively, by microarray analysis.
doi:10.1371/journal.pone.0016081.g001

3 miRNAs (miR-199a-5p, 199b*, 125-5p) was found to be similar while the expression pattern of 11 miRNAs (miR-223, 221, 24, 877, 29b, 29a, 29c, 30c, 365, 148a, and 193) was partially consistent with fibrosis grade [16]. In low graded liver fibrosis, the low expression pattern of 3 miRNAs (miR-140, 27a, and 27b) and the high expression pattern of 6 miRNAs in rat miRNAs (miR-29c*, 143, 872, 193, 122, and 146) in rat miRNA was also similar to our mouse study (GEO Series accession number GSE19865) [11] [12] [17].

The results in this study and previously completed human studies reveal that the expression level of miR-195, 222, 200c, 21, and let-7d was higher in high graded fibrotic liver tissue than in low graded fibrotic liver tissue. Additionally, the expression level of miR-301, 194, and 122 was lower in the high graded fibrotic liver tissue than in low graded fibrotic liver tissue [18] [19] [20](GEO Series accession number GSE16922). This difference in miRNA expression pattern may be contributed to (1) the difference of microarray platform, (2) difference of analytic procedure, and (3) the difference of the species (rat, mouse, and human).

The miR-199 and miR-200 families have are circumstantially related to liver fibrosis. TGFβ-induced factor (TGIF) and SMAD