D_g plot of this case is shown in Figure 5(c). Pairwise d_g's are weakly present for the site-pairs in the latter three sites. The square for the division of all the six sites into six single sites indicates strong LD in this region overall and the square for the division of the latter three sites into single sites also indicates the presence of LD.

D_g PLOTS FOR REAL DATA

A region with 22 sites was chosen from HapMap project data set [The International HapMap Consortium, 2005] and haplotype frequency was estimated with fastPhase, one of the popular haplotype inference applications for large scale data [Scheet and Stephens, 2006]. Nineteen haplotypes were inferred and their D_g plot was shown in Figure 5(d), that displayed that the pairwise triangle and the tandem triangle captured LD components of the region differently.

USAGE OF Ψ FOR HAPLOTYPE FREQUENCY INFERENCE

LIKELIHOOD FUNCTION OF GENOTYPE DATA FOR SNP PAIRS IS EXPRESSED AS A MONOVARIATE FUNCTION OF Ψ AND THE HAPLOTYPE FREQUENCY IS OBTAINED BY SOLVING THE DERIVATIVES OF UNIVARIATE FUNCTION.

Although Ψ is calculable when frequency of all haplotypes are given, the majority of LD mapping studies are based on unphased genotype data of SNPs, where the haplotype frequency has to be inferred. As described in the section "Ψ Gives a Base for Haplotype Frequency Space", $F(n)(1_{st})$ and Ψ are in one-to-one correspondence. Therefore the inference of $F(n)(1_{st})$ is equivalent to the inference of Ψ.

Consider haplotype frequency inference from unphased genotype data of a SNP pair. For two SNPs, the four haplotype frequencies are expressed with Ψ as:

$$\begin{aligned} f_1 &= \tfrac{1}{4}(\psi(2)(1_{st})+\psi(1)(1_{st})+\psi(1)(2_{nd})+\psi(0)(1_{st})),\\ f_2 &= \tfrac{1}{4}(-\psi(2)(1_{st})+\psi(1)(1_{st})-\psi(1)(2_{nd})+\psi(0)(1_{st})),\\ f_3 &= \tfrac{1}{4}(-\psi(2)(1_{st})-\psi(1)(1_{st})+\psi(1)(2_{nd})+\psi(0)(1_{st})),\\ f_4 &= \tfrac{1}{4}(\psi(2)(1_{st})-\psi(1)(1_{st})-\psi(1)(2_{nd})+\psi(0)(1_{st})). \end{aligned} \quad (7)$$

$\ln(L)$, logarithm of likelihood function to obtain a unphased genotype data is expressed as a function of f_i;

$$\ln(L) = G_1\log(f_1)+G_2\log(f_2)+G_3\log(f_3) + G_4\log(f_4)+G_5\log(f_1f_4+f_2f_3)+C,$$

where $G_i (i = 1,\ldots,4)$ represents the number of chromosomes that are deterministically known from unphased genotype data, and G_5 is the number of double heterozygotes, and C is a constant.

The EM algorithm attempts to maximize L by handling f_1, f_2, f_3 and f_4 as variables where f_1+f_2 and f_1+f_3 are fixed at the value given by method of moments. Because f_i is expressed with Ψ, $\ln(L)$ is also a function of Ψ. Although Ψ for SNP pairs has four elements, $\psi(0)(1_{st})$ is always constant and value of $\psi(1)(1_{st})$ and $\psi(1)(2_{nd})$ are known under the condition where f_1+f_2 and f_1+f_3 are given by the method of moments ($\psi(1)(1) = (f_1+f_2)-(f_3+f_4)$ and $\psi(1)(2_{nd}) = (f_1+f_3)-(f_2+f_4)$). Therefore the equations (6) are transformed to:

$$\begin{aligned} f_1 &= \tfrac{1}{4}(\psi(2)(1_{st})+c_1),\\ f_2 &= \tfrac{1}{4}(-\psi(2)(1_{st})-c_2),\\ f_3 &= \tfrac{1}{4}(-\psi(2)(1_{st})-c_3),\\ f_4 &= \tfrac{1}{4}(\psi(2)(1_{st})+c_4). \end{aligned} \quad (8)$$

where c_i denotes constant terms of frequency with appropriate signs.

It is shown that $\ln(L)$ is expressed as a monovariate function of $\psi(2)(1_{st})$. $\ln(L)$ is defined for the finite range of $\psi(2)(1_{st})$, where $0 \le f_i \le 1$, and the function is continuous and differentiable in the range. Therefore the global maximum can be obtained by solving its derivatives with conventional searching methods.

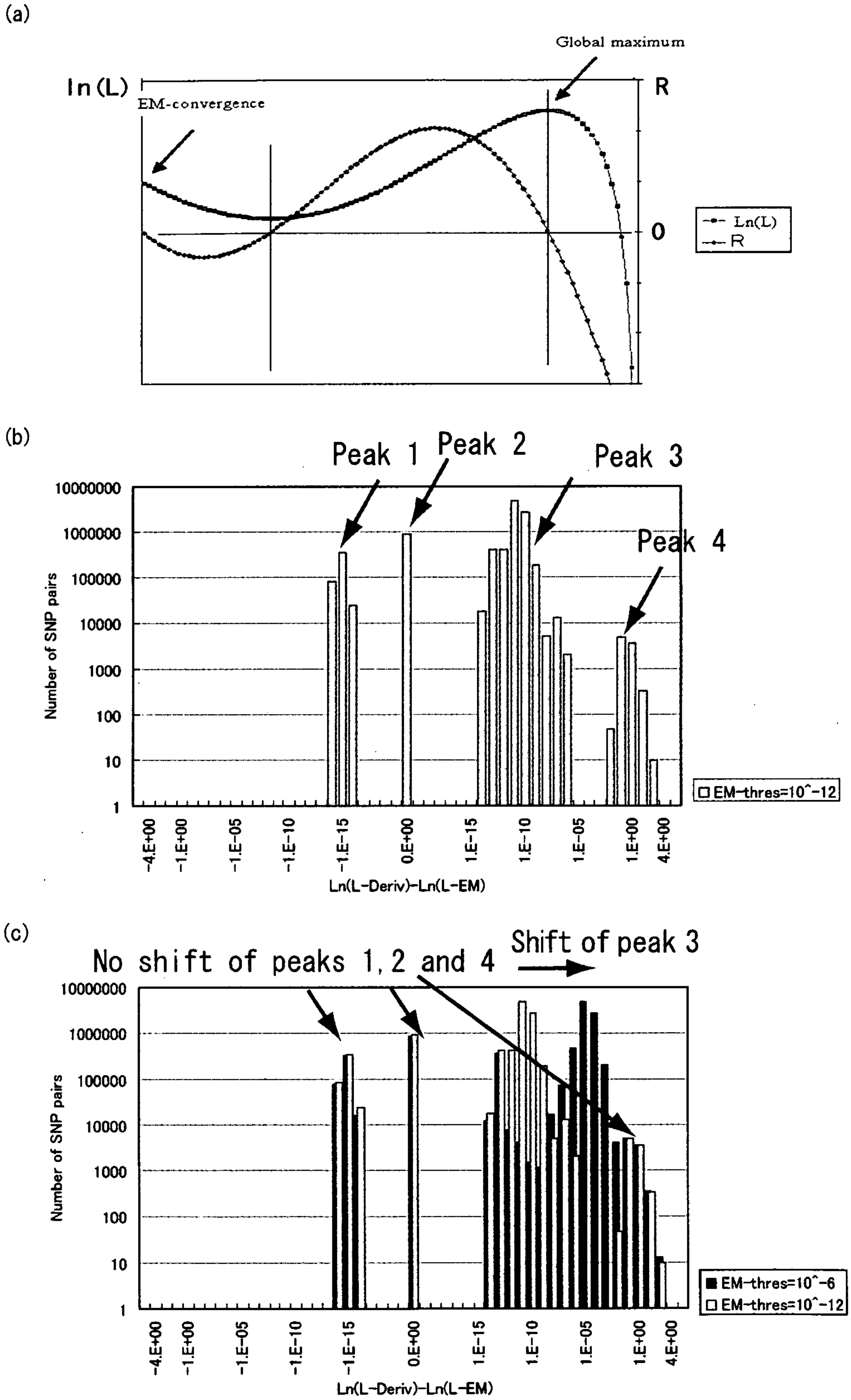
Equation transformations and its Newton-Raphson estimation of the derivatives are described in Appendix 2.

In the case of $n = 2$, maximum likelihood estimates of Ψ was obtained by solving a univariate likelihood function, as above. Similarly, when Ψ is solved for all subsets of $S(n)(1_{st})$ except for $S(n)(1_{st})$ itself where all the elements of Ψ for $S(n)(1_{st})$ but $\psi(n)(1_{st})$ are given, the likelihood function can be expressed as a univariate function of $\psi(n)(1_{st})$. Appendix 3 gives this generalization of likelihood function expressed as a univariate functionof $\psi(n)(1_{st})$ $(n = 1,2,\ldots)$.

COMPARISON WITH THE EM ALGORITHM

The EM algorithm is known to give reliable estimates of haplotype frequencies of SNP pairs in the majority of cases, but is susceptible to convergence to a local maximum [Nin, 2004]. Figure 6a shows an example of convergence to a local maximum of $\ln(L)$ for a SNP pair from the HapMap Project. We evaluated how frequently the standard EM algorithm converges to a local minimum but not to the global maximum using HapMap Project data [The International HapMap Consortum, 2005].

Ψ-based method and the EM algorithm were applied to 10 million SNP pairs of chromosome 10 within a 250 kb window with 45 unrelated Japanese of the HapMap project. The average number of iterations of EM method was 22.2, and the average number of iterations to solve five derivatives in Ψ-based method was 126.1. The results of the Ψ-based method indicated that 39.8% of the pairs did not have


(a)
Global maximum
ln(L)
R
EM-convergence
0
Ln(L)
R
(b)
Peak 1
Peak 2
Peak 3
Peak 4
10000000
1000000
100000
10000
1000
100
10
1
Number of SNP pairs
□ EM-thres=10^-12
-4.E+00
-1.E+00
-1.E-05
-1.E-10
-1.E-15
0.E+00
1.E-15
1.E-10
1.E-05
1.E+00
4.E+00
Ln(L-Deriv)-Ln(L-EM)
(c)
No shift of peaks 1, 2 and 4
Shift of peak 3
10000000
1000000
100000
10000
1000
100
10
1
Number of SNP pairs
■ EM-thres=10^-6
□ EM-thres=10^-12
-4.E+00
-1.E+00
-1.E-05
-1.E-10
-1.E-15
0.E+00
1.E-15
1.E-10
1.E-05
1.E+00
4.E+00
Ln(L-Deriv)-Ln(L-EM)

local extrema, while 61.1% of pairs had a single local extreme in the search range and 0.025% had multiple local extrema. Among the pairs with one local extreme, 81.5% of them was a local maximum, and the remainder was a local minimum. Difference of ln(L) between inferences of the two methods was shown in Fig. 6(b). Peak 1 (Fig. 6(b)) represented 4.5% of SNP pairs for which the EM algorithm gave slightly higher likelihood. The EM algorithm gave better inference due to luck to start at the best value for the majority of SNP pairs in the peak 1. Peak 2 (Fig. 6(b)) represented 9.3% of pairs and two methods gave almost identical results. Peaks 3 and 4 (Fig. 6(b)) represented 86.1% of pairs for which the Ψ-based method gave slightly better result. When we allowed the EM algorithm method to stop earlier with looser convergence threshold, peaks 1, 2 and 4 did not change but a part of peak 3 shifted to right (Fig. 6(c)). This change indicated that the EM algorithm method could give better estimate for a part of SNP pairs in peak 3 by modifying its parameters but that the EM algorithm method converged to a local maximum for SNP pairs in peak 4. However the peak 4 represented only 0.09% of total SNP pairs. More detailed characterization of SNP pairs for which the EM algorithm method converged to a local maximum were described in the Appendix 3. Conditions of inference of the standard EM algorithm and the Ψ-based algorithm are available in the Appendix 5.

DISCUSSION

In this paper, a novel tensor Ψ was introduced to quantitate genetic heterogeneity with SNPs in populations. The Ψ was consisted of 2^n elements for a sequence with n sites that were mutually transformable with 2^n values of haplotype frequency. Actually 2^n-1 non-constant variables in Ψ were the base of the haplotype frequency space with 2^n-1 dimensions. Each element of Ψ represented one of subsets of n sites and they were arranged in a structure of tensor and gave information on two types of randomness of the population, the allele frequency randomness and the inter-site randomness. As an example of utility of Ψ, we proposed a generalized LD index, D_g(Pair), between two SNPs was formulated using the elements of Ψ, and its basic feature was compared with D' and r^2. Moreover LD index for a set of multiple sites more than two, D_g(Div) was also defined as a natural extension of D_g(Pair). The components of D_g for SNP pairs were drawn in the pairwise triangle and the representative components of D_g for multiple sites were drawn in the tandem triangle. For another practical purpose, Ψ offered the absolute maximum haplotype frequency inference for SNP pairs with tolerable increase of computational burden and it overcomes the problem to converge to a local maximum by the EM algorithm method. Application of the Ψ-based haplotype inference algorithm to larger SNP sets seemed possible but modifications to limit computational burdens would be necessary.

Because populational DNA sequence heterogeneity is a product of many genetic events over years and Ψ carry complete information on heterogeneity of individual sites and inter-site dependency for any combinations of sites in the region, it is necessarily complex. In order to describe the complexity, Ψ has almost fully simplified formula. (i) It uses minimum number of variables (2^n for sequence of length n). (ii) All the variables are recurrently defined so that each element represents a subset of the set of n sites. (iii) The variables are arranged in a structure based on their mutual relations (tensor structure). Although it seems still difficult to use all the information included in Ψ in order to untangle genetic heterogeneity of species, Ψ would contribute to formulate and understand interspecies genetic heterogeneity.

ACKNOWLEDGMENTS

The authors thank all the contributors to the HapMap Project, and the members, Particularly Dr. Alexandre Vasilescu, of the Center for Genomic Medicine, Graduate School of Medicine, Kyoto University and the SNP Research Center, RIKEN for valuable discussion. This work was supported in part by the CREST program (JST), Research on Measures for Intractable Diseases and Research on Human Genome and Tissue Engineering (Ministry of Health, Labour and Welfare, Japan) and BioBankJapan project.

ELECTRONIC DATABASE INFORMATION

See also HapMap: http://www.hapmap.org/index.html; Program sources and tools to calculate D_g are available, http://www.genome.med.Kyoto-u.ac.jp/ra/statgenet/index_en.html

Fig. 6. Comparison of Ψ based-haplotype inference method and EM methods. (a) Plots of ln(L) and R for an SNP pair from the HapMap Project (See Appendix 2 for the definition of R. The pair has a genotype distribution of 10, 11, 6, 1, 12, 0, 0, 0, 0, for AABB, AABb,..., aabb, and the estimated global maximum of haplotype frequency is $h(AA) = 0.547$, $h(AB) = 0.291$, $h(aB) = 0.053$, $h(ab) = 0.109$, $D' = 0.45$. The standard EM method converges to $h(AB) = 0.438$, $h(Ab) = 0.400$, $h(aB) = 0.162$, $h(ab) = 0.00$, $D' = 1.00$. Vertical lines denote $R = 0$, at which ln(L) takes a local minimum and local maximum. (b) Distribution of difference in ln(L) between the two methods for 10^6 SNP pairs from the HapMap Project. The convergence threshold for the EM method is 10^{-12}. (c) Comparison of EM convergence thresholds (10^{-6} and 10^{-12}).

REFERENCES

Aquadro CG, Begun DJ, Kndahl EC. 1994. Selection, recombination and DNA polymorphism in Drosophila. In: Golding B, editors. Non-Neutral Evolution: Theories and Molecular Data. New York: Chapman & Hall.

Collins FS, Green ED, Guttmacher AE, Guyer MS. 2003. A vision for the future of genomics research. Nature 422:835–847.

Devlin B, Risch N. 1995. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 29:311–322.

Hartl DL, Clark AG. 1997a. Principles of Population Genetics, 3rd edition. MA: Sinauer Associates, Inc. p 294–296.

Hartl DL, Clark AG. 1997b. Principles of Population Genetics, 3rd edition. MA: Sinauer Associates, Inc. p 95–106.

Hartl DL, Clark AG. 1997c. Principles of Population Genetics, 3rd edition. MA: Sinauer Associates, Inc. p 57–61.

Kidd KK, Pakstis AJ, Speed WC, Kill JR. 2004. Understanding human DNA sequence variation. J Hered 95:406–420.

Morton NE. 2005. Linkage disequilibrium maps and association mapping. J Clin Invest 115:1425–1430.

Navarro A, Barton NH. 2003. Accumulating postzygotic isolation genes in parapartry: a new twist on chromosomal speciation. Evolution 57:447–459.

Niu T. 2004. Algorithms for inferring haplotypes. Genet Epidemiol 27:334–347.

Noor MA, Grams KL, Bertucci LA, Reiland J. 2001. Chromosomal inversions and the reproductive isolation of species. Proc Natl Acad Sci 98:12084–12088.

Nothnagel R, Furst R, Rohde K. 2002. Entropy as a measure for linkage disequilibrium over multilocus haplotype blocks. Hum Hered 54:186–198.

Rieseberg LH. 2001. Chromosomal rearrangements and speciation. Trends Ecol Evol 98:12084–12088.

Rieseberg LH, Livingstone K. 2003. Evolution. Chromosomal speciation in primates. Science 300:267–268.

Rowland T, Weisstein EW. 2006. Tensor. From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/Tensor.html.

Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 78: 629–644.

The International HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437:1299–1320.

Weisstein EW. 2006a. Power Set. From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/PowerSet.html.

Weisstein EW. 2006b. Determinant Expansion by Minors. From MathWorld—A Wolfram Web Resource. http://mathworld.wolfram.com/DeterminantExpansionbyMinors.html.

Zapata C. 2000. The D′ measure of overall gametic disequilibrium between pairs of multiallelic lici. Evolution 54:1809–1812.

APPENDIX

APPENDIX 1

Equivalence of D_g^{Pair} to conventional pair-wise linkage disequilibrium.

$$
\begin{aligned}
&(f_1-f_2-f_3+f_4)-((f_1+f_2)-(f_3+f_4))\times((f_1+f_3)-(f_2+f_4))\\
&=(f_1-f_2-f_3+(1-f_1-f_2-f_3))-(f_1+f_2-f_3\\
&\quad-(1-f_1-f_2-f_3))\times(f_1-f_2+f_3-(1-f_1-f_2-f_3))\\
&=(1-2(f_2+f_3))-(2(f_1+f_2)-1)\times(2(f_1+f_3)-1)\\
&=(1-2(f_2+f_3))-4(f_1+f_2)\times(f_1+f_3)+2((f_1+f_2)\\
&\quad+(f_1+f_3))-1\\
&=4(f_1-(f_1+f_2)\times(f_1+f_3))\\
&=4(f_1-f_1\times(f_1+f_2+f_3)-f_2\times f_3)\\
&=4(f_1\times(1-f_1-f_2-f_3)-f_2\times f_3)\\
&=4(f_1\times f_4-f_2\times f_3).
\end{aligned}
$$

APPENDIX 2

Monovariate likelihood function expressed as a function of $\psi(2)(1_{st})$ and its maximal likelihood estimation.

$$
\begin{aligned}
f_1&=\tfrac{1}{4}(\psi(2)(1_{\mathrm{st}})+c_1),\\
f_2&=\tfrac{1}{4}(-\psi(2)(1_{\mathrm{st}})-c_2),\\
f_3&=\tfrac{1}{4}(-\psi(2)(1_{\mathrm{st}})-c_3),\\
f_4&=\tfrac{1}{4}(\psi(2)(1_{\mathrm{st}})+c_4),
\end{aligned}
\qquad (A7)
$$

where c_i denote constant terms of frequency with appropriate signs.

Because

$$\frac{\mathrm{d}f_i}{\mathrm{d}\psi(2)(1_{\mathrm{st}})}=\frac{1}{4},\quad \text{for } i=1,4,$$

$$\frac{\mathrm{d}f_i}{\mathrm{d}\psi(2)(1_{\mathrm{st}})}=-\frac{1}{4},\quad \text{for } i=2,3,$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\psi(2)(1_{\mathrm{st}})}(f_1f_4+f_2f_3)=\psi(2)(1_{\mathrm{st}})$$

from the equations (7), we have

$$
\begin{aligned}
&\frac{\mathrm{d}\ln(L)}{\mathrm{d}(\psi(2)(1_{\mathrm{st}}))}(\psi(2)(1_{\mathrm{st}}))\\
&\qquad=\frac{1}{4}\left(\frac{G_1}{f_1}-\frac{G_2}{f_2}-\frac{G_3}{f_3}+\frac{G_4}{f_4}+\frac{G_5}{f_1f_4+f_2f_3}(\psi(2)(1_{\mathrm{st}}))\right).
\end{aligned}
$$

The global maximum of $\ln(L)$ is given by $\psi(2)(1_{\mathrm{st}})$ among the solutions of $[\mathrm{d}\ln(L)/\mathrm{d}(\psi(2)(1_{\mathrm{st}}))](\psi(2)(1_{\mathrm{st}}))=0$ in the defined range of $\psi(2)(1_{\mathrm{st}})$ or the two endpoints of the range. Because $\ln(L(\psi(2)(1_{\mathrm{st}}))$ and $[\mathrm{d}\ln(L)/\mathrm{d}(\psi(2)(1_{\mathrm{st}}))](\psi(2)(1_{\mathrm{st}}))$ are both continuous in the defined range where $0\le f_i\le 1$, a conventional searching algorithm gives the estimate of $\psi(2)(1_{\mathrm{st}})$ corresponding to the global maximum of $\ln(L(\psi(2)(1_{\mathrm{st}}))$. The followings are the steps to solve the derivative.

Let $\ln(L(\psi(2)(1_{\mathrm{st}}))=0$ take the form of $\ln(L(\psi(2)(1_{\mathrm{st}})))=\frac{R(\psi(2)(1_{\mathrm{st}}))}{T(\psi(2)(1_{\mathrm{st}}))}=0$, so that all the solutions of $\ln(L(\psi(2)(1_{st})) = 0$ are included in the solutions of $R(\psi(2)(1_{st})) = 0$.

$$R(\psi(2)(1_{st})) = (G_1 f_2 f_3 f_4 - G_2 f_1 f_3 f_4 - G_3 f_1 f_2 f_4 \\ + G_4 f_1 f_2 f_3)(f_1 f_4 + f_2 f_3) + (\psi(2)(1_{st}))G_5 f_1 f_2 f_3 f_4 = 0.$$

Now solutions of $R(\psi(2)(1_{st}))$ cover all candidate values of $\psi(2)(1_{st})$ as the global maximum of $\ln(L(\psi(2)(1_{st}))$. Then, R can be re-expressed as:

$$\begin{aligned} &R(\psi(2)(1_{st})) \\ &= \left(\tfrac{1}{4}\right)^5((G_1(\psi(2)(1_{st}) + c_2)(\psi(2)(1_{st}) + c_3)(\psi(2)(1_{st}) + c_4) \\ &+ G_2(\psi(2)(1_{st}) + c_1)(\psi(2)(1_{st}) + c_3)(\psi(2)(1_{st})c_4) \\ &+ G_3(\psi(2)(1_{st}) + c_1)(\psi(2)(1_{st}) + c_2)(\psi(2)(1_{st}) + c_4) \\ &+ G_4(\psi(2)(1_{st}) + c_1)(\psi(2)(1_{st}) + c_2)(\psi(2)(1_{st}) + c_3)) \\ &\times ((\psi(2)(1_{st}) + c_1)(\psi(2)(1_{st}) + c_4) + (\psi(2)(1_{st}) + c_2) \\ &\times (\psi(2)(1_{st}) + c_3)) \\ &+ 4G_5\psi(2)(1_{st})(\psi(2)(1_{st}) + c_1)(\psi(2)(1_{st}) + c_2)(\psi(2)(1_{st}) \\ &+ c_3)(\psi(2)(1_{st}) + c_4)) = 0 \end{aligned}$$

$R(\Psi(2)(1_{st}))$ is a fifth-order polynomial equation, and its first through fifth derivative equations are obtained by regular transformation. Actually the fifth derivative is given as

$$\begin{aligned} &\frac{d^5}{(d\psi(2)(1_{st}))^5} R(\psi(2)(1_{st})) \\ &= \left(\frac{1}{4}\right)^5 \times 5 \times 4 \times 3 \times 2 \times (2G_1 + 2G_2 + 2G_3 + 2G_4 + 4G_5) \\ &= \left(\frac{1}{4}\right)^5 \times 240 \times N_{\text{chromosomes}}, \end{aligned}$$

where $N_{\text{chromosomes}}$ stands for number of chromosomes in the genotype data. $[d^4/(d\psi(2)(1_{st}))^4]R(\psi(2)(1_{st})) = 0$ is a first-order function and it is solved arithmetically. Thereafter solutions of $[d^3/(d\psi(2)(1_{st}))^3]R(\psi(2)(1_{st})) = 0$, $[d^2/(d\psi(2)(1_{st}))^2]R(\psi(2)(1_{st})) = 0$, $[d/(d\psi(2)(1_{st}))]R(\psi(2)(1_{st})) = 0$ and $R(\psi(2)(1_{st})) = 0$ are obtained using the Newton-Raphson method. The value of $\ln(L)\psi(2)(1_{st})$ for all local maxima and the two endpoints are then calculated and the absolute maximum is determined.

APPENDIX 3

Generalization of likelihood function expressed as a function of $\psi(n)(1)$.

Assume n SNPs that construct $\Gamma = \{\gamma_i\}$ composite genotypes. α_i individuals are observed to have a genotype γ_i. Further, assume γ_i has n_i heterozygous sites, and let $\Theta(\gamma_i) = \{(\theta_1, \hat{\theta}_1), (\theta_2, \hat{\theta}_2), \cdots, (\theta_{nc_i}, \hat{\theta}_{nc_i})\}$ denote the set of potential haplotype pairs for γ_i, where nc_i is the number of haplotype pairs for γ_i: ($nc_i = 1$ when $n_i = 0$, and $nc_i = 2^{(n_i-1)}$) otherwise. The $\ln(L)$ for the observed genotype data is expressed as

$$\ln(L) = \sum_{\gamma_i \in \Gamma} \alpha_i \times \ln\left(\sum_{j=1}^{nc_i}(f(\theta_j)f(\hat{\theta}_j))\right) + C \quad (*)$$

where $f(\theta_i)$ denotes frequency of θ_i. When all Ψ's except for $\psi(n)(1)$ are solved, $\psi(n)(1)$ is the only unsolved variable in Ψ. Therefore all $f(\theta_j)$ and $f(\hat{\theta}_j)$ are expressed as a univariate function of $\psi^{(n)}$ and equation (*) is also a univariate function of $\psi(n)(1)$ and differentiable as follows:

$$\frac{d}{d(\psi(n)(1))}\ln(L) = \sum_{\gamma_i \in \Gamma} \alpha_i \times \frac{\frac{d}{d(\psi(n)(1))}\left(\sum_{j=1}^{nc_i}(f(\theta_j)f(\hat{\theta}_j))\right)}{\sum_{j=1}^{nc_i}(f(\theta_j)f(\hat{\theta}_j))}.$$

Denote the subset of n_i SNPs that are heterozygous in genotype γ_i by $S^{(n_i)}_{\text{hetero}}(\gamma_i)$, and let $P(S^{(n_i)}_{\text{hetero}}(\gamma_i))$ be its power set and let $S(p_i)(q_i)(S^{(n_i)}_{\text{hetero}}(\gamma_i))$ be an element of $P(S^{(n_i)}_{\text{hetero}}(\gamma_i))$. Because $[d/d(\psi(n)(1))]f(\theta_j) = \pm(1/2^n)$, numerator of an element in (*), $[d/d(\psi(n)(1))]\left(\sum_{j=1}^{nc_i}(f(\theta_j)f(\hat{\theta}_j))\right)$, can be expressed as

$$d/d(\psi(n)(1))\left(\sum_{j=1}^{nc_i}(f(\theta_j)f(\hat{\theta}_j))\right) = \frac{1}{2^n} \times 2^{(nc_i+1)} \sum_{S(p_i)(q_i)(S^{(n_i)}_{\text{hetero}}(\gamma_i)) \in P(S^{(n_i)}_{\text{hetero}}(g_i)), u \neq S^{(n_i)}_{\text{hetero}}(g_i)} v(p_i)(q_i) \times \psi(p_i)(q_i),$$

where $\psi(u)$ denotes Ψ for a subset u and $v(u)$ is the value of corresponding haplotype.

APPENDIX 4

Classification of SNP pairs for which the EM algorithm did not direct toward the global maximum.

The SNP pairs that were not affected by the tightening of the threshold can be grouped into four categories (Patterns 1–4). The SNP pairs of Pattern 1 (85.0% of unaffected pairs) had a symmetric distribution of deterministic chromosomes for only two haplotypes with double heterozygotes. Such pairs exhibited two global maximum estimates at the two ends of the range of $\psi(2)(1_{st})$. As the EM algorithm started from the symmetric haplotype frequency in LE, the solution did not move from the LE condition due to this symmetry. For pairs in Pattern 2 (10.7%), the EM algorithm converged to $D' = 1$, whereas the $D' \neq 1$ condition gave the global maximum. Pairs in Pattern 3 (4.1%) were the opposite case, where the EM method converged to $D' \neq 1$ and the Ψ-based method converged to $D' = 1$. Pairs in Pattern 4 (0.28%) had multiple local maxima and the EM converged to a local maximum that was not the global maximum.

APPENDIX 5

Settings of programs to perform the standard EM algorithm and the Ψ-based algorithm.

For the standard EM, the maximum number of iterations was set at 10^9, and the calculation was stopped when the difference in $\log_{10}L$ between iterations became less than 10^{-12}. Without limitation on the maximum number of iterations, calculation did not end due to the slowness of convergence for some cases. For the Ψ-based method, no limitation was applied on the maximum number of iterations, and the iteration was stopped only when the difference in estimated x between iterations became less than 10^{-6}. Convergence of the Newton-Raphson method was fast in this case and it was unnecessary to set a limitation on the maximum number of iterations for the Ψ-based method.

APLAR Journal of Rheumatology 2006; 9: 353–358

REVIEW ARTICLE

The genetics of systemic lupus erythematosus: differences across ethnicities

Yuta KOCHI,[1] Kenichi SHIMANE[1,2] and Kazuhiko YAMAMOTO[1,2]

[1]*Laboratory for Rheumatic Diseases, SNP Research Center, RIKEN, Yokohama,* [2]*Department of Allergy and Rheumatology, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan*

Abstract

With the acceleration in understanding of the human genome over the past decade, genetic studies have revealed several specific genes associated with a predisposition to systemic lupus erythematosus (SLE). These studies have shown that some of the genetic variants are shared with other autoimmune diseases, and the contribution of each variant to disease differs among different ethnic groups. This article summarizes recent findings from genetic analyses of SLE, with particular emphasis on ethnic differences between Asian populations and others.

Key words: case-control study, genetic polymorphism, genetic predisposition to disease, linkage analysis, systemic lupus erythematosus.

Systemic lupus erythematosus (SLE) is a systemic autoimmune disease characterized by multiple organ damage. Pathological processes are mainly composed of autoantibody production and immune complex deposition, where genetic and environmental factors enhance the immune dysregulation. The presence of a genetic predisposition to SLE has been confirmed in familial and twin studies, with the relative risk ratio for siblings of affected individuals to disease incidence in the general population (λs) elevated to 20, and the concordance rate in monozygotic twins (24–58%) much higher than that in dizygotic twins (2–5%).[1] Taken together, these data suggest multiple genetic factors are involved in disease pathogenesis. With rapid increases in the understanding of the human genome over the past decade, genetic studies including familial linkage analyses and case-control association studies have mapped multiple candidate loci for SLE and revealed specific genes predisposing to the disease. Herein, we review recent findings from genetic analyses of SLE, with particular emphasis on ethnic difference between Asian populations and others.

Correspondence: Y. Kochi, Laboratories for Rheumatic Diseases, SNP Research Center, RIKEN, Yokohama 230-0045, Japan. Email: ykochi@src.riken.jp

LINKAGE ANALYSES

To date, at least 10 linkage studies have been undertaken for whole-genome surveys of SLE susceptibility genes. Although most of these studies primarily recruited Caucasian families, some also enrolled families of other ethnic groups, including African-American, Hispanic and Asian families. However, no linkage studies have been conducted using a sufficient number of Asian families. Although many candidate loci have been identified, the results are inconsistent across studies. This inconsistency may reflect false-positive and false-negative errors partially attributable to the low statistical power of individual studies. A recent meta-analysis of these studies by Lee *et al.* identified genomic regions at 6p22.3–6p21.1 and 16p12.3–16q12.2 as strong candidate loci for SLE susceptibility with statistical significance in a genome-wide scan.[2] Other candidate loci repeatedly presented in linkage studies are summarized in Table 1.

As each locus identified by linkage analysis encompasses at least several megabases of the genomic region, fine mapping using additional markers such as single-nucleotide polymorphisms (SNPs) is needed to determine the true susceptibility genes. Alternatively, candidate gene approach analyses can be performed to target genes in

Table 1 Candidate loci of systemic lupus erythematosus linkage analyses†

Chromosome location	Candidate genes in the locus
1p36	C1q, TNFRSF1B (TNFR2)
1q22–24	FCGRs, FCRL3, CRP
2q37	PDCD1
4p15–13	
6p11–p22	HLA, C4
16q12–21	
17p13–12	TNFSF13 (APRIL)
17q11	
18q21–23	BCL-2
19q13	
20p13	

†Candidate loci that have been replicated in at least two linkage studies were listed, which is reviewed in detail in Lee and Nath.[2]

the loci exhibiting pathological relevance to the disease. Several genes such as FCGRs in 1q23,[3] PARP in 1q41–42[4] and PDCD1 in 2q37[5] have been successfully identified using this approach.

CANDIDATE GENES

Many association studies using case-control cohorts or multiplex families with affected individuals have been performed to investigate whether candidate genes of interest are associated with disease susceptibility. Both genes in the linkage loci and those outside have been examined, and several genes have shown associations with disease susceptibility (Table 2). Among these, we preferentially selected and described herein those for which associations have been confirmed in multiple independent studies.

Genes in the major histocompatibility (MHC) complex (6p21)

The MHC complex, located in a 3.6-Mb region of chromosome 6 (6p21), is one of the strongest candidate loci for SLE linkage studies, as mentioned above. This region contains at least 128 genes with predicted expression, 40% of which are estimated to function in the immune system.[6] Many autoimmune diseases, in addition to other diseases such as myocardial infarction and narcolepsy, have been associated with genes in this region.[7]

Specific alleles of the HLA-DR gene, which encodes an essential molecule in antigen presentation to T-cells, are believed to form the major determinant of SLE susceptibility in this region. While increased frequency of HLA-DR2 (DRB1*1501) and DR3 (DRB1*0301) alleles have been repeatedly observed in Caucasian SLE patients,[8]

Table 2 Candidate genes associated with systemic lupus erythematosus susceptibility†

Gene		Chromosome	Ethnicity		Association with other
Name	Function	location	Individual studies	Meta-analyses	autoimmune diseases
HLA-DR	Antigen presentation	6p21	Af, As, C, H	–	RA, AITD, T1D, MS
FASL	Apoptosis	1q23	Af, As, H	–	–
FAS	Apoptosis	10q24	As	–	–
CRP	Cell clearance	1q23	C	–	–
C4	Complement pathway	6p21	Af, As, C, H	–	Several autoimmune diseases‡
MBL2	Complement pathway	10q11.2–22	Af, As, C	Mix	–
TNF-α	Cytokine	6p21	Af, As, C	C	Several autoimmune diseases‡
IL-10	Cytokine	1q31–32	As, H	–	MS
FCGR2A	Immunoglobulin receptor	1q23	Af, As, C	C	MS
FCGR2B	Immunoglobulin receptor	1q23	As, C	As	–
FCGR3A	Immunoglobulin receptor	1q23	Af, As, C	As, C	RA
FCGR3B	Immunoglobulin receptor	1q23	As, C	As, C	MS
FCRL3	Lymphocyte coreceptor	1q23	As	–	RA, AITD
CTLA-4	Lymphocyte coreceptor	2q33	As, C	As	RA, AITD, T1D
PDCD1	Lymphocyte coreceptor	2q37	C	–	RA, T1D
PTPN22	Lymphocyte signalling	1p13	Af, C	C	RA, ATID, T1D
IRF5	Lymphocyte signalling	7q32	C	–	–
TYK2	Lymphocyte signalling	19p13	C	–	–

†Results of positive association but not negative association are presented in both individual studies and meta-analyses.
‡The association observed may be attributable to the linkage disequilibrium with HLA polymorphism.
Af, African; As, Asian; C, Caucasian; H, Hispanic; Mix, Mixture of ethnic groups; RA, rheumatoid arthritis; AITD, autoimmune thyroiditis; T1D, type 1 diabetes; MS, multiple sclerosis.

DR2 allele has commonly been associated with disease susceptibility in most Asian populations.[9] Other alleles, such as DR3 in Chinese,[10] DR4 in Japanese,[11] and DR9 (DRB1*0901) in Koreans,[12] are also reportedly increased in patients. This complicated association of the HLA-DR locus with disease may reflect a diverse distribution of HLA alleles in the populations studied. However, other genes in linkage disequilibrium with the HLA-DR gene may also be involved and influence the increased frequency of specific HLA-DR alleles.

Complement component 4 (C4) gene, which encodes a component of the complement cascade, is another candidate in this region. C4 is encoded by two tandemly duplicated genes (C4A and C4B) and is highly polymorphic, with copy-number polymorphisms (CNP) including nonexpressed alleles (designated as C4AQ*0 and C4BQ*0) for which no protein product is identifiable. Many studies of Caucasian populations have demonstrated that the C4AQ*0 null allele is associated with SLE susceptibility, with a gene dose-dependent effect.[13] As C4AQ*0 allele is in strong linkage disequilibrium with HLA-DR3 allele, separation of the relative contributions of the two genes is difficult. To solve this problem, a study of an HLA-DR3-negative cohort was performed, showing an independent effect of C4AQ*0 allele on disease.[14] This independent association of C4AQ*0 allele has also been described in Japanese and Chinese populations[15] in which frequency of DR3 allele was relatively lower than in Caucasian populations.

FCGR gene family (1q23)

The Fc fragment of IgG, low-affinity IIb, receptor (FCGR2B) gene is located in 1q23, and has been associated with SLE susceptibility in Japanese;[16] a finding further confirmed by a meta-analysis using case-control cohorts of Thai and Chinese populations.[17] This polymorphism in the transmembrane domain of FCGR2B replaces the amino acid isoleucine with threonine (I232T), resulting in a reduced association of the receptor with lipid rafts of cell membrane and reduced inhibitory function in B-cells and monocytes.[18,19] Although this mutant has rarely been observed in Caucasian populations, other functional polymorphisms in the promoter region of FCGR2B gene have been associated with disease in a Caucasian case-control cohort[20] suggesting ethnic differences exist between Asian and Caucasian populations.

Many studies have reported associations of variants in other FCGR family genes with SLE susceptibility, as reviewed in detail elsewhere.[21] More than 20 studies in several ethnic groups have examined associations between SLE susceptibility and an SNP in the extracellular domain of FCGR2A, the mutant allele of which reduces binding affinity of IgG. Although the results are inconsistent, a meta-analysis has concluded that a positive association exists in Caucasian populations[22] but not in Asian populations.[23] Similarly, an SNP in the ligand-binding domain of FCGR3A is enriched in patients with SLE and may represent a risk factor for lupus nephritis, and has been confirmed by meta-analysis in both Caucasian[24] and Asian populations.[23] Both disease susceptibility alleles in FCGR2A and FCGR3A display less affinity to IgG compared to normal alleles, suggesting that defective clearance of immune complexes may be responsible for disease pathogenesis. Moreover, CNP in FCGR3B, another member of the FCGRs, is associated with human SLE in Caucasians.[25] Interestingly, similar associations between CNP of the rat ortholog, Fcgr3, and susceptibility to lupus-like nephritis have also been observed.[25]

We have recently reported a regulatory variant in the Fc receptor-like 3 (FCRL3) gene, a homolog of classical Fcγ receptors, was associated with three autoimmune diseases in the Japanese population, including rheumatoid arthritis (RA), autoimmune thyroiditis (AITD) and SLE.[26] This polymorphism alters the binding affinity of nuclear factor (NF)κB, resulting in high expression in cells with the disease-susceptible genotype. As this association has been replicated in an independent Japanese RA cohort[27] and a Caucasian AITD cohort[28] the FCRL3 polymorphism may be a common genetic factor in different autoimmune diseases and different ethnic groups, and further confirmation is required. Although the precise function of FCRL3 is unknown, its preferential expression in the germinal centre light zone suggests it may affect the clonal selection of B-cells and augment the emergence of self-reactive cells.

MBL2 (10q11.2-22)

The protein encoded by mannose binding lectine 2 (MBL2) gene recognizes mannose and N-acetylglucosamine on bacterial pathogens, and is capable of activating the classical complement pathway. Because deficiencies of complement pathway molecules have been reportedly associated with SLE pathogenesis[29] MBL2 is another attractive candidate gene of SLE susceptibility. Serum levels of MBL2 vary from person to person, and 5–10% of populations throughout the world lacks the protein. In addition to the common form of MBL2 gene (allele A), three mutant forms of MBL2 (alleles B, C and D for each, and allele O for altogether) are produced, due to three non-synonymous SNPs in exon 1. These mutant protein forms are unstable in serum, resulting in almost undetectable levels in individuals with the O/O genotype. Moreover, two SNPs in the promoter region of MBL2 gene (alleles L and X) also reportedly affect gene expression.[30] Allele frequencies of mutant forms differ among ethnic groups, with allele B prevalent in Caucasian and Asian populations and allele C most common in African-American populations. A recent meta-analysis of 15 association studies on the MBL2 gene in Caucasian, African-American and Asian populations has shown that alleles B, L and X are associated with SLE susceptibility with a modest effect size.[31] Interestingly, another study using a prospective cohort has revealed that the MBL2 genotype could represent a prognostic factor for arterial thrombosis in SLE patients.[32]

IRF5 (7q32)

Increased production of type I interferon (IFN) and expression of IFN-inducible genes is commonly observed in SLE patients, and may be pivotal in the disease pathogenesis. Recently, using joint analysis of linkage and association in Swedish and Finnish Caucasians, SNPs in the IFN regulatory factor 5 (IRF5) gene displayed strong signals.[33] Follow-up confirmation was achieved by a replication case-control analysis of Caucasian populations in Argentina, Spain, Sweden and the US.[34] In the latter study, a common haplotype of the IRF5 gene, which was increased in affected individuals, was shown to drive elevated expression of multiple unique isoforms of IRF5. Although a clear association with the IRF5 gene exists in Caucasians, no studies have yet been performed in Asian populations.

PDCD1 (2q37)

The programmed cell death 1 (PDCD1) gene encodes an inhibitory receptor on lymphocytes, and represents one candidate gene in the linkage loci 2q37, since mice lacking Pdcd1 gene, the murine ortholog of PDCD1, develop spontaneous lupus-like disease phenotypes.[35] A regulatory SNP in PDCD1 has been associated with Hispanic and Caucasian populations, and gene expression is decreased in the disease-susceptibility genotype.[5] The association was not significantly replicated in a Taiwanese SLE cohort, but a significant association was identified in the same study using a Taiwanese RA cohort.[36] To reach a firm conclusion on the contribution of PDCD1 gene to SLE susceptibility, additional studies in several ethnic groups may be needed.

PTPN22 (1p13)

A non-synonymous SNP in the tyrosine phosphatase non-receptor type 22 (PTPN22) gene is a genetic predisposition commonly shared by most human autoimmune diseases in Caucasian populations. Association of the PTPN22 variant with disease was first reported in Caucasian patients with type I diabetes[37] followed by positive associations in studies of RA,[38] SLE,[39] AITD[40] and other autoimmune diseases.[41] The autoimmune-predisposing allele has been shown to be a gain-of-function mutant, and the encoded phosphatase is a more negative regulator of T-cells. However, this polymorphism has not been observed in East Asian populations[38,42] and thus represents another example of ethnic differences in disease predisposition.

CONCLUSIONS AND FUTURE PROSPECTS

Recent genetic studies of SLE have yielded new insights into the pathogenesis of disease. While some genetic variants are solely associated with SLE and may determine the unique phenotype of disease, others also increase the risk of other autoimmune diseases. In addition, differences among different ethnic groups in the contribution of each genetic variant to disease are also becoming apparent. Although these remarks are based on our successful research experience, a grasp of the complete picture of SLE genetics remains elusive. Several genes remain to be uncovered, and the limitations of conventional linkage-based, candidate gene approach analysis must be overcome to reveal these.

With the completion of the international HapMap project[21] and the emergence of genotyping technologies, genetic research into complex traits is entering a new era. The HapMap project provided a catalogue of common genetic variants, comprising a total of 250 000 SNPs for Asian populations to cover the whole genome. These variants can now be genotyped for an individual at once, using new technologies such as DNA array-based methods. This allows a more comprehensive approach to surveying for genes predisposing to disease. In the not-too-distant future, all the predispositions of SLE should be clarified.

REFERENCES

1 Tsao BP, Cantor RM, Kalunian KC, Wallace DJ, Hahn BH, Rotter JI (1998) The genetic basis of systemic lupus erythematosus. *Proc Assoc Am Physicians* 110, 113–7.

2 Lee YH, Nath SK (2005) Systemic lupus erythematosus susceptibility loci defined by genome scan meta-analysis. *Hum Genet* 118, 434–43.

3 Salmon JE, Millard S, Schachter LA, *et al.* (1996) Fc gamma RIIA alleles are heritable risk factors for lupus nephritis in African Americans. *J Clin Invest* 97, 1348–54.

4 Tsao BP, Cantor RM, Grossman JM, *et al.* (1999) PARP alleles within the linked chromosomal region are associated with systemic lupus erythematosus. *J Clin Invest* **103**, 1135–40.

5 Prokunina L, Castillejo-Lopez C, Oberg F, *et al.* (2002) A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nat Genet* 32, 666–9.

6 The MHC sequencing consortium. (1999) Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* **401**, 921–3.

7 Shiina T, Inoko H, Kulski JK (2004) An update of the HLA genomic region, locus information and disease associations: 2004. *Tissue Antigens* 64, 631–49.

8 Harley JB, Moser KL, Gaffney PM, Behrens TW (1998) The genetics of human systemic lupus erythematosus. *Curr Opin Immunol* **10**, 690–6.

9 Fong KY, Boey ML (1998) The genetics of systemic lupus erythematosus. *Ann Acad Med Singapore* 27, 42–6.

10 Zhang J, Ai R, Chow F (1997) The polymorphisms of HLA-DR and TNF B loci in northern Chinese Han nationality and susceptibility to systemic lupus erythematosus. *Chin Med Sci J* 12, 107–10.

11 Hirose S, Ogawa S, Nishimura H, Hashimoto H, Shirai T (1988) Association of HLA-DR2/DR4 heterozygosity with systemic lupus erythematosus in Japanese patients. *J Rheumatol* 15, 1489–92.

12 Hong GH, Kim HY, Takeuchi F, *et al.* (1994) Association of complement C4 and HLA-DR alleles with systemic lupus erythematosus in Koreans. *J Rheumatol* 21, 442–7.

13 Yang Y, Chung EK, Zhou B, *et al.* (2004) The intricate role of complement component C4 in human systemic lupus erythematosus. *Curr Dir Autoimmun* 7, 98–132.

14 Howard PF, Hochberg MC, Bias WB, Arnett FC Jr, McLean RH (1986) Relationship between C4 null genes, HLA-D region antigens, and genetic susceptibility to systemic lupus erythematosus in Caucasian and black Americans. *Am J Med* **81**, 187–93.

15 Dunckley H, Gatenby PA, Hawkins B, Naito S, Serjeantson SW (1987) Deficiency of C4A is a genetic determinant of systemic lupus erythematosus in three ethnic groups. *J Immunogenet* 14, 209–18.

16 Kyogoku C, Dijstelbloem HM, Tsuchiya N, *et al.* (2002) Fcgamma receptor gene polymorphisms in Japanese patients with systemic lupus erythematosus: contribution of FCGR2B to genetic susceptibility. *Arthritis Rheum* **46**, 1242–54.

17 Chu ZT, Tsuchiya N, Kyogoku C, *et al.* (2004) Association of Fcgamma receptor IIb polymorphism with susceptibility to systemic lupus erythematosus in Chinese: a common susceptibility gene in the Asian populations. *Tissue Antigens* 63, 21–7.

18 Kono H, Kyogoku C, Suzuki T, *et al.* (2005) FcgammaRIIB Ile232Thr transmembrane polymorphism associated with human systemic lupus erythematosus decreases affinity to lipid rafts and attenuates inhibitory effects on B cell receptor signaling. *Hum Mol Genet* 14, 2881–92.

19 Floto RA, Clatworthy MR, Heilbronn KR, *et al.* (2005) Loss of function of a lupus-associated FcgammaRIIb polymorphism through exclusion from lipid rafts. *Nat Med* **11**, 1056–8.

20 Su K, Wu J, Edberg JC, *et al.* (2004) A promoter haplotype of the immunoreceptor tyrosine-based inhibitory motif-bearing FcgammaRIIb alters receptor expression and associates with autoimmunity. I. Regulatory FCGR2B polymorphisms and their association with systemic lupus erythematosus. *J Immunol* 172, 7186–91.

21 Croker JA, Kimberly RP (2005) Genetics of susceptibility and severity in systemic lupus erythematosus. *Curr Opin Rheumatol* 17, 529–37.

22 Karassa FB, Bijl M, Davies KA, *et al.* (2003) Role of the Fcgamma receptor IIA polymorphism in the antiphospholipid syndrome: an international meta-analysis. *Arthritis Rheum* 48, 1930–8.

23 Tsuchiya N, Kyogoku C (2005) Role of Fc gamma receptor IIb polymorphism in the genetic background of systemic lupus erythematosus: insights from Asia. *Autoimmunity* 38, 347–52.

24 Karassa FB, Trikalinos TA, Ioannidis JP (2003) The Fc gamma RIIIA-F158 allele is a risk factor for the development of lupus nephritis: a meta-analysis. *Kidney Int* **63**, 1475–82.

25 Aitman TJ, Dong R, Vyse TJ, *et al.* (2006) Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* 439, 851–5.

26 Kochi Y, Yamada R, Suzuki A, *et al.* (2005) A functional variant in FCRL3, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities. *Nat Genet* 37, 478–85.

27 Ikari K, Momohara S, Nakamura T, *et al.* (2006) Supportive evidence for a genetic association of the FCRL3 promoter polymorphism with rheumatoid arthritis. *Ann Rheum Dis* 65, 671–3.

28 Simmonds MJ, Heward JM, Carr-Smith J, Foxall H, Franklyn JA, Gough SC (2006) Contribution of single nucleotide polymorphisms within FCRL3 and MAP3K7IP2 to the pathogenesis of Graves' disease. *J Clin Endocrinol Metab* 91, 1056–61.

29 Seelen MA, Roos A, Daha MR (2005) Role of complement in innate and autoimmunity. *J Nephrol* 18, 642–53.

30 Madsen HO, Garred P, Thiel S, *et al.* (1995) Interplay between promoter and structural gene variants control basal serum level of mannan-binding protein. *J Immunol* **155**, 3013–20.

31 Lee YH, Witte T, Momot T, *et al.* (2005) The mannose-binding lectin gene polymorphisms and systemic lupus erythematosus: two case-control studies and a meta-analysis. *Arthritis Rheum* 52, 3966–74.

32 Ohlenschlaeger T, Garred P, Madsen HO, Jacobsen S (2004) Mannose-binding lectin variant alleles and the risk of arterial thrombosis in systemic lupus erythematosus. *N Engl J Med* **351**, 260–7.

33 Sigurdsson S, Nordmark G, Goring HH, *et al.* (2005) Polymorphisms in the tyrosine kinase 2 and interferon

regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am J Hum Genet* 76, 528–37.

34 Graham RR, Kozyrev SV, Baechler EC, *et al.* (2006) A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat Genet* 38, 550–5.

35 Nishimura H, Nose M, Hiai H, Minato N, Honjo T (1999) Development of lupus-like autoimmune diseases by disruption of the PD-1 gene encoding an ITIM motif-carrying immunoreceptor. *Immunity* 11, 141–51.

36 Lin SC, Yen JH, Tsai JJ, *et al.* (2004) Association of a programmed death 1 gene polymorphism with the development of rheumatoid arthritis, but not systemic lupus erythematosus. *Arthritis Rheum* 50, 770–5.

37 Bottini N, Musumeci L, Alonso A, *et al.* (2004) A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet* 36, 337–8.

38 Begovich AB, Carlton VE, Honigberg LA, *et al.* (2004) A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 75, 330–7.

39 Kyogoku C, Langefeld CD, Ortmann WA, *et al.* (2004) Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. *Am J Hum Genet* 75, 504–7.

40 Criswell LA, Pfeiffer KA, Lum RF, *et al.* (2005) Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. *Am J Hum Genet* 76, 561–71.

41 Bottini N, Vang T, Cucca F, Mustelin T (2006) Role of PTPN22 in type 1 diabetes and other autoimmune diseases. *Semin Immunol* 18, 207–13.

42 Mori M, Yamada R, Kobayashi K, Kawaida R, Yamamoto K (2005) Ethnic differences in allele frequency of autoimmune-disease-associated SNPs. *J Hum Genet* 50, 264–6.

話 題

関節リウマチ感受性遺伝子FCRL3*

高 地 雄 太**

Key Words：FCRL3, rheumatoid arthritis, single-nucleotide polymorphisms, case-control association study

はじめに

関節リウマチ(rheumatoid arthritis；RA)は，遺伝因子に環境因子が重なり合うことにより発症に至ると考えられる多因子疾患である[1]．家系解析により，一卵性双生児再発危険率λ_{MZ}(＝一卵性双生児における発症一致率/集団発症率)は12～62倍，同胞再発危険率λ_{sib}(同胞発症率/集団発症率)は2～17倍と算定されており，RAの遺伝性を強く支持する[2]．また，λ_{MZ}とλ_{sib}のあいだにかなりの開きがあることは，原因となる遺伝的因子が異なった染色体上に複数存在することを示唆する．RAにおける最大の遺伝素因は，HLAクラスII遺伝子である*HLA-DR*遺伝子の多型であり，RAの遺伝素因の30～50％を占めるとされる．日本人では，*HLA-DR*多型のうち，*HLA-DRB1**0405, *0101といったサブタイプとの強い関連が報告されている[3]．HLA以外の遺伝因子の探索は，HLA遺伝子多型と比較して，その相対的に低い寄与率から，困難が予測されてきた．しかし，ヒトゲノム情報の整備および遺伝子多型のタイピング技術の進歩により，ホールゲノムを対象とした，大規模検体を用いたゲノム解析が可能となり，ここ数年，*PADI4*，*PTPN22*，*SLC22A4*といった遺伝子の多型とRA感受性の関連が相次いで報告されるに至っている[4]．著者らは，ホールゲノムに分布する一塩基多型(single-nucleotide polymorphisms；SNPs)を用いて，そのアレル頻度を患者群・対照群で比較する関連解析(case-control association study)を行い，疾患と関連する遺伝子の探索を行ってきた．今回，1番染色体長腕21-23領域(1q21-23)領域に存在する*FCRL3*遺伝子のSNPとRA感受性との関連を同定した[5]．以下に，*FCRL3*遺伝子多型と自己免疫疾患との関連について，われわれの研究成果を中心に概説する．

1q21-23領域における関連解析

これまで行われてきた連鎖解析や関連解析では，1q21-23領域に存在する遺伝子が，RA以外にも全身性エリテマトーデス(SLE)[6]，多発性硬化症[7]，乾癬[8]などの自己免疫性疾患感受性と関連することが報告されている．また，ヒト疾患のみならず，ループスモデル，コラーゲン誘導性関節炎モデル，実験的アレルギー性脳脊髄炎などの動物疾患モデルにおいても，QTL(quantitative trait locus)解析により，候補領域としてあげられてきている[9]．したがって，本領域において，ヒト・マウス共通の自己免疫疾患感受性遺伝子の存在が強く示唆された．

われわれは，まず，1q21-23領域(16Mbp)に存在する多型の連鎖不平衡状態を評価するために，日本人の代表的な遺伝子多型のデータベースであるJSNP[10]より，491 SNPsを抽出した．これらのSNPsについて，日本人対照群658人のジェノタイピングを行い，連鎖不平衡係数を算出した

* Association of a FCRL3 gene variant with rheumatoid arthritis susceptibility.

** Yuta KOCHI, M.D., Ph.D.：独立行政法人理化学研究所遺伝子多型研究センター関節リウマチ関連遺伝子研究チーム〔〒230-0045 横浜市鶴見区末広町1-7-22〕；Laboratory for Rheumatic Diseases, SNP Research Center, RIKEN, Yokohama 230-0045, JAPAN

表１　*FCRL3*遺伝子多型とRA感受性との関連

SNP		アレル	アレル１頻度		ジェノタイプ11対12＋22		
番号	位置	1/2	患者群	対照群	オッズ比(95％CI)	χ^2	p
fcrl3_3	-169	C/T	0.420	0.345	2.15(1.58-2.93)	24.3	0.00000085
fcrl3_4	-110	A/G	0.253	0.184	3.01(1.71-5.29)	16.1	0.000060
fcrl3_5	Exon2	C/G	0.419	0.346	2.05(1.51-2.78)	21.6	0.0000033
fcrl3_6	Intron3	A/G	0.422	0.344	2.02(1.49-2.75)	20.8	0.0000052

CI：confidence interval

ところ，110個の連鎖不平衡ブロックが同定された．次に，関連解析の１次スクリーニングとして，RA患者94人の検体を用いて，これら491 SNPsのジェノタイピングを行い，対照群658人とのアレル頻度の比較を行った．その結果，9 SNPsにおいて，患者・対照間での有意な差を認めた($p<0.01$)．さらにこれらのSNPsについては，２次スクリーニングとして，患者734人を追加でジェノタイピングし，対照群と比較したところ，*FCRL3*(*Fc receptor-like 3*)遺伝子のイントロンのSNPとの強い関連を認めた(オッズ比1.39，$p<0.0001$)．このことより，この多型，もしくは，連鎖不平衡状態にある多型が疾患に直接的に関与していることが考えられた．そこで，*FCRL*遺伝子群(*FCRL1～5*)が存在する２つの連鎖不平衡ブロックの詳細な解析を行った．この領域に存在する41 SNPsに対して，RA患者群830人，対照群658人のアレル頻度およびジェノタイプ頻度の比較を行ったところ，*FCRL3*遺伝子の４つのSNPがRA感受性と強い関連を示すことが明らかにされた(表１)．

プローモーター領域多型(-169C→T)の解析

RA感受性との強い関連を認めた4 SNPsはいずれも翻訳領域に存在せず，アミノ酸置換をもたらさない(図１-A)．そこで，これらの多型が，遺伝子発現になんらかの影響を与えて疾患に関与している可能性を考え，ハプロタイプ別の転写活性の評価を行った．患者群で認めた３つのハプロタイプからなるプローモーター配列を，それぞれベクターにクローニングし，ルシフェラーゼ・アッセイにより転写活性を比較したところ，SNP fcrl3_3(-169C→T)のＣアレルを含むハプロタイプで転写活性が高かった．また，-169C→Tの周辺配列を用いた転写増強活性の評価においても，-169Cアレルで有意に高かった(図１-B)．したがって，-169C→Tの周辺配列になんらかの転写因子が結合し，アレルによって，その結合能が変化するため，転写活性の違いが生じている可能性が考えられた．転写因子結合予測ソフト(TRANSFAC)を用いて解析したところ，この配列はNF-κB結合モチーフとの相同性が高いことが明らかになった．実際に，ゲルシフトアッセイにより，-169C→T周辺配列に，NF-κBが結合することが示され，さらに，-169Cアレルにおいて強く結合することが明らかになった(図１-C)．これらのことから，疾患感受性と強い関連を認めたこの多型(-169C→T)は転写因子NF-κBの結合を介して，*FCRL3*遺伝子の発現量を制御していることが考えられた．健常人のＢリンパ球のRNAを用いた定量的RT-PCR法による解析では，-169C/C，-169C/T，-169T/Tというジェノタイプの順で，*FCRL3*の発現量が高かった(図１-D)．

*FCRL3*遺伝子多型(-169C→T)と自己抗体産生の関連

次に，*FCRL3*遺伝子多型が，RAの病態になんらかの影響を与える可能性を考え，RA患者における自己抗体の産生と，*FCRL3*多型(-169C→T)との関連を調べた．RAにおける代表的な自己抗体であるリウマトイド因子(RF)および抗環状シトルリン化ペプチド抗体(抗CCP抗体)を測定し，ジェノタイプ別に評価した．リウマトイド因子の疾患経過中の最大値は，感受性アレル(-169C)の数で有意に回帰された(表２)．また，抗CCP抗体の陽性率もジェノタイプによって差があり，感受性アレル数が多いほど陽性率が高かった．疾患感受性アレル数と，*FCRL3*遺伝子の発現量および自己抗体産生が相関することをあわせて

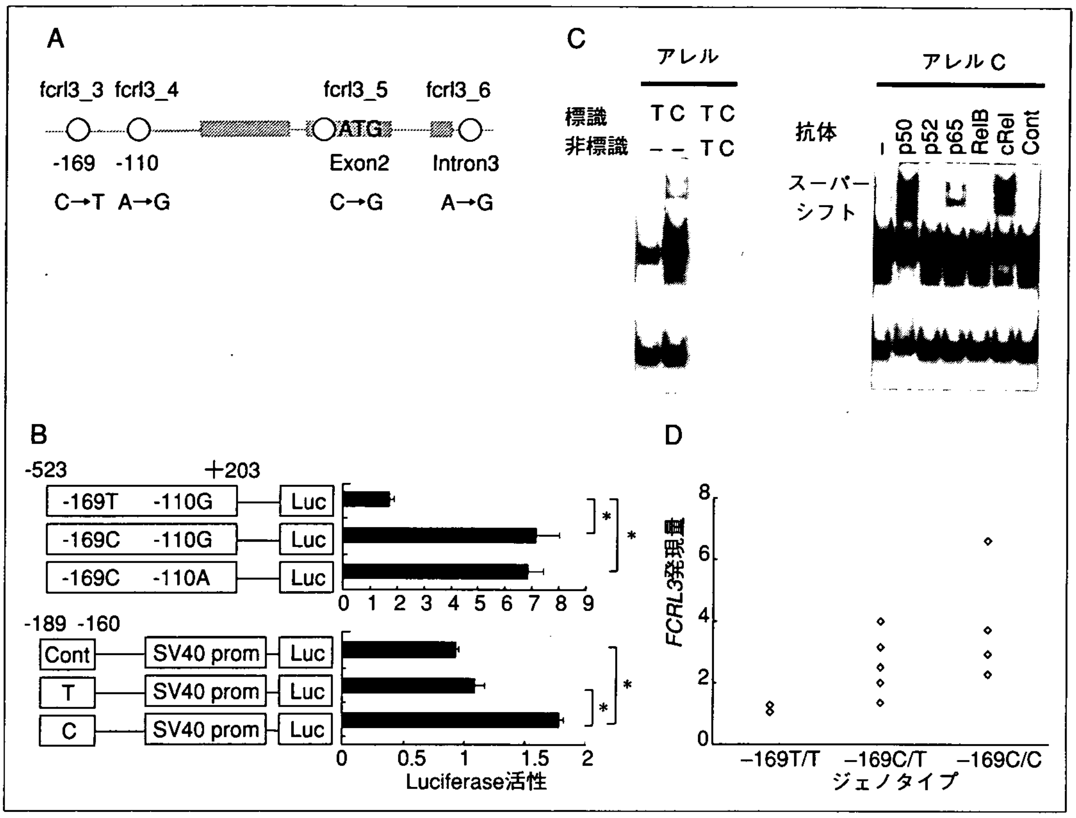


図1　RA感受性と関連する*FCRL3*遺伝子多型(-169C→T)は遺伝子発現を制御する

A：RA感受性と強い関連を認めた*FCRL3*遺伝子の4 SNPs，

B：*FCRL3*遺伝子の転写活性をルシフェラーゼ・アッセイにより評価．-169Cアレル(疾患感受性アレル)で活性が高い(*$p<0.001$)，

C：SNP周辺配列に結合する転写因子(EMSA)．-169Cアレルで核蛋白質の強い結合を認め(左パネル)，抗NF-κB抗体でスーパーシフトする(右パネル)，

D：健常人B細胞における*FCRL3*遺伝子の発現量を，SNPのジェノタイプ別に定量．　(文献[5]より一部改変)

考えると，*FCRL3*遺伝子の発現量の増加が，自己抗体産生になんらかの影響を与えるものと考えられた．

*FCRL3*遺伝子多型とほかの自己免疫疾患との関連

これまで，自己免疫疾患との関連が明らかになった遺伝子多型のうち，*CTLA4*，*PTPN22*，*SLC22A4*といった遺伝子の多型は，複数の自己免疫疾患の感受性と関連していることが報告されている．このことは，自己免疫疾患において共通の遺伝因子が存在することを示唆する．*FCRL3*遺伝子多型が，ほかの自己免疫疾患でも，疾患感受性と関連している可能性を検討するため，SLEおよび自己免疫性甲状腺疾患(AITD)における関連解析を行った．解析には，SLE患者564人，AITD患者509人(バセドウ病患者351人，橋本病患者158人)および対照群2,046人を用いた．アレル頻度比較では，AITD患者，SLE患者ともに，疾患群においてRA感受性*FCRL3*多型(-169C)の頻度が高く，有意な関連を認めた(AITD：オッズ比1.38，$p=0.0000042$，SLE：オッズ比1.17，$p=0.025$)．さらに，SLE患者の代表的な自己抗体である抗DNA抗体価を，ジェノタイプ別に評価したところ，その最大値(活動期のもの)は，-169C/Cジェノタイプ群において，それ以外のジェノタイプ(-169C/T，-169T/T)と比較して，有意に高かった(294.1IU/*ml* vs 145.5IU/*ml*，n＝120，$p<$

表2　SNP-169C→Tジェノタイプと自己抗体産生

	リウマトイド因子		抗CCP抗体	
ジェノタイプ	人数 (N=148)	血清抗体価 ±SEM(IU/ml)	人数 (N=71)	陽性率(%)
-169 C/C	29	479.9±91.3*	17	$100.0^{\dagger}$
-169 C/T	75	323.7±47.3*	35	$94.3^{\dagger}$
-169 T/T	44	216.4±44.0*	19	$73.7^{\dagger}$

* $R^2=0.049$, $p=0.0065$(回帰分析), $^{\dagger}p=0.029$(フィッシャーの直接検定)

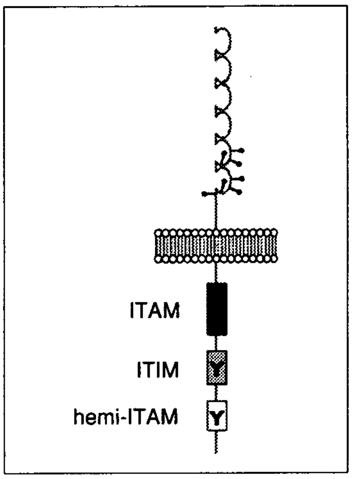


図2　FCRL3の分子構造

ITAM：immunoreceptor tyrosine-based activation motif, ITIM：immunoreceptor tyrosine-based inhibition motif　　(文献[12]より一部改変)

0.05). このことは，FCRL3多型が，RA以外の自己免疫疾患においても，疾患感受性に関連しているだけではなく，発症後の病態において，自己抗体産生を増強する因子となっている可能性を示唆する．

FCRL3分子と自己免疫

*FCRL3*遺伝子は，Fcγレセプター遺伝子との相同性が高い遺伝子群として同定されたFc receptor-like遺伝子ファミリーに属する[11]．その蛋白質構造から，膜型受容体としてのシグナル伝達機能が予測されているが，リガンド・機能ともに未知である．細胞内ドメインは，免疫細胞のレセプターに特徴的なチロシンモチーフをもつため，このレセプターはリガンドとの結合により，細胞内に正もしくは負のシグナルを伝達する可能性が考えられている(図2)[12]．*FCRL3*遺伝子は，脾臓・リンパ節・扁桃といった2次リンパ組織において発現しているが，とくに，胚中心におけるB細胞での高発現が確認されている[13]．胚中心はB細胞が，その抗原レセプターの変異を起こし，クローン選択を受ける場として知られるが，*FCRL3*遺伝子の高発現が，自己抗体産生と関連していることから，FCRL3は胚中心におけるB細胞の選択において，なんらかの影響を与え，自己応答性クローンの出現およびその活性化に寄与している可能性が考えられる．今後，FCRL3蛋白の詳細な機能解析が，自己免疫疾患の病態における役割を明らかにするものと考えられる．

おわりに

関節リウマチ(rheumatoid arthritis；RA)は，多因子疾患であることからも推測されるように，さまざまな病態が混在する疾患である．たとえば，数年の経過で全身の骨破壊をきたす患者が存在する一方で，軽度の関節症状のみで経過する患者も臨床的には経験される．近年，生物学的製剤をはじめとするRAの治療法の進歩には目を見張るものがあるが，個々の患者の予後予測に基づいた治療法の選択が必ずしもできていないのが現状である．これまでにも，*HLA-DR*遺伝子多型やリウマトイド因子の有無などがRAの予後と関連することが報告されているが[14]，本研究において，*FCRL3*遺伝子多型が自己抗体産生と関連するという事実は，*FCRL3*遺伝子多型が，予後予測因子になりうる可能性を示唆する．ゲノム解析によってもたらされる多型情報を複合的に解析することにより，今後，個人の遺伝情報に基づいた予後予測，および治療法の選択が可能になることが期待される．

文　献

1) Firestein GS. Evolving concepts of rheumatoid arthritis. Nature 2003 ; 423 : 356.

2) Seldin MF, Amos CI, Ward R, et al. The genetics revolution and the assault on rheumatoid arthritis. Arthritis Rheum 1999 ; 42 : 1071.

3) Kochi Y, Yamada R, Kobayashi K, et al. Analysis of single-nucleotide polymorphisms in Japanese rheumatoid arthritis patients shows additional susceptibility markers besides the classic shared epitope susceptibility sequences. Arthritis Rheum 2004 ; 50 : 63.

4) Gregersen PK. Pathways to gene identification in rheumatoid arthritis : PTPN22 and beyond. Immunol Rev 2005 ; 204 : 74.

5) Kochi Y, Yamada R, Suzuki A, et al. A functional variant in FCRL3, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities. Nat Genet 2005 ; 37 : 478.

6) Moser KL, Neas BR, Salmon JE, et al. Genome scan of human systemic lupus erythematosus : evidence for linkage on chromosome 1q in African-American pedigrees. Proc Natl Acad Sci USA 1998 ; 95 : 14869.

7) Dai KZ, Harbo HF, Celius EG, et al. The T cell regulator gene SH2D2A contributes to the genetic susceptibility of multiple sclerosis. Genes Immun 2001 ; 2 : 263.

8) Capon F, Semprini S, Chimenti S, et al. Fine mapping of the PSORS4 psoriasis susceptibility region on chromosome 1q21. J Invest Dermatol 2001 ; 116 : 728.

9) Marrack P, Kappler J, Kotzin BL. Autoimmune disease : why and where it occurs. Nat Med 2001 ; 7 : 899.

10) Haga H, Yamada R, Ohnishi Y, et al. Gene-based SNP discovery as part of the Japanese Millennium Genome Project : identification of 190,562 genetic variations in the human genome. Single-nucleotide polymorphism. J Hum Genet 2002 ; 47 : 605.

11) Davis RS, Dennis G Jr, Odom MR, et al. Fc receptor homologs : newest members of a remarkably diverse Fc receptor gene family. Immunol Rev 2002 ; 190 : 123.

12) Davis RS, Wang YH, Kubagawa H, et al. Identification of a family of Fc receptor homologs with preferential B cell expression. Proc Natl Acad Sci USA 2001 ; 98 : 9772.

13) Miller I, Hatzivassiliou G, Cattoretti G, et al. IRTAs : a new family of immunoglobulinlike receptors differentially expressed in B cells. Blood 2002 ; 99 : 2662.

14) Goronzy JJ, Matteson EL, Fulbright JW, et al. Prognostic markers of radiographic progression in early rheumatoid arthritis. Arthritis Rheum 2004 ; 50 : 43.

*　　*　　*

特集II　関節リウマチ研究における新たな視点

新規関節リウマチ感受性遺伝子*FCRL3*[*]

高地雄太[**]

Key Words：FCRL3, rheumatoid arthritis, single-nucleotide polymorphisms, case-control association study

はじめに

関節リウマチ(rheumatoid arthritis；RA)は，環境因子および遺伝因子が複合的に関与することによって発症する多因子疾患である[1]．遺伝因子としては，HLAクラスII遺伝子である*HLA-DR*遺伝子多型との関連が古くより知られており，RAの最大の遺伝因子である[2]．RA感受性との関連が報告されているのは，*HLA-DRB1**0405,0401,0101といった，いわゆる“shared epitope”アレル(アミノ酸配列の70～74残基にQRRAA, RRRAA, もしくはQKRAAを含むもの)である．最近，RAにおいて，その高い特異性で注目されている，抗環状シトルリン化ペプチド抗体(抗CCP抗体)の有無とshared eitope保有の有無が，高い相関を示すことが報告された[3]．この事実から，シトルリン化された蛋白が特定のHLA-DRサブタイプによって提示されやすい，といった可能性が考えられるが，RAの病態における*HLA-DR*遺伝子多型の関与を考察する上で興味深い．

一方で，HLA領域外については，複数の遺伝因子の関与が考えられているが，個々の遺伝因子の疾患への寄与度が非常に低いため，その同定は困難が予想されてきた．しかし，ここ2～3年，ヒトゲノム配列の情報整備と遺伝子多型のタイピング技術の格段の進歩とともに，大規模な検体を用いて多数の遺伝子(およびその多型)を対象とした解析が可能となってきたことにより，RAをはじめとする自己免疫疾患に関連する遺伝子の同定が相次いでなされてきた．たとえば，*CTLA4*遺伝子多型は，自己免疫性甲状腺炎や1型糖尿病の感受性と関連が報告されたが[4]，RA感受性との関連も最近報告されている．また，*PTPN22*遺伝子多型も同様にRAを含む複数の自己免疫疾患との関連が報告されている[5][6]．RAに特異的な感受性遺伝子としては，*PADI4*遺伝子の多型があげられる[7]．*PADI4*遺伝子は，蛋白質のアルギニン残基をシトルリン化する酵素をコードする．*PADI4*遺伝子多型がRA感受性と関連する事実は，シトルリン化という蛋白の翻訳後修飾が，RAの病態においてなんらかの役割を果たしていることを，遺伝学的に裏づけるものである．このように，ゲノム解析によって，RAに関連する遺伝子(およびその多型)が明らかにされ，RAの病態の新たな側面が解明されつつあるといえる．

われわれは，平成12年度より，理化学研究所遺伝子多型研究センターにおいて，ホールゲノムを対象としたSNPを用いた患者対照関連解析により，RA感受性遺伝子の探索を行ってきた．本

[*] Identification of FCRL3 as a novel susceptibility gene for rheumatoid arthritis.

[**] Yuta KOCHI, M.D., Ph.D.: 独立行政法人理化学研究所遺伝子多型研究センター関節リウマチ関連遺伝子研究チーム〔〒230-0045 横浜市鶴見区末広町1-7-22〕; Laboratory for Rheumatic Diseases, SNP Research Center, Yokohama 230-0045, JAPAN

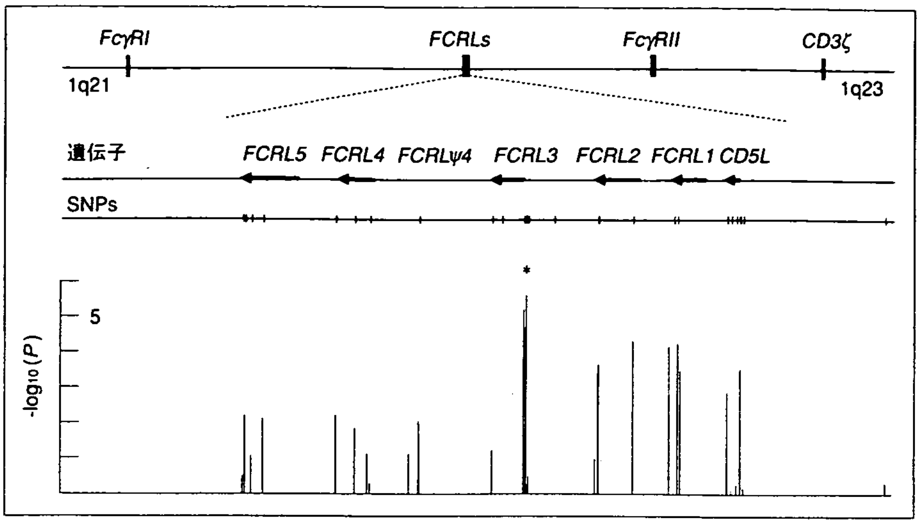


図1　FCRL遺伝子領域の関連解析

FCRL遺伝子領域に存在するSNPsを用いて，RA患者830人，対照群658人のアレル頻度比較を行った．グラフの縦軸は，x^2検定のp値の対数をとったもの．FCRL3遺伝子の4 SNPsに関連のピークを認める(*印)．

(文献[8]より引用改変)

稿では，その成果として同定された，新規自己免疫疾患感受性遺伝子*FCRL3*(Fc receptor-like 3)[8]について解説する．

候補領域1q21-23における関連解析

われわれは，ホールゲノムに分布する約10万個のSNPsをタイピングし，患者・対照間で比較することにより，RA感受性と関連する遺伝子多型の探索を行ってきたが，その結果，1番染色体長腕21-23領域(1q21-23)に，RA感受性と強い関連をもつSNPsが見出された．この領域は，ヒト自己免疫疾患および動物モデルの連鎖解析重複候補領域の一つとして知られている[9]．マウスでは1番，3番染色体領域に相当し，コラーゲン誘導性関節炎モデル，ループスモデル，アレルギー性脳脊髄炎，自己免疫性糖尿病モデルなどの候補領域となっている．また，ヒトでは，SLE,乾癬における連鎖解析の候補領域となっており，RAや多発性硬化症においても，この領域に存在する遺伝子多型との関連が報告されている．この領域における注目すべき遺伝子としては，免疫グロブリンGのFc部分に対する受容体であるFcγレセプター遺伝子群がある(図1)．疾患との関連では，*FCGR2*B,*FCGR3A*遺伝子に存在する多型と，SLEおよびRAとの関連が報告されている[10][11]．さらに，最近，この領域にFcγレセプターとの相同性の高い*FCRL*遺伝子群の存在が明らかにされており[12]，これらの遺伝子と疾患との関連が注目されている．以上のように，本領域は，自己免疫疾患共通の感受性遺伝子が存在する可能性が高い領域といえる．われわれは，本領域を詳細に解析することにより，疾患感受性遺伝子多型の同定を行った．

まず，日本人の代表的なSNPデータベースであるJSNPより，1q21-23領域(16Mbp)に存在する491 SNPsを抽出し，日本人対照群658人のジェノタイピングを行った．これらのSNP間の連鎖不平衡係数を算出することにより，連鎖不平衡状態を評価し，110個の連鎖不平衡ブロックを同定した．次に，関連解析の1段階目のスクリーニングとして，これら491 SNPsに対して，RA患者94人のジェノタイピングを行い，対照群とのアレル頻度の比較を行った．その結果，9 SNPsにおいて，患者・対照間での有意な差を認めた($p<0.01$)．さらにこれらのSNPsについては，2次スクリーニングとして，ケース734人を追加でジェノタイピングし，対照群と比較したところ，*FCRL3*遺伝子のイントロンのSNPとの強い関連を認めた(オッズ比1.39, $p<0.0001$)．このことより，この多型，もしくは，この多型と連鎖不平衡状態

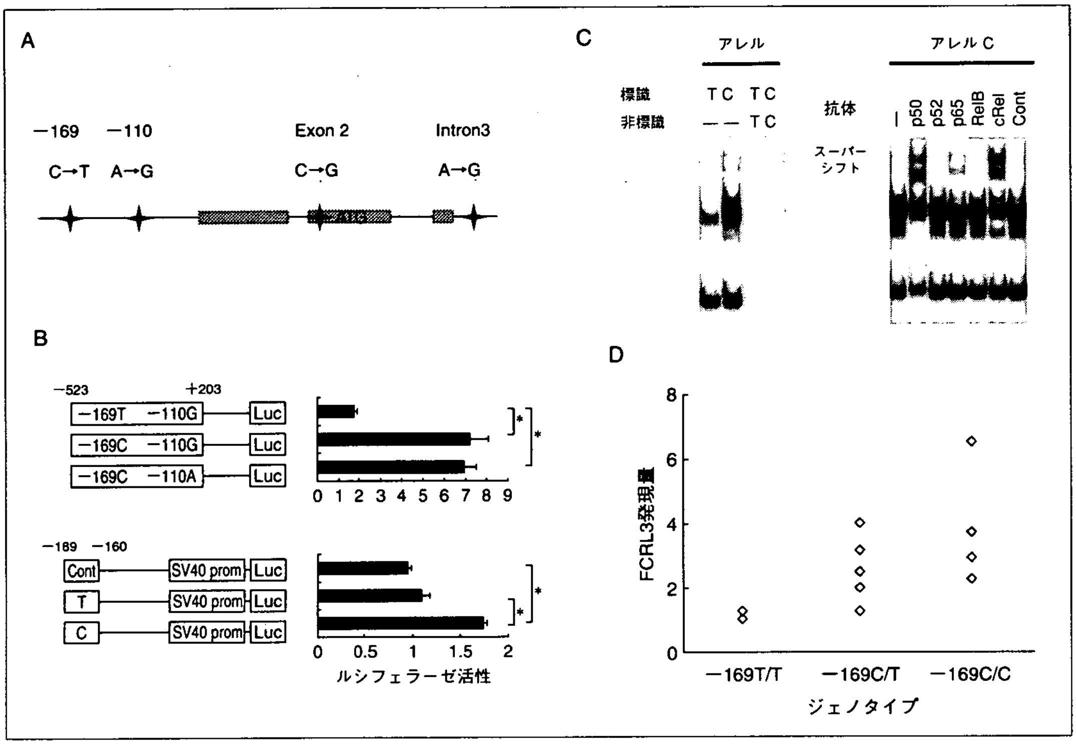


図2　RA感受性と関連するFCRL3遺伝子多型(-169C→T)は遺伝子発現を制御する

A：RA感受性と強い関連を認めたFCRL3遺伝子の4SNPs.

B：FCRL3遺伝子の転写活性をルシフェラーゼアッセイにより評価. -169Cアレル(疾患感受性アレル)で活性が高い(*p<0.001,).

C：SNP周辺配列に結合する転写因子(EMSA). -169Cアレルで核蛋白質の強い結合を認め(左パネル), 抗NF-κB抗体でスーパーシフトする(右パネル).

D：健常人B細胞におけるFCRL3遺伝子の発現量を, SNPのジェノタイプ別に定量.　(文献[8]より引用改変)

にある多型が疾患に直接的に関与していることが考えられた. そこで, FCRL遺伝子群(FCRL1～5)が存在する2つの連鎖不平衡ブロックの詳細な解析を行った. 新規に16 SNPsの同定を行い, 既知の25 SNPsとあわせて, RAケース群830人, 対照群658人のアレル頻度およびジェノタイプ頻度の比較を行った. その結果, *FCRL3*遺伝子の4つSNPがRA感受性と強い関連を示すことが明らかにされた(うち, SNP -169C→T, 劣性遺伝形式比較でのオッズ比2.15, p<0.000001)(図1).

*FCRL3*遺伝子多型の解析

1. *FCRL3*遺伝子多型と遺伝子発現

RA感受性との強い関連を認めたSNPsはいずれも翻訳領域に存在しないため(図2A), 蛋白の構造変化を介して疾患に関与している可能性は考えられない. そこで, これらの多型が遺伝子発現になんらかの影響を与える可能性を考え, ハプロタイプ別プロモーター活性の評価を, ルシフェラーゼアッセイを用いて行った(図2B). 3つのプロモーターハプロタイプのうち, SNP(-169C→T)のCアレルを含むハプロタイプで転写活性が強いことが明らかになった. -169C→T周辺配列を用いた評価においても, -169Cアレルでの転写増強活性が強く, この部位になんらかの転写因子が結合することが考えられた. 次に, 転写因子結合予測ソフト(TRANSFAC)を用いて, この部位に結合する転写因子の予測を行ったところ, この配列はNF-κB結合モチーフとの相同性が高いことが明らかになった. 実際に, ゲルシフトアッセイを用いた解析により, -169C→T周辺配列に, NF-κBのコンポーネント(p50, p65, c-