Genetic contribution to susceptibility and disease phenotype in rheumatoid arthritis

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

Rheumatoid arthritis (RA) is the most common chronic arthritis in the world. RA is characterized by inflammatory joint synovitis and a resultant joint destruction. Patients with RA often display positivity for rheumatoid factor (RF) and/or anti-citrullinated peptide antibody (ACPA). Methotrexate is an anchor oral drug to treat RA. Biological agents, targeting TNF or IL-6R, are efficient treatment to RA which prevent joint destruction in patients with RA. However, patients with RA are heterogeneous. Joint destruction develops rapidly in some patients but slowly in others. ACPA and/or RF are not positive for all patients with RA. Moreover, positivities of ACPA and RF do not always correlate with each other. About 30% of patients with RA do not respond to biological treatment. What kind of factors determines the heterogeneity of RA? Genetic and environmental effects are assumed to explain these variance. In this review, we focus on genetic components and review how much variance of susceptibility to RA or RA phenotype can be explained and determined by genetic components. Recent technological advancement has enabled us to perform genome-wide association studies to detect susceptibility loci to complex diseases with an unbiased approach. More than 100 susceptibility loci to RA have been detected so far, and functional analyses have been successfully performed for some. Autoantibody status in patients with RA is strongly associated with HLA alleles. Unfortunately, detecting markers associated with response to treatment in patients with RA have not been very successful to date.

Rec.12/5/2013, Acc.1/22/2014, pp71-77

Correspondence should be addressed to: Chikashi Terao, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, 53 Shogoin Kawahara-cho, Sakyoku, Kyoto 606-8507, Japan. Phone: +81-75-366-7405, Fax: +81-75-751-4167, E-mail: a0001101@kuhp.kyoto-u.ac.jp

Key words rheumatoid arthritis, genetics, HLA, genome-wide association study, joint destruction

Introduction
Rheumatoid arthritis (RA) is the most common cause of adult chronic arthritis, affecting around 1% of the population all over the world\(^1\). About 80% of the patients are female. Although many patients are middle-aged when they develop RA, it can also affect children as well as the elderly. Previous studies have revealed that both environmental and genetic components contribute to the disease
course as well as the onset of RA. In this review, we focus on the genetic components of RA.

**RA genetic component—how much do genetic components determine?—**

(A) Heritability

It has been well known since ancient times that genetic components are associated with RA onset. High twin concordance indicates genetic contribution for its onset. The concordance of monozygotic twins in RA was reported to be ~15%, while the concordance of dizygotic twins in RA was reported to be ~6%. The most famous study addressing heritability in RA is the study reported by MacGregor et al., which is based on two nation-wide studies for twins recruiting 246 and 203 monozygotic or dizygotic twins either of which had RA in Finland and UK, respectively. Recently, a Denmark team reported a twin study recruiting 155 twin pairs where at least one of the twins have RA. The Danish results showed that additive genetic components explain about 12% of RA variance, which is smaller than the two previous studies. Heritability in RA based on twin studies reported that 12 to 65% of the disease susceptibility should be attributed to genetic components. These values contained wide confidence intervals, and the big difference among reports could be explained by the quality of the twin studies as well as the reliability of information acquired for RA diagnosis. On this point, reports from Asian countries are quite limited. Our group in Kyoto University recently established the KURAMA (Kyoto University Rheumatoid Arthritis Management Alliance) database to store detailed clinical information and specimens from patients with arthritis and arthropathy. In collaboration with the IORRA database, the largest Japanese RA cohort in Tokyo Women’s Medical University and the largest RA database in Japan, we are collecting information of twins and we estimate that 40 to 50% of variance of RA susceptibility might be explained by genetic components (Terao et al, manuscript in preparation).

(B) The HLA locus

What kind of genes affects RA susceptibility? HLA has been shown to be strongly associated with RA susceptibility. The HLA locus is estimated to contribute about half of the entire genetic components to RA susceptibility, and HLA-DRB1 in particular is well known to be associated with RA. The shared epitope (SE) concept, consisting of allelic groups of the 70th to 74th amino acids of the HLA-DRB1 protein, is a widely accepted to explain RA susceptibility beyond ethnicity. SE includes HLA-DRB1*01:01, 04:01, 04:04, 04:05, 04:10, 10:01, 14:02, 14:06 and other minor alleles. In Caucasians, *01:01 is the most common SE and *04:05 is the most common in East Asians. HLA-DRB1 alleles other than SE are reported to be associated with RA. HLA-DRB1*09:01 is the second strongest susceptibility allele to RA in Asians following 04:05. HLA-DRB1*15 has also been shown to be a susceptibility allele to RA. These reports indicate that susceptibility alleles cannot be simply explained by combinations of the 70th to 74th amino acid residues. Recently, Raychaudhuri et al showed the five amino acid residues in HLA-DRB1, DPB1, and B, namely, 11th, 71st and 74th amino acids in HLA-DRB1, 9th amino acid in HLA-DPB1 and 9th amino acid in HLA-B, can explain the susceptibility to seropositive RA.

(C) Non-HLA gene

While approximately half of the susceptibility to RA can be explained by the HLA locus, the rest of the variance should be attributed to non-HLA genes. Genome-wide association study (GWAS) using SNPs as markers is a very powerful tool to detect common variants associated with disease susceptibility. Although more than 30 susceptibility genes were reported by the middle of 2013, functional analyses of SNPs or genes have been performed only for a fraction of the genes. PTPN22 is the strongest susceptibility non-HLA gene to RA in population with European descent. The causative variant in PTPN22 to RA (R620W), which introduce an alteration of amino acid in PTPN22 resulting in decreased signaling of T-cell receptor and B-cell receptor involved with establishment of immune tolerance, is not found in the Asian population. Menard et al showed that naïve B cells from individuals with the R620W polymorphism contained high proportion of autoreactive B cells and that B cell-related genes are overexpressed in these participants. PADI4, encoding a citrullination enzyme, is the strongest susceptibility non-HLA gene to RA in Asians and its susceptibility effect on RA has recently been reported in a population with European descent. Suzuki et al showed that mRNA transcribed from the susceptibility haplotype demonstrated prolonged half-life time of degradation, indicating that excess PADI4 protein is produced and excess citrulliation would occur in subjects carrying the risk haplotype. In fact, they showed that those who carry the risk haplotype showed higher levels of a fraction.
of ACPA. Myouzen et al recently reported that NFkBIE and RTKN2 are associated with RA. In this manuscript, they showed that variants resulting in the activation of NF-κB conferred increased risk of RA. At the same time, they showed a successful example of detecting causative variants by integrating various in-silico database commonly available. In 2011, we showed that a common variant of AIRE is associated with RA. Rs2075876, showing the strongest association with RA, displayed an association with decreasing expression of AIRE. We also showed that there is another variant in Exon 7 in AIRE which introduces amino acid alteration in SAND domain of the protein. These results suggest that the decreasing expression of AIRE possibly through disruption of transcription factor binding site or functional alteration of the AIRE protein would lead to the survival of autoreactive T-cells in the thymus which in turn leads to the development of RA. We also reported that rs2000811, a common variant of MBP on chromosome 18 is associated with RA. We showed different allelic expression according to genotypes of rs2000811. Since MBP encodes two different proteins, classical MBP expressed in CNS and Golli-MBP expressed in blood cells especially in B lymphocytes, our results suggest that Golli-MBP is involved with RA pathophysiology.

In 2010, Kochi et al reported CCR6 as a susceptibility locus to RA. In addition to finding rs3093024 showing a strong association with RA susceptibility, they showed that rs3093024 is associated with increased expression of CCR6. Furthermore, they found that dinucleotide polymorphisms in strong LD with rs3093024 showed alteration of transcription factor binding. They also showed that fraction of IL17+ cells in peripheral blood in RA patients are beautifully correlated with combination of the polymorphisms.

Meta-analysis by combining GWAS data is a powerful method to detect novel susceptibility loci. Okada et al have increased the number of RA susceptibility genes to 101. Functional analysis to analyze the causative variants and mechanisms underlying linkage between variants and disease onset is quite important. Although taking these 101 loci into account, sum of the common variants do not fully explain the variance attributable to genetic components. Could the rest of the genetic variance be explained by rare variants? Although exome sequencing by using next generation sequencer identified numerous causative genes in Mendelian diseases, genetic analyses focusing on rare variants for RA have not been successful so far. It might suggest that even if there are penetrating rare variants in RA families, these variants are localized in these families. Stahl et al applied polygenic model to large datasets of RA and other common diseases and showed that most of the variance explained by genetic components are attributable to common variants. Their calculation suggests that a set of common variants far from significance include truly associated markers and that these truly associated markers confer big variance.

2) Disease phenotype
(A) Autoantibody production
The Danish twin study reported that concordance of autoantibody production is not higher in monozygotic twins than in dizygotic twins. On the contrary, the UK twin study showed higher concordance in monozygotic twins than in dizygotic twins for seropositivity/negativity and erosion. Heterogeneity of HLA association was suggested between ACPA-positive RA and -negative RA. When we focus on only erosive ACPA-negative RA, we did not find association between SE and this subset. There are specific HLA-DRB1 alleles to ACPA-negative RA. In European population, HLA-DR3 and DR13 were suggested. In Asian population, HLA-DRB1*12:01, *15:02, *13:02 and homozygotes of HLA-DR8 were shown to be associated with ACPA-negative RA. Our group and a UK group found that ACPA-negative RA consists of two genetically distinct subsets based on RF positivity. ACPA-negative RF-positive RA showed association with SE and other HLA-DRB1 alleles associated with ACPA-positive RA. Our results also showed that susceptibility or protective HLA-DRB1 alleles associated with ACPA-negative RA are shared between these two subsets of ACPA-negative RA. This indicates that ACPA-negative RF-positive RA is similar to ACPA-positive RA, but do not belong to the same subset.

What about non-HLA genes? Are there any ACPA-negative RA-specific susceptibility genes like HLA alleles? A European GWAS focusing on ACPA-negative subsets failed to show specific susceptibility alleles to ACPA-negative RA. A US study suggests that ACPA-negative RA share common susceptibility alleles with ACPA-positive RA. The failure to identify specific non-HLA susceptibility genes to ACPA-negative RA may be explained by low power and this genetic heterogeneity in ACPA-negative RA. GWAS focusing on each of the two ACPA-negative
RA subsets would lead to the isolation of specific susceptibility non-HLA genes.

Not only the production of autoantibody, but also levels of autoantibody was associated with HLA-DRB1 alleles. A Japanese study showed that HLA-DRB1*09:01, the second strongest susceptibility allele to RA in Japanese, decreases ACPA levels in ACPA-positive RA\(^{32}\). We replicated this association using independent sets and further showed that the apparent association between SE and increasing levels of ACPA can be attributable to the association between HLA-DRB1*09:01 and decreasing of ACPA levels\(^{33}\).

(B)Joint destruction

Because joint destruction due to RA synovitis lower ADL in patients with RA, estimating genetic components affecting joint destruction would result in improving follow-up procedures for patients with RA. van der Helm-van Mil AH et al. reported smaller variation of joint destruction in monozygotic twins compared with dizygotic twins or uncorrelated pairs\(^{34}\). Iceland has unique nationwide registries which covers all populations for 10 centuries in which many populations were genome-scanned with reliability of 97%. Using this Iceland cohort, Keve et al estimated genetic components contributing to joint destruction\(^{35}\). They revealed that 45-58% of joint destruction can be attributed to genetic components. It should be noted that there are no cohorts similar to Iceland cohorts in other countries. Whether we can generalize this result in other population should be confirmed by a large-scale RA cohort which contains kinship data as a part of its cohort.

(C)Response to treatment

Determining genetic components associated with response to treatment in RA is a big challenge. Methotrexate is an anchor drug for RA among DMARDs. While biological agents brought drastic change in therapy for RA, a considerable number of patients with RA do not respond to biological agents in spite of the expensive costs. Incorporating genetic loci associated with efficacy of treatment would improve overall outcome of RA treatment. Many studies have been performed to address genetic components associated with response to anti-TNF agents. To date, none of the genetic components have satisfied GWA significant level. Previous studies have suggested \textit{TNFA}, \textit{PTPRC}, \textit{CD84}, \textit{PDZD2}, \textit{EYA2}, \textit{C9orf72}, \textit{MOBKL2B}, \textit{TLR4}, \textit{DBC1}, \textit{FOXP1}, \textit{PLA2G4A}, \textit{NUBPL}, \textit{CNTN5}, \textit{LOC10130480}\(^{36-40}\) were associated with response to anti-TNF agents. Cui et al reported that an SNP in the \textit{CD84} region strongly influencing the gene expression of \textit{CD84} is associated with etanercept response\(^{37}\).

Tocilizumab is a humanized antibody against IL6R developed in Japan. A research group in Roche company recently reported that eight loci showed suggestive associations with change in DAS28 or that in components of DAS28 in patients with RA treated with TCZ\(^{41}\). On response to IL-6R antagonist, the expression of IL6R protein encoded by \textit{IL6R} is the primary interest as a candidate of associated locus to TCZ response. Because the expression of IL-6R is strongly regulated by rs8192284\(^{42,43}\), a polymorphism in the IL6R region, rs8192284 is an excellent candidate to assess its effect on TCZ response. However, the previous study and an unpublished Japanese study (Terao et al, manuscript in preparation) did not find strong association with this polymorphism.

Response to these agents targeting TNF is different in the same patients. Even if a patient does not respond to one biological agent targeting TNF, he or she might display good response to another biological agent targeting TNF and vice versa. These indicate that analysis focusing on response to biological agents should be performed not on molecule-targeted by biological agents, but on each biological agent. It should be noted that candidate genes and polymorphisms associated with response to biological agents have not displayed overlap with any susceptibility genes to RA. In fact, although seropositivity has been reported to be associated with response to biological agents, SE has not been reported to be associated with response to biological agents. This indicates that response to biological agents is regulated in complex mechanisms.

Summary

As a conclusion, genetic components are deeply associated with RA. Analysis of susceptibility alleles or amino acid residues in the HLA locus to RA is almost established. In particular, up to 100 susceptibility loci have already been identified and almost all common variants with certain effect sizes have been detected. Common variants with small effects are believed to explain the rest of the variance. Further increase in the number of RA patients (more than 200,000) or analyses of the African population would complete a list of susceptibility genes. The application of genetic results into animal models and the identification of novel therapeutic targets are strongly desirable. Unfortunately, predictive model of RA incorporating genetic vari-
nants has not been fruitful to date. Incorporating autoantibody status or smoking would lead to a more efficient prediction\(^\text{42}\). Large parts of the genetic components for phenotypes of RA have remained uncovered. In particular, whether non-HLA genes or polymorphisms affect production of autoantibody or level of autoantibody and joint destruction are largely unknown. Further studies addressing these points would reveal basic structure of RA progression. The isolation and characterization of genetic components associated with therapeutic response is the most desirable for both clinicians and patients. However, the results have not been so productive to date, and the reasons could be due to 1. Heterogeneous evaluation according to clinicians and 2. Low contributions of genetic components to response. Objective evaluation of response to therapeutic options such as multi-biomarker disease activity (MBDA) score\(^\text{65}\) would be one option to detect significant signals in a much more efficient way.

**Acknowledgments**

No conflict interest exists.

**Source of funding**

None

**Abbreviations**


**Reference**


34) van der Helm-van Mil AH, Kern M, Gregersen PK, Huizinga TW: Variation in radiologic joint destruction in rheumatoid arthritis differs between monozygotic and


