From genetics to functional insights into rheumatoid arthritis

A. Suzuki[1], K. Yamamoto[1,2]

[1]Laboratories for Rheumatic Diseases, SNP Research Center, Yokohama, Japan; [2]Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo, Tokyo Japan.

Akari Suzuki
Kazuhiko Yamamoto

Please address correspondence to:
Akari Suzuki,
Laboratories for Rheumatic Diseases,
SNP Research Center,
RIKEN, 1-7-22, Suehiro,
Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan.
E-mail: akaris@src.riken.jp

ABSTRACT

Autoimmune diseases are caused by multiple factors. Rheumatoid Arthritis (RA) is one of the most common human systemic autoimmune diseases with a prevalence of 0.5%–1% worldwide. It is characterised by inflammation of the synovial tissue and formation of rheumatoid pannus, which erodes adjacent cartilage and bone, causing subsequent joint destruction. RA is believed to result from a combination of genetic and environmental factors. In addition to the well characterised HLA locus, a number of susceptibility genes and loci have been identified by genome-wide association studies (GWAS). However, genetic information alone does not necessarily yield insight into the understanding of the pathogenesis of RA.

We previously reported that Peptidylarginine deiminase type 4 (PADI4) is one such RA susceptibility gene. PADI4 catalyses the conversion of peptidylarginine into peptidylcitrulline, and citrulline-containing epitopes are the most specific targets of many RA-specific autoantibodies. We established that SNPs within the coding region of PADI4 are associated with the development of RA and that these RA-associated SNPs produce allelically imbalanced gene expression, which has pathological consequences. However, the individual effects of susceptibility genes are likely to be small, and it is the combination of alleles along with strong effects on the specific pathways affected by these susceptibility genes that are essential for the development of RA. To understand the role of genetics in the pathogenesis of RA, it is therefore important to understand the physiological significance of each susceptibility gene in relation to RA.

Genome-wide association studies (GWAS) involving thousands of individuals have reproducibly linked hundreds of common genetic variants to over 17 categories of diseases, including autoimmune and cardiovascular diseases (http://www.genome.gov/gwastudies). Due to a dramatic increase in sample size, recent GWAS have intensified efforts to identify additional variants that may more comprehensively explain disease heritability. Each putative risk locus identified has a relatively small effect on overall disease risk [odds ratio (OR) <1.5], explaining only a small fraction of the heritability of common diseases. GWAS of Rheumatoid Arthritis (RA) have successfully linked the risk of developing the disease to numerous loci that contain novel candidate genes, each with a small effect size (1). Some RA loci contain no predicted candidate gene (1). New approaches, including cis-expression quantitative trait loci (cis-eQTL) analysis, are being used to identify candidate genes within loci of interest. However, it remains important to determine the function of each gene identified to establish its role in the pathogenesis of RA.

Genetic risk factors involved in the development of RA

There has recently been a significant increase in our genetic understanding of RA as a result of large-scale GWAS that linked common SNPs to RA risk (minor allele frequency >5%). Numerous genetic risk factors have been identified in multiple ethnic groups (1-4). Some of these factors are shared among different ethnic groups, whereas others are unique to specific ethnic groups. Approximately 30% of causative SNPs demonstrate cis-eQTL (5). However, the function of some causative SNPs remains unclear. In common diseases caused by multiple factors, such as RA, non synonymous common substitutions are rare. The functional outcome of causative SNPs can sometimes be predicted using FuncPred (http://snpinfo.niehs.nih.gov/snpinfo/snpfunc.htm). However, the actual functional outcome of SNPs and its re-

A.

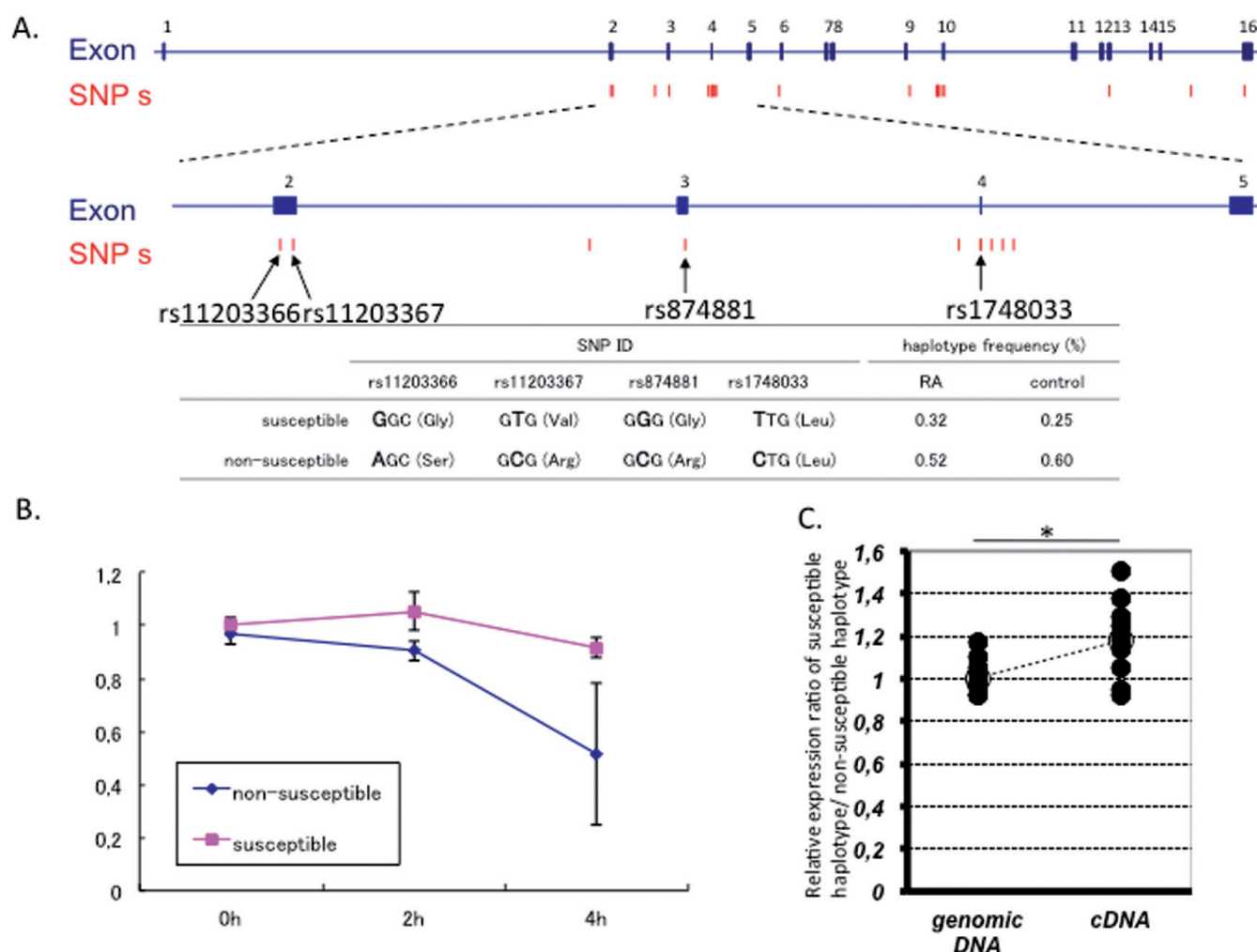|		SNP ID			haplotype frequency (%)	
	rs11203366	rs11203367	rs874881	rs1748033	RA	control
susceptible	G**G**C (Gly)	G**T**G (Val)	G**G**G (Gly)	**T**TG (Leu)	0.32	0.25
non-susceptible	A**G**C (Ser)	G**C**G (Arg)	G**C**G (Arg)	**C**TG (Leu)	0.52	0.60

B.

C.

Fig. 1. The two major haplotypes of PADI4 and its gene structure are shown (**A**). Haplotype frequencies are summarised in the table. Allelically imbalanced PADI4 expression was detected using an RNA instability assay (**B**) and an allele-specific transcriptional quantitative (ASTQ) assay using real-time TaqMan RT-PCR (**C**). (B) Transient transfection experiments were performed in a K562 cell line using expression vectors containing full-length cDNA of each PADI4 haplotype. Degradation rates were detected by real-time TaqMan RT-PCR after inhibition of transcription using actinomycin D (16). (**C**) ASTQ was performed using cDNA from peripheral blood leukocytes of heterozygous individuals (4). The TaqMan probe for rs11203366 (G/A), which was used as a marker, was located in exon 2 of PADI4. The mean signal intensity of genomic DNA after adjustment in a heterozygous individual was 1.00 with a standard deviation of 0.08 and that of cDNA after adjustment was 1.18 with a standard deviation of 0.17 ($p<0.05$, U-test).

lationship to disease must be examined using relevant cells or tissues. Thus, it is difficult to understand the reasons why certain susceptibility SNPs or genes caused diseases. The results of these studies suggest that many factors, including the effects of susceptibility genes on one or more pathways, leads to the development of RA.

Numerous genetic risk factors with modest effect sizes (<1.5) have been identified through GWAS. However, the genetic risk loci identified to date only account for approximately half of the total heritability of RA. Therefore, half of the genetic contribution to RA remains unknown (6). The remaining genetic risk profile may involve large

numbers of common SNPs with smaller effect sizes, as well as rare and structural variants, which are collectively referred to as "missing heritability" factors. Several associations between structural variations and disease have been identified, for example the link between copy number variation (CNV) and autoimmune disease (7, 8). However, the results of a recent GWAS of CNVs using eight common diseases and controls analysed by the Wellcome Trust Case Control Consortium (WTC-CC) suggest that common CNVs are unlikely to account for a large amount of the missing heritability (9). GWAS analyses are still uncovering risk variants with medium effect sizes (OR =

1.5–2) and rare variations. Identification of these variants may help resolve missing heritability.

Functions of causal RA-associated variants

A number of candidate genes have been identified by examining causal genetic variants linked to mapped SNPs that are located within susceptibility genes or their functionally-regulated regions or by identifying amino acid changes that influence the target gene function. For example, a large-scale linkage disequilibrium study revealed an RA-susceptibility variant within PADI4 in a Japanese population (10). The number of transcripts generated from the
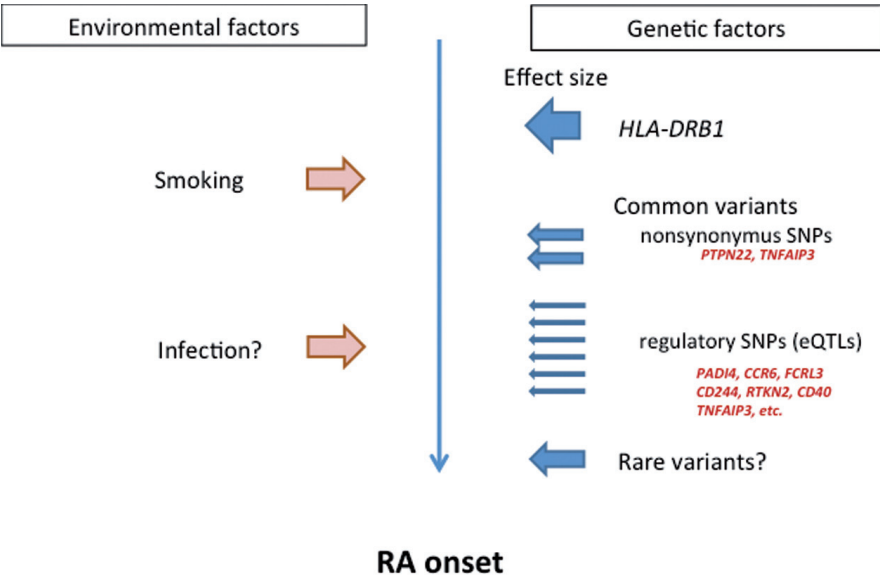
Fig. 2. A developmental model of RA. Combinations of factors differ among various populations and individuals.
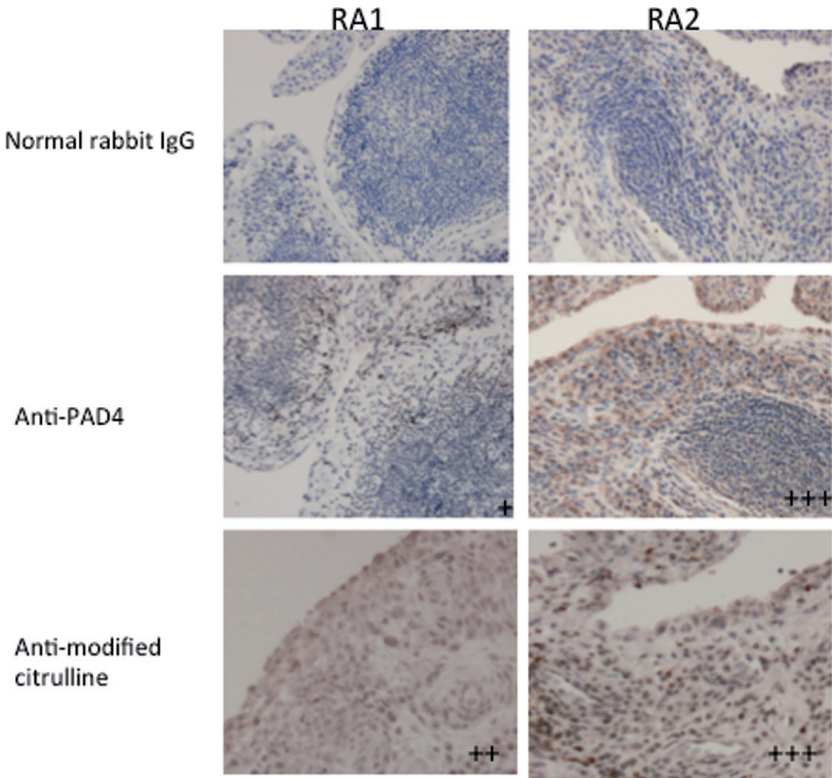

Fig. 3. Distribution of PAD4 and citrullinated peptides in synovial tissue from two patients with RA. PAD4 and citrullinated peptides were detected by anti-PAD4 and anti-modified citrulline (anti-MC) antibodies, respectively.

susceptibility haplotype was greater than that from the nonsusceptibility haplotype both *in vitro* and *in vivo* (Fig. 1) (10). A 1.2-fold change in the transcription of PADI4 was observed in peripheral blood leukocytes from individuals with the susceptibility hap-lotype. Thus, although the contribution of each genetic factor to autoimmune disease is small, a combination of genetic factors is believed to contribute to the development of RA (Fig. 2).

The phenotyping of susceptibility alleles for selected specific traits may reveal the molecular mechanism(s) that underlie these diseases; however, ultimately the functional identification of a causal gene(s) within a disease-associated locus will be more important for understanding the effects of SNP on pathophysiology of RA.

Physiological functions of susceptibility genes in RA

HLA-DRB1 gene is a well-known genetic risk factor for RA, which represents the largest effect size of all genetic risk factors for RA. The main function of HLA-DRB1 is to present antigen to T-cells. The PADI4 RA susceptibility gene is a non-HLA genetic factor that influences the development of RA through its involvement in the process of citrullination. It has been reported that HLA-DR4, an RA-associated HLA-DRB1 subtype, undergoes a high-affinity interaction with citrullinated peptides (11). This strongly suggests that PADI4 is functionally related to HLA-DRB1, because PADI4 produces antigens against which autoantibodies that specifically recognise citrullinated peptides are produced, which are highly specific to patients with RA. Protein citrullination is essential for the development of RA, and it follows that citrullinated protein-autoantibody immune complexes play a pathogenic role in the autoimmune inflammation associated with RA. Citrullinated peptides and PADI4 have been detected in the synovial tissue of patients with RA (Fig. 3). Despite the importance of these observations, the pathophysiological role of PADI4 remains obscure, because they do not directly demonstrate a relationship between PADI4 and the development of RA. Thus, it is difficult to translate identification of susceptibility gene into pathophysiological disease mechanisms.

Mouse models are powerful tools that allow the *in vivo* functions of genes identified by GWAS to be tested to better understand associations between genes and phenotypes. Although mouse is a good disease model, it is difficult to estimate the effect of disease-associated SNPs in mice because the effect size is very small. However, some knockout and transgenic mice show varied phe-
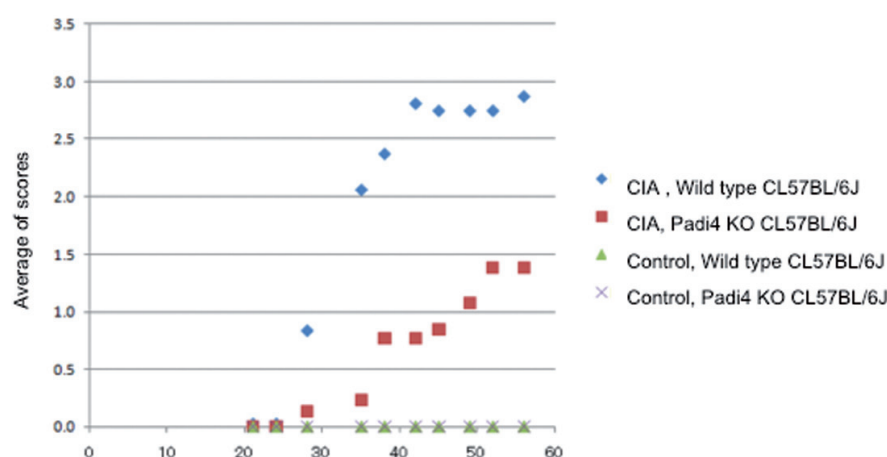
Fig. 4. Suppression of arthritis in CIA model of Padi4 knock-out CL57BL/6J mice.

notypes (12, 13). In addition, tissue-specific mouse models are also important for translating findings into useful clinical information, allowing novel therapeutic strategies and drug-tissue specificity to be established (14).

We found that Padi4 knockout mice showed reduced RA disease scores (Fig. 4), consistent with results of the eQTL study. The effect size of PADI4 is approximately 1.1–1.2, which is also the estimated OR (1, 15), and based on data from Suzuki *et al.* (10), the population attributable risk (PAR) can be calculated to be 24%. This effect is too small to find disease phenotypes using mouse models. However, mouse models that functionally test disease-associated genes can be used to identify causal genes from GWAS by yielding functional evidence. Pathophysiological mechanisms can be more easily determined in mouse models than in humans.

Conclusion: challenges in the post-GWAS era

GWAS have been extensively performed to identify common genetic risk factors that are associated with RA, leading to the discovery of novel disease genes. Unknown genetic risk factors and missing heritability are likely to be more comprehensively characterised through whole genome sequencing using next-generation sequencers. In addition, advanced bioinformatics tools will be needed to integrate functional and genomic data and to model the pathways and gene interactions involved in disease.

Most GWAS have focused on the association of SNPs with disease susceptibility, but these studies have rarely examined the functional impact of associated risk loci. Mouse models can be used to identify potential pharmacological targets by providing a platform for the functional study of susceptibility genes. Elucidating the mechanism(s) by which these genes influence the disease will have great clinical impact.

References

1. OKADA Y, TERAO C, IKARI K *et al.*: Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat Genet* 2012; 44: 511-6.
2. KURREEMAN FA, STAHL EA, OKADA Y *et al.*: Use of a multiethnic approach to identify rheumatoid-arthritis-susceptibility loci, 1p36 and 17q12. *Am J Hum Genet* 2012; 90: 524-32.
3. OROZCO G, EYRE S, HINKS A *et al.*: Association of CD40 with rheumatoid arthritis confirmed in a large UK case-control study. *Ann Rheum Dis* 2009; 69: 813-6.
4. MYOUZEN K, KOCHI Y, OKADA Y *et al.*: Functional variants in NFKBIE and RTKN2 involved in activation of the NF-kappaB pathway are associated with rheumatoid arthritis in Japanese. *PLoS Genet* 2012; 8: e1002949.
5. OKADA Y, SHIMANE K, KOCHI Y *et al.*: A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus eyrthematosus in Japanese. *PLoS Genet* 2012; 8: e1002455.
6. OROZCO G, BARRETT JC, ZEGGINI E: Synthetic associations in the context of genome-wide association scan signals. *Hum Mol Genet* 2010; 19: R137-44.
7. McKINNEY C, MERRIMAN ME, CHAPMAN PT *et al.*: Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. *Ann Rheum Dis* 2008; 67: 409-13.
8. MAMTANI M, ROVIN B, BREY R *et al.*: CCL3L1 gene-containing segmental duplications and polymorphisms in CCR5 affect risk of systemic lupus erythaematosus. *Ann Rheum Dis* 2008; 67: 1076-83.
9. CRADDOCK N, HURLES ME, CARDIN N *et al.*: Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 2010; 464: 713-20.
10. SUZUKI A, YAMADA R, CHANG X *et al.*: Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat Genet* 2003; 34: 395-402.
11. 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.
12. HUGILL A, SHIMOMURA K, ASHCROFT FM, COX RD: A mutation in KCNJ11 causing human hyperinsulinism (Y12X) results in a glucose-intolerant phenotype in the mouse. *Diabetologia* 2010; 53: 2352-6.
13. DAI X, JAMES RG, HABIB T *et al.*: A disease-associated PTPN22 variant promotes systemic autoimmunity in murine models. *J Clin Invest* 2013; 123: 2024-36.
14. CLARK RH, McTAGGART JS, WEBSTER R *et al.*: Muscle dysfunction caused by a KATP channel mutation in neonatal diabetes is neuronal in origin. *Science*; 329: 458-61.
15. IKARI K, KUWAHARA M, NAKAMURA T *et al.*: Association between PADI4 and rheumatoid arthritis: A replication study. *Arthritis Rheum* 2005; 52: 3054-7.
16. LECLERC GJ, LECLERC GM, BARREDO JC: Real-time RT-PCR analysis of mRNA decay: half-life of Beta-actin mRNA in human leukemia CCRF-CEM and Nalm-6 cell lines. *Cancer Cell Int* 2002; 2: 1.