

# Shared genetic factors and their causality in autoimmune diseases

Kazuhiko Yamamoto ,<sup>1</sup> Yukinori Okada<sup>2</sup>

Autoimmune diseases including rheumatic diseases affect 3%–10% of the world population. The frequency of incidence is dependent on the geographical area, ethnicity, sex and environmental factors.<sup>1</sup> Except for rare monogenic diseases, the majority of autoimmune diseases are defined as a complex trait disorder, where multiple genetic and environmental factors interact. For example, twin studies have estimated that the heritability of rheumatoid arthritis (RA) is ~60%.<sup>2,3</sup> RA is classified into anti-citrullinated protein antibodies (ACPA) positive and negative subtypes. The above-mentioned heritability of RA applies primarily to patients with RA who are positive for ACPAs, whereas the heritability of seronegative RA appears to be low,<sup>4</sup> suggesting the differences in the genetic background of these two conditions. Therefore, although the classification criteria are satisfied, each autoimmune disease appears to be heterogeneous. On the other hand, autoimmune diseases share several symptoms, such as Raynaud phenomenon and arthritis. In a clinical situation, the terms ‘poly-autoimmunity’ and ‘familial autoimmunity’ are used. The former suggests the presence of multiple autoimmune diseases in a single patient, and the latter implies the situation where diverse autoimmune diseases are persistent in a family.

Genes in the major histocompatibility complex (MHC) have been reported to be the strongest single genetic effect in the majority of autoimmune diseases. Although the precise mechanisms remain to be discovered, these MHC loci are the first and the most robust examples of shared genetic factors among several autoimmune diseases. Since 2007, rapid advances in genome-wide association study (GWAS) have enhanced the identification of hundreds of genetic risk factors for many complex diseases. Thereafter, one possible approach to expanding our

understanding of the genetics of autoimmune disease using GWAS data is cross-trait analyses to identify genetic correlations with other complex traits. In fact, it has been reported that genetic variants associated with one autoimmune disease likely confer risk to other diseases (ie, pleiotropy).<sup>5</sup> For example, an interesting approach for a GWAS meta-analysis examining RA and coeliac disease identified a number of loci with pleiotropic effects.<sup>6</sup>

Recently, two large studies combining several diseases were performed. Li *et al* conducted a meta-analysis of 10 paediatric autoimmune diseases.<sup>7</sup> The authors stated that among the associated loci identified, 81% were shared by at least two autoimmune diseases, indicating a relatively high rate of sharing. On the other hand, Ellinghaus *et al* performed a subset-based meta-analysis of five closely associated seronegative inflammatory diseases including ankylosing spondylitis, Crohn’s disease, psoriasis, primary sclerosing cholangitis (PSC) and ulcerative colitis.<sup>8</sup> The authors evaluated whether comorbidity is due to pleiotropy (sharing of risk alleles by multiple diseases) or heterogeneity (a subgroup of one disease has a higher load of risk alleles for another disease). Their analysis suggests that the increased comorbidity rates among patients are due to biological pleiotropy rather than heterogeneity. For example, regarding the high prevalence of inflammatory bowel disease (IBD; Crohn’s disease and ulcerative colitis) in PSC, the authors speculate that PSC-IBD is a unique disease that shares some genetic factors with ulcerative colitis but distinct from classical IBD phenotypes. These pieces of information support the presence of shared pathological pathways as the basis for clinical co-occurrence. The authors proposed that patients with concomitant syndromes are genetically distinct from patients without concomitant syndromes, indicating classical and non-classical subsets based on genetics in one disease.

In *Annals of Rheumatic Diseases*, Acosta-Herrera *et al* have reported their results from a genome-wide meta-analysis of four systemic seropositive rheumatological

immune-mediated inflammatory diseases (IMIDs) including systemic sclerosis (SSc), systemic lupus erythematosus (SLE), RA and idiopathic inflammatory myopathies (IIM).<sup>9</sup> Autoantibody production is one of the main features of these diseases. These diseases also share symptoms, modes of progression, environmental risk factors and high rates of familial aggregation. The authors combined previously published GWAS datasets from European-descent populations to find out shared genetic aetiologies. They identified 27 shared loci, five of which were new.

A subset-based meta-analysis yielded 26 single nucleotide polymorphisms (SNPs) associated with at least two systemic IMIDs. The results indicated that 85% of the associated variants were shared by at least three diseases. SLE exhibited 100% association, SSc and IIM shared the majority, but RA was associated with less than half of the 26 SNPs, suggesting that SLE was more likely to be a prototype of these four systemic IMIDs and that RA shared only a part of the features. The associated SNPs are highly enriched in the functional categories of B and T cells, natural killer cells and monocytes. These cells appear to be relevant in systemic seropositive rheumatic IMIDs. The authors further discuss the significant functional categories of skin cells. This could be related to the nature of connective tissue diseases and the possible involvement of epithelial cells transdifferentiated into mesenchymal cells and further to the fibrotic processes. The authors also focused on reasonable gene products as the target of drug repositioning in systemic IMIDs. In fact, we do not have specific treatments for these diseases except for RA. For example, the gene-product of *TYK2* is targeted by tofacitinib and baricitinib, Janus kinase inhibitors. Since an intronic SNP, rs1185725, within *TYK2* is associated with IIM, SLE and SSc, these drugs are reasonable candidates for therapy repositioning in these diseases. In fact, these drugs are now in clinical trials for SLE, SSc and dermatomyositis.

In the clinical practice of rheumatology, accurate diagnosis is the first step which is followed by the administration of proper treatments based on the diagnosis. This process has been successful for several decades. However, we have been encountering several incomprehensive situations. For example, recent biological therapies do not exhibit similar efficiencies and sometimes are ineffective in different individuals with the same diagnosis.<sup>10</sup> In this regard, the majority of diagnoses have been performed using classification criteria developed on the idea

<sup>1</sup>Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

<sup>2</sup>Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan

**Correspondence to** Dr Kazuhiko Yamamoto, Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan; kazuhiko.yamamoto@riken.jp

of 'one disease, one normal distribution'. However, the concept of 'disease' would be more plastic. As discussed above, RA does not appear to be homogenous at least in terms of ACPA positivity, and an IBD-PSC subtype could exist in IBD. Thus, it is possible to propose that disease definition in rheumatic diseases would be flexible based on several different variables including extensive clinical data, genotype, gene expression and available multiple omics.

The above-mentioned shared genetic factors among different autoimmune diseases illustrate the mechanistic basis of shared disease risks. This idea is based on the premise of a common genetic basis of immune-mediated diseases. However, it is also possible that identical risk factors might have different effects in different diseases.<sup>11</sup> For example, recent GWAS data indicate that substantial proportions of risk variants work as expression quantitative locus (eQTL). Further, this eQTL is likely to work in cell-specific manners and also vary depending on various stimulatory conditions. Therefore, the effects of risk variants can be context specific. It is also important to point out that several variants exhibit opposing risk profiles in different diseases. Perhaps, a locus with the same direction of effect for different diseases would be working in a shared mechanism. On the other hand, a locus with an opposite direction of effect might indicate any counteracting disease mechanisms. For autoimmune diseases and other immune-mediated disorders, a certain locus might be involved in 'polarisation' of the immune response such as the modulation of eQTL effects or changing the balance of several cytokines. Both directions of effect would confer important information and thereby help elucidate the complex mechanistic relationship between these diseases. In addition, assessment of a locus especially having a risk with one disease is also important to elucidate pathophysiology specific to the disease. For example, the functional *PADI4* variants have a strong risk with RA,<sup>12</sup> but not with other immune-mediated diseases. To this end, the application of the meta-analysis methodology to handle risk heterogeneity among the multiple phenotypes is recommended to comprehensively capture both shared and opposite directional effects among phenotypes.<sup>13</sup>

GWAS findings indicate that approximately 80% of autoimmune risk variants occur in non-coding regions.<sup>14</sup> As reported previously, the majority of associated variants are enriched in DNase I hypersensitivity site hotspots. These are the regions

where chromatin has lost its condensed structure and thus indicate regulatory DNA regions. In fact, disease-associated variants are enriched in enhancers, within transcription start sites and transcription factor binding sites of lymphocytes. It is also important to note that GWAS data themselves have not conclusively identified a causal gene or a causal variant due to conserved linkage disequilibrium blocks. Further genetic, as well as cellular, studies are necessary to finally define the real causal variants. In this regard, it is interesting to know that two signals in the interleukin (*IL2*)/*IL21* locus on chromosome 4q27 are distinct. Type 1 diabetes mapping to *IL2* and other diseases to *IL21*.<sup>15</sup> We note that cross-phenotype analysis such as that by Acosta-Herrera *et al* can contribute to fine-mapping of the causal variants by combining genetic data from multiple diseases.<sup>9</sup>

Regarding several different autoimmune diseases, more than 100 loci have been reported in one disease. Because genetic information implies a clear causal relationship to the disease, how do these facts imply the pathological mechanisms of diseases? Since the effect size (OR) of each risk variant is usually small (example OR=1.05–1.5), people often argue against the substantial effect of each variant. In this regard, the combination of multiple risk variants could perturb multiple pathways contributing to several alterations of immune functions. Finally, multiple such alterations influence the overall risk of disease. It is also possible that each individual variant can have a large molecular effect in the pathway of immune function but only account for a small fraction of disease susceptibility as a whole.<sup>16</sup> In fact, we have demonstrated that RA biological genes, which were identified by a trans-ethnic meta-analysis of GWAS and subsequently predicted *in silico*, are the targets of approved therapies for RA, and thus further suggested that drugs approved for other indications may be repurposed for the treatment of RA.<sup>14</sup> Acosta-Herrera *et al* also discuss the possible drug repositioning to SLE, SSc and dermatomyositis as described above. Further, as GWAS predicted that psoriasis, IBD and AS could be the leading disease indications for anti-IL-17 and anti-IL-23 treatments, several of these treatments have been in fact approved and some are now in the midst of ongoing clinical studies.<sup>17,18</sup> All these facts suggest that risk variants with small effect sizes can be really the target of therapy because of their causalities and that genetics of rheumatology should contribute to drug development.

Recently, Witoelar and others have reported genome-wide pleiotropy between Parkinson's disease and autoimmune diseases.<sup>19</sup> The authors performed a systematical investigation of pleiotropy between Parkinson's disease and type 1 diabetes, Crohn's disease, ulcerative colitis, RA, coeliac disease, psoriasis and multiple sclerosis and found 17 novel significant loci. Especially, there was considerable overlap between Parkinson's disease and Crohn's disease, ulcerative colitis and RA. These findings strengthen the hypothesis that risk genes of Parkinson's disease might contribute to the diseases through immune system disturbance. Further, it also emphasises on the importance of cross-phenotype analyses between different complex phenotypes. Through these analyses, we might find common mechanisms among several autoimmune diseases with different phenotypes, such as shared regulatory mechanisms of transcription factor operated with the Epstein-Barr virus EBNA2 protein across risk loci as described by Harley *et al*.<sup>20</sup>

Since the majority of the investigation on shared genetic factors have been conducted using European populations, it would also be important to conduct multi-ethnic cross-phenotype analyses. The distribution of genetic variants and compositions of the disease sub-phenotypes would be heterogeneous among ethnicity, and thus, such efforts should contribute fine-mapping of the causal variants as well as identification of novel risk genes and implicated pathways. Another point that could be assessed is polygenic risk score (PRS) analysis. PRS calculates a disease risk score for each individual according to thousands of the genome-wide SNPs with moderate risk identified by the GWAS. Recent studies demonstrated that PRS estimated from the large-scale GWAS results could be used as a proxy variable to predict disease susceptibility of the individuals,<sup>21</sup> suggesting its critical importance to the future of clinical medicine. Currently, PRS is mostly analysed and applied for every single phenotype separately. As a next step, parallel calculation and phenotype comparison of PRS should reflect both shared and opposite mechanisms of rheumatic diseases, as well as differential diagnosis of the affected diseases of the individuals.

**Handling editor** Josef S Smolen

**Contributors** Each author (KY and YO) wrote a part of the manuscript and has integrated the parts. Both have reviewed and approved the final manuscript.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** KY has received honoraria or research grants from AbbVie, Astellas, AYUMI, Bristol-Myers Squibb, Chugai, Eisai, Janssen, Mitsubishi Tanabe, Ono and UCB.

**Patient consent for publication** Not required.

**Provenance and peer review** Commissioned; externally peer reviewed.

© Author(s) (or their employer(s)) 2019. No commercial re-use. See rights and permissions. Published by BMJ.



**To cite** Yamamoto K, Okada Y. *Ann Rheum Dis* 2019;**78**:1449–1451.

Received 17 February 2019

Revised 23 February 2019

Accepted 24 February 2019

Published Online First 6 March 2019



► <http://dx.doi.org/10.1136/annrheumdis-2018-214127>

*Ann Rheum Dis* 2019;**78**:1449–1451.  
doi:10.1136/annrheumdis-2019-215099

**ORCID iD**

Kazuhiko Yamamoto <http://orcid.org/0000-0001-9037-3625>

**REFERENCES**

- Cooper GS, Bynum MLK, Somers EC. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. *J Autoimmun* 2009;**33**:197–207.
- MacGregor AJ, Snieder H, Rigby AS, *et al*. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 2000;**43**:30–7.
- Cho JH, Gregersen PK. Genomics and the multifactorial nature of human autoimmune disease. *N Engl J Med* 2011;**365**:1612–23.
- Padyukov L, Seielstad M, Ong RTH, *et al*. A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. *Ann Rheum Dis* 2011;**70**:259–65.
- Pickrell JK, Berisa T, Liu JZ, *et al*. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* 2016;**48**:709–17.
- Zhernakova A, Stahl EA, Trynka G, *et al*. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet* 2011;**7**:e1002004.
- Li YR, Li J, Zhao SD, *et al*. Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat Med* 2015;**21**:1018–27.
- Ellinghaus D, Jostins L, Spain SL, *et al*. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet* 2016;**48**:510–8.
- Acosta-Herrera M, Kerick M, González-Serna D, *et al*. Genome-wide meta-analysis reveals shared new loci in systemic seropositive rheumatic diseases. *Ann Rheum Dis* 2019;**78**:311–9.
- Smolen JS, Aletaha D, Barton A, *et al*. Rheumatoid arthritis. *Nat Rev Dis Primers* 2018;**4**.
- Jonkers IH, Wijmenga C. Context-specific effects of genetic variants associated with autoimmune disease. *Hum Mol Genet* 2017;**26**:R185–R192.
- 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.
- Lee CH, Eskin E, Han B. Increasing the power of meta-analysis of genome-wide association studies to detect heterogeneous effects. *Bioinformatics* 2017;**33**:i379–88.
- Okada Y, Wu D, Trynka G, *et al*. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014;**506**:376–81.
- Todd JA. Etiology of type 1 diabetes. *Immunity* 2010;**32**:457–67.
- Cotsapas C, Hafler DA. Immune-mediated disease genetics: the shared basis of pathogenesis. *Trends Immunol* 2013;**34**:22–6.
- Duerr RH, Taylor KD, Brant SR, *et al*. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006;**314**:1461–3.
- Liu Y, Helms C, Liao W, *et al*. A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. *PLoS Genet* 2008;**4**:e1000041.
- Witoelar A, Jansen IE, Wang Y, *et al*. Genome-wide pleiotropy between Parkinson disease and autoimmune diseases. *JAMA Neurol* 2017;**74**:780–92.
- Harley JB, Chen X, Pujato M, *et al*. Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat Genet* 2018;**50**:699–707.
- Khera AV, Chaffin M, Aragam KG, *et al*. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018;**50**:1219–24.