International Immunology, Vol. 33, No. 12, pp. 647–651
https://doi.org/10.1093/intimm/dxab046
Advance Access publication 27 July 2021

Functional genetics for studying the human immune system

Kazuhiko Yamamoto, Akari Suzuki and **Matteo Maurizio Guerrini**

Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Correspondence to: K. Yamamoto; E-mail: kazuhiko.yamamoto@riken.jp

Received 1 June 2021, *editorial decision* 21 July 2021; *accepted* 27 July 2021

REVIEW

Abstract

Although small numbers of immune-mediated diseases are inherited due to rare genetic mutations, most are multifactorial diseases caused by multiple elements including genetic and environmental factors. In the case of autoimmune diseases, many disease-susceptibility genes, including several in the major histocompatibility gene complex, have been reported, and over the past 10 years, genome-wide association studies (GWAS) have been used to analyze disease-susceptibility loci in representative diseases. Furthermore, many disease-susceptibility variants have been found to be related to gene expression levels. The expression of genes involved in disease pathogenesis is often cell-type-specific, and this is closely related to epigenome alterations. Genomic information is present even before the onset of a disease and has a clear causal relationship to the disease (i.e. the outcome). Therefore, it is important to establish functional genetics in human immunology to understand the pathogenesis of diseases using these pieces of information. We can then apply these results to drug discovery. Here, we will review these issues, especially focusing on autoimmune diseases, and discuss current and future directions of human immune system research.

Keywords: autoimmune diseases, genetic factors, genome function, genome-wide association analysis, human immunology

Introduction

Although there have been significant mandates to promote study of the human immune system, relatively limited methods have been offered. We propose here that human genetics should play important roles in this field because of the large numbers of human genetic variations and their functional involvements in immune functions.

Human immunology research and genome function

Immunological responses are generated by the interaction of many immunocompetent cells and molecules. Immunology as a science has several different analytical tools, mainly using mice, including gene knockouts; as a result, many new mechanistic insights continue to be discovered. However, although the mouse and the human immune systems are fundamentally similar, there are many important differences in the details. In particular, it is often the case that treatments that are effective in mouse models of immune diseases are ineffective in humans. Therefore, the importance of research that directly analyzes the human immune system has been emphasized.

In general, cause and consequence cannot be determined based solely on the observation of immune response phenomena. For example, data on gene expression, protein expression and epigenetics alone do not allow us to establish whether they are the cause or the consequence of a disease. Therefore, analytical systems that clearly identify cause and outcome, similar to gene knockouts in mice, are needed in human immunology research. Disease-susceptibility (risk) gene variations deserve attention as one means to achieve this. With a few exceptions, such as antigen receptor genes, genomic information exists prior to the onset of disease and does not change. Therefore, genetic variations that are statistically associated with a particular phenomenon indicate a clear relationship of being the cause of that phenomenon.

Furthermore, many of the disease-susceptibility variants revealed by genome-wide association studies (GWAS) have been found to be expression quantitative trait loci (eQTLs), which are involved in gene expression levels. In particular, significant overlap between disease-susceptibility variations and the location of histone modifications in promoters and activating enhancers of specific immune cell subsets has been reported (1). By using these relationships, we can qualitatively and quantitatively determine the functional molecules of immunocompetent cells in various immunological diseases by integrating genomic information, gene expression

(mRNA), epigenome and protein analysis to provide causal information on intermediate traits between genome and disease. Comprehensive understanding of these traits will enable the development of genomic functional studies that provide detailed information of an individual's immune status, including causal and consequential factors (2, 3).

Autoimmune diseases and genetic factors

Among immunological diseases, autoimmune diseases have a high incidence of disease in families, and the concordance rate of disease among monozygotic twins is higher than that of dizygotic twins, which is higher than the prevalence in the general population. For example, the disease concordance rate for rheumatoid arthritis (RA) in monozygotic twins varies from report to report, but is roughly 12–15%, while the concordance rate for dizygotic twins or siblings who share 50% of their genes is 2–4%. On the other hand, the overall prevalence of RA has been reported to be 0.5–1%, although it varies somewhat depending on ethnicity. Naturally, there is a possibility that factors such as infectious agents, differences in economic status and environmental factors may be involved in family clustering, but even taking these into account, the involvement of genetic factors is considered to be certain.

The importance of the major histocompatibility complex (MHC), also known as the human leukocyte antigen (HLA), system in immune responses has been widely reported. The association between HLA genetic polymorphisms and autoimmune diseases has been reported not only in RA but also in many other diseases such as systemic lupus erythematosus (SLE), ankylosing spondylitis, Behcet's disease, Graves' disease and type 1 diabetes. The fact that HLA class I and class II molecules have the ability to present antigens to T cells makes it easy to speculate about the relationship between these molecules and the immune response, but it is not clear whether these associations can really be explained by the antigen presentation function alone.

For example, we, together with an international research consortium analyzed the genomic data of ~60 000 RA patients from Japanese, Asian and European populations using the HLA imputation method, which comprehensively analyzes HLA gene sequences on a supercomputer (4). We found that the *HLA-DOA* gene, one of the non-classical HLA genes that is involved in peptide loading on HLA class II molecules, is independently involved in the pathogenesis of RA, in addition to the classical HLA genes such as *HLA-DRB1*, *HLA-DPB1* and *HLA-B* that have been reported so far. Furthermore, unlike conventional classical HLA genes, this non-classical HLA gene has been shown to be involved in disease pathogenesis through changes in its expression levels (4).

Concept and current status of genome-wide association analysis

Since the complete deciphering of the human genome, the possibility of analyzing disease-susceptibility gene variations in autoimmune diseases and other frequent diseases (common diseases, synonymous with multifactorial diseases) has been discussed. Among the variations, single-nucleotide polymorphisms (SNPs) have been the focus of attention as the target of analysis. A polymorphism is a variation with a frequency of 1% or more in a population and is often referred to as a genetic variation, including those with low frequencies. On the other hand, internationally, the HapMap project has analyzed haplotype blocks (linkage disequilibrium blocks) and selected tag SNPs to detect them (5). With the development of SNP typing technology on microarrays, the methodology of GWAS in its current form was finalized (6). Since 2007, GWAS of common diseases using commercial microarrays has been conducted worldwide, and many risk variants of immunological diseases, especially autoimmune diseases, have been reported (7).

In 2014, we integrated previously conducted GWAS data related to RA and analyzed 30 000 RA patients and 70 000 controls, with a particular focus on Asian and European populations (8). As a result, 42 new regions were found to be associated with RA, and a total of 101 gene regions, including those previously reported, were found to be disease-susceptibility loci for RA. However, each region contains a large number of variants and several genes in linkage disequilibrium. Therefore, we integrated various databases to estimate the genetic variants and genes that are most likely to be implicated in RA in each gene region. Then, we investigated the relationship between the target genes of therapeutic drugs for various diseases registered in the drug discovery database and found that the RA-associated genes are related to the target genes of therapeutic drugs for RA through a network of protein–protein interactions. These discoveries demonstrate that genetic analysis not only provides important information for understanding the pathogenesis of RA, but is also useful for future drug discovery (genomic drug discovery).

Although some of these reported genetic variants are involved in the production of qualitatively different protein molecules with amino acid mutations, it is becoming clear that many (estimated 80–90%) of the disease-susceptible genetic variants are eQTLs, and the accumulation of differences in gene function and expression due to such genetic variants is now thought to be associated with the development of disease.

From GWAS to genetic function studies

As mentioned above, since the majority of risk genetic variants in autoimmune diseases are eQTLs, the importance of studying genetic functions on the basis of this information has been recognized. We collected peripheral blood mononuclear cells from 105 healthy individuals, separated them into major immune cell subsets such as CD4-positive T cells, CD8-positive T cells, B cells, natural killer (NK) cells and monocytes using a cell sorter, and quantified the gene expression levels of these cells by next-generation sequencing. Furthermore, the association with genetic variations known from GWAS was comprehensively analyzed and an eQTL catalog was created. We used healthy subjects in order to minimize as much as possible the bias of disease on eQTL. Whereas most of the previous studies have analyzed total blood leukocytes, this study has created a database of how genetic variations affect the expression levels of particular

genes in particular immune cell subsets. Since expression levels vary among cell subsets because of various mechanisms, including the epigenome, for example, these data can be used to understand which cells are affected by risk variants and which cells are affected by individual differences in expression levels. Such cell subset-specific information may explain the genetic mechanisms of these diseases. For example, the function of risk variants for RA can be assessed with the eQTL catalog to understand which genes, in which immune cells, are a risk for RA (9).

Next, we developed a new method to further analyze the pathogenesis of immune diseases by applying this eQTL catalog. This method focuses on the direction of gene expression (increase or decrease in expression level) caused by risk mutations. Normally, most risk variants are genetically calculated to have no significant influence on disease pathogenesis on their own. Therefore, independent evaluation of individual risk variants is unlikely to provide a complete picture of the genetic mechanisms underlying immune diseases. The new method, however, allows us to comprehensively evaluate and interpret the effects of multiple risk variants by aggregating them into a single pathway activity. Specifically, by using genetic information from RA patients and healthy controls, we were able to estimate that the involvement of 176 genes is genetically influenced in CD4-positive T cells, and that activation of the intracellular pathway from the tumor necrosis factor (TNF) receptors has an important role in the pathogenesis of RA. It is interesting to note that these results were calculated based solely on genetic information. Although TNF inhibitors are one of the effective treatments for RA and it was clear that TNF is important in the pathogenesis of RA, the genetic mechanism of its involvement in RA had not been understood. In this regard, it turned out to be clear that the activation of pathways from TNF receptors in a CD4-positive T cell plays an important role (9). A similar approach, called transcriptome-wide association study (TWAS), which integrates GWAS and gene expression data sets, has been proposed (10).

It is not clear by what mechanism genetic risk variants for each disease lead to individual differences in gene expression. GWAS is a system to obtain information covering all chromosomes by typing a limited number of tag SNPs representing the linkage disequilibrium blocks. However, since there are many genetic variants in each linkage disequilibrium block that behave in the same fashion, it is not possible to narrow down the responsible variant from this information alone. Initially, it was assumed that the variant most strongly associated with the disease, i.e. with the lowest *P*-value, was responsible, but it has become clear that this is not always the case (Fig. 1). Therefore, various methods have been tried to identify the functionally responsible variants.

Most of the disease-associated eQTLs are located in non-coding regions and exert their functions in a cell subset-specific manner. The first step is to study the genetic variation for promoter and enhancer functions. If we can identify the location of promoters and enhancers that are working in a cell-specific manner, we can presume and test whether disease-risk variants near or within those locations are affecting gene expression, for example, through differential binding of transcription factors. In this regard, allele-specific expression (ASE) of genes could be used to identify *cis*-regulatory effects of variants (11). More recently, the involvement of long non-coding RNAs (lncRNAs) in particular has been reported in several immunological diseases. In fact, however, methods to detect lncRNAs have not been established, and known lncRNA information is currently limited to the study of each disease. Therefore, the analysis of such cell-specific and stimulus-specific promoters, enhancers and lncRNAs is becoming an important future research area.

Epigenome research

As mentioned above, the probability that the identical twin of a patient suffering from an autoimmune disease develops the disease is not 100%, clearly indicating that factors other than genetic factors are in play. In this regard, it has been suggested that the epigenome, which regulates gene expression through environmental influences, plays an important role in autoimmune diseases. Indeed, differences in DNA methylation and gene expression have been reported in monozygotic twins discordant for SLE (12). Further, it is important to point out that the epigenome regulates gene expression in a cell-type-specific manner. In fact, several approaches such as stratified linkage disequilibrium score regression (stratified LDSC), which annotate the cell-type specificity of GWAS risk variants, have been reported (13).

With regard to the epigenome, SLE has been studied earlier than other diseases and, since the 1990s, it has been shown that the level of methylation in peripheral T cells is lower in active SLE than in inactive SLE. Furthermore, genome-wide analysis of DNA methylation in CD4-positive T cells has reported changes in methylation levels with the development of SLE (14). A more comprehensive study compared methylation of ~46 000 CpG sites in CD4-positive T cells, CD19-positive B cells and CD14-positive monocytes in SLE and healthy subjects. According to this analysis, hypomethylation was more pronounced in SLE, especially in the vicinity of the gene bodies. The degree of hypomethylation was evident in the order of T cells, B cells and monocytes (15). In terms of histone modifications, hypoacetylation of histones H3 and H4 and hypomethylation of H3K9 in CD4-positive T cells have been reported (16).

In RA, as in SLE, DNA of T cells is demethylated, thereby possibly inducing activation of autoreactive T cells. On the other hand, aggressive granulation tissue is also thought to be induced in synovial fibroblast-like cells (17). Suppression of the activity of DNA methyl-transferase 1 (DNMT1), which is involved in maintenance of methylation, is thought to lead to demethylation. For example, abnormal methylation of the CpG site of the MMP13 gene causes increased expression of MMP13, which leads to degradation of type II collagen in cartilage. In terms of histone modifications, it has been shown that the histone deacetylase (HDAC) has different activities in different cells. For example, it has been reported that HDAC expression is up-regulated in peripheral blood and synovial fibroblast-like cells, although HDAC activity is low in synovial tissue as a whole (17). In synovial tissue, hypomethylation of histone H3K9 and hyperacetylation of histones H3 and H4 have been observed (18).

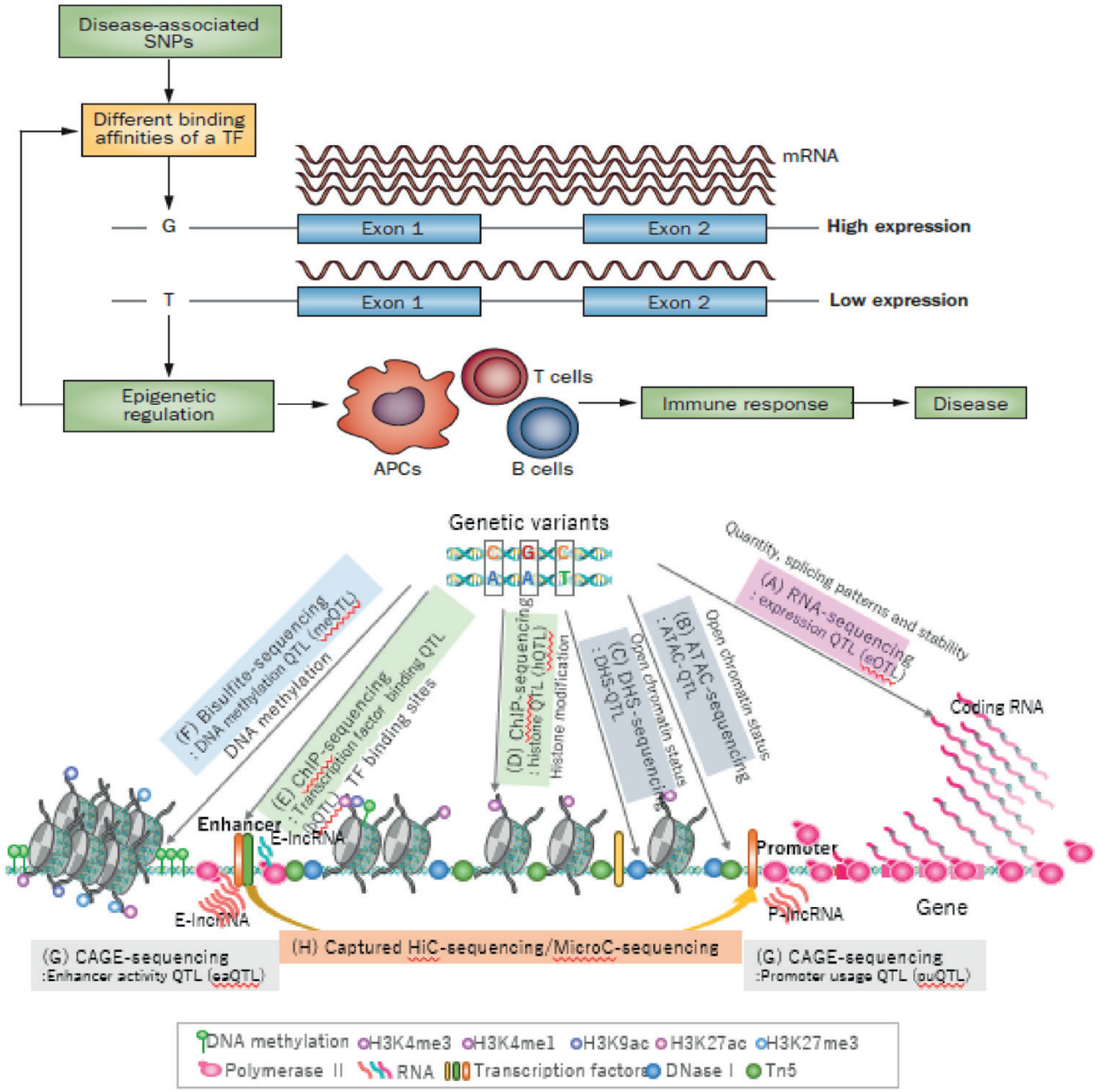

Fig. 1. Genetic variants and QTL. Genetic variants can affect different types of transcriptional and post-transcriptional regulatory factors, including gene expression level detected by RNA-sequencing (A), open chromatin status by ATAC-sequencing (B) or DHS-sequencing (C), histone modifications by ChIP-sequencing (D), transcription factor binding sites by ChIP-sequencing (E), DNA methylation by Bisulfite-sequencing (F), promoter usage and enhancer activity by CAGE-sequencing (G) (19) and chromatin conformation by capture HiC-sequencing/Micro-sequencing (H) (20) (partially modified from references (21, 22)). APC, antigen-presenting cells; ATAC, assay for transposase-accessible chromatin; CAGE, cap analysis gene expression; ChIP, chromatin immunoprecipitation; DHS,DNase I hypersensitive site; HiC, high-throughput chromosome conformation capture.

By combining the results of the previous analysis of epigenomic changes in each disease with the genome functional analysis, including GWAS information described above, we can understand that the analysis of epigenomic changes such as DNA methylation and histone modification around disease-susceptible genetic regions in each cell subset is becoming more important. By integrating this information, we can proceed to the accurate identification of the responsible risk variations and the analysis of the mechanisms of functional changes mediated by these variations. In addition, we can identify the environmental factors that cause cell-specific and site-specific epigenetic changes, and analyze the interaction between genetic and environmental factors. In this way, we can expect to expand our understanding of the human immune system (Fig. 1).

Conclusions

We have discussed the relationship of genetic polymorphisms and immune cell functions, focusing on functional genomics based on GWAS analysis. Considering the enormous variety of human genetic variations and the differences in individual immune functions suggested by these variations, we propose that study of functional genetics is an important direction to better understand the human immune system.

Conflicts of interest statement: the authors declared no conflicts of interest.

References

1 Trynka, G., Sandor, C., Han, B. *et al.* 2013. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* 45:124.
2 Ye, C. J., Feng, T., Kwon, H. K. *et al.* 2014. Intersection of population variation and autoimmunity genetics in human T cell activation. *Science* 345:1254665.
3 Farh, K. K., Marson, A., Zhu, J. *et al.* 2015. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518:337.
4 Okada, Y., Suzuki, A., Ikari, K. *et al.* 2016. Contribution of a non-classical HLA gene, HLA-DOA, to the risk of rheumatoid arthritis. *Am. J. Hum. Genet.* 99:366.
5 International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature* 437:1299.
6 Hirschhorn, J. N. and Daly, M. J. 2005. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* 6:95.
7 Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661.
8 Okada, Y., Wu, D., Trynka, G. *et al.*; RACI Consortium; GARNET Consortium. 2014. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 506:376.
9 Ishigaki, K., Kochi, Y., Suzuki, A. *et al.* 2017. Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. *Nat. Genet.* 49:1120.
10 Wainberg, M., Sinnott-Armstrong, N., Mancuso, N. *et al.* 2019. Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* 51:592.
11 Gutierrez-Arcelus, M., Baglaenko, Y., Arora, J. *et al.*; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium. 2020. Allele-specific expression changes dynamically during T cell activation in HLA and other autoimmune loci. *Nat. Genet.* 52:247.
12 Javierre, B. M., Fernandez, A. F., Richter, J. *et al.* 2010. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res.* 20:170.
13 Finucane, H. K., Bulik-Sullivan, B., Gusev, A. *et al.*; ReproGen Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; RACI Consortium. 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* 47:1228.
14 Jeffries, M. A., Dozmorov, M., Tang, Y. *et al.* 2011. Genome-wide DNA methylation patterns in CD4$^+$ T cells from patients with systemic lupus erythematosus. *Epigenetics* 6:593.
15 Absher, D. M., Li, X., Waite, L. L. *et al.* 2013. Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4$^+$ T-cell populations. *PLoS Genet.* 9:e1003678.
16 Hu, N., Qiu, X., Luo, Y. *et al.* 2008. Abnormal histone modification patterns in lupus CD4$^+$ T cells. *J. Rheumatol.* 35:804.
17 Nakano, K., Whitaker, J. W., Boyle, D. L. *et al.* 2013. DNA methylome signature in rheumatoid arthritis. *Ann. Rheum. Dis.* 72:110.
18 Karouzakis, E., Gay, R. E., Michel, B. A. *et al.* 2009. DNA hypomethylation in rheumatoid arthritis synovial fibroblasts. *Arthritis Rheum.* 60:3613.
19 Murata, M., Nishiyori-Sueki, H., Kojima-Ishiyama, M. *et al.* 2014. Detecting expressed genes using CAGE. *Methods Mol. Biol.* 1164:67.
20 Krietenstein, N., Abraham, S., Venev, S. V. *et al.* 2020. Ultrastructural details of mammalian chromosome architecture. *Mol. Cell* 78:554.
21 Yamamoto, K., Okada, Y., Suzuki, A. *et al.* 2015. Genetics of rheumatoid arthritis in Asia—present and future. *Nat. Rev. Rheumatol.* 11:375.
22 Suzuki, A., Guerrini, M. M. and Yamamoto, K. 2021. Functional genomics of autoimmune disease. *Ann. Rheum. Dis.* 80:689.