The Journal of Rheumatology 2020;47:1597–9
doi:10.3899/jrheum.191404

The Journal of Rheumatology

Editorial

DNA Methylation of the MHC Region in Rheumatoid Arthritis: Perspectives and Challenges

Wenqing Qiu[1] and Yun Liu[1]

The MHC, which covers a region about 4 Mb at 6p21.3, is one of the most polymorphic regions in the human genome. With a high density of more than 200 genes, most of which are directly involved in the immune response to self or non-self antigens, MHC genes have long been associated with a wide range of complex human diseases, including autoimmune or inflammatory diseases and cancer. Rheumatoid arthritis (RA) is a systemic autoimmune disease; recent investigations in large genome-wide association studies using single-nucleotide polymorphisms have confirmed the correlation of classic HLA genes and non-classic HLA genes with RA in many populations[1,2]. Functional and structural analyses indicate that these genetic variants reside in the peptide-binding groove, which may affect the binding affinity of the citrullinated peptides, and eventually lead to the development of RA[1,3]. However, because of the heterogeneity among ethnic groups and clinical subtypes, the major RA-risk allele in MHC is quite different among different populations. For example, HLA-DRB1*04 includes the *04:01 and *04:04 alleles, which are the dominant RA-risk alleles in whites[4], whereas DRB1*04:05/*0901 are the major RA-risk alleles in Asians, specifically for RA patients positive for anticitrullinated protein antibodies (ACPA)[5,6]. A recent study conducted target deep sequencing of the entire MHC region for ACPA-positive RA in a Han Chinese population, and it found that instead of *HLA-DRB1* alleles, HLA-DQA1:160D confers the greatest independent genetic risk[7]. Meanwhile, other genes, such as *HLA-B* and *HLA-DPB1*, have also been identified[8]. Despite the recent efforts to determine the genetic risk factors for RA, the identified risk loci to date account for only a small fraction of the total susceptibility, indicating that other factors are involved in the disease risk.

Many studies have shown that DNA methylation plays an important role in the development and progression of many common diseases[9]. It can act as a mediator of genetic susceptibility and disease phenotype[10,11] and as an integrator of internal genetic and external environmental risk factors in disease development[12,13]. Considering the association of the MHC genetic variants with susceptibility to RA, the potential role of DNA methylation changes in the MHC region has raised much attention. Genome-scale DNA methylation analysis of peripheral blood mononuclear cells based on a Swedish population showed that 9 differential methylation positions (DMP) within the MHC region can mediate the genetic risk for RA[11]. A replication study based on 2 different populations supports the finding that at least 1 DMP in the MHC region is associated with RA[14]. Because DNA methylation is relatively dynamic and can be influenced by many factors (such as infection, drugs, race, and sex), whether the previously reported DMP found within the MHC region are in accord among different ethnic groups still warrants further investigation.

In this issue of *The Journal*, Anaparti, *et al* report on DNA methylation changes within MHC of patients with RA from an indigenous North American (INA) population that is known to have prevalent RA[15]. They conducted whole blood targeted bisulfite sequencing of an MHC locus of about 3.8 Mb using a bacterial artificial chromosome clone-based target enrichment technology[16]. Different from the popular array-based platform (such as the Illumina Infinium HumanMethylation450 BeadChip arrays) for interrogating DNA methylation in the preset CpG sites (a cytoside base followed by a guanine base), this method retains the advantage of recording all of the CpG within the desired region with single-base resolution. By comparing ACPA-positive RA patients with matched ACPA-negative first-degree relatives (ACPA−/FDR), they identified 74 DMP

This work was supported by the National Natural Science Foundation of China (No. 31771451 to Yun Liu).

[1]W. Qiu, MS, Y. Liu, PhD, MOE Key Laboratory of Metabolism and Molecular Medicine, Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, and Zhongshan Hospital, Fudan University, Shanghai, China.

Address correspondence to W. Qiu or Y. Liu, 130 Dong An Road, Shanghai, China. Email: qiuwq@fudan.edu.cn; yliu39@fudan.edu.cn.

See MHC-specific DNA methylation in RA, *page 1614*

within the MHC region, with 32 loci mapped to 21 annotated genes, while most of the rest of the DMP were located within the intronic regions. They compared the DMP found in this research with previously published data[11] and verified 4 exactly matched DMP. The remaining 70 DMP are in close proximity to the previously reported DMP, possibly owing to the different methods and sample size used. To find out the functional relationship between the DMP and the relative gene expression, they selected 5 genes with multiple DMP located either within the gene body or promoter 1 Kb regions, and they found that the relative expression of 2 genes (*C6orf10* and *HCG18*) was significantly influenced by the related DMP[15].

This study confirmed the presence of known DMP and found novel DMP within the MHC region of INA patients with RA, and integrated the RNA expression data to partly reveal the functional role of DNA methylation changes in RA[15]. A recent study of multiple sclerosis integrated the HLA variants with allele-specific DNA methylation and allele-specific RNA expression data. It found that the major risk haplotype DRB1*15:01 is hypomethylated and expressed at a higher level compared to other haplotypes[17]. Another study analyzed the allele-specific methylation of CpG sites between carriers of the type 1 diabetes risk haplotypes HLA-DR3-DQ2 and HLA-DR4-DQ8 and found a marked difference in their methylation status and transcript expression[18]. These studies demonstrate that DNA methylation in the MHC region may mediate the genetic risk for disease development and highlight the importance of integrating multiomic data to elucidate the molecular mechanisms underlying disease susceptibility. Even so, they bring new challenges. The present enrichment approaches for the MHC region are mainly based on collection strategies with probes or PCR, which rely on the known HLA alleles and may result in biased coverage. Additionally, sequencing reads are aligned to the reference genome, but owing to extreme polymorphisms and high levels of sequence homology of HLA genes, sequencing reads with low mapping rate are lost and reads with multiple alignment are abandoned, which leads to ambiguous results. One potential strategy for this problem is to map reads to an HLA personalized haplotype. Several approaches that use known HLA alleles (the IMGT/HLA database)[19,20,21] or a population-based reference graph[22] have been developed to infer the HLA reference closest to individual haplotypes. However, new methods with higher accuracy and low cost are required to obtain HLA personalized haplotypes for downstream analysis. Technologies such as long-read and linked-reads sequencing combined with high molecular weight DNA molecular enrichment technology[23,24] may further advance this research.

REFERENCES

1. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet 2012;44:291-6.
2. Terao C, Ohmura K, Kochi Y, Ikari K, Maruya E, Katayama M, et al. A large-scale association study identified multiple HLA-DRB1 alleles associated with ACPA-negative rheumatoid arthritis in Japanese subjects. Ann Rheum Dis 2011;70:2134-9.
3. Margulies DH. How MHC molecules grab citrullinated peptides to foster rheumatoid arthritis. J Biol Chem 2018;293:3252-3.
4. Lee HS, Korman BD, Le JM, Kastner DL, Remmers EF, Gregersen PK, et al. Genetic risk factors for rheumatoid arthritis differ in Caucasian and Korean populations. Arthritis Rheum 2009;60:364-71.
5. Shimane K, Kochi Y, Suzuki A, Okada Y, Ishii T, Horita T, et al. An association analysis of HLA-DRB1 with systemic lupus erythematosus and rheumatoid arthritis in a Japanese population: effects of *09:01 allele on disease phenotypes. Rheumatology 2013;52:1172-82.
6. Lee HS, Lee KW, Song GG, Kim HA, Kim SY, Bae SC. Increased susceptibility to rheumatoid arthritis in Koreans heterozygous for HLA-DRB1*0405 and *0901. Arthritis Rheum 2004;50:3468-75.
7. Guo J, Zhang T, Cao H, Li X, Liang H, Liu M, et al. Sequencing of the MHC region defines HLA-DQA1 as the major genetic risk for seropositive rheumatoid arthritis in Han Chinese population. Ann Rheum Dis 2019;78:773-80.
8. Eyre S, Bowes J, Diogo D, Lee A, Barton A, Martin P, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet 2012;44:1336-40.
9. Jin Z, Liu Y. DNA methylation in human diseases. Genes Dis 2018;5:1-8.
10. Hong X, Hao K, Ladd-Acosta C, Hansen KD, Tsai HJ, Liu X, et al. Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children. Nat Commun 2015;6:6304.
11. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol 2013;31:142-7.
12. Meng W, Zhu Z, Jiang X, Too CL, Uebe S, Jagodic M, et al. DNA methylation mediates genotype and smoking interaction in the development of anti-citrullinated peptide antibody-positive rheumatoid arthritis. Arthritis Res Ther 2017;19:71.
13. Marabita F, Almgren M, Sjöholm LK, Kular L, Liu Y, James T, et al. Smoking induces DNA methylation changes in multiple sclerosis patients with exposure-response relationship. Sci Rep 2017;7:14589.
14. van Steenbergen HW, Luijk R, Shoemaker R, Heijmans BT, Huizinga TW, van der Helm-van Mil AH. Differential methylation within the major histocompatibility complex region in rheumatoid arthritis: a replication study. Rheumatology 2014;53:2317-18.
15. Anaparti V, Agarwal P, Smolik I, Mookherjee N, El-Gabalawy H. Whole blood targeted bisulfite sequencing and differential methylation in *C6orf10* gene of patients with rheumatoid arthritis. J Rheumatol 2020;47:1614-23.
16. Day K, Song J, Absher D. Targeted sequencing of large genomic regions with CATCH-Seq. PLoS One 2014;9:e111756.
17. Kular L, Liu Y, Ruhrmann S, Zheleznyakova G, Marabita F, Gomez-Cabrero D, et al. DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis. Nat Commun 2018;9:2397.
18. Kindt ASD, Fuerst RW, Knoop J, Laimighofer M, Telieps T, Hippich M, et al. Allele-specific methylation of type 1 diabetes susceptibility genes. J Autoimmun 2018;89:63-74.
19. Aguiar VRC, César J, Delaneau O, Dermitzakis ET, Meyer D. Expression estimation and eQTL mapping for HLA genes with a personalized pipeline. PLoS Genet 2019;15:e1008091.
20. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 2019;37:907-15.

21. Ka S, Lee S, Hong J, Cho Y, Sung J, Kim HN, et al. HLAscan: genotyping of the HLA region using next-generation sequencing data. BMC Bioinformatics 2017;18:258.
22. Dilthey A, Cox C, Iqbal Z, Nelson MR, McVean G. Improved genome inference in the MHC using a population reference graph. Nat Genet 2015;47:682-8.
23. Gabrieli T, Sharim H, Fridman D, Arbib N, Michaeli Y, Ebenstein Y. Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). Nucleic Acids Res 2018;46:e87.
24. Bennett-Baker PE, Mueller JL. CRISPR-mediated isolation of specific megabase segments of genomic DNA. Nucleic Acids Res 2017;45:e165.