Thynn HN, Chen XF, Dong SS, Guo Y, Yang TL. Commentary: An Allele-Specific Functional SNP Associated with Two Systemic Autoimmune Diseases Modulates IRF5 Expression by Long-Range Chromatin Loop Formation. J Immunological Sci. (2020); 4(1): 6-9

Journal of Immunological Sciences

Journal of Immunological Sciences

Commentary Article Open Access

Commentary: An Allele-Specific Functional SNP Associated with Two Systemic Autoimmune Diseases Modulates IRF5 Expression by Long-Range Chromatin Loop Formation

Hlaing Nwe Thynn, Xiao-Feng Chen, Shan-Shan Dong, Yan Guo, Tie-Lin Yang*

Key Laboratory of Biomedical Information Engineering of Ministry of Education, Biomedical Informatics & Genomics Center, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, P. R. China

Article Info

Article Notes
Received: February 14, 2020
Accepted: March 20, 2020

***Correspondence:**
Dr. Tie-Lin Yang, Ph.D., Key Laboratory of Biomedical Information Engineering of Ministry of Education, Biomedical Informatics & Genomics Center, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, P. R. China; Telephone No: 86-29-82668463; Email: yangtielin@xjtu.edu.cn.

Systemic Lupus Erythematosus (SLE) and Systemic Sclerosis (SSc) are two typical inflammatory systemic autoimmune diseases sharing similar pathogenic features. Genetic factors play important roles in the pathogenesis of both diseases[1]. We have witnessed huge success of genome-wide association studies (GWASs) in identifying hundreds of susceptibility genetic variants associated with SLE and SSc, however, over 90% of which are located in noncoding regions. It is challenging and important to translate GWAS findings into biological insights towards clinical applications. Currently, compared with traditional genetic association studies, functional studies characterizing causal functional variants and downstream molecular mechanisms at disease susceptibility loci are still orders of magnitude fewer.

With the emergence of multiple omics technologies including genomics, epigenomics, transcriptomics, proteomics and metabolomics, the deposits of omics databases strongly speed up the clarification of molecular mechanisms associated with human complex diseases[2]. Our recent review[3] highlighted the advantages and challenges of multiple omics data. The methodologies using single-omics data explain limited insights into the biological mechanisms of a disease. The advanced methodologies incorporating with multi-omics approaches followed by functional experiments comprehensively capture more biological insights, which result in better understanding of molecular mechanisms and pathogenesis of diseases[3]. Therefore, integrated multi-omics approaches become powerful tools for interpretation of GWAS susceptibility loci into clinical application. In our publication[4], we provided a mechanistic insight that a noncoding GWAS variant associated with both SLE and SSc acted as an allele-specific strong enhancer to directly regulate the distal gene *IRF5* expression mediated by transcription factor EVI1 through the current robust strategies, including integrated multi-omics approaches and follow-up functional experiments.

The *IRF5-TNPO3* locus at 7q32.1 possesses one of the strongest association signals with both SLE and SSc[5,6], in which *IRF5* (interferon regulatory factor 5) is a well-known immunologic gene[7], whereas the immunologically functional role of *TNPO3* (transportin 3) is still unknown. It is interestingly motivated to investigate the regulatory mechanisms underlying *TNPO3*. Recent emerging studies at many GWAS loci have demonstrated that noncoding variants within potential regulatory elements regulate their distal target

genes associated with diseases via chromatin looping interactions[8-11]. Our study indicated that the nearest gene of disease-associated SNPs or the gene harboring disease-associated SNPs may not be the true target genes in many GWAS loci. This investigation might fulfill the gap between GWAS findings and clinical application of diseases.

Prioritizing variants within GWAS-associated regions is the first crucial step of current research to provide insights into the conversion of statistical associations into target genes and disease biology. As the numbers of GWAS samples become enormous, the association signals at GWAS loci are also increasing. Stepwise conditional analysis is a comprehensive strategy to identify additional multiple association signals at a previously identified GWAS locus[12]. Bayesian fine-mapping followed by functional epigenomics annotation is also a potential strategy to select and prioritize candidate causal variants within GWAS-associated regions[13]. Several previous studies applying stepwise conditional analysis and Bayesian fine-mapping successfully detected additional molecular mechanisms underlying GWAS loci. For example, Galarneau et al. identified seven independent SNPs at *BCL11A*, *HBS1L-MYB* and β-globin loci, using stepwise conditional analysis and fine-mapping, which could explain heritable variation in hemoglobin levels from 38.6% to 49.5%[14]. With high-resolution Bayesian fine-mapping at immune-related loci, 48 new multiple sclerosis susceptibility variants were identified, which enhanced the catalogue of multiple sclerosis risk variants[15]. In our study, we first uncovered that 7q32.1 locus encompassed several independently associated SLE risk variants in either *IRF5* or *TNPO3* region. We could further prioritize a potentially functional independent variant rs13239597 located in *TNPO3* promoter region through the strategies combining stepwise conditional analysis, Bayesian genomic fine-mapping and functional epigenomics annotation. It was also found that the SNP rs13239597 resided within or near putative regulatory enhancer elements in three immune-related cell lines. Here, the comprehensive strategies incorporating with genomics and epigenomics data predicted a newly functional GWAS variant associated with both SLE and SSc, illuminating the potential application of our strategies for more GWAS loci. This approach could be used to investigate more functional causal variants associated with other complex diseases.

The disease-associated SNPs located in noncoding regions of the human genome do not have their protein functions to influence, however, they could affect the expression of their target genes associated with disease phenotypes[16]. The identification of SNPs associated with gene expression levels is known as expression quantitative trait locus (eQTL), which is a critical step towards the better mechanistic understanding of the functional role of phenotype-associated SNPs in GWAS. In context of improving the substantial databases, eQTL analysis represents one of the most straightforward approaches to the identification of candidate susceptibility genes at risk loci in relevant cell/tissue types, which provides evidence of allele-specific functional impacts for risk SNPs[16]. In our study, *cis*-eQTL analysis using various datasets consistently demonstrated that *IRF5* instead of the nearby gene *TNPO3* was the distal target gene (~118 kb) of rs13239597. In this case, we implemented high-throughput chromatin interaction (Hi-C) analysis, topologically associating domain (TAD) analysis and chromosome conformation capture (3C) assay to corroborate the long-range chromatin interaction between rs13239597 and its distal target gene *IRF5*. Hi-C analysis is a unique and powerful tool to reveal the ultimate connectivity between the genomic sequence and spatial conformation, and specific long-range contacts between distant genomic elements such as genes and regulatory elements[17]. Besides, TAD is a self-interacting genomic region, which means that DNA sequences within a TAD boundary physically interact with each other more frequently than with sequences outside the TAD[18]. In our study, Hi-C and TAD analyses were performed using the robust databases, including Capture Hi-C data[19], Hi-C data from 4D Genome databases[20,21], ChIA-PET data from UCSC ENCODE databases[22] and TAD data from GEO databases[23]. Moreover, 3C assay was performed as a follow-up functional experiment, which is a pioneer technique to investigate the three-dimensional structure of chromatin and to analyze long-range looping interactions between any pair of selected genomic loci[24]. When integrating multi-omics data and employing follow-up functional assay, the long-range chromatin looping interaction between rs13239597 and its distal target gene *IRF5* could be convincingly validated.

Furthermore, another follow-up functional experiments were also performed to validate the allele-specific regulation between rs13239597 and *IRF5*. We performed dual-luciferase reporter assay that is a widely used tool to study gene expression at transcription level, and CRISPR-Cas9 that is currently the simplest, most versatile and precise method of genetic manipulation to edit parts of the genome. We also validated whether the detected regulatory activity of rs13239597 on the distal gene *IRF5* was fictitious due to the intermediary effect of the nearby gene *TNPO3* by gene-silencing. Taken together, these experimental results prominently revealed that rs13239597 acted as an allele-specific enhancer regulating *IRF5* expression independently of *TNPO3*. To further vigorously reinforce the allele-specific functionality of rs13239597 on *IRF5*, we encourage to use the base-editing techniques and genome-edited mouse models that were not available in our study.

In addition, to explore the functional mechanisms underlying rs13239597 as a strong allele-specific enhancer on *IRF5*, we further investigated the transcription factors

binding to rs13239597. Chromatin immunoprecipitation (ChIP) assay is widely used as a powerful and versatile technique to evaluate the association between transcription factors and their specific genomic regions involved in the regulatory activities of gene expression within the natural chromatin context of the cells[25]. Through various bioinformatics analyses and functional experiments including allele-specific ChIP assay, dual-luciferase reporter assay in EVI1 suppressed cells and gene knockdown assay, it was detected that the transcription factor EVI1 allele-specifically bound to rs13239597 and ameliorated the enhancer activity to augment *IRF5* expression. EVI1 is crucial for hematopoietic stem cells giving rise to production of human lymphoid cells[26]. Previous studies also demonstrated that many IRF family members play important roles in the differentiation of hematopoietic cells[27], which was consistent with our findings. Finally, to evaluate the role of EVI1 in long-range chromatin interactions between rs13239597 and *IRF5* promoter, chromosome conformation capture (3C) assay was performed in EVI1 knockdown cells. 3C interaction frequencies were significantly lower in EVI1-suppressed cells than in non-treated wild-type cells, highlighting the role of EVI1 in long-range transcriptional regulatory activity of rs13239597 on *IRF5*. These results also supported that transcription factors might participate in long-range chromatin interactions to explain the complex molecular mechanisms underlying GWAS functional variants associated with other complex diseases.

Taken together, our findings uncovered a new long-range regulatory mechanistic insight of a noncoding functional variant rs13239597 acting as an allele-specific enhancer to directly modulate *IRF5* expression with the reinforcement of EVI1, via long-range chromatin loop formation. This study is also the first attempt to address the molecular mechanisms underlying a long-range regulatory SNP associated with both SLE and SSc autoimmune diseases. Our approach by integrating multi-omics analyses and follow-up functional experiments could be applied for the investigation of functional mechanisms underlying noncoding disease risk variants for more human complex diseases, which would accelerate fulfilling the current issues towards the understanding of complex genetic architectures and the promising therapeutic target for precision medicine.

Conflict of Interest Statement

The authors declare no competing interests.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (31871264 and 31970569); Innovative Talent Promotion Plan of Shaanxi Province for Young Sci-Tech New Star (2018KJXX-010); Zhejiang Provincial Natural Science Foundation of China (LGF18C060002); and the Fundamental Research Funds for the Central Universities.

References

1. Scherlinger M, Guillotin V, Truchetet ME, et al. Systemic lupus erythematosus and systemic sclerosis: All roads lead to platelets. Autoimmun Rev. 2018; 17(6): 625-635.
2. Sun YV, Hu YJ. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. Advances in genetics. 2016; 93: 147-190.
3. Yang TL, Shen H, Liu A, et al. A road map for understanding molecular and genetic determinants of osteoporosis Nature reviews. Endocrinology. 2020; 16(2): 91-103.
4. Thynn HN, Chen XF, Hu WX, et al. An Allele-Specific Functional SNP Associated with Two Systemic Autoimmune Diseases Modulates IRF5 Expression by Long-Range Chromatin Loop Formation. The Journal of investigative dermatology. 2020; 140(2): 348-360.e11.
5. Graham RR, Kozyrev SV, Baechler EC, et al. A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. Nat Genet. 2006; 38(5): 550-5.
6. Radstake TR, Gorlova O, Rueda B, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet. 2010; 42(5): 426-9.
7. Ban T, Sato GR, Nishiyama A, et al. Lyn Kinase Suppresses the Transcriptional Activity of IRF5 in the TLR-MyD88 Pathway to Restrain the Development of Autoimmunity. Immunity. 2016; 45(2): 319-32.
8. Gao P, Xia JH, Sipeky C, et al. Biology and Clinical Implications of the 19q13 Aggressive Prostate Cancer Susceptibility Locus. Cell. 2018; 174(3): 576-589.e18.
9. Palstra RJ, de Crignis E, Röling MD, et al. Allele-specific long-distance regulation dictates IL-32 isoform switching and mediates susceptibility to HIV-1. Sci Adv. 2018; 4(2): e1701729.
10. Sokhi UK, Liber MP, Frye L, et al. Dissection and function of autoimmunity-associated TNFAIP3 (A20) gene enhancers in humanized mouse models. Nature communications. 2018; 9(1): 658-658.
11. Chen XF, Zhu DL, Yang M, et al. An Osteoporosis Risk SNP at 1p36.12 Acts as an Allele-Specific Enhancer to Modulate LINC00339 Expression via Long-Range Loop Formation. Am J Hum Genet. 2018; 102(5): 776-793.
12. Cannon ME, Mohlke KL. Deciphering the Emerging Complexities of Molecular Mechanisms at GWAS Loci. Am J Hum Genet. 2018; 103(5): 637-653.
13. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet. 2018; 19(8): 491-504.
14. Galarneau G, Palmer CD, Sankaran VG, et al., Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. Nature genetics. 2010; 42(12): 1049-1051.
15. International Multiple Sclerosis Genetics Consortium (IMSGC), Beecham AH, Patsopoulos NA, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nature genetics. 2013; 45(11): 1353-1360.
16. eGTEx Project. Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. Nat Genet. 2017; 49(12): 1664-1670.
17. Belton JM, McCord RP, Gibcus JH, et al. Hi-C: a comprehensive

technique to capture the conformation of genomes. Methods. 2012; 58(3): 268-76.

18. Krijger PH, de Laat W. Regulation of disease-associated gene expression in the 3D genome. Nat Rev Mol Cell Biol. 2016; 17(12): 771-782.

19. Mifsud B, Tavares-Cadete F, Young AN, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet. 2015; 47(6): 598-606.

20. Jin F, Li Y, Dixon JR, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature. 2013; 503(7475): 290-4.

21. Teng L, He B, Wang J, et al. 4DGenome: a comprehensive database of chromatin interactions. Bioinformatics. 2015; 31(15): 2560-4.

22. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 2012; 22(9): 1760-74.

23. Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012; 485(7398): 376-80.

24. Davison LJ, Wallace C, Cooper JD, et al. Long-range DNA looping and gene expression analyses identify DEXI as an autoimmune disease candidate gene. Hum Mol Genet. 2012; 21(2): 322-33.

25. Huang Q, Whitington T, Gao P, et al. A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. Nature Genetics. 2014; 46(2): 126.

26. Goyama S, Yamamoto G, Shimabe M, et al. Evi-1 is a critical regulator for hematopoietic stem cells and transformed leukemic cells. Cell Stem Cell. 2008; 3(2): 207-20.

27. Tamura T, Yanai H, Savitsky D, et al. The IRF family transcription factors in immunity and oncogenesis. Annu Rev Immunol. 2008; 26: 535-84.