Journal of Autoimmunity 74 (2016) 161–175

Contents lists available at ScienceDirect

Journal of Autoimmunity

journal homepage: www.elsevier.com/locate/jautimm

Review article

The genetic basis of systemic lupus erythematosus: What are the risk factors and what have we learned

Maria Teruel [a, **], Marta E. Alarcón-Riquelme [a, b, *]

[a] Center for Genomics and Oncological Research, GENYO, Pfizer/University of Granada/Andalusian Government, PTS, Granada, 18016, Spain
[b] Institute of Environmental Medicine, Karolinska Institute, Stockholm, 171 67, Sweden

ARTICLE INFO

Article history:
Received 26 July 2016
Accepted 2 August 2016
Available online 10 August 2016

Keywords:
Association studies
SNPs
Functional variants
Co-autoimmunity
Populations
Next-generation sequencing

ABSTRACT

The genome-wide association study is a free-hypothesis approach based on screening of thousands or even millions of genetic variants distributed throughout the whole human genome in relation to a phenotype. The relevant role of the genome-wide association studies in the last decade is undisputed because it has permitted to elucidate multiple risk genetic factors associated with the susceptibility to several human complex diseases. Regarding systemic lupus erythematosus (SLE) this approach has allowed to identify more than 60 risk *loci* for SLE susceptibility across populations to date, increasing our understanding on the pathogenesis of this disease. We present the latest findings in the genetic of SLE across populations using genome-wide approaches. These studies revealed that most of the genetic risk is shared across borders and ethnicities. Finally, we focus on describing the most important risk loci for SLE attempting to cover the genetic findings in relation to functional polymorphisms, such as missense single nucleotide polymorphisms (SNPs) or regulatory variants involved in the development of the disease. The functional studies try to identify the causality of some GWAS-associated variants, many of which fall in non-coding regions of the genome, suggesting a regulatory role. Many loci show an environmental interaction, another aspect revealed by the studies of epigenetic modifications and those associated with genetic variants. Finally, new-generation sequencing technologies can open other paths in the research on SLE genetics, the role of rare variants and the detailed identification of causal regulatory variation. The clinical relevance of the genetic factors will be shown when we are able to use them or in combination with other molecular measurements to re-classify a heterogeneous disease such as SLE.

Contents

** Corresponding author.
* Corresponding author. Center for Genomics and Oncological Research, GENYO, Pfizer/University of Granada/Andalusian Government, PTS, Granada, 18016, Spain.
E-mail addresses: maria.teruel@genyo.es (M. Teruel), marta.alarcon@genyo.es, marta.alarcon@ki.se (M.E. Alarcón-Riquelme).

http://dx.doi.org/10.1016/j.jaut.2016.08.001
0896-8411/

1. Introduction

Technological advances have propitiated spectacular advances in the identification of the genetic contribution to risk for a large number of complex phenotypes and particularly of systemic lupus erythematosus (SLE). What was once considered an impossible task has now been fulfilled by and large through the use of genotyping arrays and the willingness of researchers to collaborate in World wide networks and join forces to share samples of thousands of patients.

SLE (MIM 152700) is an inflammatory, autoimmune, and multisystemic disorder of the connective tissue characterized by autoantibody production and tissue injury, in which environment triggers its development in susceptible individuals.

SLE is not a very common disease, and in fact, the prevalence has never been established properly outside of the European population where it is only 0.05% [1]. The disease in African-Americans, Asians, and individuals of admixed Native American and European (primarily Spanish) ancestry (wrongly known as Hispanics) is much more severe with higher frequencies of kidney inflammation and relapses of disease activity, but not necessarily more prevalent [2–8]. The familial aggregation of SLE is very high [2], supporting an important contribution of genetics.

Despite the advances in genetics, the clinical impact of the genetic data has not reached the bedside, and still much needs to be done. Nevertheless, nearly 30 years of attempts in identifying candidate genes through isolated candidate gene hypothesis-driven studies have been overcome in less than 10 years by genome-wide hypothesis-free association studies (GWAS), with some important exceptions: the HLA (Human leukocyte antigen) locus [9–13], the Fc gamma receptors (FcG) [14–17], *PTPN22* [18–20], first described in type I diabetes, *STAT4* that was described in a rheumatoid arthritis (RA)-SLE combined study [21], and two interferon type-I related genes, *IRF5* and *TYK2*, that were discovered thanks to the longstanding and perseverant belief of the role of type I interferons in the development of the disease [22]. This study, however already utilized methodology later used for high throughput genotyping, and was, in a sense, the beginning of a fructiferous period of discovery. These *loci* have been basically the only ones discovered prior to GWAS studies and were clearly confirmed by these. Finally, *PDCD1*, identified following genome-wide linkage studies in extended pedigrees and multicase families from Iceland and Sweden [23] has been replicated in isolated studies and not in others, and some variants have shown different population frequencies [24–28], but not detected in GWAS, primarily because the gene is located at the end of the long arm of chromosome 2 and few SNPs are found within the gene, and none of those showing the association. In our view, this genetic association remains unclear.

The HLA locus, and in particular the association with HLA class I alleles was identified in the seventies [9–11,29]. Association with class II followed, and eventually it was observed that association was also observed with the complement component genes C2 and C4 also found within the region of the class III genes [12,13,30–34], associations observed due to the strong linkage disequilibrium (LD) within the *locus*. However, the strong and extended linkage disequilibrium, particularly characteristic of the European population, within the Major Histocompatibility Complex (MHC) have made it almost impossible to discern the precise genetic contribution of individual alleles and loci. The study of thousands of individuals and in particular of several ancestries has opened the possibility of finding recombinants within the HLA locus. Such studies reveal the true architecture of the genetic association, and in particular the functional effects of genetic variants in that *locus*, as well as the origin of the HLA risk alleles. Other genes, in particular rare variants appear also to have an important effect in SLE, mainly with the identification of immune phenotypes resembling SLE in individuals who are carriers of rare monogenic diseases, but in this paper we are not considering these (see Table 1), as recent extensive reviews are available [35–39].

In this review, we focus on describing the latest findings in the genetics of SLE across populations using genome-wide association arrays. We describe the most important risk *loci* for SLE attempting to cover the genetic findings in relation to the functional polymorphisms. Many of the GWAS-associated variants fall in non-coding regions of the genome suggesting the presence of regulatory elements and the role of epigenetics. Many recent publications support this hypothesis. We also consider the latest findings on genetics using new-generation sequencing technologies.

2. Genome-wide association studies (GWAS)

2.1. European GWAS

The genetic variants primarily investigated in the past 5–10 years have dealt with common variants, where a relatively elevated allele frequency allows for the study of genetic association using a case-control design. The first two genome-wide association studies were performed in European populations and were published simultaneously in *Nature Genetics* [40] and *The New England Journal of Medicine* [41]. The GWAS done by the International Lupus Genetics Consortium (SLEGEN) used 730 women who were individual members of families with multiple cases of SLE [40]. The study was complemented with genotyped and out-of-study controls making an overall, after quality control of 720 SLE female cases and 2337 female controls. The study identified several loci that have been in their majority, confirmed. Some of the loci have been more carefully fine mapped and, in most cases, more than one gene with independent association has been defined within each *locus*. The work implicated a large number of collaborators, a requirement to gain sufficient power for the identification of new genetic loci and consolidate their genetic association [42]. The study used the Human Hap300 array for whole genome genotyping. Results were validated using a set of patients and controls known as the Lupus Large Association Study (LLAS1) consisting of 1840 SLE patients and 1825 controls, all of European or European American ancestry.