REVIEW

An update on the genetics of systemic lupus erythematosus

Nina Oparina[a], Manuel Martínez-Bueno[b], and Marta E. Alarcón-Riquelme[a,b]

Purpose of review

The aim of this study is to update on the most recent findings on the genetics of systemic lupus erythematosus.

Recent findings

Our overview focuses particularly on results from expression quantitative trait loci, exome sequencing, and rare variants and their impact on disease.

Summary

Systemic lupus erythematosus is a systemic autoimmune disease for which a significant number of susceptibility genes have been identified. Several genome-wide association studies were recently published in different populations that provide a better picture of the molecular mechanisms. It is becoming clear that the genetic architecture of lupus is quite well established but more information is required on the role of rare variants.

Keywords

autoimmunity, exome sequencing, genome-wide association study, low-frequency variation, rare variation, systemic lupus erythematosus

INTRODUCTION

Systemic lupus erythematosus (SLE MIM [152700]) is a systemic autoimmune disease (SAD) with a complex multifactorial etiology and a broad spectrum of clinical manifestations [1,2]. The complexity resides in the combination of various environmental and genetic factors in the initiation and progression toward disease development that occurs with time.

In this review, we provide an update of the genetics of SLE focusing on genetic association studies and fine mapping of known genetic variants affecting gene expression, but also on rare and de-novo variants and their potential role in familial aggregation and clinical features of the disease.

GENETIC ASSOCIATIONS AND GENOME-WIDE ASSOCIATION STUDIES

SLE is a complex SAD that affects every organ and system in the body and with varying clinical and serological manifestations. Basically, the loss of tolerance in the immune system leads to the presence of autoantibodies and the deposition of immune complexes in various tissues, causing a great diversity of symptoms [3]. SLE has a strong genetic component supported by twin and family studies [4–6]. Multiple genome-wide and candidate-gene association studies identified over 80 SLE susceptibility loci, explaining about 30% of narrow-sense SLE heritability [7–9]. Although the overall heritability of complex diseases is difficult to estimate, a classical study in European (EUR) population did estimate a heritability of 66 plus-minus 11% for SLE [10], indicating that there was still more than 50% of heritability estimations missing for SLE from current genome-wide association studies (GWAS). However, the estimates of genetic influence on SLE susceptibility vary between studies and populations, reaching up to 44% in a Taiwanese study [11]. Even so,

[a]Unit of Chronic Inflammatory Diseases, Institute for Environmental Medicine IMM, Karolinska Institutet, Solna, Sweden and [b]GENYO, Center for Genomics and Oncological Research Pfizer/University of Granada/Andalusian Regional Government, Parque Tecnológico de la Salud, Granada, Spain

Correspondence to Marta E. Alarcón-Riquelme, GENYO, Av de la Ilustración, 114, Parque Tecnológico de la Salud, 18016, Granada, Spain. Tel: +34 958 715 500x113; e-mail: marta.alarcon@genyo.es

Curr Opin Rheumatol 2019, 31:659–668

DOI:10.1097/BOR.0000000000000654

1040-8711 Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

KEY POINTS

- Currently, genetic studies have allowed the identification of more than 80 risk *loci* for SLE susceptibility, more than 50 of which were independently replicated.
- In general, there is the enrichment of *loci* SLE-susceptibility lying within transcription factors.
- Majority of established SLE risk loci, identified in independent GWAS studies correspond to adaptive and innate immunity, but there several stable risk associations remain unexplained.
- Many of SLE associated loci are also expression QTLs, regulating expression of these or neighboring genes.
- Availability of large-scale population sequencing databases allowed performing more comprehensive studies, on the basis of the genotype imputation both for common and rare variant SLE genetic studies.
- The most recent progress was achieved in the role of rare and de-novo variants in SLE susceptibility, with ~100 candidate loci identified, but further research is needed for their replication, validation, and causal mechanisms uncovering.

one-third of the heritability of SLE would still remain to be explained.

Most of the genetic studies have identified common variants showing consistent and robust genetic association after the strong advent of microarray technology and despite stringent Bonferroni correction. The main genes, associated with SLE susceptibility are summarized in Figure 1 and Table 1. The first three GWAS for SLE were simultaneously published [12–14] and addressed between 100,000 and 300,000 single nucleotide polymorphisms (SNPs). Nowadays, following the genotyping of millions of SNP per individual, the combined analysis of genotyping and sequencing can be used for imputation of missing SNVs onto the experimental genotype data to increase resolution. Both new whole-genome and whole-exome sequencing and population-specific databases such as the 1000 Genomes or the EUR ancestry-centered Haplotype Reference Consortium were shown to be useful for this purpose [15–17]. In general, the majority of the genes identified as susceptibility to SLE correspond to innate and adaptive immune system pathways.

The various GWAS studies identified several genes for SLE such as *BLK*, *ITGAM-ITGAX*, and *PXK* [12–14] and confirmed the human leucocyte antigens (*HLA*) and *FCGR2A*. An updated list of all known and replicated loci for SLE is found in Table 1. New knowledge on the function of some of the genes has been attained, for instance, the role of BANK1 in MyD88-TRAF6 innate immune signaling in B cells and the importance of its exon-2 TIR-domain where the genetic association described is located [13,18]. *IRF5* remains one of the best SLE-associated loci along with other interferon regulators *IRF7* and *IRF8*. A more recent GWAS in EURs analyzed over 8274 cases and 23,579 controls and identified 10 new associations [19]. These new signals increased the explained proportion of genetic variance in SLE from 8 to 16.3%. The validity of these SLE association signals was supported by meta-analysis using Bayesian methods. Among SLE risk variants from publications prior to 2017, 45 (34% of all tested) were under applied false-positive probability estimation approaches [20]. The new analysis approaches increased efficiency even in smaller case-control datasets. Thus a study including 907 cases and 1524 EUR controls followed by genotype imputation and metaanalysis using a larger sample was able to reveal new SLE-associated loci *GRB2*, *SMYD3*, *ST8SIA4*, *LAT2*, and *ARHGAP27* and supported 51 of 52 of the previously known [21]. Machine learning was applied on a targeted ImmunoChip genotyping study providing 15 new SLE-associated candidate loci and increasing SLE risk prediction accuracy, especially for lupus nephritis (area under the curve 0.91) [22].

The *HLA* class II region is undoubtedly associated with SLE susceptibility in all populations, but the tight linkage disequilibrium that characterizes the region has made it difficult to identify independent signals. Particularly, the DR15 haplotype containing classical alleles HLA-DRB1*15:01, HLA-DQB1*06:02, HLA-DQA1*01:02 [23], and HLA-DRB5*01:01 [24] is consistently replicated. Various signals within *HLA* class I and class III were found to independently contribute to SLE genetic susceptibility [25]. Targeted next-generation sequencing (NGS) of the major histocompatibility complex (MHC) region showed a haplotype with regulatory polymorphisms associated with changes in the gene expression of the HLA class II molecules [26]. Proper analysis of the highly variable *MHC* locus across populations is challenging. Interestingly, using local ancestry analysis on an Amerindian–EUR mixed ancestries' sample, it was found that the *HLA* risk-alleles had a EUR origin, whereas protective alleles or haplotypes were Native American [27]. A combined analysis of risk alleles of the HLA performed in EURs and African Americans showed that the long-range SLE associated signals in EUR corresponded to narrow independent peaks in African Americans data. Population-specific independent associations were detected. Nevertheless the majority of risk alleles, despite the great diversity of the

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

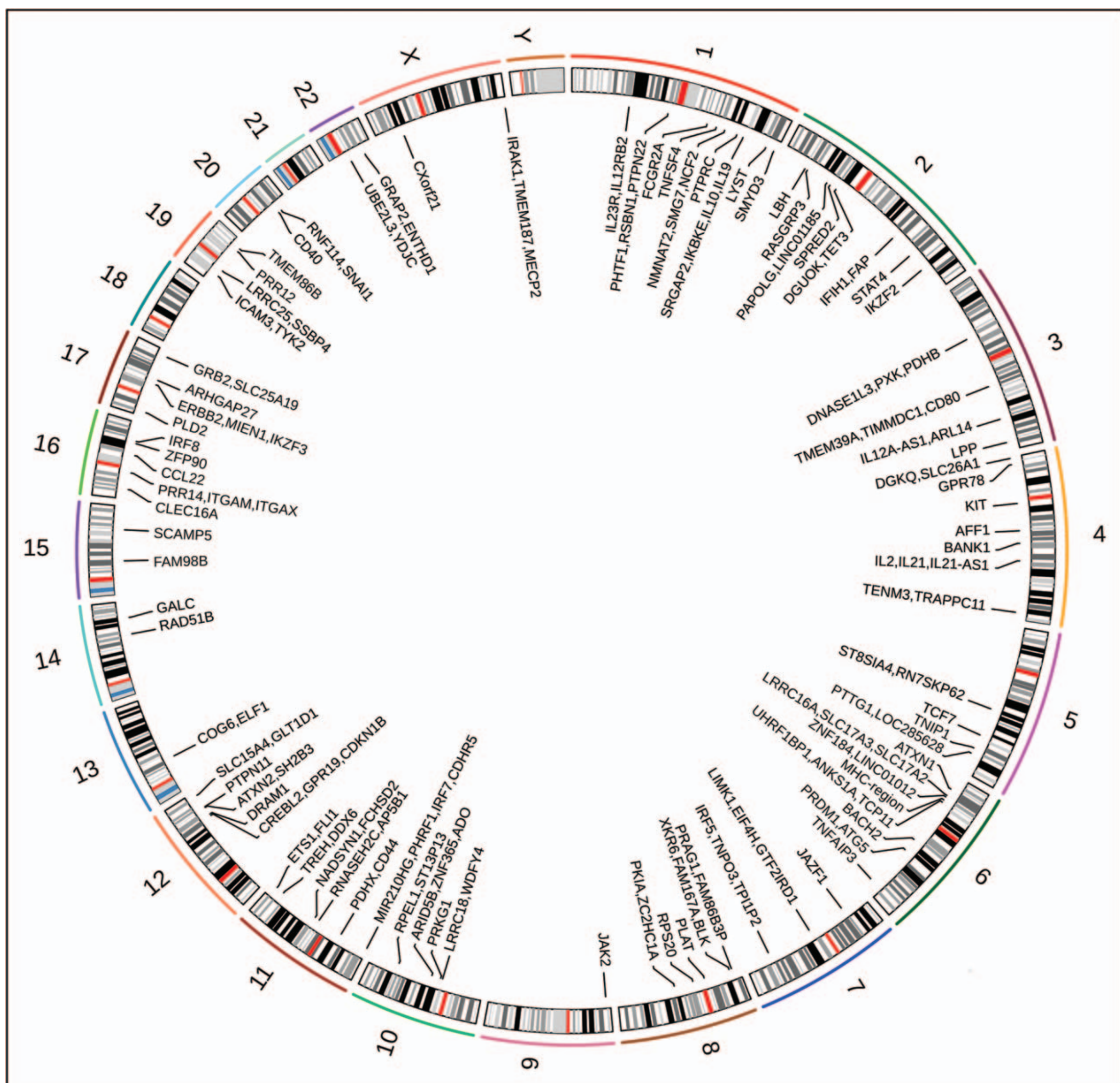

FIGURE 1. Schematic representation depicting genes located within the systemic lupus erythematosus risk loci (88 in total) according to their genomic position. The full set of variants and loci for this plot is summarized in Supplementary Table 1, http://links.lww.com/COR/A45. The annotation of genes and cytogenetic bands correspond to the hg19 assembly. The red block in each chromosome marks its centromere. This figure was made using R package 'Rcircos' [82].

African ancestry population, demonstrated similarity to EUR *HLA* SLE association and significant concordance in the direction of the allele effects [28]. Another study characterizing SLE risk in the HLA region across EUR, African, and Hispanic ancestries described risk-allele heterogeneity within DQA1/DQB1 and DRB1 [9]. HLA-DRB1∗15:01 and HLA-DQB1∗06:02 were associated to SLE risk in Asian ancestry populations [29,30]. The contribution of HLA alleles and specific amino-acid residues to the risk of SLE and the presence of specific autoantibodies has recently been dissected in a large East Asian sample [31].

Many of the SLE-associated genes identified in EURs have been replicated and confirmed in other populations, particularly the Chinese [8,3–38,39▪,40–43]. Targeted loci evaluation analysis confirmed *SPRED2*, *IKZF2*, *IL12A*, *TCF7*, *PLD2*, and others [39▪]. A study of a large cohort of East Asian patients (11,656 cases versus 23,968 controls) revealed five new SLE loci: *MYNN*, *ATG16L2*, *CCL22*, *ANKS1A*, and *RNASEH2C* [8]. A transancestral study of sporadic SLE in EURs and East Asian datasets of 12,280 cases and 18,828 controls further revealed *ST3AGL4*, *MFHAS1*, and *CSNK2A2*. The most pronounced results were obtained for *CSNK2A2* kinase,

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Table 1. List of the 88 genomic regions annotated with the SLE trait which include associations with P values less than 8×10^{-8} according to GWAS catalog (https://www.ebi.ac.uk/gwas/)

N locus	CytoBand	Reported gene(s)	Ancestry	*N* studies
1	1p31.3	IL23R, IL12RB2	AFR, EUR, HIS	3
2	1p13.2	PHTF1, RSBN1, PTPN22	AFR, EAS, EUR, HIS	5
3	1q23.3	FCGR2A	AFR, EAS, EUR, HIS	3
4	1q25.1	TNFSF4	AFR, EAS, EUR, HIS	10
5	1q25.3	NMNAT2, SMG7, NCF2	AFR, EAS, EUR, HIS	2
6	1q31.3	PTPRC	EAS, EUR	1
7	1q32.1	SRGAP2, IKBKE, IL10, IL19	AFR, EAS, EUR, HIS	4
8	1q42.3	LYST	EUR	1
9	1q44	SMYD3	EUR	1
10	2p23.1	LBH	AFR, EUR, HIS	2
11	2p22.3	RASGRP3	AFR, EAS, EUR, HIS	1
12	2p16.1	PAPOLG, LINC01185	AFR, EUR, HIS	1
13	2p14	SPRED2	EAS, EUR	1
14	2p13.1	DGUOK, TET3	EAS, EUR	2
15	2q24.2	IFIH1, FAP	AFR, EAS, EUR, HIS	3
16	2q32.3	STAT4	AFR, EAS, EUR, HIS	15
17	2q34	IKZF2	EAS, EUR	2
18	3p14.3	DNASE1L3, PXK, PDHB	AFR, EUR, HIS	5
19	3q13.33	TMEM39A, TIMMDC1, CD80	AFR, EAS, EUR, HIS	3
20	3q25.33	IL12A-AS1, ARL14	AFR, EAS, EUR, HIS	3
21	3q28	LPP	EAS, EUR	1
22	4p16.3	DGKQ, SLC26A1	AFR, EUR, HIS	2
23	4p16.1	GPR78	EAS	1
24	4q12	KIT	EAS	1
25	4q21.3	AFF1	EAS	1
26	4q24	BANK1	AFR, EAS, EUR, HIS	3
27	4q27	IL2, IL21, IL21-AS1	AFR, EUR, HIS	2
28	4q35.1	TENM3, TRAPPC11	EAS	2
29	5q21.1	ST8SIA4, RN7SKP62	AFR, EUR, HIS	2
30	5q31.1	TCF7	AFR, EUR, HIS	3
31	5q33.1	TNIP1	AFR, EAS, EUR, HIS	8
32	5q33.3	PTTG1, LOC285628	AFR, EUR, HIS	5
33	6p22.3	ATXN1	AFR, EAS, EUR, HIS	1
34	6p22.2	LRRC16A, SLC17A3, SLC17A2	AFR, EUR, HIS	1
35	6p22.1	ZNF184, LINC01012	EUR	1
36	6p21.32–33	MHC region	AFR, EAS, EUR, HIS	12
37	6p21.31	UHRF1BP1, ANKS1A, TCP11	EAS, EUR	4
38	6q15	BACH2	EAS, EUR	1
39	6q21	PRDM1, ATG5	AFR, EAS, EUR, HIS	7
40	6q23.3	TNFAIP3	AFR, EAS, EUR, HIS	7
41	7q15.1	JAZF1	AFR, EAS, EUR, HIS	4
42	7q11.23	LIMK1, EIF4H, GTF2IRD1	AFR, EAS, EUR, HIS	6
43	7q32.1	IRF5, TNPO3, TPI1P2	AFR, EAS, EUR, HIS	15
44	8p23.1	PRAG1, FAM86B3P	AFR, EUR, HIS	1
45	8p23.1	XKR6, FAM167A, BLK	AFR, EAS, EUR, HIS	15
46	8p11.21	PLAT	AFR, EUR, HIS	1
47	8q12.1	RPS20	AFR, EUR, HIS	1

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Table 1 (*Continued*)

***N* locus**	**CytoBand**	**Reported gene(s)**	**Ancestry**	***N* studies**
48	8q21.12	PKIA, ZC2HC1A	AFR, EUR, HIS	1
49	9p24.1	JAK2	EAS, EUR	1
50	10q11.23	LRRC18, WDFY4	AFR, EAS, EUR, HIS	4
51	10q21.1	PRKG1	AFR, EAS, EUR, HIS	1
52	10q21.2	ARID5B, ZNF365, ADO	EAS, EUR, AFR, HIS	3
53	10q24.33	RPEL1, ST13P13	EUR, HIS	1
54	11p15.5	MIR210HG, PHRF1, IRF7, CDHR5	AFR, EAS, EUR, HIS	6
55	11p13	PDHX, CD44	AFR, EAS, EUR, HIS	4
56	11q13.1	RNASEH2C, AP5B1	AFR, EAS, EUR, HIS	1
57	11q13.4	NADSYN1, FCHSD2	EUR, EAS	2
58	11q23.3	TREH, DDX6	AFR, EUR, HIS	1
59	11q24.3	ETS1, FLI1	AFR, EAS, EUR, HIS	6
60	12p13	CREBL2, GPR19, CDKN1B	EAS	1
61	12q23.2	DRAM1	EAS	1
62	12q24.12	ATXN2, SH2B3	EAS, EUR	2
63	12q24.13	PTPN11	EUR	1
64	12q24.33	SLC15A4, GLT1D1	AFR, EAS, EUR, HIS	4
65	13q14.11	COG6, ELF1	EAS, EUR	2
66	14q24.1	RAD51B	AFR, EUR, HIS	2
67	14q31.3	GALC	AFR, EUR, HIS	1
68	15q31.3	FAM98B	EUR	1
69	15q24.2	SCAMP5	EAS, EUR	2
70	16p13.13	CLEC16A	AFR, EAS, EUR, HIS	2
71	16p11.2	PRR14, ITGAM, ITGAX	AFR, EAS, EUR, HIS	10
72	16q13	CCL22	AFR, EUR, HIS	1
73	16q22.1	ZFP90	AFR, EAS, EUR, HIS	1
74	16q24.1	IRF8	AFR, EAS, EUR, HIS	5
75	17p13.2	PLD2	EUR	1
76	17q12–q21.1	ERBB2, MIEN1, IKZF3	EAS, EUR	2
77	17q21.31	ARHGAP27	EUR	1
78	17q25.1	GRB2, SLC25A19	AFR, EUR, HIS	2
79	19p13.2	ICAM3, TYK2	AFR, EAS, EUR, HIS	3
80	19p13.11	LRRC25, SSBP4	AFR, EUR, HIS	1
81	19q13.33	PRR12	EUR	1
82	19q13.42	TMEM86B	AFR, EUR, HIS	1
83	20q13.12	CD40	AFR, EUR, HIS	1
84	20q13.13	RNF114, SNAI1	EUR	1
85	22q11.21	UBE2L3, YDJC	AFR, EAS, EUR, HIS	6
86	22q13.1	GRAP2, ENTHD1	AFR, EUR, HIS	1
87	Xp21.2	CXorf21	EUR	1
88	Xq28	IRAK1, TMEM187, MECP2	EAS, EUR, HIS	2

'*N* studies' refers to the number of references compiled in GWAS-catalog describing association to SLE for each locus. These references are listed by variant in Supplementary Table 1, http://links.lww.com/COR/A45. 'Ancestry' indicates the populations in which association to SLE has been described in the GWAS-catalog reference list for each locus. AFR, African American; EAS, East Asian; EUR, European; HIS, Hispanic or Native American; MHC, major histocompatibility complex; SLE, systemic lupus erythematosus.

1040-8711 Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

showing a B-lymphocyte-specific regulatory effect of the associated risk variants [42]. A GWAS in Koreans identified a new *locus* in chromosome 11q14 (*ATG16L2*, *FCHSD2*, *SIGLEC12*, and *P2RY2*) and confirmed many other [44]. Interestingly, *ATG16L2* was originally described in Crohn's disease [45], showing a tighter link across several autoimmune diseases, where some genes are implied in the tissue-specific regulation or in the systemic inflammation still needs to be investigated.

Despite the prevailing immune-relevant functions among the established SLE loci, for some of them, such as *JAZF1*, *XKR6*, *UHRF1BP1*, or *WDFY4*, there are no known studies of their role in disease development. For some described loci despite lack of known function the expression pattern confirms presence of corresponding mRNAs in relevant tissues such as spleen, lymph nodes or blood cells. In general, the enrichment of susceptibility *loci* lying within transcription factors was a major finding EURs [19]. Fig. 1 and Table 1 show the 88 genomic regions annotated with the 'SLE' trait which we include associations with p value less than 8×10^{-8} according to the GWAS-catalog (https://www.ebi.ac.uk/gwas/). All significant associations are shown in Supplementary Table 1, http://links.lww.com/COR/A45.

GENE EXPRESSION REGULATION

Genetic variants can affect disease susceptibility through modifying genetic expression levels [expression quantitative trait loci (eQTLs)]. Genotype and gene expression data are used combined for this analysis. Publicly available eQTLs datasets help evaluate whether the detected SLE-risk SNPs influence transcript levels of genes [46–50]. These approaches initially showed that trait-associated SNPs were more likely to be eQTLs [50]. Noteworthy, the STAT4 locus, significantly associated with SLE in all studied populations, remained unexplained in the terms of expression regulation. SLE risk variants lay in intronic areas not genetically linked to the STAT4 promoter. A study based on fine-mapping and eQTL analysis, supported the importance of STAT4-located SLE risk rs11889341 variant on expression of neighboring STAT1, but not STAT4 expression increase in B cells [51]. The SLE protective role of *SIGLEC12* described in East Asians is mediated by eQTLs enhancing the expression of the gene [52]. Ten of the Immunochip-based identified variants in East Asians altered gene expression [30]. Another East Asian GWAS supported regulatory variation in 82 genes and overrepresentation of p53, MEF2A, and E2F1 transcription-factor-binding sites [8]. A large-scale exome-wide study in a Han Chinese sample identified an intergenic variant with a *cis*-eQTL effect reducing TPCN2 expression in immune cells from SLE patients [41]. Usually, eQTL-mapping studies identify variants affecting gene expression of nearby genes (*cis*-eQTLs). However, *trans*-eQTLs have also been detected. For example, the SLE SNP rs4917014 acts in *cis* on IKZF1 and in *trans* for C1QB and five type I interferon genes, both hallmarks of SLE [53]. Similarly, the major SLE-risk SNPs in BANK1 are eQTLs, and risk associated with increased gene expression, connecting the regulatory landscape of BANK1 with cotranscriptional splicing of exon-2 [54]. The use of RNA-seq to assess genome-wide transcription abundance provides information on allele-specific expression and RNA-isoform expression, which is not available from gene expression microarrays [55]. It has been shown that conventional gene expression quantification underestimated the identification of causal *cis*-eQTLs [56▪▪]. Integrating eQTL data derived from both microarray and RNA-Seq experiments with GWAS results in SLE identified new susceptibility genes (e.g., *NADSYN1* and *TCF7*). In addition, this procedure allows the identification of novel SLE associated splicing events and noncoding RNAs contributing to the better understanding of the functional consequence of regulatory variants [57▪▪].

RARE AND LOW-FREQUENCY VARIANTS

Individual effects of common variation on susceptibility are small, with odds ratios ranging from as little as 1.01–2.5 at the most. Description of the effect of rare variation on complex phenotypes is an open field. In this regard, few examples exist of low-frequency or rare genetic variants in SLE related to the genes identified in GWAS studies. One example is *BLK*, where a low-frequency variant (a nonsynonymous change from alanine to threonine) was identified [58]. Another example is *NCF2*, where a nonsynonymous coding mutation in exon 12, a histidine to glutamine substitution in the PB1 domain of NCF2 protein reducing its binding efficiency to the guanine nucleotide exchange factor Vav1 [59].

A recent study of Sardinia SLE and multiple sclerosis patients revealed a new insertion–deletion variant in the *TNFSF13B* gene encoding the B-cell activating factor (BAFF). Serum levels of BAFF are frequently elevated in SLE patients. This variant resulted in the alternative polyadenylation of the transcript, lacking the binding site for inhibitory microRNAs and consequently BAFF upregulation [60▪▪]. This mutation propagated becoming more frequent possibly because of its protective role against malaria.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

DNASE1-like deoxyribonucleases are well-established loci, known for their rare loss-of-function variants in familial SLE. *DNASE1L3* located in the close vicinity of *PXK* contains rare heterozygous variants enriched in sporadic SLE [61▪]. A later study performed quantitative PCR analysis of copy number variations in *DNASE1L3* and *DNASE2* and found that although the variants were universally rare in most populations with no copy loss-or-gain events, these were more frequent in EURs and Asian SLE [62].

The abundance of articles describing rare genetic variants, historically those on complement deficiencies and lately on interferonopathies where rare mutations found in monogenic diseases can be found in some SLE patients [63–66,69▪▪] suggest that such mutations may explain the relatively elevated frequency of familial cases of SLE [5] (8–12%), or alternatively, that SLE is a much more heterogeneous collection of individuals. Possibly, mutations causing forms of monogenic diseases having ill-described autoimmune manifestations are much more prevalent than previously suspected, contributing to the overall SLE phenotype. It is therefore potentially relevant to identify the possible genes involved and perform searches for autoimmunity among rare diseases to help further determine the rainbow of pathways that lead to disease. However, and relevant, whereas monogenic lupus may imply the contribution of mutations with high penetrance, in complex phenotypes such as SLE, we would expect rare variation with small effects, very hard to identify using genetic association [66].

In the first large SLE study of its kind exome sequencing of 30 parent-affected-offspring trios identified using SLE-associated burden analysis three functional de-novo loci: *DNMT3A*, *PRKDC*, and *C1QTNF4* [67▪▪]. In another study, a set of 71 SLE patients and their healthy parents was studied using whole genome sequencing and the contribution of ultrarare missense and nonsense variants in genes known to be causal for monogenic SLE was estimated. Enrichment of ultrarare variants was found for 22 genes. Interestingly, for *C1S*, *DNASE1L3*, *DNASE1*, *IFIH1*, and *RNASEH2A* disrupting rare variants were shown to be significantly associated with SLE mostly in the heterozygous state [68].

By performing exome sequencing on the most distantly related affected individuals from two large Icelandic families, multiple rare and likely pathogenic variants in 19 genes cosegregating with disease through multiple generations were identified [69▪▪]. The data was supplemented with information on coexpressed and protein–protein interacting partners. This unsupervised functional analysis showed enrichment in the gene ontology categories of immune system development, lymphocyte activation, DNA repair, and VDJ T and B-cell receptor gene recombination. Further support was found using a very stringent aggregate association analysis in sporadic cases for the *FAM71E1/EMC10* locus. The *EMC10* gene (endoplasmic reticulum membrane complex subunit 10) codes for a protein involved in endoplasmic reticulum-associated degradation and lipid transport. Another interesting gene was *DCLRE1C*. *DCLRE1C* is involved in double-strand break repair, cellular response to DNA damage stimuli, and chromosome organization protein Artemis. Recessive mutations in this gene cause Omenn syndrome, a severe combined immunodeficiency associated with increased cellular radiosensitivity because of a defect in V(D)J recombination leading to early arrest of B and T-cell maturation [70]. A recent functional study demonstrated that Artemis-deficient cells have type I and type III interferon signatures because of the chronic accumulation of DNA [71].

Owing to the difficulty in defining the role of single rare or low-frequency mutations in association analyses, a very stringent imputation-based approach for the whole genome rare variant enrichment analysis in SLE patients of EUR ancestry [72] was applied. This approach uncovered 98 top candidate loci. The loci were prioritized by two independent approaches: highly stringent sequence kernel association test and a case-control burden test. Several of these loci had immune-relevant functions, whereas for others their role in SLE remains obscure. Significant overrepresentation of Online Mendelian Inheritance in Man (OMIM) disease genes was observed, suggesting that such variants could be involved in the generation of combined SLE phenotypes, which looks reasonable considering the broad spectrum of SLE clinical manifestations and the large number of anecdotal descriptions of rare phenotypes in SLE patients [1,2]. Proper analysis of rare variants may be overcome with the numerous computational approaches being published (reviewed in [73]). In addition to the importance of avoiding sequencing errors and analysis artifacts, there is no consensus for the best strategy for rare variants analysis. Sequencing of targeted genes is an optimal procedure for detecting mutations. Thus, exome sequencing of SLE-risk genes in an European ancestry sample with 69 SLE affected and 97 healthy controls showed that rare BLK and BANK1 missense variants contributed to risk [74]. Resequencing of the SLE associated *ITGAM* gene in 73 SLE cases identified two case-specific nonsynonymous variants, F941V and G1145S that significantly impaired phagocytosis [75]. Similarly, targeted resequencing of coding

1040-8711 Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

and conserved regulatory regions in a set of autoimmunity-realed loci in a Swedish SLE cohort identified a rare regulatory variant rs200395694:G>T in intron 4 of *MEF2D* associated with the presence of Raynaud's phenomenon, anti-ribonucleoprotein and anti-Sm antibodies [76]. In-depth whole-genome sequencing is the most comprehensive approach for measuring rare variation in both coding and noncoding regions. However, its application for large-scale cohorts is still limited. A tentative alternative would be to combine genotype imputation with targeted sequencing in a gene-centered strategy as a first glimpse followed by directed sequencing of the resulting best hits and a final corroboration of the association with SLE through an aggregated effect analysis. However, the reliability of the imputation of rare variation is still under discussion [72].

As in some of the first candidate gene and GWAS studies, modern rare and de-novo variant-based articles mostly analyze coding regions searching for missense and nonsense variants [60■■,69■■,76]. Several articles addressed the impact of the regulatory role of noncoding mutations showing association with extreme gene expression [77–80]. Overall, it was demonstrated that rare variants contribute to large gene expression changes across tissues providing an integrative method for interpretation of rare variants effects [81].

The heterogeneity of SLE loci identified in rare and de-novo variant studies could reside in the diversity of pathways, involved in SLE pathogenesis, without ruling out other additional explanations such as local population differences, different experimental designs and computational approaches.

In summary, rare and de-novo variant analysis has shown its applicability for SLE genetics research and has uncovered several candidate SLE loci for further studies.

CONCLUSION

During the last decades of SLE susceptibility studies, we came from preliminary candidate genes to genome-wide scans of thousands of SLE cases and controls from different populations. Initially, EUR biased research was spread to Asian, Amerindian, and African ancestry, supporting the universal role of many SLE loci. Currently, genetic studies have focused on the identification of causal variants, their mechanisms of action, and the involvement of rare variation (see Fig. 1 and Table 1). Further progress of the field is based on multiomic studies including data on gene expression, epigenetics, gene-gene interaction studies, and the analysis of rare and de-novo variants and copy number variations. Rare variants are already identified as important components of the genetic context of SLE and future studies can clarify their functional impact and the combined role of common and rare variant in the individual SLE risk.

Acknowledgements

None.

Financial support and sponsorship

Funding for the preparation of this manuscript has been received from the grant of the Spanish Ministry of Economy and Competitiveness (MINECO) SAF2016-78631-P to M.E.A.R. and from the Fundación Pública Andaluza Progreso y Salud.

Conflicts of interest

There are no conflicts of interest.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:
■ of special interest
■■ of outstanding interest

1. Crampton SP, Morawski PA, Bolland S. Linking susceptibility genes and pathogenesis mechanisms using mouse models of systemic lupus erythematosus. Dis Model Mech 2014; 7:1033–1046.
2. Leffers HC, Lange T, Collins C, *et al.* The study of interactions between genome and exposome in the development of systemic lupus erythematosus. Autoimmun Rev 2019; 18:382–392.
3. Alarcon-Segovia D, Alarcon-Riquelme ME. Etiopathogenesis of systemic lupus erythematosus: the itineraries of three troikas. In: Lahita RG, editor. Systemic Lupus Erythematosus, Fourth Edition. Academic Press; 2004. pp. 93–107.
4. Delgado-Vega A, Sanchez E, Lofgren S, *et al.* Recent findings on genetics of systemic autoimmune diseases. Curr Opin Immunol 2011; 22:698–705.
5. Alarcon-Segovia D, Alarcon-Riquelme ME, Cardiel MH, *et al.* Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. Arthritis Rheum 2005; 52:1138–1147.
6. Deng Y, Tsao BP. Advances in lupus genetics and epigenetics. Curr Opin Rheumatol 2014; 26:482–492.
7. Morris DL, Sheng Y, Zhang Y, *et al.* Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. Nat Genet 2016; 48:940–946.
8. Molineros JE, Yang W, Zhou XJ, *et al.* Confirmation of five novel susceptibility loci for systemic lupus erythematosus (SLE) and integrated network analysis of 82 SLE susceptibility loci. Hum Mol Genet 2017; 26:1205–1216.
9. Langefeld CD, Ainsworth HC, Cunninghame Graham DS, *et al.* Transancestral mapping and genetic load in systemic lupus erythematosus. Nat Commun 2017; 8:16021.
10. Lawrence JS, Martins CL, Drake GL. A family survey of lupus erythematosus. 1. Heritability. J Rheumatol 1987; 14:913–921.
11. Kuo CF, Grainge MJ, Valdes AM, *et al.* Familial aggregation of systemic lupus erythematosus and coaggregation of autoimmune diseases in affected families. JAMA Intern Med 2015; 175:1518–1526.
12. Harley JB, Alarcon-Riquelme ME, Criswell LA, *et al.* Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet 2008; 40:204–210.
13. Kozyrev SV, Abelson AK, Wojcik J, *et al.* Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nat Genet 2008; 40:211–216.
14. Hom G, Graham RR, Modrek B, *et al.* Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med 2008; 358:900–909.
15. International HapMap 3 Consortium. Altshuler DM, Gibbs RA, *et al.* Integrating common and rare genetic variation in diverse human populations. Nature 2010; 467:52–58.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

16. 1000 Genomes Project Consortium. Abecasis GR, Auton A, *et al.* An integrated map of genetic variation from 1,092 human genomes. Nature 2012; 491:56–65.
17. Iglesias AI, van der Lee SJ, Bonnemaijer PW, Höhn R, *et al.* Haplotype reference consortium panel: practical implications of imputations with large reference panels. Hum Mutat 2017; 38:1025–1032.
18. Georg I, Díaz-Barreiro A, Morell M, *et al.* BANK1 interacts with TRAF6 and MyD88 in innate immune signaling in B cells. Cell Mol Immunol 2019.
19. Bentham J, Morris DL, Cunninghame Graham DS, *et al.* Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat Genet 2015; 47:1457–1464.
20. Jeong DY, Lee SW, Park YH, *et al.* Genetic variation and systemic lupus erythematosus: a field synopsis and systematic meta-analysis. Autoimmun Rev 2018; 17:553–566.
21. Julià A, López-Longo FJ, Pérez Venegas JJ, *et al.* Genome-wide association study meta-analysis identifies five new loci for systemic lupus erythematosus. Arthritis Res Ther 2018; 20:100.
22. Almlöf JC, Alexsson A, Imgenberg-Kreuz J, *et al.* Novel risk genes for systemic lupus erythematosus predicted by random forest classification. Sci Rep 2017; 7:6236.
23. Morris DL, Taylor KE, Fernando MM, *et al.* Unraveling multiple MHC gene associations with systemic lupus erythematosus: model choice indicates a role for HLA alleles and non-HLA genes in Europeans. Am J Hum Genet 2012; 91:778–793.
24. Louthrenoo W, Kasitanon N, Wichainun R, *et al.* The genetic contribution of HLA-DRB5*01:01 to systemic lupus erythematosus in Thailand. Int J Immunogenet 2013; 40:126–130.
25. Fernando MM, Stevens CR, Sabeti PC, *et al.* Identification of Two Independent Risk Factors for Lupus within the MHC in United Kingdom Families. PLoS Genet 2007; 3:e192.
26. Raj P, Rai E, Song R, *et al.* Regulatory polymorphisms modulate the expression of HLA class II molecules and promote autoimmunity. Elife 2016; 5; pii: e12089. doi: 10.7554/eLife.12089.
27. Alarcon-Riquelme ME, Ziegler JT, Molineros J, *et al.* GWAS in an Amerindian ancestry population reveals novel systemic lupus erythematosus risk loci and the role of European admixture. Arthritis Rheumatol 2016; 68:932–943.
28. Hanscombe KB, Morris DL, Noble JA, *et al.* Genetic fine mapping of systemic lupus erythematosus MHC associations in Europeans and African Americans. Hum Mol Genet 2018; 27:3813–3824.
29. Kim K, Bang SY, Lee HS, Okada Y, *et al.* The HLA-DRbeta1 amino acid positions 11-13-26 explain the majority of SLE-MHC associations. Nat Commun 2014; 5:5902.
30. Sun C, Molineros JE, Looger LL, *et al.* High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. Nat Genet 2016; 48:323–330.
31. Molineros JE, Looger LL, Kim K, Okada Y, *et al.* Amino acid signatures of HLA Class-I and II molecules are strongly associated with SLE susceptibility and autoantibody production in Eastern Asians. PLoS Genet 2019; 15:e1008092.
32. Yang W, Ng P, Zhao M, *et al.* Population differences in SLE susceptibility genes: STAT4 and BLK, but not PXK, are associated with systemic lupus erythematosus in Hong Kong Chinese. Genes Immun 2009; 10:219–226.
33. Cen H, Leng RX, Wang W, *et al.* Association study of IFIH1 rs1990760 polymorphism with systemic lupus erythematosus in a Chinese population. Inflammation 2013; 36:444–448.
34. Chang YK, Yang W, Zhao M, *et al.* Association of BANK1 and TNFSF4 with systemic lupus erythematosus in Hong Kong Chinese. Genes Immun 2009; 10:414–420.
35. Han JW, Zheng HF, Cui Y, *et al.* Genome-wide association study in Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genet 2009; 41:1234–1237.
36. Qin L, Lv J, Zhou X, *et al.* Association of IRF5 gene polymorphisms and lupus nephritis in a Chinese population. Nephrology 2010; 15:710–713.
37. Siu HO, Yang W, Lau CS, *et al.* Association of a haplotype of IRF5 gene with systemic lupus erythematosus in Chinese. J Rheumatol 2008; 35:360–361.
38. Zhang Z, Zhu KJ, Xu Q, *et al.* The association of the BLK gene with SLE was replicated in Chinese Han. Arch Dermatol Res 2010; 302:619–624.
39. ▪ Zhang YM, Zhou XJ, Nath SK, *et al.* Evaluation of 10 SLE susceptibility loci in Asian populations, which were initially identified in European populations. Sci Rep 2017; 7:41399.

Update on loci identified in Asians.

40. Akizuki S, Ishigaki K, Kochi Y, *et al.* PLD4 is a genetic determinant to systemic lupus erythematosus and involved in murine autoimmune phenotypes. Ann Rheum Dis 2019; 78:509–518.
41. Wen L, Zhu C, Zhu Z, *et al.* Exome-wide association study identifies four novel loci for systemic lupus erythematosus in Han Chinese population. Ann Rheum Dis 2018; 77:417.
42. Wang YF, Zhang Y, Zhu Z, *et al.* Identification of ST3AGL4, MFHAS1, CSNK2A2 and CD226 as loci associated with systemic lupus erythematosus (SLE) and evaluation of SLE genetics in drug repositioning. Ann Rheum Dis 2018; 77:1078–1084.
43. Ghodke-Puranik Y, Imgruet M, Dorschner JM, *et al.* Novel genetic associations with interferon in systemic lupus erythematosus identified by replication and fine-mapping of trait-stratified genome-wide screen. Cytokine 2019. doi: 10.1016/j.cyto.2018.12.014.
44. Lessard CJ, Sajuthi S, Zhao J, *et al.* Identification of a systemic lupus erythematosus risk locus spanning ATG16L2, FCHSD2, and P2RY2 in Koreans. Arthritis Rheumatol 2016; 68:1197–1209.
45. Yang SK, Hong M, Zhao W, *et al.* Genome-wide association study of Crohn's disease in Koreans revealed three new susceptibility loci and common attributes of genetic susceptibility across ethnic populations. Gut 2014; 63:80–87.
46. Stranger BE, Forrest MS, Clark AG, *et al.* Genome-wide associations of gene expression variation in humans. PLoS Genet 2005; 1:e78.
47. Dixon AL, Liang L, Moffatt MF, *et al.* A genome-wide association study of global gene expression. Nat Genet 2007; 39:1202–1207.
48. Dimas AS, Deutsch S, Stranger BE, *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 2009; 325:1246–1250.
49. Yang TP, Beazley C, Montgomery SB, *et al.* Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics 2010; 26:2474–2476.
50. Nicolae DL, Gamazon E, Zhang W, *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet 2010; 6:e1000888.
51. Patel ZH, Lu X, Miller D, *et al.* A plausibly causal functional lupus-associated risk variant in the STAT1-STAT4 locus. Hum Mol Genet 2018; 27:2392–2404.
52. Flores R, Zhang P, Wu W, *et al.* Siglec genes confer resistance to systemic lupus erythematosus in humans and mice. Cell Mol Immunol 2019; 16:154–164.
53. Westra HJ, Peters MJ, Esko T, *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 2013; 45:1238–1243.
54. Martínez-Bueno M, Oparina N, Dozmorov MG, *et al.* Trans-ethnic mapping of BANK1 identifies two independent SLE-risk linkage groups enriched for co-transcriptional splicing marks. Int J Mol Sci 2018; 19; pii: E2331.
55. Sun W, Hu Y. eQTL mapping using RNA-seq data. Stat Biosci 2013; 5:198–219.
56. ▪▪ Odhams CA, Cunninghame Graham DS, Vyse TJ. Profiling RNA-Seq at multiple resolutions markedly increases the number of causal eQTLs in autoimmune disease. PLoS Genet 2017; 13:e1007071.

Recent work on the identification of several eQTLs.

57. ▪▪ Odhams CA, Cortini A, Chen L, *et al.* Mapping eQTLs with RNA-seq reveals novel susceptibility genes, noncoding RNAs and alternative-splicing events in systemic lupus erythematosus. Hum Mol Genet 2017; 26:1003–1017.

European eQTLs and other important genomic events altered by SLE eQTLs.

58. Delgado-Vega AM, Dozmorov MG, Quiros MB, *et al.* Fine mapping and conditional analysis identify a new mutation in the autoimmunity susceptibility gene BLK that leads to reduced half-life of the BLK protein. Ann Rheum Dis 2012; 71:1219–1226.
59. Jacob CO, Eisenstein M, Dinauer MC, *et al.* Lupus-associated causal mutation in neutrophil cytosolic factor 2 (NCF2) brings unique insights to the structure and function of NADPH oxidase. Proc Nat Acad Sci USA 2012; 109:E59–67.
60. ▪▪ Steri M, Orrù V, Idda ML, *et al.* Overexpression of the cytokine BAFF and autoimmunity risk. N Engl J Med 2017; 376:1615–1626.

An insertion-deletion found in the gene coding for the autoreactive B-cell cytokine BAFF.

61. ▪ Almlöf JC, Nystedt S, Leonard D, *et al.* Whole-genome sequencing identifies complex contributions to genetic risk by variants in genes causing monogenic systemic lupus erythematosus. Hum Genet 2019; 138:141–150.

The identification of new of low-frequency variants causal for monogenic lupus.

62. Ueki M, Fujihara J, Kimura-Kataoka K, *et al.* Low genetic heterogeneity of copy number variations (CNVs) in the genes encoding the human deoxyribonucleases 1-like 3 and II potentially relevant to autoimmunity. PLoS One 2019; 14:e0215479.
63. Alperin JM, Ortiz-Fernández L, Sawalha AH. Monogenic lupus: a developing paradigm of disease. Front Immunol 2018; 9:2496.
64. Costa-Reis P, Sullivan KE. Monogenic lupus: it's all new! Curr Opin Immunol 2017; 49:87–95.
65. Teruel M, Alarcón-Riquelme ME. The genetic basis of systemic lupus erythematosus: what are the risk factors and what have we learned. J Autoimmun 2016; 74:161–175.
66. Manolio TA, Collins FS, Cox NJ, *et al.* Finding the missing heritability of complex diseases. Nature 2009; 461:747–753.
67. ▪▪ Pullabhatla V, Roberts AL, Lewis MJ, *et al.* De novo mutations implicate novel genes in systemic lupus erythematosus. Hum Mol Genet 2018; 27:421–429.

The first study of its kind in SLE that identified novel de novo mutations with NGS.

68. Almlöf JC, Nystedt S, Leonard D, *et al.* Whole-genome sequencing identifies complex contributions to genetic risk by variants in genes causing monogenic systemic lupus erythematosus. Hum Genet 2019; 138:141–150.

1040-8711 Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

69. ■■ Delgado-Vega AM, Martinez-Bueno M, Oparina NY, *et al.* Whole exome sequencing of patients from multicase families with systemic lupus erythematosus identifies multiple rare variants. Sci Rep 2018; 8:8775.
A first study on multicase families and exome sequencing.
70. Ege M, Ma Y, Manfras B, *et al.* Omenn syndrome due to ARTEMIS mutations. Blood 2005; 105:4179–4186.
71. Gul E, Sayar EH, Gungor B, *et al.* Type I IFN-related NETosis in ataxia telangiectasia and Artemis deficiency. J Allergy Clin Immunol 2018; 142:246–257.
72. Martínez-Bueno M, Alarcón-Riquelme ME. Exploring impact of rare variation in systemic lupus erythematosus by a genome wide imputation approach. Front Immunol 2019; 10:258.
73. Nicolae DL. Association tests for rare variants. Annu Rev Genomics Hum Genet 2016; 17:117–130.
74. Jiang SH, Athanasopoulos V, Ellyard JI, *et al.* Functional rare and low frequency variants in BLK and BANK1 contribute to human lupus. Nat Commun 2019; 10:2201.
75. Roberts AL, Thomas ER, Bhosle S, *et al.* Resequencing the susceptibility gene, ITGAM, identifies two functionally deleterious rare variants in systemic lupus erythematosus cases. Arthritis Res Ther 2014; 16:R114.
76. Farias FH, Dahlqvist J, Kozyrev SV, *et al.* A rare regulatory variant in the MEF2D gene affects gene regulation and splicing and is associated with a SLE sub-phenotype in Swedish cohorts. Eur J Hum Genet 2019; 27:432–441.
77. Montgomery SB, Lappalainen T, Gutierrez-Arcelus M, Dermitzakis ET. Rare and common regulatory variation in population-scale sequenced human genomes. PLoS Genet 2011; 7:e1002144.
78. Zhao J, Akinsanmi I, Arafat D, *et al.* A burden of rare variants associated with extremes of gene expression in human peripheral blood. Am J Hum Genet 2016; 98:299–309.
79. Zeng Y, Wang G, Yang E, *et al.* Aberrant gene expression in humans. PLoS Genet 2015; 11:e1004942.
80. Li X, Battle A, Karczewski KJ, *et al.* Transcriptome sequencing of a large human family identifies the impact of rare noncoding variants. Am J Hum Genet 2014; 95:245–256.
81. Li X, Kim Y, Tsang EK, *et al.* The impact of rare variation on gene expression across tissues. Nature 2017; 550:239–243.
82. Gu Z, Gu L, Eils R, *et al.* Circlize implements and enhances circular visualization in R. Bioinformatics 2014; 30:2811–2812.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.