Seminars in Immunopathology
https://doi.org/10.1007/s00281-021-00900-w

REVIEW

Recent advances in understanding the genetic basis of systemic lupus erythematosus

Eunji Ha[1,2] · **Sang-Cheol Bae**[3,4] · **Kwangwoo Kim**[1,2]

Received: 24 June 2021 / Accepted: 14 October 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Systemic lupus erythematosus (SLE) is a polygenic chronic autoimmune disease leading to multiple organ damage. A large heritability of up to 66% is estimated in SLE, with roughly 180 reported susceptibility loci that have been identified mostly by genome-wide association studies (GWASs) and account for approximately 30% of genetic heritability. A vast majority of risk variants reside in non-coding regions, which makes it quite challenging to interpret their functional implications in the SLE-affected immune system, suggesting the importance of understanding cell type–specific epigenetic regulation around SLE GWAS variants. The latest genetic studies have been highly fruitful as several dozens of SLE loci were newly discovered in the last few years and many loci have come to be understood in systemic approaches integrating GWAS signals with other biological resources. In this review, we summarize SLE-associated genetic variants in both the major histocompatibility complex (MHC) and non-MHC loci, examining polygenetic risk scores for SLE and their associations with clinical features. Finally, variant-driven pathogenetic functions underlying genetic associations are described, coupled with discussion about challenges and future directions in genetic studies on SLE.

Keywords Systemic lupus erythematosus · Genetics · Genome-wide association study · Genetic variant · Polygenic risk score

Introduction

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease, which damages multiple tissues and organs, resulting from the production of autoantibodies to nuclear antigens [1]. The clinical manifestations of SLE are highly heterogeneous, including cutaneous, musculoskeletal, renal, hematologic, neurologic, and other diverse symptoms [1]. There have been continuous efforts to diagnose and characterize patients with SLE based on diverse diagnostic criteria since 1971 [2–4]. The 1997 revised American College of Rheumatology (ACR) and the 2012 Systemic Lupus International Collaborating Clinics (SLICC) criteria [2, 3] have broadly been used for the classification of SLE patients. These criteria for SLE are highly useful but showed suboptimal performance in terms of sensitivity and specificity [3, 5]. The European League Against Rheumatism (EULAR)/ACR criteria for SLE were newly developed for better classification of SLE in 2019, with improved performance with a sensitivity of 96.1% and a specificity of 93.4% [4].

The prevalence and incidence rates of SLE vary widely in the literature depending on ethnicity, geographic differences, and sex [6–8]. The highest prevalence and incidence rates recorded were in Afro-Caribbean people living in the UK (517.5 per 100,000 people; 31.5 per 100,000 person-years), while SLE patients in certain countries have been rarely observed (e.g., 3.2 per 100,000 individuals in India;

This article is a contribution to the special issue on: Genetics and functional genetics of Autoimmune diseases - Guest Editors: Yukinori Okada & Kazuhiko Yamamoto

✉ Sang-Cheol Bae
scbae@hanyang.ac.kr

✉ Kwangwoo Kim
kkim@khu.ac.kr

[1] Department of Biology, Kyung Hee University, Seoul, Republic of Korea

[2] Department of Biomedical and Pharmaceutical Sciences, Kyung Hee University, Seoul, Republic of Korea

[3] Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Republic of Korea

[4] Hanyang University Institute for Rheumatology Research, Seoul, Republic of Korea

Published online: 03 November 2021

0.3 per 100,000 person-years in Ukraine) [6]. These variations may result from various aspects involved in case identification, data collection, the structure and policy of health-care systems, socioeconomic inequalities and so on [6]. Despite such variations in the reported prevalence and incidence rates of SLE, it seems apparent that the incidence and prevalence rates in people of African, Asian, and Aboriginal origins are higher than those in people of white ancestry [6–8]. About 90% of patients with SLE are women of reproductive ages [6, 7, 9].

SLE develops in genetically susceptible individuals exposed to environmental, sex-related, or endogenous triggers [10, 11]. SLE-risk factors induce in concert abnormalities of the immune system, which generally include (1) the hyperactivation of innate immunity by cellular and external nucleic acids, (2) increased sensitivity to nucleic acids and the production of anti-nuclear autoantibodies in adaptive immunity, (3) reduced induction of regulatory T-cells, and (4) ineffective clearance of immune complexes and apoptotic cells [12]. Significant evidence of environmental risk factors has been documented with respect to exposures, such as cigarette smoking, ultraviolet radiation, Epstein–Barr virus (EBV) infection, and silica dust, which have contributed to both an onset of the disease and lupus flares [10, 11, 13]. Considering that the majority of SLE patients are women, it is likely that sex plays an important role in SLE pathogenesis, possibly mediated by the altered expression of genes escaping from X chromosome inactivation and hormonal effects [14, 15].

Disease severity can be influenced by genetic factors, environmental exposures, and socioeconomic status, including the genetic burden of SLE risk, ethnicity, sex, onset age, income level, education, health insurance, social support system, and treatment compliance rates [6, 16, 17]. As compared with in white populations, severity and mortality rates are higher in African-American populations [6]. Severe phenotypes such as lupus nephritis tend to be more frequent among male patients and patients who experience a childhood onset of their disease; these individuals have been known to show a greater SLE genetic burden, relative to that in female patients and adult-onset patients, respectively [16, 17]. Poverty, inadequacy of education, lack of health insurance, poor social support, and poor medication compliance rates are all associated with disease outcomes, in conjunction with the influence of ancestry or not [6].

To date, nearly 180 genomic loci have been identified as associated with SLE susceptibility in genetic studies in multiple ancestries [18–52] (Fig. 1A), accounting for up to 30% of liability in SLE patients [35, 36, 50]. Such genetic findings can be used to estimate a degree of genetic risk for SLE and to provide more effective drug targets [35, 36, 42, 53–59] (Fig. 1D). Despite remarkable advances in genetic studies on SLE in the past few decades, we learned that the genetic variance explained by the identified SLE variants is still far less than the known genetic heritability (h^2) of SLE [60, 61], which we refer to as the missing heritability. In addition, the vast majority of identified risk variants are present in the non-coding regions, which makes it difficult to interpret their functions and disease-relevant genes in SLE susceptibility loci, and suggest the importance of the allele-specific regulatory effects of disease genes. Recent genetic studies leverage diverse cell type–specific epigenetic resources and other biological resources to draw better pictures of the disease pathogenesis, thus updating the genetic architecture of SLE [62–70]. This review attempts to provide the most updated catalog of SLE-associated variants and focuses on the recent advances in integrative genetic studies, with discussions of current challenges and prospects.

Early findings in genetic association studies in SLE

The estimated heritability of SLE ranges from 44 to 66% in family studies [60, 61]. High sibling risk ratios ($8 < \lambda_s < 29$) and high concordance rates between monozygotic twins (20–40%) relative to dizygotic ones and non-twin full siblings (2–5%) were observed in family-based cohort analyses [71, 72], suggesting a strong contribution of genetic factors to SLE and the importance of conducting genetic studies on SLE.

Before the development of SNP-based genome-wide association technology, nine genes were known to be causal of SLE family-based approaches or candidate gene studies [73]. All these genes have crucial roles in immune-related functions [73] and include human leukocyte antigen (*HLA*), *C2*, *C4*, *C1q*, *FCGR2A*, *FCGR3A*, *PDCD1*, *PTPN22*, and *IRF5*. For example, the deficiency of complement component genes, including *C2*, *C4*, and *C1q*, are also deeply involved in the inefficient recognition of immune complexes, reduced clearance of cell debris, and prolonged immune stimulation [74, 75]. Most of the known SLE-risk variants within these genes have low frequencies and large effect sizes on the risk of SLE [73].

Genetic associations of the major histocompatibility complex (MHC) region

Genetic variants within the MHC region at the short-arm band 21.3 of chromosome 6 have shown strong associations with SLE in multiple ethnic groups [73, 76]. The human MHC region is highly polymorphic, containing the highest-density genetic variants, like SNPs, indels, and copy number variations, under extremely extensive

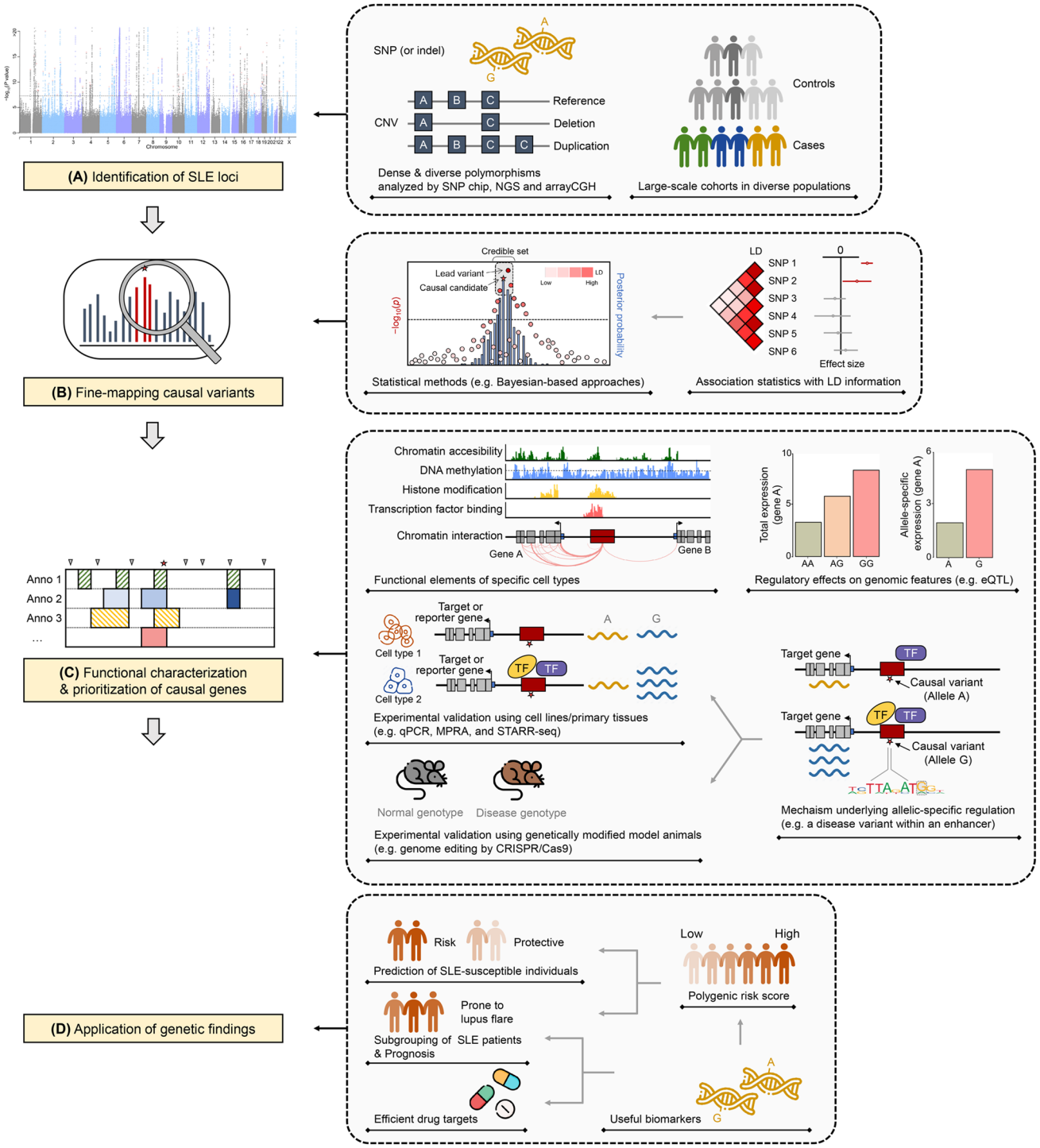

Fig. 1 Overview of approaches in identification of SLE loci, prioritization of causal variants/genes, and clinical application. *CNV* copy number variation, *NGS* next generation sequencing, *LD* linkage disequilibrium, *eQTL* expression quantitative trait loci, *Anno* annotation, *TF* transcription factor, *qPCR* quantitative polymerase chain reaction, *MPRA* massively parallel reporter assay, *STARR-seq* self-transcribing active regulatory region sequencing, *CRISPR* clustered regularly interspaced short palindromic repeats, *Cas9* CRISPR-associated protein 9

linkage disequilibrium (LD). There are more than 120 MHC-located genes including MHC class I genes (*HLA-A, -B, -C, -E, -F,* and *-G*), class II genes (*HLA-DP, -DM, -DO, -DQ,* and *-DR*), and class III genes (complement components and others), most of which encode key members in both innate and adaptive immune responses [77, 78]. The most significant SLE association in the MHC region has been shown around and within the *HLA-DRB1* gene encoding an HLA-DR β-chain of HLA-DR protein that would play an important role in determining the immune tolerance to self-antigens [11].

Extremely dense variants within HLA genes in the MHC region in high LD construct long-range haplotypes (so-called classical alleles) that produce qualitatively and quantitatively distinct HLA molecules. Haplotype-based approaches have discovered several *HLA-DRB1* classical alleles, including HLA-DRB1*03:01 and HLA-DRB1*15:01, involved in associations with SLE susceptibility in multiple ancestries [34, 36, 38, 79–85], with a high degree of allelic heterogeneity among different ethnicities in terms of allelic frequencies and statistical significance [11]. For instance, HLA-DRB1*03:01 is the most SLE-risk allele in European ancestries [82], but not common enough to detect its association in East Asian SLE cohorts [34, 36, 38, 83]. An amino acid–level fine-mapping analysis of *HLA-DRB1* using HLA imputation identified that the amino acid positions 11, 13, and 26 at the epitope-binding pocket of HLA-DRB1 provide a better HLA–SLE association model, explaining the previous association results at a classical allele level in diverse populations [84]. A recent follow-up study in six East Asian cohorts replicated the association of the haplotype of amino acid residues at positions 11–13–37 (or 11–13–26) [85].

Upon conditioning on the *HLA-DRB1* associations, other HLA genes (*HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-A*, and *HLA-B*) and non-HLA genes (*C4*, *MICB*, *NOTCH4*, *TNXB*, and *SLC44A4*) have been suggested to have secondary independent association signals contributing to susceptibility to SLE [31, 38, 82, 85, 86], implying the complex SLE association architecture of the MHC region. From a stepwise conditional regression analysis of residues of HLA molecules in East Asian populations, the following six amino acid positions were independently associated with SLE: *DRB1*-13, *DRB1*-11, *DRB1*-26 or *DRB1*-37, *A*-70, *DPB1*-35, *DQB1*-37, and *B*-9 (presented in order of their significance) [85]. SLE-risk residues at these positions were mainly large, hydrophobic, and negatively charged, which might facilitate interaction with many positively charged SLE autoantigens with T-cell receptors in SLE patients [85]. As expected, some SLE-risk amino acid positions in *HLA-DRB1* and *HLA-DPB1* were associated with the positivity of autoantibodies to nRNP, Ro/La, ACL, or Sm [85].

With respect to non-HLA MHC genes, the copy numbers of *C4A* and *C4B* genes imputed from SNP data showed shared effects on SLE-risk in large-size European and African American cohorts, when adjusting for the HLA-DRB1*03:01 allele [86]. A conditional analysis revealed that HLA-DRB1*03:01 is no longer significant after controlling for the *C4* alleles, taking advantage of an African cohort where the HLA-DRB1*03:01 and *C4* alleles were in a very low LD [86]. Lower copy numbers of *C4A* and *C4B* exhibited the strongest SLE risk in the human genome, explaining sex-biased vulnerability in SLE [86]. The secondary associations independent of the *C4* alleles localized rs2105898 near *HLA-DRB1*, a known eQTL for multiple neighboring genes involved in SLE pathogenesis [86]. However, it needs to further investigate *C4*-conditioned *HLA-DRB1* association signals in an amino acid or multi-allelic association model for *HLA-DRB1* in SLE because there remain strong residual association signals within the MHC region independent of the *C4* alleles, the SLE-risk eQTL is in a high LD with HLA-DRB1*15:01, and the association of multi-allelic *HLA-DRB1* should be evaluated at a gene level to understand disease-associated effects of all *HLA-DRB1* classical alleles in a single association model [86].

Updates of non-HLA susceptibility loci for SLE from recent genome-wide association studies (GWASs)

High-throughput, cost-effective, genome-wide genotyping technologies and well-defined landscapes of genetic variants and LD in the human genome have enabled researchers to analyze associations in population-scale cohorts, shifting the research focus toward common variants with small effect sizes on the risk of common complex diseases like SLE [1]. Since 2007, there have been remarkable achievements made in genetic studies on SLE [18–52], bringing the number of non-HLA SLE susceptibility loci to 179. The association summary statistics of the lead variants in SLE loci are provided in Table 1. The reported SLE loci have modest effect sizes and explain about 30% and, at most, 24% of total phenotypic variance in European and East Asian studies, respectively [35, 36, 50].

Among the 179 non-HLA loci associated with SLE, almost half have been reported by recent genetic studies performed mostly in East Asian populations since 2018 [40–52], implying an importance of GWASs involving understudied non-European ancestries with relatively high prevalence and severity. The latest international collaboration effort was described in the largest-ever SLE genetic

Table 1 List of 179 non-HLA loci associated with SLE

Variant	Chr	Pos	EA	Reported gene	OR	PMID	Pop	Type
rs12093154	1	1,243,545	A	*C1QTNF12*	0.84	33536424	EAS + EUR	Protein-altering
rs3795310	1	8,371,547	T	*RERE*	0.88	33536424	EAS + EUR	Non-coding
rs28411034	1	37,811,325	A	*MTF1*	0.86	33536424	EAS + EUR	Synonymous
rs6702599	1	67,359,716	C	*IL12RB2*	0.84	33536424	EAS + EUR;EUR;EUR + AFR + AMR	Non-coding
rs2476601	1	113,834,946	A	*AL137856.1, PHTF1, PTPN22, RSBN1*	1.43	26502338	EAS + EUR;EUR	Protein-altering
rs9651076	1	116,500,680	A	*CD58, NAP1L4P1*	1.12	33272962	EAS;EAS + EUR	Non-coding
rs116785379	1	157,138,367	C	*ETV3*	1.21	33272962	EAS	Non-coding
rs11264750	1	157,527,370	G	*FCRL5*	0.75	33536424	EAS;EAS + EUR	Protein-altering
rs76107698	1	161,600,039	C	*AL590385.2, FCGR2A, FCGR2C*	0.79	33272962	EAS;EAS + EUR;EUR;EUR + AFR + AMR	Non-coding
rs2205960	1	173,222,336	T	*AL645568.1, AL645568.2, AL645568.3, LOC100506023, TNFSF4*	1.37	33272962	AMR;EAS;EAS + EUR;EAS + SAS;EUR;EUR + AFR + AMR	Non-coding
rs549669428	1	174,925,885	G	*RABGAP1L*	0.84	33536424	EAS + EUR	Non-coding
rs13306575	1	183,563,302	A	*NCF2, NMNAT2, SMG7*	1.31	33272962	AMR;EAS;EAS + EUR;EUR;EUR + AFR + AMR	Protein-altering
rs10911628	1	184,680,369	A	*AL713852.1, EDEM3*	1.95	24871463	EUR	Non-coding
rs1547624	1	192,574,707	T	*AL390957.1*	1.17	33536424	EAS	Non-coding
rs4143303	1	198,701,340	A	*AL157402.1, PTPRC*	0.88	33272962	EAS;EAS + EUR	Non-coding
rs3806357	1	202,010,327	A	*ELF3*	1.11	33272962	EAS	Non-coding
rs4844538	1	206,469,377	A	*AL591846.1, IKBKE, IL10, IL19, SRGAP2*	1.11	33272962	EAS;EAS + EUR;EUR;EUR + AFR + AMR	Non-coding
rs9782955	1	235,876,577	C	*LYST*	1.16	26502338	EUR	Non-coding
rs1780813	1	246,280,780	T	*SMYD3*	1.82	29848360	EUR	Non-coding
rs75362385	2	7,432,948	T	*LOC100506274*	0.89	33272962	EAS	Non-coding
rs7579944	2	30,222,160	T	*H3P5, LBH*	0.88	33272962	EAS;EAS + EUR;EUR + AFR + AMR	Non-coding
rs13385731	2	33,476,823	T	*RASGRP3*	1.29	33272962	EAS	Non-coding
rs1432296	2	60,841,032	A	*LINC01185*	1.18	28714469	EUR + AFR + AMR	Non-coding
rs11126034	2	65,353,087	T	*SPRED2*	1.12	33272962	EAS;EUR	Non-coding
rs10207954	2	73,989,388	A	*AS1, DGUOK, TET3*	1.15	33272962	EAS;EAS + EUR;EAS + SAS	Non-coding
rs73954925	2	111,119,597	C	*BCL2L11*	1.17	33272962	EAS	Non-coding
rs218174	2	135,900,775	A	*DARS, LCT*	1.12	33272962	EAS	Synonymous
rs2381401	2	143,263,405	T	*ARHGAP15*	1.15	33536424	EAS + EUR	Non-coding
rs11679244	2	162,225,885	A	*FAP, IFIH1*	1.12	33272962	EAS;EUR;EUR + AFR + AMR	Non-coding
rs9630991	2	190,567,413	A	*AC108047.1*	0.85	33536424	EAS + EUR;EUR	Non-coding
rs11889341	2	191,079,016	T	*STAT4*	1.41	33272962	AMR;EAS;EAS + EUR;EAS + SAS;EUR;EUR + AFR + AMR;SAS	Non-coding
rs7572733	2	198,065,082	T	*PLCL1*	1.14	33272962	EAS	Protein-altering
rs3087243	2	203,874,196	A	*CTLA4, ICOS*	0.89	33536424	EAS + EUR	Non-coding
rs7565158	2	212,729,246	T	*AC093865.1, ERBB4, IKZF2*	1.10	33272962	EAS;EAS + EUR;EUR	Non-coding
rs438613	3	28,030,595	T	*CMC1, LINC01967*	0.92	33272962	EAS;EAS + EUR	Non-coding
rs9852465	3	58,479,456	G	*AC098479.1, AC116036.2, PDHB, PXK*	1.10	28714469	EUR;EUR + AFR + AMR	Non-coding
rs7637844	3	72,176,765	A	*LINC00870*	0.88	33272962	EAS	Non-coding
rs144104218	3	119,518,879	A	*CD80, TIMMDC1, TMEM39A*	0.83	33272962	EAS;EAS + SAS;EUR;EUR + AFR + AMR	Protein-altering
rs564976	3	160,011,272	C	*AS1, IL12A*	1.14	26502338	EUR	Non-coding
rs1317082	3	169,779,797	A	*LRRC34, MYNN*	1.10	33272962	EAS	Protein-altering
rs6762714	3	188,752,450	T	*LPP*	1.16	27399966	EAS + EUR	Non-coding
rs13101828	4	971,932	A	*DGKQ*	0.91	33272962	EAS;EUR;EUR + AFR + AMR	Synonymous
rs231694	4	2,699,117	T	*FAM193A, TNIP2*	1.11	33272962	EAS;EAS + EUR	Non-coding
rs13116227	4	8,556,539	T	*AC105345.1, GPR78*	1.34	29494758	EAS	Non-coding
rs113284964	4	40,305,570	G	*LINC02265*	1.13	33272962	EAS	Non-coding

Table 1 (continued)

Variant	Chr	Pos	EA	Reported gene	OR	PMID	Pop	Type
rs2855772	4	54,682,309	C	*KIT*	1.40	29494758	EAS	Non-coding
rs6533951	4	78,723,125	A	*LINC01094*	1.11	33272962	EAS	Non-coding
rs6841907	4	83,225,843	T	*COQ2*	0.91	33272962	EAS	Non-coding
rs116940334	4	87,023,100	T	*AFF1*	0.83	33272962	EAS	Non-coding
rs4643809	4	101,834,942	T	*BANK1*	0.85	33272962	EAS;EUR;EUR + AFR + AMR	Synonymous
rs58107865	4	108,140,462	C	*LEF1*	0.80	33272962	EAS	Non-coding
rs11724582	4	122,470,309	A	*IL2, IL21*	1.14	28714469	EUR	Non-coding
rs10018951	4	183,688,220	T	*TRAPPC11*	1.31	29494758	EAS	Non-coding
rs7725218	5	1,282,299	A	*TERT*	1.13	33272962	EAS	Non-coding
rs6871748	5	35,885,880	C	*AC112204.3, IL7R*	0.89	33536424	EAS + EUR	Protein-altering
rs2544920	5	100,805,670	A	*RN7SKP62, ST8SIA4*	1.12	33272962	EAS;EAS + EUR;EUR;EUR + AFR + AMR	Non-coding
rs74989671	5	128,398,268	G	*FBN2*	1.54	32771030	SAS	Non-coding
rs370449198	5	131,784,646	A	*FNIP1*	0.72	33272962	EAS	Non-coding
rs2549002	5	132,493,886	A	*IRF1*	0.91	33272962	EAS	Synonymous
rs6874758	5	134,093,501	C	*AC008608.1, TCF7*	1.24	33272962	EAS;EAS + EUR;EUR	Non-coding
rs10036748	5	151,078,585	T	*TNIP1*	1.19	33272962	AMR;EAS;EAS + EUR;EUR;EUR + AFR + AMR	Non-coding
rs2421184	5	159,459,931	A	*LINC01845*	1.11	33272962	EAS	Non-coding
rs2431697	5	160,452,971	T	*MIR3142, MIR3142HG*	1.24	33272962	EAS;EAS + EUR;EUR;EUR + AFR + AMR	Non-coding
rs9503037	6	243,302	A	*AL035696.1, AL365272.1, LOC285766*	0.88	33272962	EAS	Non-coding
rs17603856	6	16,630,667	T	*ATXN1*	1.14	27399966	EAS + EUR	Non-coding
rs35789010	6	25,513,951	A	*CARMIL1*	1.46	28714469	EUR	Synonymous
rs36014129	6	25,884,291	A	*H2AC3P, H2BP5*	1.50	28714469	EUR	Synonymous
rs10946940	6	27,592,808	A	*471P, CD83P1, RNU6*	1.45	24871463	EUR	Non-coding
rs6457796	6	34,860,776	T	*ANKS1A, PPARD, UHRF1BP1*	0.81	33272962	EAS;EAS + EUR;EAS + SAS;EUR;EUR + AFR + AMR	Protein-altering
rs34868004	6	36,747,254	CA	*CPNE5*	1.10	33272962	EAS	Non-coding
rs597325	6	90,292,775	A	*BACH2*	0.91	33272962	EAS;EAS + EUR	Non-coding
rs548234	6	106,120,159	T	*ATG5, PRDM1*	0.82	33272962	EAS;EAS + EUR;EUR;EUR + AFR + AMR	Non-coding
rs9488914	6	116,369,686	T	*DSE*	0.86	33272962	EAS	Non-coding
rs148314165	6	137,908,901	G	*AL356234.2, AL591468.1, BTF3L4P3, LINC02528, TNFAIP3*	1.71	33272962	EAS;EAS + EUR;EAS + SAS;EUR;EUR + AFR + AMR	Protein-altering
rs9322454	6	154,249,517	A	*IPCEF1*	1.09	33272962	EAS	Non-coding
rs702814	7	28,133,113	A	*JAZF1*	1.14	26502338	EUR	Non-coding
rs4598207	7	50,218,883	A	*AC020743.2, AC020743.3, C7orf72, IKZF1*	1.33	33272962	EAS;EUR;EUR + AFR + AMR	Non-coding
rs13238909	7	67,611,386	A	*ST3AGL4*	0.85	29625966	EAS	Non-coding
rs150518861	7	74,152,347	A	*EIF4H, LIMK1*	1.66	29848360	EUR	Non-coding
rs117026326	7	74,711,703	T	*AC211433.2, AC211433.3, GTF2IRD1, LOC101926943*	2.14	33272962	EAS;EAS + EUR;EUR + AFR + AMR;SAS	Non-coding
rs77009341	7	75,559,377	C	*HIP1*	2.01	29724251	EAS	Non-coding
rs3757387	7	128,936,032	T	*AC011005.2, AC018639.1, AC025594.1, AC025594.2, IRF5, TNPO3*	0.69	33272962	AMR;EAS;EAS + EUR;EAS + SAS;EUR;EUR + AFR + AMR	Non-coding
rs2955587	8	8,240,557	G	*ALG1L13P, FAM86B3P, PRAG1*	1.11	28714469	EUR;EUR + AFR + AMR	Non-coding
rs2428	8	8,783,635	T	*MFHAS1*	1.13	29625966	EAS	Non-coding

Table 1 (continued)

Variant	Chr	Pos	EA	Reported gene	OR	PMID	Pop	Type
rs7819602	8	10,869,332	C	*AC011008.2, XKR6*	1.15	28714469	EUR	Non-coding
rs2736332	8	11,482,456	C	*AF131216.5, BLK*	1.36	33272962	EAS;EAS + EUR;EAS + SAS;EUR;EUR + AFR + AMR;SAS	Non-coding
rs2272736	8	42,319,645	A	*IKBKB, PLAT*	0.82	33272962	EAS;EUR + AFR + AMR	Protein-altering
rs2953898	8	56,068,244	C	*RPS20*	1.19	28714469	EUR + AFR + AMR	Synonymous
rs142937720	8	70,417,931	A	*NCOA2*	0.89	33272962	EAS	Non-coding
rs17374162	8	71,982,724	A	*AS1, MSC*	0.92	33272962	EAS	Non-coding
rs4739134	8	78,643,913	T	*AC068700.2*	1.12	28714469	EUR + AFR + AMR	Non-coding
rs2445610	8	127,184,843	G	*CASC19, PCAT1*	0.89	33536424	EAS	Non-coding
rs16902895	8	128,413,347	A	*LINC00824*	1.12	33272962	EAS;EAS + EUR	Non-coding
rs1887428	9	4,984,530	C	*JAK2*	0.92	33272962	EAS;EAS + EUR	Non-coding
rs7858766	9	21,267,088	T	*IFNA22P*	1.14	33272962	EAS	Synonymous
rs1405209	9	99,823,263	C	*AL162394.1, AL359710.1, AS1, NR4A3, STX17*	1.11	33536424	EAS + EUR;EUR + AFR + AMR	Synonymous
rs77448389	10	5,868,783	A	*ANKRD16*	0.86	33272962	EAS	Non-coding
rs7097397	10	48,817,351	A	*AC060234.3, LRRC18, PCDH15, WDFY4*	0.81	33272962	EAS;EAS + EUR;EAS + SAS;EUR;EUR + AFR + AMR	Protein-altering
rs7902146	10	62,041,271	T	*ARID5B*	0.90	33272962	EAS;EAS + SAS;EUR	Non-coding
rs10995261	10	62,651,528	T	*AC024598.1, AC067752.1, ZNF365*	0.91	33272962	EAS;EAS + EUR	Non-coding
rs10823829	10	71,706,952	T	*CDH23*	0.91	33272962	EAS;EAS + EUR	Synonymous
rs4917385	10	103,243,964	T	*RPEL1, ST13P13*	0.72	26606652	AMR	Non-coding
rs111447985	10	103,918,153	A	*STN1*	1.17	33272962	EAS	Non-coding
rs58164562	10	110,904,356	T	*BBIP1*	0.89	33272962	EAS	Non-coding
rs1131665	11	613,208	T	*CDHR5, IRF7, MIR210HG, PHRF1*	1.19	28714469	EAS + EUR;EUR;EUR + AFR + AMR	Protein-altering
rs3750996	11	4,091,970	A	*STIM1*	1.17	33272962	EAS	Non-coding
rs77885959	11	18,340,835	T	*GTF2H1*	1.69	33272962	EAS	Protein-altering
rs2785198	11	35,071,482	A	*AL356215.1, LOC100507144, PDHX*	1.18	33272962	EAS;EAS + EUR;EUR;EUR + AFR + AMR	Non-coding
rs10896045	11	65,788,053	A	*AP5B1, OVOL1*	1.17	33272962	EAS;EAS + EUR	Non-coding
rs4930642	11	69,048,902	A	*TPCN2*	1.15	33272962	EAS	Non-coding
rs3794060	11	71,476,633	C	*NADSYN1*	1.23	26502338	EUR	Protein-altering
rs77971648	11	72,929,435	T	*AP002761.2, FCHSD2*	1.29	33272962	EAS	Protein-altering
rs377392985	11	118,780,114	CAAAAAAAAA	*AP002954.1, DDX6*	1.16	33272962	EAS	Non-coding
rs9736939	11	128,435,976	A	*AP001122.1, ETS1, LINC02098*	1.27	33272962	EAS;EAS + SAS;EUR;EUR + AFR + AMR	Non-coding
rs2540119	12	4,031,710	T	*PARP11*	1.09	33272962	EAS	Non-coding
rs4251697	12	12,721,528	A	*AC008115.2, CDKN1B, CREBL2, GPR19, GPR19/CDKN1B*	0.64	33272962	EAS;EAS + SAS	Non-coding
rs4622329	12	101,928,157	A	*DRAM1*	1.12	33272962	EAS;EAS + SAS	Non-coding
rs6539078	12	103,522,302	T	*AC084364.3, AC084364.4, LOC105369945*	0.89	33272962	EAS;EAS + EUR	Non-coding
rs77465633	12	111,495,741	A	*ATXN2*	1.34	33272962	EAS;EUR;EUR + AFR + AMR	Non-coding
rs3999421	12	120,930,715	A	*CABP1, XLOC_009911*	0.91	33272962	EAS;EAS + EUR	Non-coding
rs11059928	12	128,811,558	A	*SLC15A4*	0.82	33272962	EAS;EUR;EUR + AFR + AMR	Protein-altering
rs200521476	12	132,463,596	G	*FBRSL1*	0.88	33272962	EAS	Non-coding
rs57141708	13	41,001,255	A	*ELF1*	1.18	33272962	EAS;EAS + SAS	Non-coding
rs76725306	13	49,603,317	A	*AL135901.1, RCBTB1*	1.16	33536424	EAS + EUR	Non-coding
rs1885889	13	99,439,046	G	*AL136961.1, TM9SF2*	0.87	33536424	EAS;EAS + EUR	Non-coding
rs911263	14	68,286,876	C	*RAD51B*	0.89	28714469	EUR;EUR + AFR + AMR	Non-coding
rs11845506	14	87,916,691	C	*GALC*	5.00	28714469	AFR	Non-coding

Table 1 (continued)

Variant	Chr	Pos	EA	Reported gene	OR	PMID	Pop	Type
rs12148050	14	102,797,451	G	*TRAF3*	0.91	33536424	EAS + EUR	Non-coding
rs2819426	14	104,945,922	C	*AHNAK2, AHNAK2/PLD4*	0.82	33272962	EAS	Protein-altering
rs7170151	15	38,554,477	T	*FAM98B, RASGRP1*	1.11	33272962	EAS;EUR	Non-coding
rs11553760	15	74,798,906	T	*CSK, SCAMP5*	1.11	33272962	EAS;EAS + EUR;EUR	Synonymous
rs869310	15	77,537,964	G	*AC046168.1, AC046168.2*	0.88	33536424	EAS + EUR	Non-coding
rs8023715	15	97,064,451	A	*LINC02253, RN7SKP181*	1.81	24871463	EUR	Non-coding
rs35985016	15	100,988,807	A	*LRRK1*	0.84	33272962	EAS	Protein-altering
rs34361002	16	11,096,177	T	*CIITA, CLEC16A*	1.14	33272962	EAS;EUR;EUR + AFR + AMR	Synonymous
rs79401250	16	23,860,136	T	*PRKCB*	1.17	33272962	EAS	Non-coding
rs534645300	16	30,802,134	A	*AC093249.1, PRR14, ZNF629*	0.81	33272962	EAS	Non-coding
rs34572943	16	31,261,032	A	*ITGAM, ITGAX*	1.68	28714469	AMR;EAS + EUR;EUR;EUR + AFR + AMR	Protein-altering
rs11288784	16	50,055,296	G	*HEATR3*	0.90	33272962	EAS	Protein-altering
rs669763	16	57,356,566	C	*AC108081.1, CCL22*	1.12	33272962	EAS;EUR + AFR + AMR	Non-coding
rs2731783	16	58,219,556	A	*CSNK2A2*	1.12	29625966	EAS	Non-coding
rs28410471	16	68,520,852	A	*36P, RNU4, ZFP90*	1.13	33272962	EAS;EAS + EUR;EUR + AFR + AMR	Non-coding
rs11376510	16	79,711,775	G	*MAFTRR*	0.90	33272962	EAS	Non-coding
rs11117432	16	85,985,665	A	*AC092723.3, AC092723.4, AC092723.5, IRF8*	0.73	33272962	EAS;EAS + EUR;EUR;EUR + AFR + AMR	Non-coding
rs933717	16	87,381,644	T	*MAP1LC3B*	0.13	29044928	EAS	Non-coding
rs2286672	17	4,809,322	T	*PLD2*	1.25	26502338	EUR	Protein-altering
rs61759532	17	7,337,072	T	*AC026954.1, ACAP1*	1.24	33272962	EAS;EUR	Non-coding
rs35966917	17	16,936,587	A	*TNFRSF13B*	0.91	33272962	EAS	Non-coding
rs4252665	17	39,729,130	A	*ERBB2, IKZF3, MIEN1*	1.46	28714469	EUR;EUR + AFR + AMR	Protein-altering
rs114038709	17	45,379,362	T	*AC003070.2, ARHGAP27*	1.16	29848360	EUR	Non-coding
rs2671655	17	49,390,658	T	*LOC102724596*	1.09	33272962	EAS	Non-coding
rs8072449	17	75,316,103	A	*AC011933.4, GRB2, SLC25A19*	1.19	28714469	EUR;EUR + AFR + AMR	Non-coding
rs113417153	17	78,377,098	T	*PGS1*	0.89	33272962	EAS	Non-coding
rs1788097	18	69,876,452	T	*CD226*	1.10	33272962	EAS	Protein-altering
rs118075465	18	79,626,912	A	*LOC284241*	1.14	33272962	EAS	Non-coding
rs2238577	19	948,532	T	*ARID3A*	0.89	33272962	EAS	Non-coding
rs4807205	19	2,167,879	G	*DOT1L*	1.12	33493351	EAS + EUR	Non-coding
rs5826945	19	6,697,077	A	*C3*	0.84	33272962	EAS	Non-coding
rs55882956	19	10,359,243	A	*TYK2*	0.67	33272962	EAS;EAS + EUR;EUR;EUR + AFR + AMR	Protein-altering
rs2362475	19	16,329,024	A	*AC020917.3, KLF2*	0.85	33493351	EAS	Non-coding
rs11673604	19	18,430,178	T	*IQCN, JUND, LRRC25, SSBP4*	1.14	33272962	EAS;EAS + EUR;EUR + AFR + AMR	Synonymous
rs12461589	19	32,581,862	T	*ANKRD27, PDCD5*	0.90	33272962	EAS;EAS + EUR	Non-coding
rs33974425	19	49,348,489	CCAGCTGCAT	*SLC6A16, TEAD2*	1.12	33272962	EAS;EAS + EUR	Protein-altering
rs7251	19	49,659,652	C	*IRF3*	0.88	32,719,713	EAS	Protein-altering
rs4801882	19	51,623,800	A	*SIGLEC5*	0.88	33272962	EAS	Protein-altering
rs10419308	19	55,228,445	A	*AC010327.5, AC010327.6, TMEM86B*	0.84	33536424	EAS + EUR;EUR + AFR + AMR	Synonymous
rs6074813	20	1,561,106	T	*AL049634.2*	1.12	33536424	EAS + EUR	Protein-altering
rs4810485	20	46,119,308	A	*CD40*	1.43	28714469	EUR + AFR + AMR	Non-coding
rs11697848	20	49,958,778	T	*147P, KRT18P4, RNU6*	2.12	24871463	EUR	Non-coding
rs4819670	22	18,166,589	T	*USP18*	1.15	33272962	EAS	Protein-altering

Table 1 (continued)

Variant	Chr	Pos	EA	Reported gene	OR	PMID	Pop	Type
rs4821116	22	21,619,030	T	*CCDC116, UBE2L3, YDJC*	1.24	33272962	EAS;EAS + EUR;EUR;EUR + AFR + AMR	Protein-altering
rs9611155	22	39,343,182	T	*SYNGR1*	1.14	33272962	EAS	Non-coding
rs137956	22	39,897,459	C	*ENTHD1, GRAP2*	1.14	28714469	EUR + AFR + AMR	Non-coding
rs6641111	X	12,821,671	C	*PRPS2*	1.19	33272962	EAS	Non-coding
rs887369	X	30,559,729	C	*CXorf21*	1.15	26502338	EUR	Synonymous
rs13440883	X	53,072,295	C	*GPR173*	1.16	29724251	EAS	Non-coding
rs5914012	X	56,882,269	T	*NBDY*	1.10	33272962	EAS	Synonymous
rs143181706	X	150,504,983	T	*MAMLD1*	1.50	30679154	EAS	Non-coding
rs1059702	X	154,018,741	A	*IRAK1, MECP2, TMEM187*	1.36	33272962	AMR;EAS;EUR	Protein-altering

All information on the SLE variants was retrieved from GWAS catalog (https://www.ebi.ac.uk/gwas/) and recent association studies on SLE [18–52] when surpassing the genome-wide significance threshold of $p < 5 \times 10^{-8}$. SLE loci were defined after merging significant SLE variants within 300 kb around each variant. The most significant variant in the largest cohort study for each locus is provided in the table. We excluded variants in the extended major histocompatibility complex (MHC) region on 28–34 Mb in chromosome 6 in the human genome assembly hg38. A locus containing any protein-altering LD proxies ($r^2 > 0.8$) of a lead variant in any reported populations is defined as protein-altering. A locus containing only synonymous or non-coding LD proxies ($r^2 > 0.8$) in any reported populations is defined as synonymous. A locus without any LD proxies ($r^2 > 0.8$) in coding sequences in any reported populations is defined as non-coding. *Chr* chromosome, *Pos* chromosomal position (hg38), *EA* effect allele, *OR* odds ratio estimated in the genetic study with the largest sample size, *PMID* PubMed ID of the largest study, *Pop* populations where significant associations were detected in all previous studies (Populations analyzed in different studies were separated by semicolons. Populations meta-analyzed in a single study were denoted by plus signs.), *Type* annotation of each locus

association study of East Asian populations (n = 208,370) consisting of Korean, Chinese, and Japanese participants that led to the identification of dozens of novel SLE loci (up to 46 loci) [50]. Another East Asian study published 1 month later was also very successful, discovering more than 30 SLE susceptibility loci [52], many of which overlapped with those detected in the aforementioned East Asian study [50].

Genetic association analyses followed by functional annotation and statistical analyses for gene prioritization suggested plenty of genes that potentially play pathogenic roles in aberrant immunity and cellular processes in SLE. Genes involved in the positive regulation of the type I IFN (IFN1) pathway (e.g., *STAT4*, *IRF3*, *IRF5*, *IRAK1*, and *TNFA1P3*) have been reported as plausible causal genes for SLE-risk [20, 21, 23, 33, 47]. For instance, genetic analysis pinpointed *IRF3* to be most likely causal because the SLE-associated variant rs7251 was annotated as having both protein-altering and expression regulatory effects on *IRF3* [47]. In addition, the same variant was associated with lupus nephritis, indicating that *IRF3* may play a key role in the development of SLE and its manifestations possibly by upregulating the IFN1 pathway [47].

Similarly, other genetic elements in lymphocyte signaling (e.g., *PTPN22*, *BLK*, *BANK1*, and *LRRK1*) were also suggested to be SLE-driving genes [18, 20, 23, 50]. Among them, *LRRK1*, encoding a multiple-domain leucine-rich repeat kinase, contributes to the pathogenesis of SLE by deteriorating the function of B-cells and modulating the B-cell receptor–mediated NF-kB signaling pathway [50, 87]. By association fine-mapping of SLE-associated variants based on a Bayesian statistical method (Fig. 1B), Yin et al. successfully prioritized the missense variant (rs35985016) in *LRRK1* with a remarkably highly posterior probability to be causal in the locus [50].

Genes involved in clearing apoptotic cells and immune complexes (e.g., *FCGR2A*, *ITGAM*, and *NCF1*) have been well confirmed concerning their genetic associations with SLE [19, 31, 39]. Of particular noteworthiness is the missense variant rs201802880 in *NCF1* that was identified as one of the largest-effect SLE-risk variants through an immune-loci genotyping array (known as Immunochip) in multi-ethnic populations (odds ratio: 2.0–3.8) [39]. The SLE-risk allele of rs201802880 in *NCF1* resulted in the reduced production of reactive oxygen species (ROS) [39, 88]. Despite the great size of its genetic effect on SLE, the association of the *NCF1* locus was discovered somewhat recently because most previous GWAS arrays did not cover the region with a genetically complex segmental duplication [89]. Moreover, a gene deletion in *NCF1* was reported to increase the risk of SLE in East Asian and European subjects, but more than three copies of the gene showed a protective effect against developing SLE in various populations, consistently supporting the role of lowered ROS production in the pathogenesis of SLE [39].

An expression enrichment analysis of the genes within SLE-associated loci highlighted lymphoid immune cells where SLE-locus genes were expressed significantly more [36, 90]. In addition, significant enrichments were observed in non-immune tissues including musculoskeletal, digestive,

respiratory, and stomatognathic tissues. The genetic evidence of the involvement of non-immune tissues implies various manifestations of SLE in multiple organs [90]. For gene sets, immune-related pathways mediated by cytokines, Toll-like receptors, and B- and T-cell receptors showed the greatest enrichment of SLE-locus genes [36, 52, 90].

Polygenic risk score (PRS) and clinical application

As the reported GWAS variants confer small to moderate increases in the risk of SLE, PRS for individuals might be an informative way to estimate the individual-level genetic burden in translational research and clinical application [91, 92] (Fig. 1D). PRS is usually calculated as a sum of the actual numbers of risk alleles weighted by reported log odds ratios of the corresponding risk alleles [91, 92].

A few studies have assessed PRS for SLE to investigate the association between cumulative PRS and disease manifestations or disease severity [35, 55–59]. An early study on the relationship between PRS and SLE phenotypes demonstrated that several traits, including autoantibody production and age at diagnosis, were associated with high PRS in a Caucasian population [56]. Consistently, significantly higher PRS was observed in childhood-onset compared with adult-onset SLE patients in Korean and multi-ancestry cohorts [57, 59]. In addition, patients with early SLE onset are prone to showing more severe symptoms, such as proteinuria, malar rash, anti-double-stranded DNA antibody, hemolytic anemia, arthritis, and leucopenia regardless of their ethnicity, sex, or disease duration [55]. Recent large-cohort PRS studies reconfirmed that an individual with a high PRS for SLE appeared more likely to have severe SLE phenotypes involving increased anti–double-stranded DNA, and higher prevalence of organ damage, including end-stage renal disease (ESRD) and proliferative nephritis [58, 59]. A survival analysis showed that overall mortality was elevated with increasing PRS [58]. In addition, the mean survival until the first organ damage, cardiovascular event, and ESRD onset was decreased in the patients with higher PRS [58].

Despite shared susceptibility loci across populations, an estimated effect size of a single lead GWAS variant could be heterogeneous because the GWAS variant is not necessarily a causal variant and may differently correlate with actual casual variants in different populations [93]. Such inconsistent effects of GWAS variants can generate biased PRS estimates in various ethnicities, returning the suboptimal predictive power of PRS. Therefore, pinpointing the true shared signals is important for improving the trans-ancestry portability of PRS.

A recent study suggested a method of leveraging cell type–specific regulatory elements to prioritize shared functional variants [94]. Amariuta et al. could prioritize the most likely causal variants based on their per-variant heritability and localization to cell type–specific transcription factor–binding motifs in both European and East Asian populations [94]. PRS from the prioritized variants on functional annotations was better performed in a cross-validation analysis in the trans-ancestry PRS model (trained in a European cohort, tested in an East Asian one) as compared with PRS from unprioritized, lead variants [94]. Besides capturing shared causal variants, the predictive power of PRS might be enhanced by the dissection of PRS rather than summing up all risk variants. In a Swedish cohort, the potential advantage of using pathway-based PRS was demonstrated to stratify patients with SLE [95].

Functional implications of non-coding SLE variants

Recent association studies have provided substantial updates of the genetic architecture of SLE with insights into mechanisms underlying the development of SLE. However, functional and pathological interpretations from genetics associations are challenging by the given nature of SLE variants' locations. Only a minor portion of total risk loci (34/179 loci; 19.0%) contains at least one protein-altering variant genetically correlated with lead variants ($r^2 > 0.8$; Fig. 2). Most of the SLE loci explain the disease association using only non-coding variants.

Changes in the expression level of causal genes by disease-causal non-coding variants are supposed to be observed in SLE-relevant cell types if an analysis is performed in an extremely large (infinite-size) cohort. However, relatively small sample sizes in most transcriptomics studies may not be statistically powerful enough to identify a weak to moderate regulatory effect of a disease variant, especially the variant that indirectly regulates gene expression by inducing complex epigenetic changes [96]. Indeed, evidence does not have exist for many non-coding SLE variants about direct correlations with neighboring gene expression levels [97, 98], suggesting indirect (or mediated), hidden regulatory functions mediated by various expression regulators. For example, a recent study employing a massive parallel reporter assay systemically examined the regulatory effects of 3073 GWAS variants in 91 SLE loci; the flanking sequences around 482 variants showed enhancing activity in a B-cell line and 51 variants in only 27 risk loci (e.g., rs3101018 in *C4A*) led to differential expression according to allele dosage [67].

Moreover, it has been reported that such regulations by genetic variants on the gene expression occur frequently in the distal regions of promoters [62, 65]. According to Su et al., 8.5% of SLE-SNPs in open chromatin regions

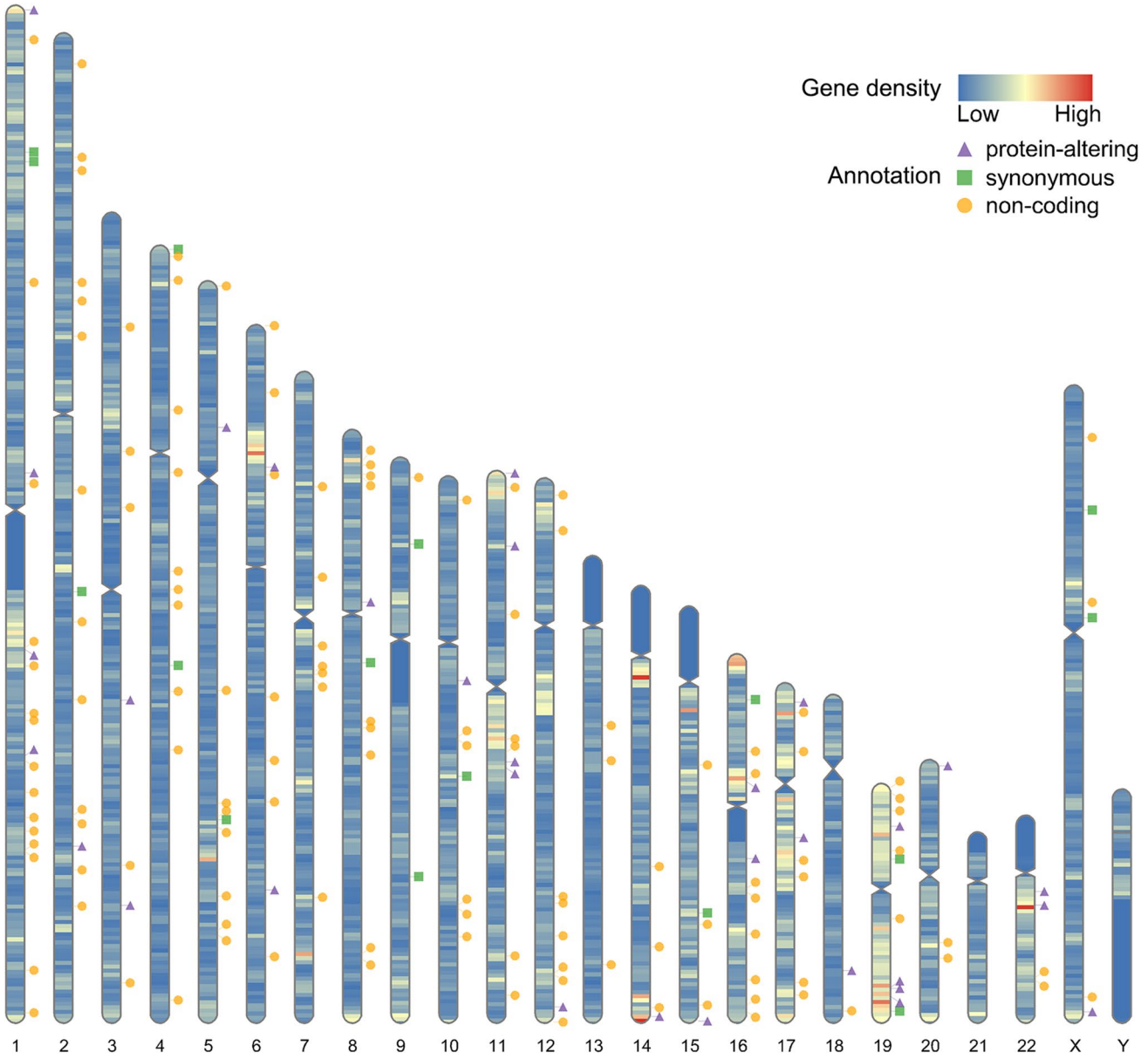

Fig. 2 Genomic distribution and functional annotation of SLE association signals. Non-HLA loci listed in Table 1 are shown in the human chromosome ideogram with the information on functional annotations of SLE-associated variants. The gene density is visualized by color gradation using the R package *RIdeogram*. A locus containing any protein-altering LD proxies ($r^2 > 0.8$) of a lead variant in any reported populations is marked as *protein-altering*. A locus containing only synonymous or non-coding LD proxies ($r^2 > 0.8$) in any reported populations is marked as *synonymous*. A locus without any LD proxies ($r^2 > 0.8$) in coding sequences in any reported populations is marked as *non-coding*

interacted only with the nearest genes, and more than 60% of accessible SLE-SNPs connected with distant genes instead of the nearest one [65]. Similarly, Chandra et al. reported that promoter-interacting eQTLs distant from their targets genes were more frequent than promoter eQTLs in transcription-activating sites (H3K27ac marks) of immune cells, even leading to cell-type specificity in expression [62]. To dissect the molecular and cellular consequences stratified by disease variants within regulatory elements, up-to-date genetic studies on SLE have used various approaches, retrieving other biological resources including gene expression, DNA methylation, histone modifications, chromatin accessibility, and microRNA [62–64, 90] (Fig. 1C).

By comprehensive genetic profiling using epigenomic analysis, three-dimensional chromatin structure analysis, and genome-editing perturbation, a non-coding variant, rs2431697, on chromosome 5 was recently reported as likely causal for SLE [63]. This study identified that the variant

located 15 kb upstream of the miR-146a gene overlaps with a $CD14^+$ monocyte-specific epigenetic feature, including chromatin immunoprecipitation followed by sequencing (ChIP-seq) peaks for H3K4me1 and H3K27ac and assay for transposase-accessible chromatin using sequencing (ATAC-seq) peaks [63]. The enhancing activity of the region around the SLE-risk variant was validated in genome-editing experiments and transcription-activating/-inhibiting dCAS9 systems. Indeed, the variant-located enhancer formed cognate loops with an miR-146a promoter, and the SLE-risk rs2431697 allele down-regulated the expression of miR-146a by differentially modulating the regional chromatin state and NF-kB binding affinity to attenuate the activation of the IFN1 pathway in SLE patients [63].

The importance of the IFN1 pathway, already long-established in SLE pathogenesis, was re-emphasized in another recent DNA methylation quantitative trait loci (meQTL) study [64]. DNA methylation profile analysis in 548 SLE patients and 587 controls identified that methylation levels at IFN1-signature genes were significantly decreased in SLE, showing the largest difference between cases and controls [64]. Moreover, meQTLs for such differentially methylated elements were enriched within SLE GWAS loci, including *PTPRC* (CD45), *MHC-class III*, *UHRF1BP1*, *IRF5*, *IRF7*, *IKZF3*, and *UBE2L3* [64]. For instance, a non-coding variant rs7444 in the 3'UTR of *UBE2L3* was identified to have a meQTL effect on DNA methylation in the *UBE2L3* promoter that may mediate the altered expression of the gene indirectly in an allele-specific manner [64]. Similar findings were reported by $CD4^+$ T-cell inter-omics study in rheumatoid arthritis, which shares a large portion of risk alleles with other autoimmune diseases such as SLE [99]; Ha et al. observed a considerably high enrichment of SNP-based heritability of rheumatoid arthritis on the methylated regions correlated with rheumatoid arthritis–specific gene expression levels in $CD4^+$ T-cells, which suggests that disease variants may shape the rheumatoid arthritis–specific transcriptomic features by the mediation of allele-specific DNA methylation [99].

In the same context, integration of omics data with genome-wide association statistics appears also to be helpful in prioritizing effector genes at SLE loci when multiple genes are closely located around the SLE-associated variant. Recently, high-resolution mapping for SLE variant accessibility and gene connectivity was accomplished by a promoter-focused capture-C analysis in follicular helper T-cells (T_{FH}), which play a crucial role in the production of anti-nuclear antibodies [65]. For example, rs527619, a proxy SNP of an SLE variant in *AP002954.1*, interacted exclusively with the promoter of *CXCR5* instead of other genes, such as *BCL9L*, in the same locus [65]. It is also possible that a single functional variant regulates multiple genes. The risk allele of rs34330 on *CDKN1B* modulated the expression of multiple neighboring genes (including *CDKN1B*, *APOLD1*, and *DDX47*) by influencing the binding of histone marks, RNA pol II, and the key immune regulator IRF-1 [100]. The rs34330-deleted cell lines presented the elevated level of proliferation derived by cell type–specific regulation of *CDKN1B* and nearby genes, implying an impact of the locus on cell cycle progression [100].

Remarkable occupation of EBV-encoding EBNA2 protein at SLE loci

EBV infection is a strong SLE-risk factor implicated in the epidemiology of the disease, increasing the prevalence of childhood SLE by as much as 50-fold [101–104]. However, pathophysiological mechanisms underlying the interaction between EBV infection and host genetic factors have remained unclear. Surprisingly, ChIP-seq in B-cells, the target cell types of EBV infection, revealed that 26 out of 52 European SLE loci contained the binding sites of both an EBV-encoding protein EBNA2 and many human TFs [66]. In particular, the sequences intersecting EBNA2 ChIP-seq peaks were largely occupied by NF-κB components such as RELA, RELB, NFKB1, and NFKB2 [66], constituting super-enhancers able to proliferate and activate EBV-infected B-cells [105]. Allele-specific differential binding of the EBNA2-mediated protein complexes by SLE-associated genetic variants was validated through ChIP-seq and quantitative polymerase chain reaction (qPCR) analyses in EBV-infected B-cells, which resulted in allele-specific expressions of nearby genes involving *IKZF2*, *CLEC16A*, *BLK*, *MIR3142* and *HLA-DQB1* [66]. Consistently, an East Asian SLE GWAS revealed a significant enrichment of EBNA2 binding sites in 17 loci among 46 newly identified loci [90], confirming a cellular role of EBV infection in B-cells in SLE pathogenesis.

Challenges and future directions of SLE genetics

Unraveling the complex etiology of SLE with highly heterogenous manifestations is the ultimate goal of population-based SLE genetics, aiming eventually to develop better strategies for the identification of SLE-susceptible individuals and the clinical application for various precision medicine topics, especially considering effective therapeutic drugs, prognosis, and lupus flare (Fig. 1). Although geneticists have undertaken enormous efforts to increase the sample sizes in GWASs [18–52], it is skeptical that GWASs will eventually explain the entire heritability in SLE, considering the imperfect coverages of GWAS arrays; potential non-additive effects; and the insufficient statistical powers,

especially for small effect sizes of risk alleles with low to rare frequencies [93]. A Bayesian inference analysis predicted the existence of additional hundreds of risk variants with small effect sizes in other polygenic autoimmune diseases [106]. Indeed, we are observing that recent large-scale GWASs continue to identify new SLE variants mostly with common frequencies and only small effect sizes (risk-allele odds ratio < 1.2) around the genome-wide significance threshold, with an almost negligible addition to the reported SNP-based heritability in SLE [42, 43, 50–52].

However, the efforts in GWAS must be recognized and continued to identify more SLE variants. SLE-risk variants explain many clinically relevant disease pathways, drug targets, and cell types in actual human patients with SLE. A small increase in GWAS variants can dramatically increase the statistical power to understand the disease's biology. For instance, a gene-set enrichment analysis using the Reactome pathway in a recent East Asian GWAS (n = 208,370) could provide better enrichment results for known SLE pathways including IFN1 signatures than those of exactly the same gene-set enrichment analysis in a recent trans-ancestral GWAS (n = 35,369), even identifying novel SLE-related pathways related to interleukins, type II IFN signatures, TRAF6-mediated IRF7 activation, and so on [52, 90].

Open international collaboration networks deploying secured interoperable analysis platforms are much needed to maximize the sample sizes with existing data or association summary statistics. It is also important to analyze understudied populations where SLE-risk alleles may be more common, possibly enough to be detectable. In addition, there might be opportunities to apply new statistical methods for genetic association testing to better control false-negative and false-positive findings, accounting for potential confounding factors. For example, two recent methods, GWAX and LT-FH, preserve valuable family histories by merging unaffected individuals with a family history of disease into disease cases and estimating disease liability as a continuous value based on both the case–control status and the configuration of family history, increasing statistical power [107, 108].

Current SNP genotyping for genetic association studies has been performed primarily by SNP arrays due to relatively low cost, high accuracy, and high genome coverage [93], and the decreasing sequencing cost of next-generation sequencing technologies will allow researchers to investigate entire variations in the human genome including SNP, indel, and copy number variations in genetically complex loci, with an almost perfect genome coverage even in the recombination hotspots.

Future GWAS will be much more strongly required to integrate with multiple omics data for dissecting heterogeneity of immune-related cells and unraveling their regulatory functions. As valuable biological resources and new technologies accumulate, we will have better opportunities to explore the intricate nature of SLE, although there are many technical challenges in the integration of different data types. For instance, one of the latest large-scale eQTL studies illustrated dynamics in eQTL effects in the context of both cell types and immunological conditions, which could be a useful resource to understand immunogenetic mechanisms in SLE [109]. Similarly, a few single-cell transcriptomic analyses in SLE have been performed to dissect signature cells involved in SLE inflammation [110, 111]. A single cell–based analysis combining with epigenetic and phenotypic data was able to identify two distinct subpopulations of low-density granulocytes, which are correlated with several clinical features such as renal function and proteinuria, confirming previous results as a pathogenic neutrophil subset [110]. IFN-stimulated genes were over-expressed in SLE, especially with high disease activity, by an expansion of IFN-expressing immune cell subpopulations [111], supporting a genetic association of IFN1-related loci. Further genetic studies with single-cell transcriptomics technology will be conducted to evaluate the effects of GWAS variants in each type of cells in the relevant organs including immune and nonimmune organs.

Many immune-mediated diseases share many risk loci with the same directional effect, exhibiting high genetic correlations [112–115]. A recent study by Peyrot et al. devised a new method called case–case GWAS (CC-GWAS) to test differences in allele frequencies and measure a genetic distance between similar diseases using summary statistics [116]. These authors illustrated the advantage of CC-GWAS using GWAS results in several psychiatric disorders, identifying 72 novel case–case loci [116]. A cross-disease association meta-analysis and heterogeneity analysis among autoimmune diseases could provide new insights regarding disease-shared and disease-specific biology in SLE.

There is growing evidence of dysbiosis in autoimmunity [117–119] and significant tissue-specific expressions of SLE GWAS genes in gastrointestinal tissues [90]. Human gut microbiota may play a role in the onset stage of SLE, contributing to inflammation, hyperactivity of gut-associated lymphoid tissues, abnormal T-cell differentiation, and the loss of self-tolerance. To understand how microbiota contribute to the etiology of SLE, intestinal microbial profiles, such as microbial diversity, disease-specific taxa, and microbial metabolites should be examined in SLE. There have been a few attempts to characterize the microbial composition in SLE patients. The lower richness of the gut microbiome has been consistently observed in patients with SLE [119–121]. An increased abundance of genera in *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, and *Proteobacteria* was reported in SLE, whereas counts of several organisms belonging to genera in *Firmicutes* were significantly depleted [120, 121]. However, these alterations

of microbiota in autoimmune patients could be a disease-driving source, a mediator, or just a secondary outcome of a disease. Further investigation in SLE would be required to evaluate the interaction between host genetic factors and the gut microbiome.

In recent years, we have witnessed considerable progress in the genetics of SLE, including in the areas of identifying numerous susceptibility loci, fine-mapping causal signals, and elucidating the functions of disease variants in biological systems. Even more active and well-designed genetic works with the integration of various omics data in disease-driving tissues are underway and will drive the next breakthroughs in SLE genetics, enhancing our understanding of the pathogenetic mechanisms of the disease and improving our ability to apply precision medicine strategies for patients with SLE.

Funding This research was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1A6A1A03038899).

Availability of data and material Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Firestein GS, Gabriel SE, McInnes IB et al (2017) Kelley and Firestein's textbook of rheumatology
2. Hochberg MC (1997) Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 40(9):1725. https://doi.org/10.1002/art.1780400928
3. Petri M, Orbai AM, Alarcon GS et al (2012) Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 64(8):2677–2686. https://doi.org/10.1002/art.34473
4. Aringer M, Costenbader K, Daikh D et al (2019) 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus. Arthritis Rheumatol 71(9):1400–1412. https://doi.org/10.1002/art.40930
5. Aringer M, Dorner T, Leuchten N et al (2016) Toward new criteria for systemic lupus erythematosus-a standpoint. Lupus 25(8):805–811. https://doi.org/10.1177/0961203316644338
6. Carter EE, Barr SG, Clarke AE (2016) The global burden of SLE: prevalence, health disparities and socioeconomic impact. Nat Rev Rheumatol 12(10):605–620. https://doi.org/10.1038/nrrheum.2016.137
7. Rees F, Doherty M, Grainge MJ et al (2017) The worldwide incidence and prevalence of systemic lupus erythematosus: a systematic review of epidemiological studies. Rheumatology (Oxford) 56(11):1945–1961. https://doi.org/10.1093/rheumatology/kex260
8. Rees F, Doherty M, Grainge M et al (2016) The incidence and prevalence of systemic lupus erythematosus in the UK, 1999–2012. Ann Rheum Dis 75(1):136–141. https://doi.org/10.1136/annrheumdis-2014-206334
9. Tomic-Lucic A, Petrovic R, Radak-Perovic M et al (2013) Late-onset systemic lupus erythematosus: clinical features, course, and prognosis. Clin Rheumatol 32(7):1053–1058. https://doi.org/10.1007/s10067-013-2238-y
10. Mak A, Tay SH (2014) Environmental factors, toxicants and systemic lupus erythematosus. Int J Mol Sci 15(9):16043–16056. https://doi.org/10.3390/ijms150916043
11 Kwon YC, Chun S, Kim K et al (2019) Update on the genetics of systemic lupus erythematosus: genome-wide association studies and beyond. Cells 8(10):1180. https://doi.org/10.3390/cells8101180
12. Longo D, Fauci A, Kasper D et al (2011) Harrison's Principles of Internal Medicine, 18th Edition. Mcgraw-hill
13. Parks CG, de Souza Espindola A, Santos MB et al (2017) Understanding the role of environmental factors in the development of systemic lupus erythematosus. Best Pract Res Clin Rheumatol 31(3):306–320. https://doi.org/10.1016/j.berh.2017.09.005
14. Moulton VR (2018) Sex hormones in acquired immunity and autoimmune disease. Front Immunol 92279.https://doi.org/10.3389/fimmu.2018.02279
15. Laffont S, Rouquie N, Azar P et al (2014) X-chromosome complement and estrogen receptor signaling independently contribute to the enhanced TLR7-mediated IFN-alpha production of plasmacytoid dendritic cells from women. J Immunol 193(11):5444–5452. https://doi.org/10.4049/jimmunol.1303400
16. Schwartzman-Morris J, Putterman C (2012) Gender differences in the pathogenesis and outcome of lupus and of lupus nephritis. Clin Dev Immunol 2012604892.https://doi.org/10.1155/2012/604892
17. Boddaert J, Huong DLT, Amoura Z et al (2004) Late-onset systemic lupus erythematosus: a personal series of 47 patients and pooled analysis of 714 cases in the literature. Medicine (Baltimore) 83(6):348–359. https://doi.org/10.1097/01.md.0000147737.57861.7c
18. Kozyrev SV, Abelson AK, Wojcik J et al (2008) Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nat Genet 40(2):211–216. https://doi.org/10.1038/ng.79
19. G. International Consortium for Systemic Lupus Erythematosus, Harley JB, Alarcon-Riquelme ME et al (2008) Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet 40(2):204–10. https://doi.org/10.1038/ng.81
20. Hom G, Graham RR, Modrek B et al (2008) Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med 358(9):900–909. https://doi.org/10.1056/NEJMoa0707865
21. Graham RR, Cotsapas C, Davies L et al (2008) Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nat Genet 40(9):1059–1061. https://doi.org/10.1038/ng.200
22. Han JW, Zheng HF, Cui Y et al (2009) Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genet 41(11):1234–1237. https://doi.org/10.1038/ng.472
23. Gateva V, Sandling JK, Hom G et al (2009) A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nat Genet 41(11):1228–1233. https://doi.org/10.1038/ng.468

24. Yang W, Shen N, Ye DQ et al (2010) Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. PLoS Genet 6(2):e1000841. https://doi.org/10.1371/journal.pgen.1000841
25. Yang J, Yang W, Hirankarn N et al (2011) ELF1 is associated with systemic lupus erythematosus in Asian populations. Hum Mol Genet 20(3):601–607. https://doi.org/10.1093/hmg/ddq474
26. Okada Y, Shimane K, Kochi Y et al (2012) A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus eyrthematosus in Japanese. PLoS Genet 8(1):e1002455. https://doi.org/10.1371/journal.pgen.1002455
27. Lee YH, Bae SC, Choi SJ et al (2012) Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis. Mol Biol Rep 39(12):10627–10635. https://doi.org/10.1007/s11033-012-1952-x
28. Yang W, Tang H, Zhang Y et al (2013) Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as associated with systemic lupus erythematosus in Asians. Am J Hum Genet 92(1):41–51. https://doi.org/10.1016/j.ajhg.2012.11.018
29. Martin JE, Assassi S, Diaz-Gallo LM et al (2013) A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. Hum Mol Genet 22(19):4021–4029. https://doi.org/10.1093/hmg/ddt248
30. Armstrong DL, Zidovetzki R, Alarcon-Riquelme ME et al (2014) GWAS identifies novel SLE susceptibility genes and explains the association of the HLA region. Genes Immun 15(6):347–354. https://doi.org/10.1038/gene.2014.23
31. Bentham J, Morris DL, Graham DSC et al (2015) Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat Genet 47(12):1457–1464. https://doi.org/10.1038/ng.3434
32. Demirci FY, Wang X, Kelly JA et al (2016) Identification of a new susceptibility locus for systemic lupus erythematosus on chromosome 12 in individuals of European ancestry. Arthritis Rheumatol 68(1):174–183. https://doi.org/10.1002/art.39403
33. Alarcon-Riquelme ME, Ziegler JT, Molineros J et al (2016) Genome-wide association study in an Amerindian ancestry population reveals novel systemic lupus erythematosus risk loci and the role of European admixture. Arthritis Rheumatol 68(4):932–943. https://doi.org/10.1002/art.39504
34. Lessard CJ, Sajuthi S, Zhao J et al (2016) Identification of a systemic lupus erythematosus risk locus spanning ATG16L2, FCHSD2, and P2RY2 in Koreans. Arthritis Rheumatol 68(5):1197–1209. https://doi.org/10.1002/art.39548
35. Morris DL, Sheng Y, Zhang Y et al (2016) Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. Nat Genet 48(8):940–946. https://doi.org/10.1038/ng.3603
36. Sun C, Molineros JE, Looger LL et al (2016) High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. Nat Genet 48(3):323–330. https://doi.org/10.1038/ng.3496
37. Marquez A, Vidal-Bralo L, Rodriguez-Rodriguez L et al (2017) A combined large-scale meta-analysis identifies COG6 as a novel shared risk locus for rheumatoid arthritis and systemic lupus erythematosus. Ann Rheum Dis 76(1):286–294. https://doi.org/10.1136/annrheumdis-2016-209436
38. Langefeld CD, Ainsworth HC, Cunninghame Graham DS et al (2017) Transancestral mapping and genetic load in systemic lupus erythematosus. Nat Commun 816021. https://doi.org/10.1038/ncomms16021
39. Zhao J, Ma J, Deng Y et al (2017) A missense variant in NCF1 is associated with susceptibility to multiple autoimmune diseases. Nat Genet 49(3):433–437. https://doi.org/10.1038/ng.3782
40. Julia A, Lopez-Longo FJ, Perez Venegas JJ et al (2018) Genome-wide association study meta-analysis identifies five new loci for systemic lupus erythematosus. Arthritis Res Ther 20(1):100. https://doi.org/10.1186/s13075-018-1604-1
41. Qi YY, Zhou XJ, Nath SK et al (2018) A rare variant (rs933717) at FBXO31-MAP1LC3B in Chinese is associated with systemic lupus erythematosus. Arthritis Rheumatol 70(2):287–297. https://doi.org/10.1002/art.40353
42. Wang YF, Zhang Y, Zhu Z et al (2018) 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 77(7):1078–1084. https://doi.org/10.1136/annrheumdis-2018-213093
43. Zhang H, Zhang Y, Wang YF et al (2018) Meta-analysis of GWAS on both Chinese and European populations identifies GPR173 as a novel X chromosome susceptibility gene for SLE. Arthritis Res Ther 20(1):92. https://doi.org/10.1186/s13075-018-1590-3
44. Wen L, Zhu C, Zhu Z et al (2018) Exome-wide association study identifies four novel loci for systemic lupus erythematosus in Han Chinese population. Ann Rheum Dis 77(3):417. https://doi.org/10.1136/annrheumdis-2017-211823
45. Akizuki S, Ishigaki K, Kochi Y et al (2019) PLD4 is a genetic determinant to systemic lupus erythematosus and involved in murine autoimmune phenotypes. Ann Rheum Dis 78(4):509–518. https://doi.org/10.1136/annrheumdis-2018-214116
46. Acosta-Herrera M, Kerick M, Gonzalez-Serna D et al (2019) Genome-wide meta-analysis reveals shared new loci in systemic seropositive rheumatic diseases. Ann Rheum Dis 78(3):311–319. https://doi.org/10.1136/annrheumdis-2018-214127
47. Zhang F, Wang YF, Zhang Y et al (2020)Independent replication on genome-wide association study signals identifies IRF3 as a novel locus for systemic lupus erythematosus.Front Genet 11600.https://doi.org/10.3389/fgene.2020.00600
48. Qi YY, Zhai YL, Liu XR et al (2020) Single nucleotide polymorphisms in PPARD associated with systemic lupus erythematosus in Chinese populations. J Immunol Res 20207285747.https://doi.org/10.1155/2020/7285747
49. Tangtanatakul P, Thumarat C, Satproedprai N et al (2020) Meta-analysis of genome-wide association study identifies FBN2 as a novel locus associated with systemic lupus erythematosus in Thai population. Arthritis Res Ther 22(1):185. https://doi.org/10.1186/s13075-020-02276-y
50. Yin X, Kim K, Suetsugu H et al (2020) Meta-analysis of 208370 East Asians identifies 113 susceptibility loci for systemic lupus erythematosus. Ann Rheum Dis. https://doi.org/10.1136/annrheumdis-2020-219209
51. Song Q, Lei Y, Shao L et al (2021) Genome-wide association study on Northern Chinese identifies KLF2, DOT1L and STAB2 associated with systemic lupus erythematosus. Rheumatology (Oxford). https://doi.org/10.1093/rheumatology/keab016
52. Wang YF, Zhang Y, Lin Z et al (2021) Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. Nat Commun 12(1):772. https://doi.org/10.1038/s41467-021-21049-y
53 Fang H, U.-D. Consortium, De Wolf H et al (2019) A genetics-led approach defines the drug target landscape of 30 immune-related traits. Nat Genet 51(7):1082–1091. https://doi.org/10.1038/s41588-019-0456-1
54. Plenge RM (2019) Priority index for human genetics and drug discovery. Nat Genet 51(7):1073–1075. https://doi.org/10.1038/s41588-019-0460-5

55. Webb R, Kelly JA, Somers EC et al (2011) Early disease onset is predicted by a higher genetic risk for lupus and is associated with a more severe phenotype in lupus patients. Ann Rheum Dis 70(1):151–156. https://doi.org/10.1136/ard.2010.141697
56. Taylor KE, Chung SA, Graham RR et al (2011) Risk alleles for systemic lupus erythematosus in a large case-control collection and associations with clinical subphenotypes. PLoS Genet 7(2):e1001311. https://doi.org/10.1371/journal.pgen.1001311
57. Joo YB, Lim J, Tsao BP et al (2018) Genetic variants in systemic lupus erythematosus susceptibility loci, XKR6 and GLT1D1 are associated with childhood-onset SLE in a Korean cohort. Sci Rep 8(1):9962. https://doi.org/10.1038/s41598-018-28128-z
58. Reid S, Alexsson A, Frodlund M et al (2020) High genetic risk score is associated with early disease onset, damage accrual and decreased survival in systemic lupus erythematosus. Ann Rheum Dis 79(3):363–369. https://doi.org/10.1136/annrheumdis-2019-216227
59. Chen L, Wang YF, Liu L et al (2020) Genome-wide assessment of genetic risk for systemic lupus erythematosus and disease severity. Hum Mol Genet 29(10):1745–1756. https://doi.org/10.1093/hmg/ddaa030
60. Lawrence JS, Martins CL, Drake GL (1987) A family survey of lupus erythematosus. 1. Heritability. J Rheumatol 14(5):913–21
61. Kuo CF, Grainge MJ, Valdes AM et al (2015) Familial aggregation of systemic lupus erythematosus and coaggregation of autoimmune diseases in affected families. JAMA Intern Med 175(9):1518–1526. https://doi.org/10.1001/jamainternmed.2015.3528
62. Chandra V, Bhattacharyya S, Schmiedel BJ et al (2021) Promoter-interacting expression quantitative trait loci are enriched for functional genetic variants. Nat Genet 53(1):110–119. https://doi.org/10.1038/s41588-020-00745-3
63. Hou G, Harley ITW, Lu X et al (2021) SLE non-coding genetic risk variant determines the epigenetic dysfunction of an immune cell specific enhancer that controls disease-critical microRNA expression. Nat Commun 12(1):135. https://doi.org/10.1038/s41467-020-20460-1
64. Imgenberg-Kreuz J, Carlsson Almlof J, Leonard D et al (2018) DNA methylation mapping identifies gene regulatory effects in patients with systemic lupus erythematosus. Ann Rheum Dis 77(5):736–743. https://doi.org/10.1136/annrheumdis-2017-212379
65. Su C, Johnson ME, Torres A et al (2020) Mapping effector genes at lupus GWAS loci using promoter Capture-C in follicular helper T cells. Nat Commun 11(1):3294. https://doi.org/10.1038/s41467-020-17089-5
66. Harley JB, Chen X, Pujato M et al (2018) Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. Nat Genet 50(5):699–707. https://doi.org/10.1038/s41588-018-0102-3
67. Lu X, Chen X, Forney C et al (2021) Global discovery of lupus genetic risk variant allelic enhancer activity. Nat Commun 12(1):1611. https://doi.org/10.1038/s41467-021-21854-5
68. Molineros JE, Singh B, Terao C et al (2019) Mechanistic characterization of RASGRP1 variants identifies an hnRNP-K-regulated transcriptional enhancer contributing to SLE susceptibility. Front Immunol 101066.https://doi.org/10.3389/fimmu.2019.01066
69. Jones SA, Cantsilieris S, Fan H et al (2019) Rare variants in non-coding regulatory regions of the genome that affect gene expression in systemic lupus erythematosus. Sci Rep 9(1):15433. https://doi.org/10.1038/s41598-019-51864-9
70. Thynn HN, Chen XF, Hu WX et al (2020) An allele-specific functional SNP associated with two systemic autoimmune diseases modulates IRF5 expression by long-range chromatin loop formation. J Invest Dermatol 140(2):348-360 e11. https://doi.org/10.1016/j.jid.2019.06.147
71. Alarcon-Segovia D, Alarcon-Riquelme ME, Cardiel MH et al (2005) Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. Arthritis Rheum 52(4):1138–1147. https://doi.org/10.1002/art.20999
72. Deapen D, Escalante A, Weinrib L et al (1992) A revised estimate of twin concordance in systemic lupus erythematosus. Arthritis Rheum 35(3):311–318. https://doi.org/10.1002/art.1780350310
73. Moser KL, Kelly JA, Lessard CJ et al (2009) Recent insights into the genetic basis of systemic lupus erythematosus. Genes Immun 10(5):373–379. https://doi.org/10.1038/gene.2009.39
74. Fielder AH, Walport MJ, Batchelor JR et al (1983) Family study of the major histocompatibility complex in patients with systemic lupus erythematosus: importance of null alleles of C4A and C4B in determining disease susceptibility. Br Med J (Clin Res Ed) 286(6363):425–428. https://doi.org/10.1136/bmj.286.6363.425
75. Botto M, Walport MJ (2002) C1q, autoimmunity and apoptosis. Immunobiology 205(4–5):395–406. https://doi.org/10.1078/0171-2985-00141
76. Schur PH (1995) Genetics of systemic lupus erythematosus. Lupus 4(6):425–437. https://doi.org/10.1177/096120339500400603
77. Horton R, Wilming L, Rand V et al (2004) Gene map of the extended human MHC. Nat Rev Genet 5(12):889–899. https://doi.org/10.1038/nrg1489
78. Trowsdale J, Knight JC (2013) Major histocompatibility complex genomics and human disease. Annu Rev Genomics Hum Genet 14301–23.https://doi.org/10.1146/annurev-genom-091212-153455
79. Jia X, Han B, Onengut-Gumuscu S et al (2013) Imputing amino acid polymorphisms in human leukocyte antigens. PLoS ONE 8(6):e64683. https://doi.org/10.1371/journal.pone.0064683
80. Dilthey AT, Moutsianas L, Leslie S et al (2011) HLA*IMP—an integrated framework for imputing classical HLA alleles from SNP genotypes. Bioinformatics 27(7):968–972. https://doi.org/10.1093/bioinformatics/btr061
81. Lim J, Bae SC, Kim K (2019) Understanding HLA associations from SNP summary association statistics. Sci Rep 9(1):1337. https://doi.org/10.1038/s41598-018-37840-9
82. Morris DL, Taylor KE, Fernando MM et al (2012) 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 91(5):778–793. https://doi.org/10.1016/j.ajhg.2012.08.026
83. Bang SY, Choi JY, Park S et al (2016) Brief Report: influence of HLA-DRB1 susceptibility alleles on the clinical subphenotypes of systemic lupus erythematosus in Koreans. Arthritis Rheumatol 68(5):1190–1196. https://doi.org/10.1002/art.39539
84. Kim K, Bang SY, Lee HS et al (2014) The HLA-DRbeta1 amino acid positions 11-13-26 explain the majority of SLE-MHC associations. Nat Commun 55902.https://doi.org/10.1038/ncomms6902
85. Molineros JE, Looger LL, Kim K et al (2019) Amino acid signatures of HLA Class-I and II molecules are strongly associated with SLE susceptibility and autoantibody production in Eastern Asians. PLoS Genet 15(4):e1008092. https://doi.org/10.1371/journal.pgen.1008092
86. Kamitaki N, Sekar A, Handsaker RE et al (2020) Complement genes contribute sex-biased vulnerability in diverse disorders. Nature 582(7813):577–581. https://doi.org/10.1038/s41586-020-2277-x
87. Morimoto K, Baba Y, Shinohara H et al (2016) LRRK1 is critical in the regulation of B-cell responses and CARMA1-dependent

NF-kappaB activation. Sci Rep 625738.https://doi.org/10.1038/srep25738
88. Linge P, Arve S, Olsson LM et al (2020) NCF1-339 polymorphism is associated with altered formation of neutrophil extracellular traps, high serum interferon activity and antiphospholipid syndrome in systemic lupus erythematosus. Ann Rheum Dis 79(2):254–261. https://doi.org/10.1136/annrheumdis-2019-215820
89. Peoples R, Franke Y, Wang YK et al (2000) A physical map, including a BAC/PAC clone contig, of the Williams-Beuren syndrome—deletion region at 7q11.23. Am J Hum Genet 66(1):47–68. https://doi.org/10.1086/302722
90. Yin X, Kim K, Suetsugu H et al (2020) Meta-analysis of 208,370 East Asians identifies 113 genomic loci and yields new non-immune cell relevant biological insights for systemic lupus erythematosus. medRxiv 2020.08.22.20178939. https://doi.org/10.1101/2020.08.22.20178939
91. Lewis CM, Vassos E (2020) Polygenic risk scores: from research tools to clinical instruments. Genome Med 12(1):44. https://doi.org/10.1186/s13073-020-00742-5
92. Torkamani A, Wineinger NE, Topol EJ (2018) The personal and clinical utility of polygenic risk scores. Nat Rev Genet 19(9):581–590. https://doi.org/10.1038/s41576-018-0018-x
93. Tam V, Patel N, Turcotte M et al (2019) Benefits and limitations of genome-wide association studies. Nat Rev Genet 20(8):467–484. https://doi.org/10.1038/s41576-019-0127-1
94. Amariuta T, Ishigaki K, Sugishita H et al (2020) Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. Nat Genet 52(12):1346–1354. https://doi.org/10.1038/s41588-020-00740-8
95. Sandling JK, Pucholt P, Hultin Rosenberg L et al (2021) Molecular pathways in patients with systemic lupus erythematosus revealed by gene-centred DNA sequencing. Ann Rheum Dis 80(1):109–117. https://doi.org/10.1136/annrheumdis-2020-218636
96. Fritz MS, Mackinnon DP (2007) Required sample size to detect the mediated effect. Psychol Sci 18(3):233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x
97. Javierre BM, Burren OS, Wilder SP et al (2016) Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell 167(5):1369-1384 e19. https://doi.org/10.1016/j.cell.2016.09.037
98. Chun S, Casparino A, Patsopoulos NA et al (2017) Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. Nat Genet 49(4):600–605. https://doi.org/10.1038/ng.3795
99. Ha E, Bang SY, Lim J et al (2021) Genetic variants shape rheumatoid arthritis-specific transcriptomic features in CD4(+) T cells through differential DNA methylation, explaining a substantial proportion of heritability. Ann Rheum Dis. https://doi.org/10.1136/annrheumdis-2020-219152
100. Singh B, Maiti GP, Zhou X et al (2021) Lupus susceptibility region containing CDKN1B rs34330 mechanistically influences expression and function of multiple target genes, also linked to proliferation and apoptosis. Arthritis Rheumatol. https://doi.org/10.1002/art.41799
101. James JA, Kaufman KM, Farris AD et al (1997) An increased prevalence of Epstein-Barr virus infection in young patients suggests a possible etiology for systemic lupus erythematosus. J Clin Invest 100(12):3019–3026. https://doi.org/10.1172/JCI119856
102. Hanlon P, Avenell A, Aucott L et al (2014) Systematic review and meta-analysis of the sero-epidemiological association between Epstein-Barr virus and systemic lupus erythematosus. Arthritis Res Ther 16(1):R3. https://doi.org/10.1186/ar4429
103. McClain MT, Heinlen LD, Dennis GJ et al (2005) Early events in lupus humoral autoimmunity suggest initiation through molecular mimicry. Nat Med 11(1):85–89. https://doi.org/10.1038/nm1167
104. Harley JB, James JA (2006) Epstein-Barr virus infection induces lupus autoimmunity. Bull NYU Hosp Jt Dis 64(1–2):45–50
105. Zhou H, Schmidt SC, Jiang S et al (2015) Epstein-Barr virus oncoprotein super-enhancers control B cell growth. Cell Host Microbe 17(2):205–216. https://doi.org/10.1016/j.chom.2014.12.013
106. Schaid DJ, Chen W, Larson NB (2018) From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet 19(8):491–504. https://doi.org/10.1038/s41576-018-0016-z
107. Liu JZ, Erlich Y, Pickrell JK (2017) Case-control association mapping by proxy using family history of disease. Nat Genet 49(3):325–331. https://doi.org/10.1038/ng.3766
108. Hujoel MLA, Gazal S, Loh PR et al (2020) Liability threshold modeling of case-control status and family history of disease increases association power. Nat Genet 52(5):541–547. https://doi.org/10.1038/s41588-020-0613-6
109. Ota M, Nagafuchi Y, Hatano H et al (2021) Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. Cell 184(11):3006-3021 e17. https://doi.org/10.1016/j.cell.2021.03.056
110. Mistry P, Nakabo S, O'Neil L et al (2019) Transcriptomic, epigenetic, and functional analyses implicate neutrophil diversity in the pathogenesis of systemic lupus erythematosus. Proc Natl Acad Sci U S A 116(50):25222–25228. https://doi.org/10.1073/pnas.1908576116
111. Nehar-Belaid D, Hong S, Marches R et al (2020) Mapping systemic lupus erythematosus heterogeneity at the single-cell level. Nat Immunol 21(9):1094–1106. https://doi.org/10.1038/s41590-020-0743-0
112. Cotsapas C, Hafler DA (2013) Immune-mediated disease genetics: the shared basis of pathogenesis. Trends Immunol 34(1):22–26. https://doi.org/10.1016/j.it.2012.09.001
113. Parkes M, Cortes A, van Heel DA et al (2013) Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nat Rev Genet 14(9):661–673. https://doi.org/10.1038/nrg3502
114. Li YR, Li J, Zhao SD et al (2015) Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. Nat Med 21(9):1018–1027. https://doi.org/10.1038/nm.3933
115. Lim J, Kim K (2019) Genetic variants differentially associated with rheumatoid arthritis and systemic lupus erythematosus reveal the disease-specific biology. Sci Rep 9(1):2739. https://doi.org/10.1038/s41598-019-39132-2
116. Peyrot WJ, Price AL (2021) Identifying loci with different allele frequencies among cases of eight psychiatric disorders using CC-GWAS. Nat Genet 53(4):445–454. https://doi.org/10.1038/s41588-021-00787-1
117. Kwon YC, Lim J, Bang SY et al (2020) Genome-wide association study in a Korean population identifies six novel susceptibility loci for rheumatoid arthritis. Ann Rheum Dis 79(11):1438–1445. https://doi.org/10.1136/annrheumdis-2020-217663
118. Yamamoto EA, Jorgensen TN (2019) Relationships between vitamin D, gut microbiome, and systemic autoimmunity. Front Immunol 103141.https://doi.org/10.3389/fimmu.2019.03141
119. Rahbar Saadat Y, Hejazian M, Bastami M et al (2019) The role of microbiota in the pathogenesis of lupus: dose it impact lupus nephritis? Pharmacol Res 139191–198.https://doi.org/10.1016/j.phrs.2018.11.023

120. Hevia A, Milani C, Lopez P et al (2014) Intestinal dysbiosis associated with systemic lupus erythematosus. mBio 5(5):e01548-14. https://doi.org/10.1128/mBio.01548-14
121. He Z, Shao T, Li H et al (2016) Alterations of the gut microbiome in Chinese patients with systemic lupus erythematosus. Gut Pathog 864.https://doi.org/10.1186/s13099-016-0146-9

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.