DOI:
10.1097/BOR.0000000000000411

Document Version
Peer reviewed version

Current Opinion in Rheumatology

Genetic Advances in SLE: an update

--Manuscript Draft--

Manuscript Number:	
Full Title:	Genetic Advances in SLE: an update
Article Type:	Review Article
Corresponding Author:	Timothy Vyse, Ph.D. M.D. UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	
Corresponding Author's Secondary Institution:	
First Author:	Lingyan Chen, MSc
First Author Secondary Information:	
Order of Authors:	Lingyan Chen, MSc
	David Lester Morris, PhD
	Timothy Vyse, Ph.D. M.D.
Order of Authors Secondary Information:	

Genetic Advances in SLE: an update

Lingyan Chen[1], David L Morris[1], Timothy J Vyse [1,2*]

1 Division of Genetics and Molecular Medicine, 2 Division of Immunology, Infection & Inflammatory Disease, King's College London, Guy's Hospital, London, SE1 9RT, UK.

Correspondence to Timothy J Vyse, PhD, FRCP, FMedSci, Professor of Molecular Medicine, King's College London, Consultant in Rheumatology, Guy's and St Thomas' NHS Trust. Tel: +44 20 7848 8517; email: tim.vyse@kcl.ac.uk.

Abstract

Purpose of review – More than 80 loci are now reported to show robust genetic association with Systemic Lupus Erythematosus (SLE). The differential functional effects of the risk alleles for the majority of these loci remain to be defined. Here, we review current SLE association findings and the recent progress in the annotation of non-coding regions of the human genome as well as the new technologies and statistical methods that can be applied to further the understanding of SLE genetics.

Recent findings – Genome-wide association studies (GWAS) have markedly expanded the catalogue of genetic signals contributing to SLE development; we can now explain more than 50% of the disease's heritability. Expression quantitative trait loci (eQTL) mapping with co-localisation analysis of GWAS results help to identify the underlying causal genes. The ENCODE, Roadmap Epigenome and the Blueprint Epigenome projects have jointly annotated more than 80% of the noncoding genome, providing a wealth of information (from healthy individuals) to define the functional elements within the risk loci. Technologies, such as next-generation sequencing, chromatin structure determination and genome editing, will help elucidate the actual mechanisms that underpin SLE risk alleles.

Summary – Gene expression and epigenetic databases provide a valuable resource to interpret genetic association in SLE. Expansion of such resources to include disease and multiple ancestries will further aid the exploration of the biology underlying the genetics.

Keywords: Systemic lupus erythematosus; GWAS; expression quantitative trait loci; epigenome; causal variants

Introduction

Systemic lupus erythematosus (SLE) is a chronic inflammatory autoimmune disease associated with a wide range of signs and symptoms varying among affected individuals and can involve many organs and systems, including the skin, joints, kidneys, lungs, central nervous system, and hematopoietic system. The population prevalence varies with ancestry, being more prevalent in non-European populations with a significant gender disparity towards women (9:1) during the years between menarche and menopause [1]. Although the exact etiology of lupus is not fully understood, a strong genetic link has been identified through the application of family and large-scale genome-wide association studies (GWAS). The concordance rate in monozygotic twins (24%) is approximately 10 fold higher than in dizygotic twins (2%) [2,3]. A recent study from Taiwan reported that the heritability was 43.9% and the proportion of phenotypic variance explained by shared and non-shared environmental factors was 25.8% and 30.3%, respectively, suggesting non-heritable factors may play a considerable role in disease pathogenesis [4].

There are now more than 80 loci reported to be associated with the susceptibility of SLE. Here, we review current SLE association findings and the recent progress in the annotation of

the non-coding region of the human genome as well as new technologies and statistical methods, in order to apply this knowledge to the understanding of SLE genetics.

Insights from GWAS

Genetic linkage analysis and candidate gene association studies identified several SLE susceptibility loci (e.g. HLA-DR2/DR3) [5]. Nevertheless, the advent and application of GWAS dramatically advanced knowledge of the genetic aetiology of SLE.

There have been seven SLE GWAS in European population [6–12], six Asian GWAS [13–18], subsequent meta-analysis and large-scale replication studies [19–22], published since 2008. Currently, 84 genetic loci are implicated as SLE risk (Figure 1: The CIRCOS plot [23] and supplementary Table 1), which, in order to avoid likely spurious associations, includes genetic associations with a P value less than 5×10^{-8} tested in a total sample size of at least 1000 individuals. The interactive version of a continually updated resource with details on SLE associations can be access through the following link: http://insidegen.com/insidegen-LUPUS-Associations.html.

With the caveat that the majority of mechanisms remain to be elucidated, it appears that the risk loci associated with SLE influence immune cell function. Although functional studies are designed with a *priori* hypotheses in mind, key pathogenic pathways that are likely influenced by SLE-associated gene products include: immune complex processing and phagocytosis; DNA degradation, apoptosis and clearance of cellular debris; neutrophil and monocytes signalling; Toll-like receptor and/or type I interferon signalling; nuclear factor kappaB activation; B and T-cell function and signalling. Some genes associated with SLE may act through several pathways. For example, *TNFAIP3*, encoding the ubiquitin-editing

enzyme A20, is a key regulator of nuclear factor-kappa-B (NF-kB)-derived pro-inflammatory responses, which is involved in both adaptive and innate immune pathways [24,25]. These SLE susceptibility loci contain predominantly common (frequency of > 0.1% in the general population) associated variants that have been confirmed among multiple ancestries, suggesting shared mechanisms in disease aetiology [26–28].

European GWAS

The largest European GWAS of SLE conducted by our group [11], comprised 7,219 SLE cases and 15,991 controls of European decent, provided considerable power to detect disease risk loci. Notably, the study identified 43 susceptibility loci, ten of which were novel loci: *SPRED2*, *IKZF2*, *IL12A*, *TCF7-SKP1*, *DHCR7-NADSYN1*, *SH2B3*, *RAD51B*, *CIITA-SOCS1*, *PLD2*, and *CXorf21*. One of the great challenges posed by interpreting GWAS data is determining the causal genes implicated by the genetic association data. As will be discussed and amplified below, we put some considerable effort into this process before naming the genes in the above list. Irrespective of the underlying causal genes, we can conclude that the heritability explained by the risk alleles mapped at these loci is 15.3%, which is a large increase over the 8.7% reported by So et al [29] in 2011 using the same measure.

Asian GWAS

An extensive large-scale fine mapping study using Immunochip conducted in 4,478 SLE cases and 12,656 controls from six East Asian cohorts identified 10 novel loci [18] in Asians, encompassing *GTF2IRD1-GTF2I*, *DEF6*, *IL12B*, *TCF7*, *TERT*, *CD226*, *PCNXL3*, *RASGRP1*, *SYNGR1*, and *SIGLEC6*. Some of these were previously reported to be associated in

Europeans, for example, *DEF6* and *TCF7*. The identification of these risk loci increased the explained heritability to 24% in Asian SLE.

Trans ancestry meta analyses of GWAS

A comparison of genetic association signals across the genome in European and Asian populations suggested that SLE susceptibility loci were shared extensively between both populations [21]. This motivated a trans ancestral approach at the genome-wide level to provide evidence of shared genetic components in the two populations and search for additional SLE associated loci. The study by Morris and Sheng et al [21], that combined three GWAS from two ethnicities: Chinese (1659 cases and 3,398 controls) and European (4,044 cases + 6,959 controls), found evidence of considerable commonality in terms of SLE association signals as well as mapping novel susceptibility loci, including *CD45, IKBKE, LBH, LPP-TPRG1-AS1, ATXN1, BACH2, GTF2I, JAK2, RNASEH2C,* and *ZFP90*. Notably, this study suggested that the increased prevalence of SLE in non-European (including Asians) has a genetic basis by comparison of genetic risk scores (GRS) between populations (Figure 2) [21] . Moreover, by using all genotyped SNPs (DNA chip) to calculate heritability explained, the explained variation (Vg) increase to 28% in Chinese subjects and 27% Europeans using the GCTA algorithm [30]. While there are still some uncertainties in the methodology for calculating heritability explained, this shows very strong evidence that we are making progress on the understanding of SLE heritability.

The latest large-scale trans ancestral study using Immunochip [31], comprising three ancestries: European (EA: 6,748,cases and 11,516 controls), African-American (AA: 2,970 case and 2,452 controls), and Hispanic Amerindian (HA: 1,872 cases and 2,016 controls), have identified nine novel loci for EA (*TMEM39A-TIMMDC1, DGKQ, LRRC16A, SLC17A4, OLIG3-LOC100130476, GTF2IRD1-GTF2I , FAM86B3P, PKIA-ZC2HC1A,* and *GRB2*), two for AA (*PTTG1-MIR146A* and *PLAT*) and two for HA (*GALC* and *CLEC16A*). By comparing results across different populations, both ancestry-dependent and ancestry-

independent contributions to SLE risk are identified with the caveat of unequal cohort sizes. The study reveals evidence of sharing of genetic risk loci between ancestries as well as evidence that each individual population carries unique genetic risk factors at the locus level and at the allelic level.

Missing heritability

In summary, the chip heritability identified by the latest GWAS have explained around 28% of the disease heritability: a marked improvement on 8.3% calculated in 2011 [29]. However, there is still one third of heritability left to explain, if we assume that the total estimated heritability is 43.9%. Explanations for the missing heritability, including larger numbers of variants of smaller effect, rarer variants (possibly with larger effects) that are not present on genotyping arrays or structural variants poorly captured by existing arrays, as well as epigenetic modifications, have been suggested [32]. Innovations in genotyping and sequencing technologies, like the Immuno-chip platform [18,31] and next generation sequencing (NGS, as described below) will advance the investigation into common and rare variants and potential effects on the immune system, enhancing our understanding of the genetic risk of SLE.

The LD that exists in the human genome facilitates the mapping of risk loci by reducing the number of genetic variants required for GWAS; however, the same correlation between genetic polymorphisms at these susceptibility loci then bedevils attempts to identify the actual causal allele(s) at risk loci. Bayesian fine mapping approaches had been proposed to derive smaller sets of SNPs (termed 'credibility sets') as the most likely causal variants at risk loci [33]. Nevertheless, statistical methods are inadequate to fully resolve the problem caused by LD. In order to further pursue likely causal SNPs within any given credibility set,

the functional effect of SNPs can be studied in silico. As the majority of variants within causal credibility sets are non-coding [34,35], function is inferred using gene transcript expression data and epigenetic modification data (as described below) (Figure 3 and Figure 4).

Application of eQTL mapping to GWAS results

Assisted by dense genome coverage of the reference panel from the 1000 Genome project [36], imputation and Bayesian inference provided evidence for missense variants underpinning association for eight genes, including *PTPN22, FCGR2A, NCF2, IFIH1, WDFY4, ITGAM, PLD2*, and *TYK2* [11]. However, as mentioned above, the majority (85% ~ 90%) of disease associated loci in SLE are located outside of protein-coding regions, suggesting that the underlying mechanism is likely regulatory, and so might exert their function through altering gene expression rather than by altering protein structure. Of note, an over-representation (n=16) of transcription factors among the 43 SLE susceptibility genes have been annotated in our recent European GWAS [11], further indicating that perturbed gene regulation was a major functional risk factor for SLE. Expression quantitative trait loci (eQTLs) mapping, which combines genome-wide expression profiling and genome-wide marker-based genotyping, takes advantage of the heritability of gene expression profiles to identify genetic variants that correlated with changes in gene expression. eQTLs can be classified as "in cis" (locally) or "in trans" (at a distance) based on their physical distance from the regulated gene.

Some studies [18,22] used public databases, such as the whole blood eQTL browser (http://genenetwork.nl/bloodeqtlbrowser/) [37] and tissue-specific GTEx portal (http://www.gtexportal.org/home/) [38], to determine whether the disease-associated SNP is

a significant eQTL. Of note, there exists some limitations when applying eQTL analysis to the GTEx whole blood datasets, as in autoimmunity, we seek eQTLs in specific immune cell subsets. In order to highlight the potential causal genes at the susceptibility loci robustly, it is essential to integrate the disease association and eQTL data using a co-localisation approach. That is, to establish that the same genetic variants that underlie the disease association also underlie the eQTL. The presence of LD in the genome can readily obfuscate this overlap. Co-localisation methods, like the regulatory trait concordance (RTC) [39], conditional analysis [30], and Bayesian co-localisation [40], can be employed to infer that the disease association and eQTL have the same allelic basis. As many variants have weak eQTL effects, erroneous conclusions will be made if analyses for co-localisation are not performed. An example of co-localisation analysis of eQTL and GWAS is shown in Figure 3.

Recent studies by Morris et al [11,21] and Odhams et al [41] examined the functional outcome of SLE associated variants through the integration of GWAS and eQTL data from various cell types ex vivo, involving T cells, B cells, NK cells, stimulated and resting monocytes, as well as lymphoblastoid cell lines (LCL). By integrating the results of eQTL and RTC analysis, they found evidence to support the role of causal genes as candidates at a given locus. For example, *SOCS1* (Suppressor of Cytokine Signalling 1) was found to be a suggestive causal gene at the locus tagged by the SNP rs9652601 (with a RTC score higher than 0.9), rather than *CLEC16A* (C-Type Lectin Domain Family 16 Member A), even though the risk variant resides within the latter one - a gene previously reported as relating to other autoimmune diseases [42]. Moreover, the Odhams et al's study [41] illustrated the benefits of using RNA-seq as opposed to microarrays for eQTL mapping, due to more informative data generated by RNA-seq. With RNA-seq, transcript profiling can be done on the gene-

level, exon-level and splice-junction-level, which is more effective in explaining potential regulatory mechanisms.

Nevertheless, we believe that many eQTLs related to SLE risk alleles remain unidentified, data from diverse stimulations and time points will be required, as well as gene expression data from patient material, to reveal the full eQTL landscape of SLE genetics.

Epigenetics to annotate functional / regulatory variants

An approach that is complementary to eQTL analyses, to examine the regulatory function of non-coding genetic variants, is to study gene regulation with epigenetics. Epigenetic modifications, a term coined to describe genome-wide chromatin modification, including DNA methylation, histone modifications, chromatin accessibility, microRNA regulations, and 2D chromatin interactions [43], constitute an additional layer of genomic regulation, and may serve as a dynamic link between genotype and phenotype. Such changes in DNA and chromatin structure correlate with changes in chromatin accessibility and transcription factor binding.

The Encyclopedia of DNA elements (ENCODE) project (https://www.encodeproject.org/) [44] has systematically mapped regions of transcription, transcription factor association, chromatin structure and histone modification, and assigns biochemical functions for 80% of the genome, in particular outside of the protein-coding regions. Overall, the project has provided an expansive resource to define the functional DNA elements for biomedical research, although the available cell types or cell lines are limited. The cells of closest immune relevance in ENCODE Tier 1 and Tier 2 are LCLs (GM12878), B cells (CD20+) and monocytes (CD14+), as well as T cells (CD4+) and peripheral blood mononuclear cell (PBMC) in Tier 3. A recent ImmunoChip study in Asians [18] took advantage of ENCODE

data to map the underlying loci. For example, one of the signals (rs73366469) identified in this study was located between two 'general transcription factor' genes, *GTF2I* and *GTF2IRD1*. By integrating the ENCODE data, they found that an indel SNP rs587608058 (r^2=0.81), ~1000bp from rs73366469, lay within conserved enhancer, active chromatin and transcription factor binding sites in LCLs and CD4+ T cells. In addition, this region was found to overlap the transcription start sites for *GTF2I* and *VCF* through chromatin interacting analysis and chromosome confirmation capture (Hi-C) analysis, providing evidence for the potential causal variants and genes at this locus for further study.

The Roadmap epigenomics project (http://www.roadmapepigenomics.org/) [46] integrated analysis of 111 reference human epigenomes to obtain a comprehensive map of the human epigenomic landscape across a large collection of primary cells, including immune cells, and tissues. This map is extremely useful for studies of genome interpretation, gene regulation, cellular differentiation, genome evolution, genetic variation and human disease. In our meta GWAS analysis of Chinese and European data [21], the histone modification markers, including acetylation markers (H3K27ac, H3K9ac) and methylation markers (H3K27me3 and H3K9me3), from blood cell types were used to investigate the potential regulatory function of the target risk loci. For example, there are several genes, including *SRGAP2*, *SRGAP2D*, *IKBKE*, *RASSF5*, *EIF2D* and *DYRK3*, located within ±200kb of the lead GWAS SNP rs2297550. The GWAS SNP was also found to be a putative eQTL for *IKBKE*, with the SLE risk allele correlated with reduced expression in CD4+ T cells [47], CD19+ B cells [48] and NK cells (data unpublished), but with increased expression in CD14+ monocytes [49]. *IKBKE* encodes a noncanonical I-kappa-B kinase (IKK) that is essential in regulating inflammatory responses to viral infection by activating the type I interferon, NF-kB and STAT signalling pathways, suggesting *IKBKE* might be the potential causal gene. Moreover,

there is an intense histone acetylation peak around the associated SNP rs2297550, indicating that rs2297550 may be a potential causal variant [21]. Figure 4 shows an example of fine mapping causal SNPs by integrating genetics and epigenetics.

Another recent completed large-scale epigenomic project, the Blueprint project (http://www.blueprint-epigenome.eu/) [50–52], has impressively shown how epigenetic information and analyses can help to study the cellular mechanisms associated with complex human diseases. Moreover, the Blueprint consortium generated three comprehensive reference panels, including genome (whole genome sequencing), transcriptome (RNA-seq), and epigenome (DNA methylation and histone modification), in three immune cells (Neutrophils, monocytes and T cells) from nearly 200 individuals to characterize the contributions of diverse genomic inputs to transcriptional variation. Summary data from these panels can be accessed through http://blueprint-dev.bioinfo.cnio.es/WP10/ .
High-resolution maps of promoter interactions [51] generated by 'Promoter capture Hi-C' (PChi-C) make it possible to study the long range regulatory in the three-dimensional nuclear space. By integrating PCHi-C data with disease-associated SNPs generated by GWAS, we can prioritize the putative target genes for the risk loci. The promoter interactomes map may serve as a more robust method to define cis-eQTLs rather than by distance, revealing insights into genomic regulatory mechanisms of diseases.

Next generation sequencing (NGS) in the genome research

With the development of NGS, high-throughput technologies that are now widely used in genome research, any part of the genome can be sequenced. Based on the coverage of the genome, NGS strategies can be classified by scale: target region sequencing, whole-exome

sequencing (WES), and whole-genome sequencing (WGS). Targeted resequencing of risk loci in disease cohorts may facilitate the identification of rare variants at common-allele-associated loci [53]. WES captures all coding exons covering 1~2% of the genome. Nevertheless, as mentioned above, approximately 85~90% of the risk loci associated with SLE are located outside of the coding-regions. Compared to WES, WGS can capture the majority of the genome, which facilitates delineation of exon duplications and gene fusions, and non-coding regions that might be missing by WES. However, the higher cost and time consuming bioinformatics analyses restrict the application of WGS [54]. In future, with the decreasing cost of sequencing and newly developed computation algorithms, WGS will be increasingly utilised.

Incorporating with a wide range of chromatin profiling experiments, NGS is applied to investigate chromatin biology by identifying genomic loci that are occupied by nucleosomes, bound to transcription factors, or accessible to nuclease cleavage [55]. Technologies such as ChIP-seq [56], FAIRE-seq, DNase-seq [57,58], Hi-C [59], and ATAC-seq [60] enable genome-wide investigations of a broad range of chromatin phenomena in both qualitative and quantitative ways. Moreover, when introducing NGS to the transcriptome level (RNA-seq), it can be used to detect changes in gene expression, as discussed earlier in this review [37,61,62].

Conclusion

Linkage analysis and GWAS studies fail to fully explain disease heritability and do not address the causal nature of risk variants. NGS continues to fuel the discovery of disease-associated common and rare variants. The advances in analysis tools, such as Bayesian fine

mapping approaches and high performance computation algorithms, help to make full use of the current massive data to uncover relationships and infer the causality among complex data. Comprehensive sets of functional annotations (ENCODE, Roadmap and Blueprint projects) in the context of complex genomic structure can be used to predict function and guide experimentation, such as precision genome editing with the CRISPR-Cas (Clustered regulatory interspaced short palindromic repeats/CRISPR-associated) [63,64], to address the long standing question of disease mechanism and heterogeneity. Nonetheless, we still have not yet fully exploited analysis of GWAS data, such as 1) genetic studies in non-EU populations with different LD, especially important in SLE given the prevalence; 2) eQTL and epigenetic data in cells from non-EU populations for functional annotation; 3) epigenetic data in larger cohorts to look at inter-individual variation; 4) eQTL and epigenetic data from disease cohorts, to look for disease specific effects [65]. Studies based on these cohorts will advance our understanding of the disease mechanism, and ultimately speed up the arrival of the era of personalized medicine with genomic data incorporated into diagnosis, prognosis, and treatment in clinics.

Key points

1. The discovery of SLE-associated risk variants has accelerated in the past two years with huge sample size genome-wide and meta-analysis studies revealing novel loci in both coding and non-coding regions of the genome.
2. eQTL mapping incorporating co-localisation analysis of GWAS results help to identify the underlying causal genes.
3. The ENCODE, Roadmap and Blueprint projects which annotate non-coding regions have created comprehensive maps of the human genome.

4. SLE associated risk loci can be analysed bioinformatically, in the context of functional annotation to predict biological impact.
5. Functional validation is required for designating variants as 'causal variants', and facilitated by the availability of genome editing tools such as CRISPR technology to artificially create the variant in a model system relevant for disease.

Acknowledgements

None.

Financial support and sponsorship

China Scholarship Council (CSC) Funding: 201406380127

Conflicts of interest

There are no conflicts of interest.

Figure titles and legends

Supplementary Table 1. A summary of SLE risk loci.

Figure 1. SLE risk loci in genomic context

The CIRCOS plot [23] shows genes located within the SLE risk loci (84 in total) according to their genomic position. The full list of variants and locus genes for this plot is summarized in

supplementary Table 1. The red block in each chromosome indicates the centromere of the chromosome. Each chromosome arm is divided into cytogenetic bands of hg19.

Figure 2. Box plots of GRS across the five major population groups.

There are standard box plots showing medians, interquartile ranges and whiskers indicating 1.5 times the interquartile range (Tukey box plots) [21]. EUR, European, N498; AMR, Amerindian, N=347; SAS, South Asian, N=487; EAS, East Asian, N=503; AFR, African, N=657; from the 1000 Genome phase 3 release. The dashed line represents the increase in prevalence with the rank order (R1 represents the lowest prevalence, and R4 the highest).

Figure 3. Overview of co-localisation analysis of GWAS and eQTL.

This figure shows an example of eQTL analysis and the application of RTC for the causality inference. Firstly, we subset the genes within the cis-window (+/- 1Mb) of the disease-associated locus (rs2736340) and perform linear regression against the genotypes of the SNP. Co-localisation analysis of the GWAS signal and the eQTL signal was performed by calculating the RTC score. SNP-expression pairs with RTC > 0.9 were considered causal.

Figure 4. Schematic overview of fine mapping causal SNPs by integrating genetics and epigenetics.

This figure illustrates the functional annotation approach by an example, *BLK* (data unpublished). The epigenetic data of two histone markers (H3K27ac and H3K9ac) from three primary cell types (B cell, T cell and monocytes) (Roadmap Project) are represented for the target locus. This region contains 17 SNPs derived from 99% Bayesian credibility set of the risk locus. Rs2736340 is associated with SLE (Figure 3). rs922483 overlaps H3K27ac in all three cell types while it overlaps the H3K9ac peak in B cells only. Furthermore, rs922483

is in strong linkage disequilibrium (LD) ($r^2 = 0.98$) with rs2736340, indicating that there is transitive evidence due to the LD that rs922483 is also associated with SLE and is an eQTL. Therefore, rs922483 is the most likely functional SNP in this risk locus.

References:

1. Johnson AE, Gordon C, Palmer RG, Bacon PA. **The prevalence and incidence of systemic lupus erythematosus in Birmingham, England: Relationship to ethnicity and country of birth**. *Arthritis Rheum.* 1995, **38**:551–558.
2. Block SR, Winfield JB, Lockshin MD, et al. **Studies of twins with systemic lupus erythematosus. A review of the literature and presentation of 12 additional sets**. *Am. J. Med.* 1975, **59**:533–552.
3. Deafen D, Escalante A, Weinrib L, et al. **A revised estimate of twin concordance in systemic lupus erythematosus**. *Arthritis Rheum.* 1992, **35**:311–318.
4. Kuo C-F, 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.
5. Gaffney PM, Kearns GM, Shark KB, et al. **A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families**. *Proc. Natl. Acad. Sci. U. S. A.* 1998, **95**:14875–9.
6. 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.
7. Kozyrev S V, Abelson A-K, Wojcik J, et al. **Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus.** *Nat. Genet.* 2008, **40**:211–216.
8. Graham RR, Cotsapas C, Davies L, et al. **Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus.** *Nat. Genet.* 2008, **40**:1059–61.

9. Harley JB, Alarcón-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.

10. Morris DL, Taylor KE, Fernando MMA, 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.** The American Journal of Human Genetics. 2012 Nov 2;91(5):778-93.

11. Bentham J, Morris DL, Graham DSC, et al. **Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus**. Nat. Genet. 2015 Dec;47(12):1457-64.

* This is the largest SLE GWAS study in European population, comprising 7,219 cases and 15,991 controls, which provided considerable power to identify SLE risk loci. Ten novel loci were identified in European ancestry and an over-presentation of transcription factors were found among the SLE susceptibility genes, suggesting the regulatory roles of disease-associated variants.

12. Demirci FY, Wang X, Kelly JA, et al. **Identification of a New Susceptibility Locus for Systemic Lupus Erythematosus on Chromosome 12 in Individuals of European Ancestry**. Arthritis & Rheumatology. 2016 Jan 1;68(1):174-83.

13. Han J-W, Zheng H-F, Cui Y, et al. **Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus.** *Nat. Genet.* 2009, **41**:1234–1237.

14. Yang W, Shen N, Ye DQ, et al. **Genome-wide association study in asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus**. *PLoS Genet.* 2010, **6**.

15. Okada Y, Shimane K, Kochi Y, et al. **A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus eyrthematosus in Japanese**. *PLoS Genet* 2012, **8**:e1002455.

16. Lee H-S, Kim T, Bang SY, et al. **Ethnic specificity of lupus-associated loci identified in a genome-wide association study in Korean women.** *Ann. Rheum. Dis.* 2014, **73**:1240–5.

17. 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.

18. 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 Mar;48(3):323-30.

* This is the first SLE Immunochip study in Asian population, which identified 10 new loci for Asian population, increase the explained heritability of SLE to 24%.

19. Kaiser R, Taylor KE, Deng Y, et al. **Single-nucleotide polymorphisms in VKORC1 are risk factors for systemic lupus erythematosus in Asians**. *Arthritis Rheum.* 2013, **65**:211–215.

20. Zhang J, Zhang L, Zhang Y, et al. **Gene-Based Meta-Analysis of Genome-Wide Association Study Data Identifies Independent Single-Nucleotide Polymorphisms in ANXA6 as Being Associated With Systemic Lupus Erythematosus in Asian Populations**. Arthritis & Rheumatology. 2015 Nov 1;67(11):2966-77.

21. 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.

* A comparison of genetic association signals across the genome in European and Asian populations suggested that SLE susceptibility loci were shared extensively between both populations, motivating the first genome-wide trans-ancestral study. Meta-analysis of 3 GWASs in 2 populations identified 10 novel loci. Notably, this study suggested that the increased prevalence of SLE in non-European (including Asians) has a genetic basis by comparison of genetic risk scores (GRS) between populations

22. Molineros JE, Yang W, Zhou X, et al. **Confirmation of five novel susceptibility loci for Systemic Lupus Erythematosus (SLE) and integrated network analysis of 82 SLE susceptibility loci.** Human molecular genetics. 2017 Jan 20.

23. Gu Z, Gu L, Eils R, et al. **Circlize implements and enhances circular visualization in R**. *Bioinformatics* 2014, **30**:2811–2812.

24. Adrianto I, Wen F, Templeton A, et al. **Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus.** *Nat. Genet.* 2011, **43**:253–8.

25. Hitotsumatsu O, Ahmad RC, Tavares R, et al. **The Ubiquitin-Editing Enzyme A20 Restricts Nucleotide-Binding Oligomerization Domain Containing 2-Triggered Signals**. *Immunity* 2008, **28**:381–390.

26. Harley B. **Recent insights into the genetic basis of systemic lupus erythematosus**. *Genes Immun.* 2009, **10**:373–379.

27. Cho JH, Gregersen PK. **Genomics and the multifactorial nature of human autoimmune disease.** *N. Engl. J. Med.* 2011, **365**:1612–23.

28. 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.

29. So HC, Gui AHS, Cherny SS, Sham PC. **Evaluating the heritability explained by known susceptibility variants: A survey of ten complex diseases**. *Genet. Epidemiol.* 2011, **35**:310–317.

30. Yang J, Lee SH, Goddard ME, Visscher PM. **GCTA: A tool for genome-wide complex trait analysis**. *Am. J. Hum. Genet.* 2011, **88**:76–82.

31. Langefeld CD, Ainsworth HC, Cunninghame Graham DS, et al. **Transancestral mapping and genetic load in systemic lupus erythematosus**. *Nat. Commun.* 2017, **In press**.

* This study is the latest large-scale trans ancestral study in SLE using Immunochip, which included three ancestries (EA, AA, and HA) and have identified nine novel loci for EA, two for AA and two for HA, revealing evidence of sharing genetic risk loci between ancestries and evidence that each individual population carries unique genetic risk factors at the locus level and at the allelic level.

32. Manolio TA, Collins FS, Cox NJ, et al. **Finding the missing heritability of complex diseases.** *Nature* 2009, **461**:747–53.

33. Maller JB, McVean G, Byrnes J, et al. **Bayesian refinement of association signals for 14 loci in 3 common diseases.** *Nat. Genet.* 2012, **44**:1294–301.

34. Dixon AL, Liang L, Moffatt MF, et al. **A genome-wide association study of global gene expression**. *Nat Genet* 2007, **39**:1202–1207.

35. Choy E, Yelensky R, Bonakdar S, et al. **Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines**. *PLoS Genet* 2008, **4**:e1000287.

36. The 1000 Genomes Project Consortium: **A global reference for human genetic variation**. *Nature* 2015, **526**:68–74.

37. Westra H-JJ, Peters MJ, Esko T, et al. **Systematic identification of trans eQTLs as putative drivers of known disease associations**. *Nat Genet* 2013, **45**:1238–1243.

38. Consortium TGte, Ardlie KG, Deluca DS, et al. **The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans**. *Science. 2015 May 8;348(6235):648-60.*

39. Nica AC, Montgomery SB, Dimas AS, et al. **Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations**. *PLoS Genet.* 2010, **6**.

40. Giambartolomei C, Vukcevic D, Schadt EE, et al. **Bayesian test for colocalisation between pairs of genetic association studies using summary statistics**. *PLoS Genet* 2014, **10**:e1004383.

41. Odhams CA, Cortini A, Chen L, et al. **Mapping eQTLs with RNA-Seq Reveals Novel Susceptibility Genes, Non-Coding RNAs, and Alternative-Splicing Events in Systemic Lupus Erythematosus**. *Hum. Mol. Genet.* 2017, **0**:ddw417.

* This study illustrated the benefits of using RNA-seq as opposed to microarrays for eQTL mapping, due to more informative data generated by RNA-seq, suggesting that transcript profiling can be done on different layers for better explaining potential regulatory mechanisms.

42. Zoledziewska M, Costa G, Pitzalis M, et al. **Variation within the CLEC16A gene shows consistent disease association with both multiple sclerosis and type 1 diabetes in Sardinia**. *Genes Immun.* 2009, **10**:15–4879.

43. Sarda S, Hannenhalli S. **Next-generation sequencing and epigenomics research: a hammer in search of nails**. *Genomics Inf.* 2014, **12**:2–11.

44. Encode Consortium. **An integrated encyclopedia of DNA elements in the human genome**. *Nature* 2012, **489**:57–74.

45. Consortium RE, Kundaje A, Meuleman W, et al. **Integrative analysis of 111 reference human epigenomes**. Nature 2015, **518**:317–330.

** The largest collection of human epigenomes for primary cells and tissues generated by the Roadmap Epigenomics Consortium establishes global maps of regulatory elements, including modification patterns, DNA accessibility, DNA methylation and RNA expression, providing a comprehensive resource for studying relevant cell types for different traits and interpreting molecular basis of human disease.

46. Raj T, Rothamel K, Mostafavi S, et al. **Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes**. Science (80-.). 2014, **344**:519–23.

47. Fairfax BP, Makino S, Radhakrishnan J, et al. **Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles**. Nat Genet 2012, **44**:502–510.

48. Fairfax BP, Humburg P, Makino S, et al. **Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression**. Science (80-.). 2014, **343**:1246949.

49. Astle WJ, Elding H, Jiang T, et al. **The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease**. Cell 2016, **167**:1415–1429.e19.

50. Javierre BM, Sewitz S, Cairns J, et al. **Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters**. Cell 2016, **167**:1369–1384.e19.

** This study is part of the IHEC consortium, which deploys a promoter Hi-C approach to generate high-resolution maps of promoter interactions in 17 human blood cell types, providing a wealthy resource for studying promoter interaction patterns of disease associated variants in immune cell types to reveal insights into genomic regulatory mechanisms.

51. Chen L, Ge B, Casale FP, et al. **Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells**. Cell 2016, **167**:1398–1414.e24.

** As part of the IHEC consortium, this study generates genome, transcriptome, and epigenome reference panels in three immune cell types (neutrophils, monocytes, and T cells) from 200 individuals, providing a resourceful database for functionally mapping disease risk variants by

integrating genetics and epigenetics. Summary statistics of QTLs (eQTL, methylation QTL (meQTL) and histone QTLs (hQTLs) from this study can be accessed through the IHEC web portal.

52. 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.

53. Royer-Bertrand B, Rivolta C. **Whole genome sequencing as a means to assess pathogenic mutations in medical genetics and cancer**. Cell. Mol. Life Sci. 2015, **72**:1463–1471.

54. Meyer CA, Liu XS. **Identifying and mitigating bias in next-generation sequencing methods for chromatin biology.** Nat. Rev. Genet. 2014, **15**:709–21.

55. Barski A, Cuddapah S, Cui K, et al. **High-Resolution Profiling of Histone Methylations in the Human Genome**. Cell 2007, **129**:823–837.

56. Boyle AP, Song L, Lee BK, et al. **High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells**. Genome Res. 2011, **21**:456–464.

57. Neph S, Vierstra J, Stergachis AB, et al. **An expansive human regulatory lexicon encoded in transcription factor footprints**. Nature 2012, **488**:83–90.

58. Lieberman-aiden E, Berkum NL Van, Williams L, et al. **Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome**. Science (80-.). 2009, **326**:289–293.

59. Buenrostro JD, Giresi PG, Zaba LC, et al. **Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.** Nat. Methods 2013, **10**:1213–8.

60. Degner JF, Marioni JC, Pai AA, et al. **Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data**. Bioinformatics 2009, **25**:3207–3212.

61. Rozowsky J, Abyzov A, Wang J, et al. **AlleleSeq: analysis of allele-specific expression and binding in a network framework.** Mol. Syst. Biol. 2011, **7**:522.

62. Wu X, Scott DA, Kriz AJ, et al. **Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells**. Nat Biotechnol 2014, **32**:670–676.

63. Zhou Y, Zhu S, Cai C, et al. **High-throughput screening of a CRISPR/Cas9 library for**

functional genomics in human cells. Nature 2014, **509**:487–91.

64. Jacquemin C, Schmitt N, Contin-Bordes C, et al. **OX40 Ligand Contributes to Human Lupus Pathogenesis by Promoting T Follicular Helper Response**. Immunity 2015, **42**:1159–1170.

** This study demonstrates that OX40L is expressed by myeloid antigen-presenting cells in active SLE patients, indicating that the expression of particular disease associated gene is context-specific, i.e. cell types and the disease status in this case.

Figure 1. SLE risk loci in genomic context_Original

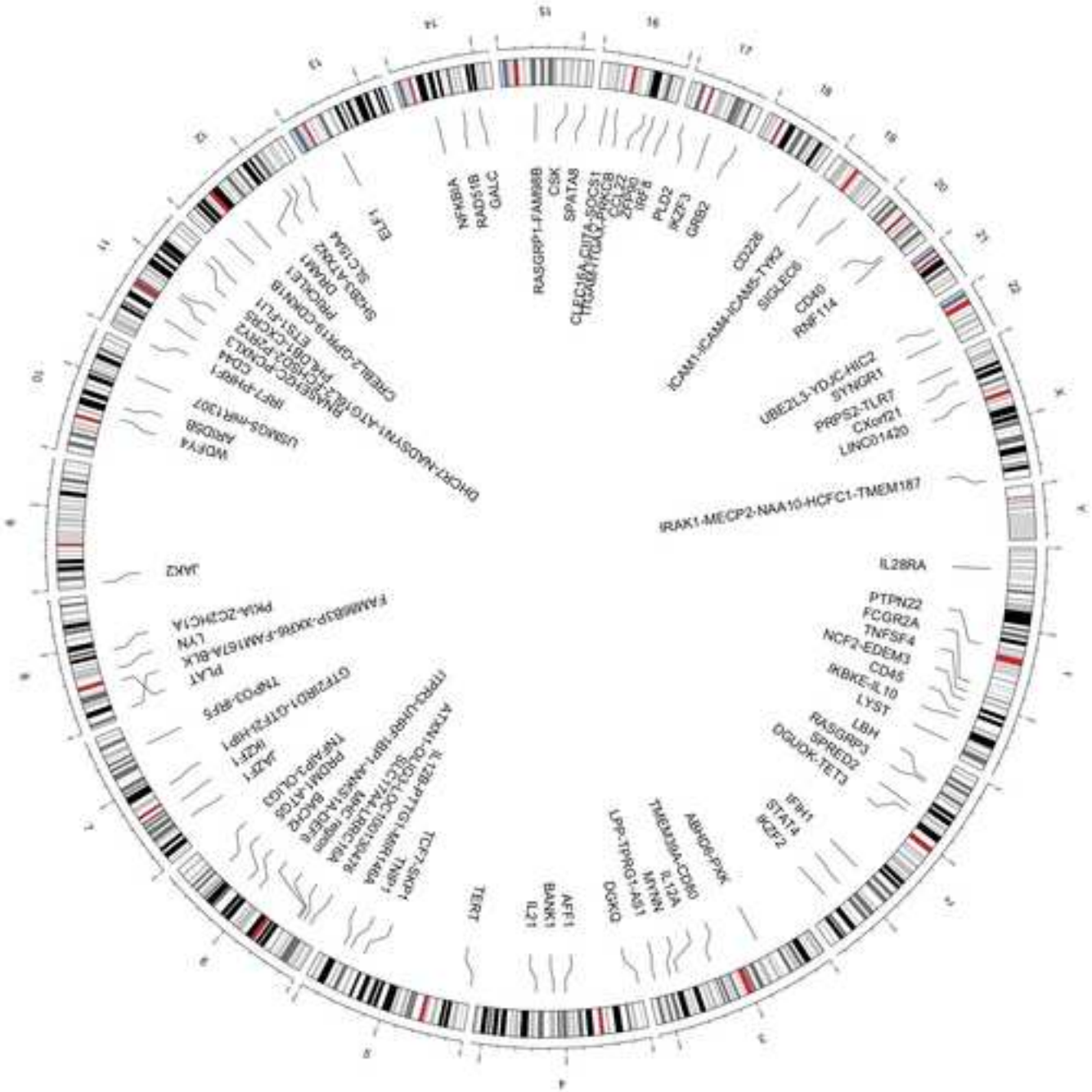


Figure 2. Box plots of GRS across the five major population groups_Previously published

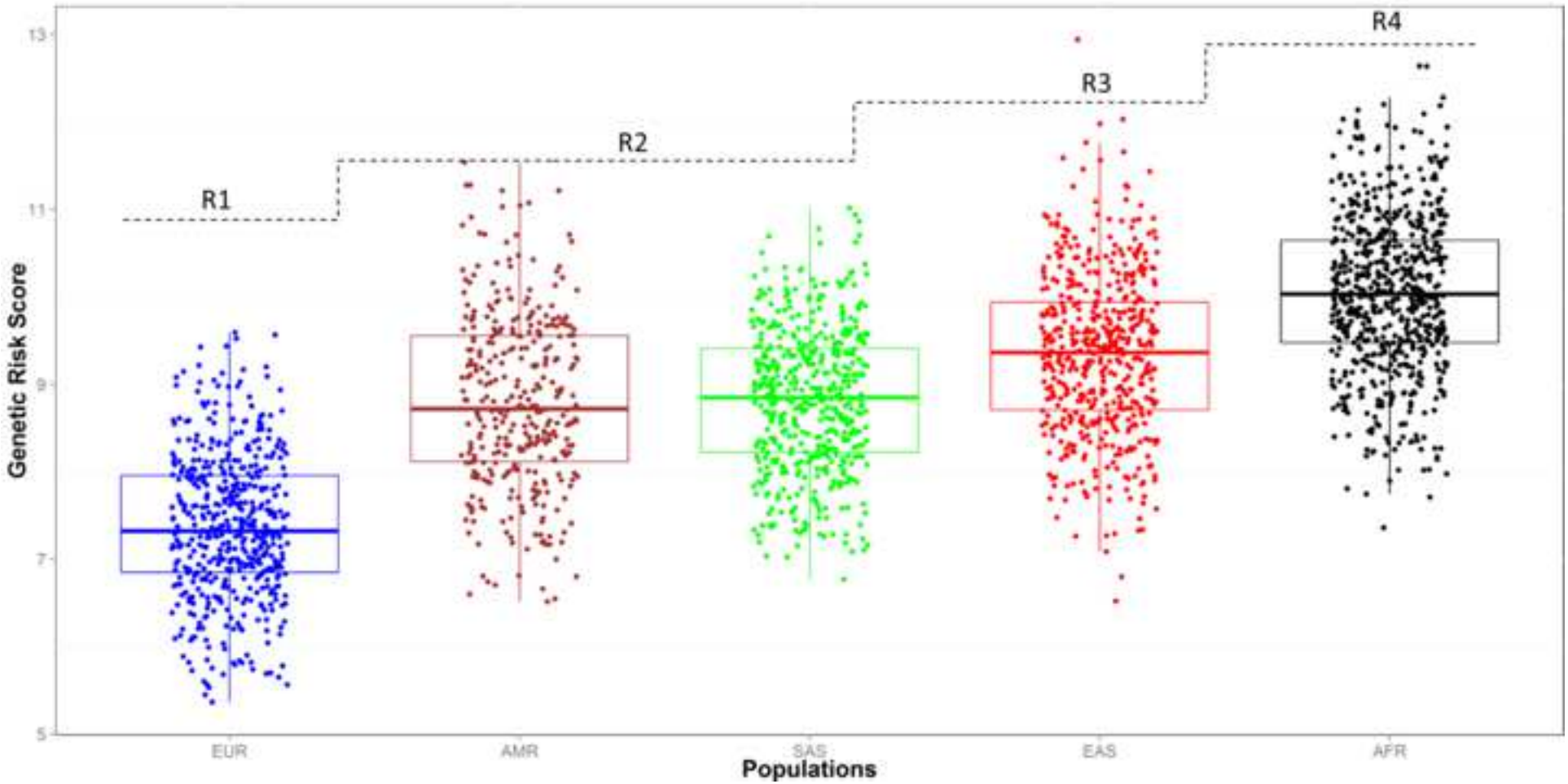


Figure 3. Overview of co-localisation analysis of GWAS and eQTL_Original

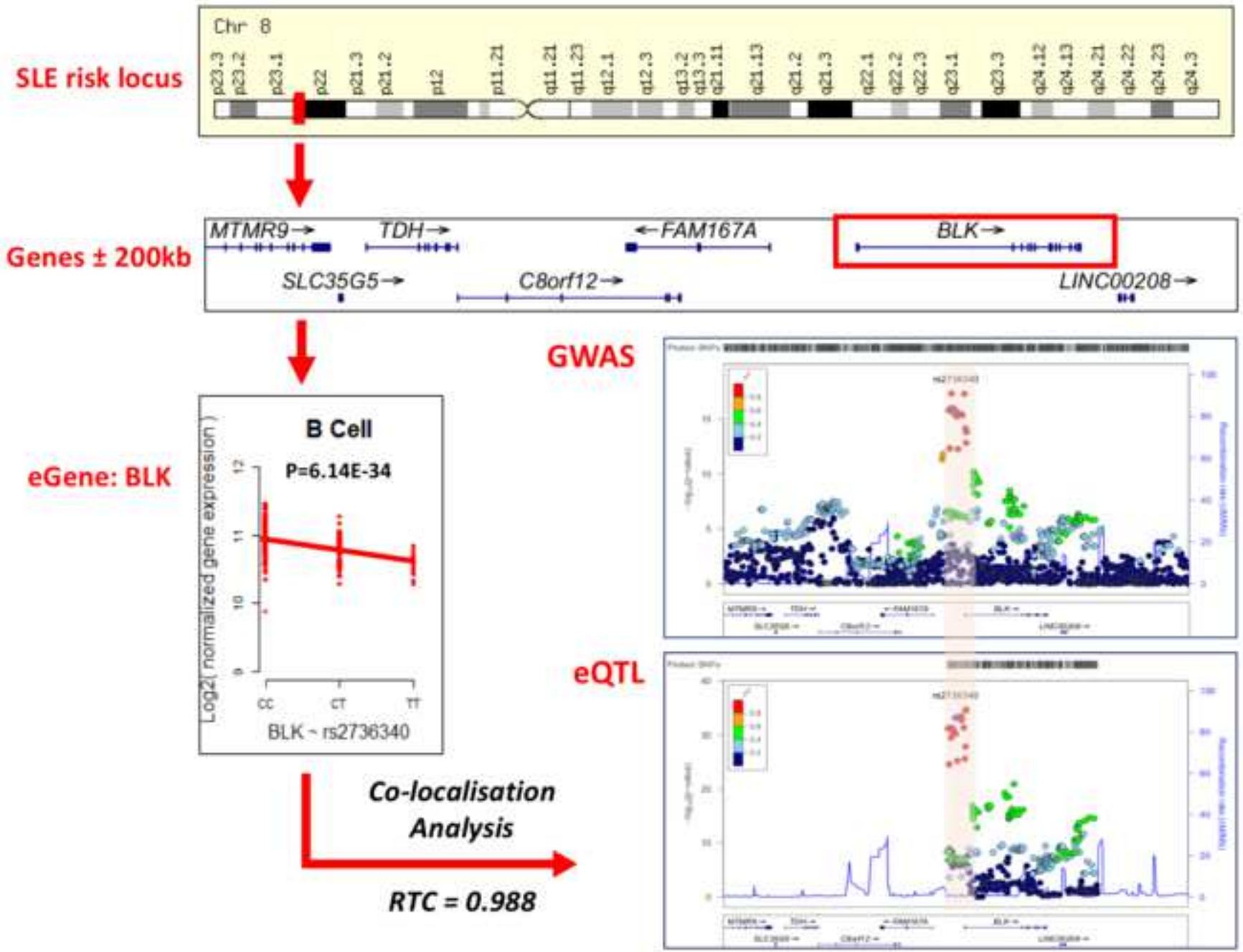
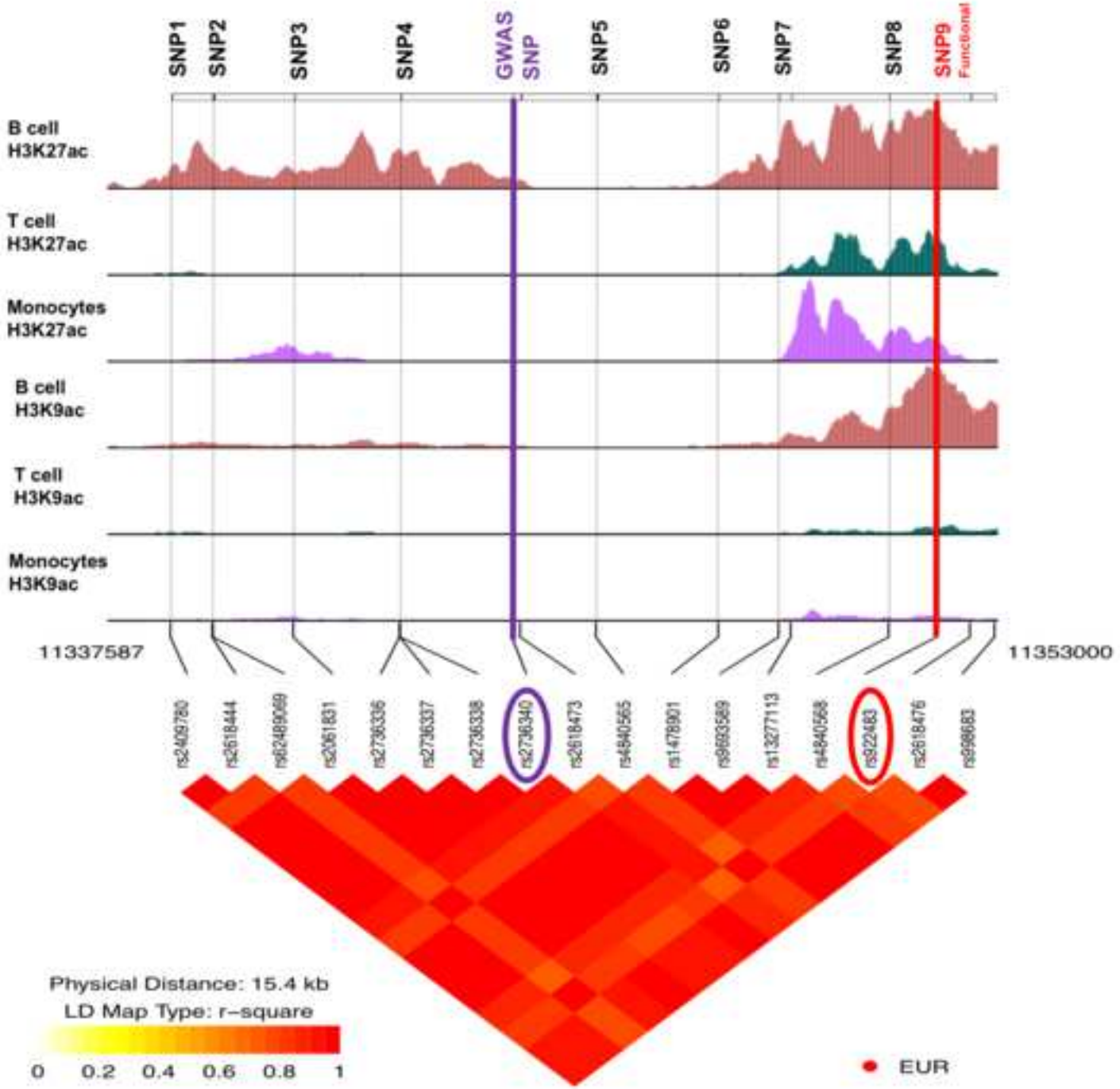


Figure 4. Schematic overview of fine mapping causal SNPs by integrating genetics and epigenetics_Original



Supplementary Table 1. A summary of SLE risk loci