



King's Research Portal

DOI:

[10.1097/BOR.0000000000000411](https://doi.org/10.1097/BOR.0000000000000411)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Chen, L., Morris, D. L., & Vyse, T. J. (2017). Genetic advances in systemic lupus erythematosus: an update. *Current Opinion in Rheumatology*, 29(5), 423-433. <https://doi.org/10.1097/BOR.0000000000000411>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

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 Chen¹, David L Morris¹, 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.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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

1 the non-coding region of the human genome as well as new technologies and statistical
2 methods, in order to apply this knowledge to the understanding of SLE genetics.
3
4
5
6
7
8
9

10 **Insights from GWAS**

11 Genetic linkage analysis and candidate gene association studies identified several SLE
12 susceptibility loci (e.g. HLA-DR2/DR3) [5]. Nevertheless, the advent and application of
13 GWAS dramatically advanced knowledge of the genetic aetiology of SLE.
14
15
16
17
18
19
20
21

22 There have been seven SLE GWAS in European population [6–12], six Asian GWAS [13–
23 18], subsequent meta-analysis and large-scale replication studies [19–22], published since
24 2008. Currently, 84 genetic loci are implicated as SLE risk (Figure 1: The CIRCOS plot [23]
25 and supplementary Table 1), which, in order to avoid likely spurious associations, includes
26 genetic associations with a P value less than 5×10^{-8} tested in a total sample size of at least
27 1000 individuals. The interactive version of a continually updated resource with details on
28 SLE associations can be access through the following link: [http://insidegen.com/insidegen-
29 LUPUS-Associations.html](http://insidegen.com/insidegen-LUPUS-Associations.html).
30
31
32
33
34
35
36
37
38
39
40
41

42 With the caveat that the majority of mechanisms remain to be elucidated, it appears that the
43 risk loci associated with SLE influence immune cell function. Although functional studies
44 are designed with a *priori* hypotheses in mind, key pathogenic pathways that are likely
45 influenced by SLE-associated gene products include: immune complex processing and
46 phagocytosis; DNA degradation, apoptosis and clearance of cellular debris; neutrophil and
47 monocytes signalling; Toll-like receptor and/or type I interferon signalling; nuclear factor
48 kappaB activation; B and T-cell function and signalling. Some genes associated with SLE
49 may act through several pathways. For example, *TNFAIP3*, encoding the ubiquitin-editing
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 enzyme A20, is a key regulator of nuclear factor-kappa-B (NF-kB)-derived pro-inflammatory
2 responses, which is involved in both adaptive and innate immune pathways [24,25]. These
3
4 SLE susceptibility loci contain predominantly common (frequency of > 0.1% in the general
5
6 population) associated variants that have been confirmed among multiple ancestries,
7
8
9 suggesting shared mechanisms in disease aetiology [26–28].
10
11
12
13

14 European GWAS

15
16 The largest European GWAS of SLE conducted by our group [11], comprised 7,219 SLE
17
18 cases and 15,991 controls of European decent, provided considerable power to detect disease
19
20 risk loci. Notably, the study identified 43 susceptibility loci, ten of which were novel loci:
21
22 *SPRED2*, *IKZF2*, *IL12A*, *TCF7-SKP1*, *DHCR7-NADSYN1*, *SH2B3*, *RAD51B*, *CIITA-SOCS1*,
23
24 *PLD2*, and *CXorf21*. One of the great challenges posed by interpreting GWAS data is
25
26 determining the causal genes implicated by the genetic association data. As will be
27
28 discussed and amplified below, we put some considerable effort into this process before
29
30 naming the genes in the above list. Irrespective of the underlying causal genes, we can
31
32 conclude that the heritability explained by the risk alleles mapped at these loci is 15.3%,
33
34 which is a large increase over the 8.7% reported by So et al [29] in 2011 using the same
35
36 measure.
37
38
39
40
41
42
43
44
45

46 Asian GWAS

47
48 An extensive large-scale fine mapping study using ImmunoChip conducted in 4,478 SLE
49
50 cases and 12,656 controls from six East Asian cohorts identified 10 novel loci [18] in Asians,
51
52 encompassing *GTF2IRD1-GTF2I*, *DEF6*, *IL12B*, *TCF7*, *TERT*, *CD226*, *PCNXL3*, *RASGRP1*,
53
54 *SYNGR1*, and *SIGLEC6*. Some of these were previously reported to be associated in
55
56
57
58
59
60
61
62
63
64
65

1 Europeans, for example, *DEF6* and *TCF7*. The identification of these risk loci increased the
2 explained heritability to 24% in Asian SLE.
3

4 Trans ancestry meta analyses of GWAS

5 A comparison of genetic association signals across the genome in European and Asian
6 populations suggested that SLE susceptibility loci were shared extensively between both
7 populations [21]. This motivated a trans ancestral approach at the genome-wide level to
8 provide evidence of shared genetic components in the two populations and search for
9 additional SLE associated loci. The study by Morris and Sheng et al [21], that combined
10 three GWAS from two ethnicities: Chinese (1659 cases and 3,398 controls) and European
11 (4,044 cases + 6,959 controls), found evidence of considerable commonality in terms of SLE
12 association signals as well as mapping novel susceptibility loci, including *CD45*, *IKBKE*,
13 *LBH*, *LPP-TPRG1-AS1*, *ATXN1*, *BACH2*, *GTF2I*, *JAK2*, *RNASEH2C*, and *ZFP90*. Notably,
14 this study suggested that the increased prevalence of SLE in non-European (including Asians)
15 has a genetic basis by comparison of genetic risk scores (GRS) between populations (Figure
16 2) [21]. Moreover, by using all genotyped SNPs (DNA chip) to calculate heritability
17 explained, the explained variation (Vg) increase to 28% in Chinese subjects and 27%
18 Europeans using the GCTA algorithm [30]. While there are still some uncertainties in the
19 methodology for calculating heritability explained, this shows very strong evidence that we
20 are making progress on the understanding of SLE heritability.
21

22 The latest large-scale trans ancestral study using ImmunoChip [31], comprising three
23 ancestries: European (EA: 6,748 cases and 11,516 controls), African-American (AA: 2,970
24 case and 2,452 controls), and Hispanic Amerindian (HA: 1,872 cases and 2,016 controls),
25 have identified nine novel loci for EA (*TMEM39A-TIMMDC1*, *DGKQ*, *LRRC16A*, *SLC17A4*,
26 *OLIG3-LOC100130476*, *GTF2IRD1-GTF2I*, *FAM86B3P*, *PKIA-ZC2HC1A*, and *GRB2*),
27 two for AA (*PTTG1-MIR146A* and *PLAT*) and two for HA (*GALC* and *CLEC16A*). By
28 comparing results across different populations, both ancestry-dependent and ancestry-
29

1 independent contributions to SLE risk are identified with the caveat of unequal cohort sizes.

2 The study reveals evidence of sharing of genetic risk loci between ancestries as well as
3
4 evidence that each individual population carries unique genetic risk factors at the locus level
5
6 and at the allelic level.
7

8 9 10 11 **Missing heritability**

12 In summary, the chip heritability identified by the latest GWAS have explained around 28%
13
14 of the disease heritability: a marked improvement on 8.3% calculated in 2011 [29].
15
16

17 However, there is still one third of heritability left to explain, if we assume that the total
18
19 estimated heritability is 43.9%. Explanations for the missing heritability, including larger
20
21 numbers of variants of smaller effect, rarer variants (possibly with larger effects) that are not
22
23 present on genotyping arrays or structural variants poorly captured by existing arrays, as well
24
25 as epigenetic modifications, have been suggested [32]. Innovations in genotyping and
26
27 sequencing technologies, like the Immuno-chip platform [18,31] and next generation
28
29 sequencing (NGS, as described below) will advance the investigation into common and rare
30
31 variants and potential effects on the immune system, enhancing our understanding of the
32
33 genetic risk of SLE.
34
35
36
37
38
39
40
41
42
43

44 The LD that exists in the human genome facilitates the mapping of risk loci by reducing the
45
46 number of genetic variants required for GWAS; however, the same correlation between
47
48 genetic polymorphisms at these susceptibility loci then bedevils attempts to identify the
49
50 actual causal allele(s) at risk loci. Bayesian fine mapping approaches had been proposed to
51
52 derive smaller sets of SNPs (termed ‘credibility sets’) as the most likely causal variants at
53
54 risk loci [33]. Nevertheless, statistical methods are inadequate to fully resolve the problem
55
56 caused by LD. In order to further pursue likely causal SNPs within any given credibility set,
57
58
59
60
61
62
63
64
65

1 the functional effect of SNPs can be studied in silico. As the majority of variants within
2 causal credibility sets are non-coding [34,35], function is inferred using gene transcript
3 expression data and epigenetic modification data (as described below) (Figure 3 and Figure
4
5
6
7 4).

11 **Application of eQTL mapping to GWAS results**

12
13
14 Assisted by dense genome coverage of the reference panel from the 1000 Genome project
15 [36], imputation and Bayesian inference provided evidence for missense variants
16
17 underpinning association for eight genes, including *PTPN22*, *FCGR2A*, *NCF2*, *IFIH1*,
18
19 *WDFY4*, *ITGAM*, *PLD2*, and *TYK2* [11]. However, as mentioned above, the majority (85%
20
21 ~ 90%) of disease associated loci in SLE are located outside of protein-coding regions,
22
23 suggesting that the underlying mechanism is likely regulatory, and so might exert their
24
25 function through altering gene expression rather than by altering protein structure. Of note,
26
27 an over-representation (n=16) of transcription factors among the 43 SLE susceptibility genes
28
29 have been annotated in our recent European GWAS [11], further indicating that perturbed
30
31 gene regulation was a major functional risk factor for SLE. Expression quantitative trait loci
32
33 (eQTLs) mapping, which combines genome-wide expression profiling and genome-wide
34
35 marker-based genotyping, takes advantage of the heritability of gene expression profiles to
36
37 identify genetic variants that correlated with changes in gene expression. eQTLs can be
38
39 classified as “in cis” (locally) or “in trans” (at a distance) based on their physical distance
40
41 from the regulated gene.
42
43
44
45
46
47
48
49
50

51
52
53 Some studies [18,22] used public databases, such as the whole blood eQTL browser
54
55 (<http://genenetwork.nl/bloodeqtlbrowser/>) [37] and tissue-specific GTEx portal
56
57 (<http://www.gtexportal.org/home/>) [38], to determine whether the disease-associated SNP is
58
59
60
61
62
63
64
65

1 a significant eQTL. Of note, there exists some limitations when applying eQTL analysis to
2 the GTEx whole blood datasets, as in autoimmunity, we seek eQTLs in specific immune cell
3 subsets. In order to highlight the potential causal genes at the susceptibility loci robustly, it is
4 essential to integrate the disease association and eQTL data using a co-localisation approach.
5 That is, to establish that the same genetic variants that underlie the disease association also
6 underlie the eQTL. The presence of LD in the genome can readily obfuscate this overlap.
7 Co-localisation methods, like the regulatory trait concordance (RTC) [39], conditional
8 analysis [30], and Bayesian co-localisation [40], can be employed to infer that the disease
9 association and eQTL have the same allelic basis. As many variants have weak eQTL
10 effects, erroneous conclusions will be made if analyses for co-localisation are not performed.
11 An example of co-localisation analysis of eQTL and GWAS is shown in Figure 3.
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

29 Recent studies by Morris et al [11,21] and Odhams et al [41] examined the functional
30 outcome of SLE associated variants through the integration of GWAS and eQTL data from
31 various cell types ex vivo, involving T cells, B cells, NK cells, stimulated and resting
32 monocytes, as well as lymphoblastoid cell lines (LCL). By integrating the results of eQTL
33 and RTC analysis, they found evidence to support the role of causal genes as candidates at a
34 given locus. For example, *SOCS1* (Suppressor of Cytokine Signalling 1) was found to be a
35 suggestive causal gene at the locus tagged by the SNP rs9652601 (with a RTC score higher
36 than 0.9), rather than *CLECI6A* (C-Type Lectin Domain Family 16 Member A), even though
37 the risk variant resides within the latter one - a gene previously reported as relating to other
38 autoimmune diseases [42]. Moreover, the Odhams et al's study [41] illustrated the benefits
39 of using RNA-seq as opposed to microarrays for eQTL mapping, due to more informative
40 data generated by RNA-seq. With RNA-seq, transcript profiling can be done on the gene-
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 level, exon-level and splice-junction-level, which is more effective in explaining potential
2 regulatory mechanisms.
3

4 Nevertheless, we believe that many eQTLs related to SLE risk alleles remain unidentified,
5 data from diverse stimulations and time points will be required, as well as gene expression
6 data from patient material, to reveal the full eQTL landscape of SLE genetics.
7
8
9

10 **Epigenetics to annotate functional / regulatory variants**

11
12
13
14
15

16 An approach that is complementary to eQTL analyses, to examine the regulatory function of
17 non-coding genetic variants, is to study gene regulation with epigenetics. Epigenetic
18 modifications, a term coined to describe genome-wide chromatin modification, including
19 DNA methylation, histone modifications, chromatin accessibility, microRNA regulations,
20 and 2D chromatin interactions [43], constitute an additional layer of genomic regulation, and
21 may serve as a dynamic link between genotype and phenotype. Such changes in DNA and
22 chromatin structure correlate with changes in chromatin accessibility and transcription factor
23 binding.
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 The Encyclopedia of DNA elements (ENCODE) project (<https://www.encodeproject.org/>)
41 [44] has systematically mapped regions of transcription, transcription factor association,
42 chromatin structure and histone modification, and assigns biochemical functions for 80% of
43 the genome, in particular outside of the protein-coding regions. Overall, the project has
44 provided an expansive resource to define the functional DNA elements for biomedical
45 research, although the available cell types or cell lines are limited. The cells of closest
46 immune relevance in ENCODE Tier 1 and Tier 2 are LCLs (GM12878), B cells (CD20+) and
47 monocytes (CD14+), as well as T cells (CD4+) and peripheral blood mononuclear cell
48 (PBMC) in Tier 3. A recent ImmunoChip study in Asians [18] took advantage of ENCODE
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 data to map the underlying loci. For example, one of the signals (rs73366469) identified in
2 this study was located between two ‘general transcription factor’ genes, *GTF2I* and
3
4 *GTF2IRD1*. By integrating the ENCODE data, they found that an indel SNP rs587608058
5
6 ($r^2=0.81$), ~1000bp from rs73366469, lay within conserved enhancer, active chromatin and
7
8 transcription factor binding sites in LCLs and CD4+ T cells. In addition, this region was
9
10 found to overlap the transcription start sites for *GTF2I* and *VCF* through chromatin
11
12 interacting analysis and chromosome conformation capture (Hi-C) analysis, providing
13
14 evidence for the potential causal variants and genes at this locus for further study.
15
16
17
18
19
20
21

22 The Roadmap epigenomics project (<http://www.roadmapepigenomics.org/>) [46] integrated
23
24 analysis of 111 reference human epigenomes to obtain a comprehensive map of the human
25
26 epigenomic landscape across a large collection of primary cells, including immune cells, and
27
28 tissues. This map is extremely useful for studies of genome interpretation, gene regulation,
29
30 cellular differentiation, genome evolution, genetic variation and human disease. In our meta
31
32 GWAS analysis of Chinese and European data [21], the histone modification markers,
33
34 including acetylation markers (H3K27ac, H3K9ac) and methylation markers (H3K27me3 and
35
36 H3K9me3), from blood cell types were used to investigate the potential regulatory function
37
38 of the target risk loci. For example, there are several genes, including *SRGAP2*, *SRGAP2D*,
39
40 *IKBKE*, *RASSF5*, *EIF2D* and *DYRK3*, located within ± 200 kb of the lead GWAS SNP
41
42 rs2297550. The GWAS SNP was also found to be a putative eQTL for *IKBKE*, with the SLE
43
44 risk allele correlated with reduced expression in CD4+ T cells [47], CD19+ B cells [48] and
45
46 NK cells (data unpublished), but with increased expression in CD14+ monocytes [49].
47
48
49
50
51
52
53
54 *IKBKE* encodes a noncanonical I-kappa-B kinase (IKK) that is essential in regulating
55
56 inflammatory responses to viral infection by activating the type I interferon, NF-kB and
57
58 STAT signalling pathways, suggesting *IKBKE* might be the potential causal gene. Moreover,
59
60
61
62
63
64
65

1 there is an intense histone acetylation peak around the associated SNP rs2297550, indicating
2 that rs2297550 may be a potential causal variant [21]. Figure 4 shows an example of fine
3 mapping causal SNPs by integrating genetics and epigenetics.
4
5
6
7
8

9 Another recent completed large-scale epigenomic project, the Blueprint project
10 (<http://www.blueprint-epigenome.eu/>) [50–52], has impressively shown how epigenetic
11 information and analyses can help to study the cellular mechanisms associated with complex
12 human diseases. Moreover, the Blueprint consortium generated three comprehensive
13 reference panels, including genome (whole genome sequencing), transcriptome (RNA-seq),
14 and epigenome (DNA methylation and histone modification), in three immune cells
15 (Neutrophils, monocytes and T cells) from nearly 200 individuals to characterize the
16 contributions of diverse genomic inputs to transcriptional variation. Summary data from
17 these panels can be accessed through <http://blueprint-dev.bioinfo.cnio.es/WP10/> .
18
19 High-resolution maps of promoter interactions [51] generated by ‘Promoter capture Hi-C’
20 (PChi-C) make it possible to study the long range regulatory in the three-dimensional nuclear
21 space. By integrating PChi-C data with disease-associated SNPs generated by GWAS, we
22 can prioritize the putative target genes for the risk loci. The promoter interactomes map may
23 serve as a more robust method to define cis-eQTLs rather than by distance, revealing insights
24 into genomic regulatory mechanisms of diseases.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 **Next generation sequencing (NGS) in the genome research**

50 With the development of NGS, high-throughput technologies that are now widely used in
51 genome research, any part of the genome can be sequenced. Based on the coverage of the
52 genome, NGS strategies can be classified by scale: target region sequencing, whole-exome
53
54
55
56
57
58
59
60
61
62
63
64
65

1 sequencing (WES), and whole-genome sequencing (WGS). Targeted resequencing of risk
2 loci in disease cohorts may facilitate the identification of rare variants at common-allele-
3 associated loci [53]. WES captures all coding exons covering 1~2% of the genome.
4
5 Nevertheless, as mentioned above, approximately 85~90% of the risk loci associated with
6
7 SLE are located outside of the coding-regions. Compared to WES, WGS can capture the
8
9 majority of the genome, which facilitates delineation of exon duplications and gene fusions,
10
11 and non-coding regions that might be missing by WES. However, the higher cost and time
12
13 consuming bioinformatics analyses restrict the application of WGS [54]. In future, with the
14
15 decreasing cost of sequencing and newly developed computation algorithms, WGS will be
16
17 increasingly utilised.
18
19

20
21
22
23
24 Incorporating with a wide range of chromatin profiling experiments, NGS is applied to
25
26 investigate chromatin biology by identifying genomic loci that are occupied by nucleosomes,
27
28 bound to transcription factors, or accessible to nuclease cleavage [55]. Technologies such as
29
30 ChIP-seq [56], FAIRE-seq, DNase-seq [57,58], Hi-C [59], and ATAC-seq [60] enable
31
32 genome-wide investigations of a broad range of chromatin phenomena in both qualitative and
33
34 quantitative ways. Moreover, when introducing NGS to the transcriptome level (RNA-seq),
35
36 it can be used to detect changes in gene expression, as discussed earlier in this review
37
38
39
40
41 [37,61,62].
42
43
44
45
46
47
48
49

50 **Conclusion**

51
52
53
54 Linkage analysis and GWAS studies fail to fully explain disease heritability and do not
55
56 address the causal nature of risk variants. NGS continues to fuel the discovery of disease-
57
58 associated common and rare variants. The advances in analysis tools, such as Bayesian fine
59
60
61
62
63
64
65

1 mapping approaches and high performance computation algorithms, help to make full use of
2 the current massive data to uncover relationships and infer the causality among complex data.
3
4 Comprehensive sets of functional annotations (ENCODE, Roadmap and Blueprint projects)
5
6 in the context of complex genomic structure can be used to predict function and guide
7
8 experimentation, such as precision genome editing with the CRISPR-Cas (Clustered
9
10 regulatory interspaced short palindromic repeats/CRISPR-associated) [63,64], to address the
11
12 long standing question of disease mechanism and heterogeneity. Nonetheless, we still have
13
14 not yet fully exploited analysis of GWAS data, such as 1) genetic studies in non-EU
15
16 populations with different LD, especially important in SLE given the prevalence; 2) eQTL
17
18 and epigenetic data in cells from non-EU populations for functional annotation; 3) epigenetic
19
20 data in larger cohorts to look at inter-individual variation; 4) eQTL and epigenetic data from
21
22 disease cohorts, to look for disease specific effects [65]. Studies based on these cohorts will
23
24 advance our understanding of the disease mechanism, and ultimately speed up the arrival of
25
26 the era of personalized medicine with genomic data incorporated into diagnosis, prognosis,
27
28 and treatment in clinics.
29
30
31
32
33
34
35
36
37
38
39
40

41 **Key points**

- 42
43
44
45 1. The discovery of SLE-associated risk variants has accelerated in the past two years
46
47 with huge sample size genome-wide and meta-analysis studies revealing novel loci in
48
49 both coding and non-coding regions of the genome.
50
51
- 52
53 2. eQTL mapping incorporating co-localisation analysis of GWAS results help to
54
55 identify the underlying causal genes.
56
- 57
58 3. The ENCODE, Roadmap and Blueprint projects which annotate non-coding regions
59
60 have created comprehensive maps of the human genome.
61
62

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
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.

16 **Acknowledgements**

17
18
19
20
21
22
23
24
25

None.

26 **Financial support and sponsorship**

27
28
29
30
31
32
33
34
35

China Scholarship Council (CSC) Funding: 201406380127

36 **Conflicts of interest**

37
38
39
40
41
42
43
44
45

There are no conflicts of interest.

46 **Figure titles and legends**

47
48
49
50
51
52
53
54

Supplementary Table 1. A summary of SLE risk loci.

55
56

Figure 1. SLE risk loci in genomic context

57
58
59
60
61
62
63
64
65

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

1 supplementary Table 1. The red block in each chromosome indicates the centromere of the
2 chromosome. Each chromosome arm is divided into cytogenetic bands of hg19.
3
4
5
6

7 Figure 2. Box plots of GRS across the five major population groups.
8

9 There are standard box plots showing medians, interquartile ranges and whiskers indicating
10 1.5 times the interquartile range (Tukey box plots) [21]. EUR, European, N=498; AMR,
11 Amerindian, N=347; SAS, South Asian, N=487; EAS, East Asian, N=503; AFR, African,
12 N=657; from the 1000 Genome phase 3 release. The dashed line represents the increase in
13 prevalence with the rank order (R1 represents the lowest prevalence, and R4 the highest).
14
15
16
17
18
19
20
21
22
23

24 Figure 3. Overview of co-localisation analysis of GWAS and eQTL.
25

26 This figure shows an example of eQTL analysis and the application of RTC for the causality
27 inference. Firstly, we subset the genes within the cis-window (+/- 1Mb) of the disease-
28 associated locus (rs2736340) and perform linear regression against the genotypes of the SNP.
29
30 Co-localisation analysis of the GWAS signal and the eQTL signal was performed by
31 calculating the RTC score. SNP-expression pairs with $RTC > 0.9$ were considered causal.
32
33
34
35
36
37
38
39
40
41

42 Figure 4. Schematic overview of fine mapping causal SNPs by integrating genetics and
43 epigenetics.
44

45 This figure illustrates the functional annotation approach by an example, *BLK* (data
46 unpublished). The epigenetic data of two histone markers (H3K27ac and H3K9ac) from
47 three primary cell types (B cell, T cell and monocytes) (Roadmap Project) are represented for
48 the target locus. This region contains 17 SNPs derived from 99% Bayesian credibility set of
49 the risk locus. Rs2736340 is associated with SLE (Figure 3). rs922483 overlaps H3K27ac in
50 all three cell types while it overlaps the H3K9ac peak in B cells only. Furthermore, rs922483
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 is in strong linkage disequilibrium (LD) ($r^2 = 0.98$) with rs2736340, indicating that there is
2 transitive evidence due to the LD that rs922483 is also associated with SLE and is an eQTL.
3
4 Therefore, rs922483 is the most likely functional SNP in this risk locus.
5
6
7
8
9
10
11
12

13 **References:**

- 14 1. Johnson AE, Gordon C, Palmer RG, Bacon PA. **The prevalence and incidence of systemic**
15 **lupus erythematosus in Birmingham, England: Relationship to ethnicity and country of**
16 **birth.** *Arthritis Rheum.* 1995, **38**:551–558.
- 17 2. Block SR, Winfield JB, Lockshin MD, et al. **Studies of twins with systemic lupus**
18 **erythematosus. A review of the literature and presentation of 12 additional sets.** *Am. J.*
19 *Med.* 1975, **59**:533–552.
- 20 3. Deafen D, Escalante A, Weinrib L, et al. **A revised estimate of twin concordance in**
21 **systemic lupus erythematosus.** *Arthritis Rheum.* 1992, **35**:311–318.
- 22 4. Kuo C-F, Grainge MJ, Valdes AM, et al. **Familial Aggregation of Systemic Lupus**
23 **Erythematosus and Coaggregation of Autoimmune Diseases in Affected Families.** *JAMA*
24 *Intern. Med.* 2015, **175**:1518–1526.
- 25 5. Gaffney PM, Kearns GM, Shark KB, et al. **A genome-wide search for susceptibility genes in**
26 **human systemic lupus erythematosus sib-pair families.** *Proc. Natl. Acad. Sci. U. S. A.* 1998,
27 **95**:14875–9.
- 28 6. Hom G, Graham RR, Modrek B, et al. **Association of Systemic Lupus Erythematosus with**
29 **C8orf13–BLK and ITGAM–ITGAX.** *N Engl J Med* 2008, **358**:900–909.
- 30 7. Kozyrev S V, Abelson A-K, Wojcik J, et al. **Functional variants in the B-cell gene BANK1**
31 **are associated with systemic lupus erythematosus.** *Nat. Genet.* 2008, **40**:211–216.
- 32 8. Graham RR, Cotsapas C, Davies L, et al. **Genetic variants near TNFAIP3 on 6q23 are**
33 **associated with systemic lupus erythematosus.** *Nat. Genet.* 2008, **40**:1059–61.
- 34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
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, PXX, 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 erythematosus 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.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
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.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
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.

1
2
3
45. Consortium RE, Kundaje A, Meuleman W, et al. **Integrative analysis of 111 reference human epigenomes**. Nature 2015, **518**:317–330.

4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

** 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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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.

- 1
2 64. Jacquemin C, Schmitt N, Contin-Bordes C, et al. **OX40 Ligand Contributes to Human**
3
4 **Lupus Pathogenesis by Promoting T Follicular Helper Response.** Immunity 2015,
5
6 **42**:1159–1170.
7

8
9 ** This study demonstrates that OX40L is expressed by myeloid antigen-presenting cells in active
10
11 SLE patients, indicating that the expression of particular disease associated gene is context-specific,
12
13 i.e. cell types and the disease status in this case.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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

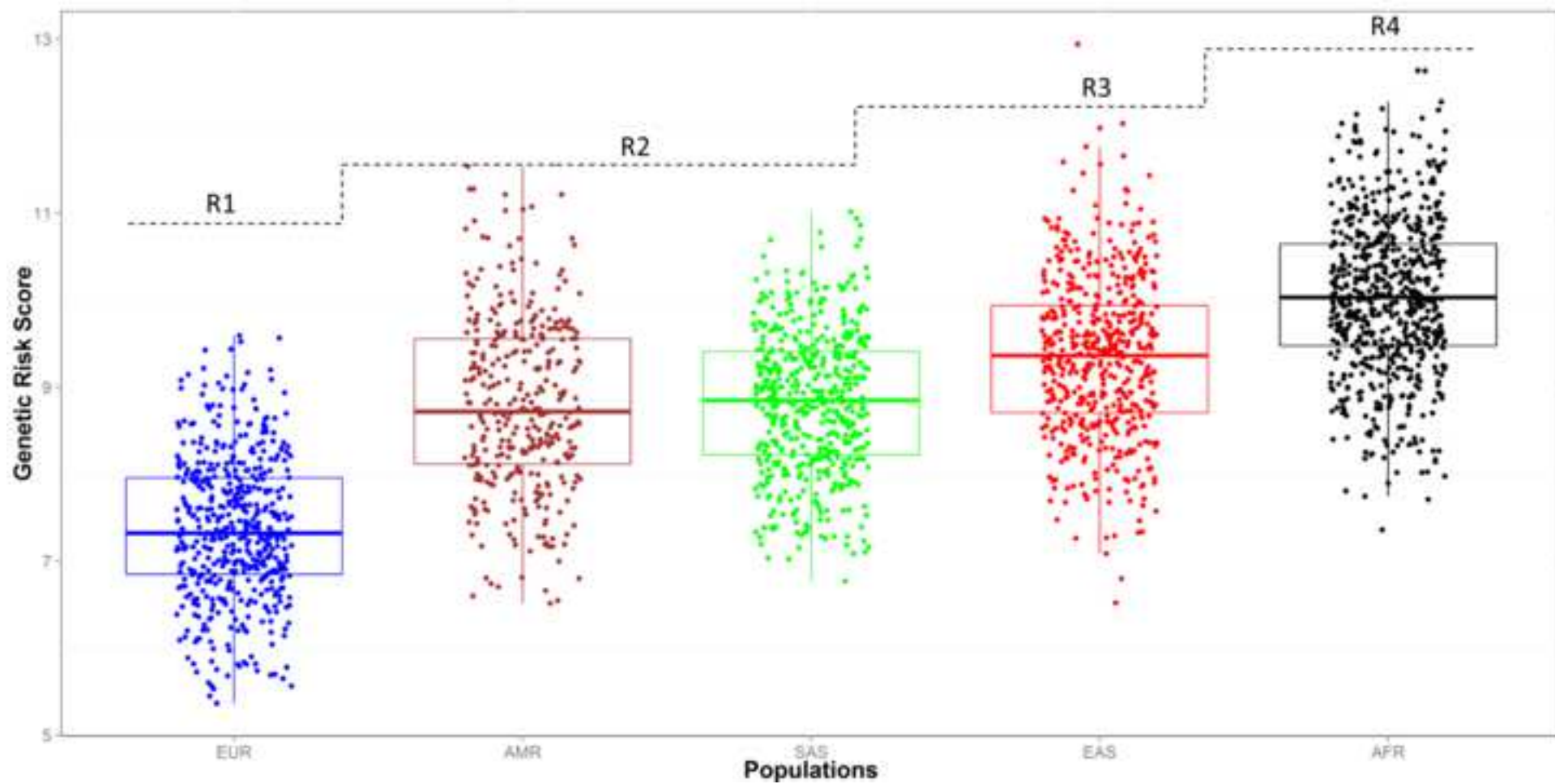
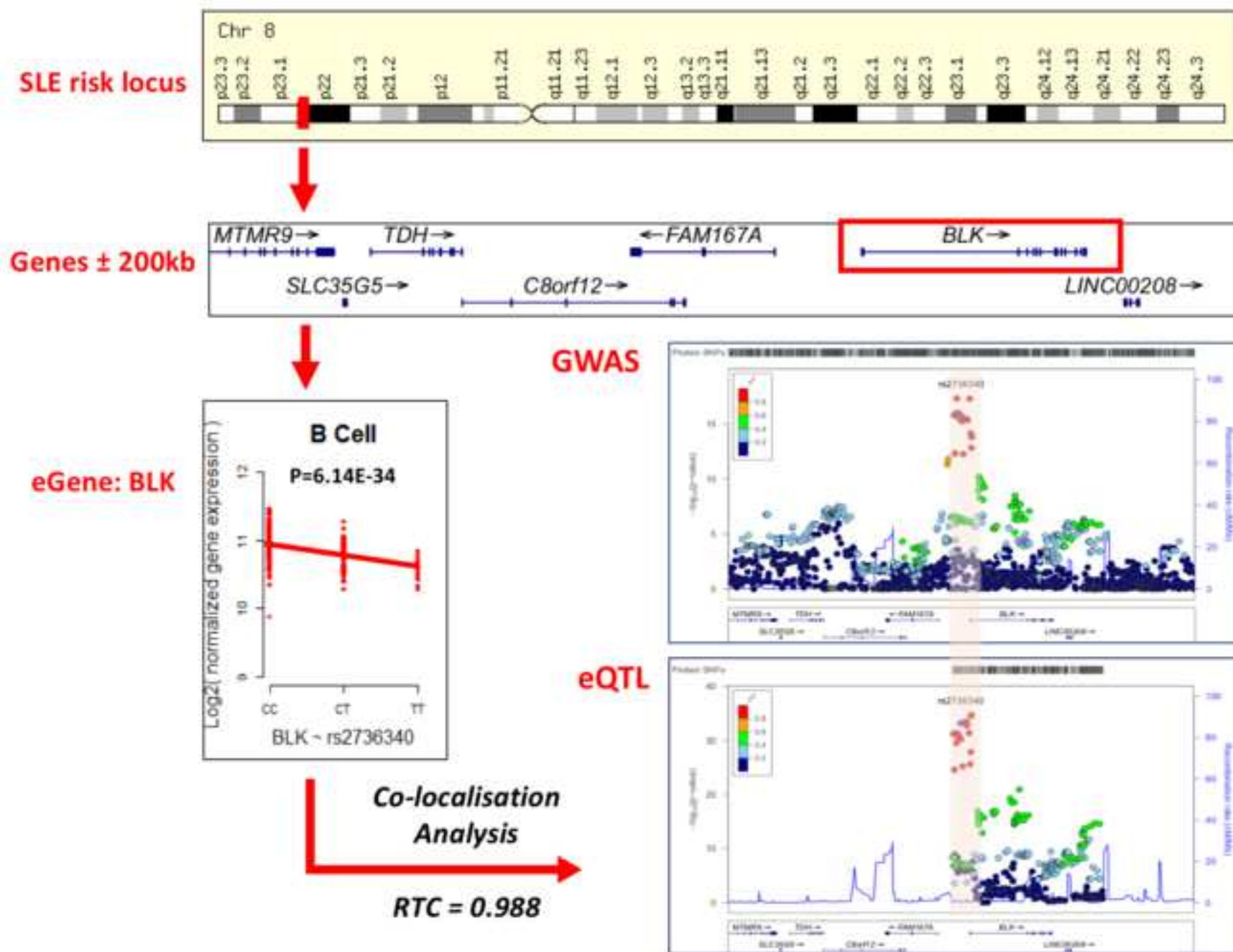


Figure 3. Overview of co-localisation analysis of GWAS and eQTL_Original





[Click here to access/download](#)

Supplemental Data File (.doc, .tif, pdf, etc.)

Supplementary Table 1. A summary of SLE risk loci.xlsx

