

# Regulatory polymorphisms modulate the expression of HLA class II molecules and promote autoimmunity

Prithvi Raj<sup>1†</sup>, Ekta Rai<sup>1,2†</sup>, Ran Song<sup>1</sup>, Shaheen Khan<sup>1</sup>, Benjamin E Wakeland<sup>1</sup>, Kasthuribai Viswanathan<sup>1</sup>, Carlos Arana<sup>1</sup>, Chaoying Liang<sup>1</sup>, Bo Zhang<sup>1</sup>, Igor Dozmorov<sup>1</sup>, Ferdicia Carr-Johnson<sup>1</sup>, Mitja Mitrovic<sup>3</sup>, Graham B Wiley<sup>4</sup>, Jennifer A Kelly<sup>4</sup>, Bernard R Lauwerys<sup>5</sup>, Nancy J Olsen<sup>6</sup>, Chris Cotsapas<sup>3</sup>, Christine K Garcia<sup>7,8</sup>, Carol A Wise<sup>8,9,10,11</sup>, John B Harley<sup>12,13</sup>, Swapan K Nath<sup>4</sup>, Judith A James<sup>4</sup>, Chaim O Jacob<sup>14</sup>, Betty P Tsao<sup>15</sup>, Chandrashekhar Pasare<sup>1</sup>, David R Karp<sup>16</sup>, Quan Zhen Li<sup>1</sup>, Patrick M Gaffney<sup>4</sup>, Edward K Wakeland<sup>1\*</sup>

<sup>1</sup>Department of Immunology, University of Texas Southwestern Medical Center, Dallas, United States; <sup>2</sup>School of Biotechnology, Shri Mata Vaishno Devi University, Katra, India; <sup>3</sup>Department of Neurology, Yale School of Medicine, New Haven, United States; <sup>4</sup>Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, United States; <sup>5</sup>Pole de pathologies rhumatismales, Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, Bruxelles, Belgium; <sup>6</sup>Division of Rheumatology, Department of Medicine, Penn State Medical School, Hershey, United States; <sup>7</sup>Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, United States; <sup>8</sup>Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, United States; <sup>9</sup>Department of Orthopaedic Surgery, University of Texas Southwestern Medical Center, Dallas, United States; <sup>10</sup>Sarah M. and Charles E. Seay Center for Musculoskeletal Research, Texas Scottish Rite Hospital for Children, Dallas, United States; <sup>11</sup>Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, United States; <sup>12</sup>Cincinnati VA Medical Center, Cincinnati, United States; <sup>13</sup>Cincinnati Children's Hospital Medical Center, Cincinnati, United States; <sup>14</sup>Department of Medicine, University of Southern California, Los Angeles, United States; <sup>15</sup>Department of Medicine, University of California, Los Angeles, Los Angeles, United States; <sup>16</sup>Rheumatic Diseases Division, Department of Medicine, University of Texas Southwestern Medical Center, Dallas, United States

\*For correspondence: edward.wakeland@utsouthwestern.edu

†These authors contributed equally to this work

**Competing interests:** The authors declare that no competing interests exist.

**Funding:** See page 42

**Received:** 05 October 2015

**Accepted:** 13 February 2016

**Published:** 15 February 2016

**Reviewing editor:** Jonathan Flint, Wellcome Trust Centre for Human Genetics, United Kingdom

© Copyright Raj et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

**Abstract** Targeted sequencing of sixteen SLE risk loci among 1349 Caucasian cases and controls produced a comprehensive dataset of the variations causing susceptibility to systemic lupus erythematosus (SLE). Two independent disease association signals in the HLA-D region identified two regulatory regions containing 3562 polymorphisms that modified thirty-seven transcription factor binding sites. These extensive functional variations are a new and potent facet of HLA polymorphism. Variations modifying the consensus binding motifs of IRF4 and CTCF in the XL9 regulatory complex modified the transcription of HLA-DRB1, HLA-DQA1 and HLA-DQB1 in a chromosome-specific manner, resulting in a 2.5-fold increase in the surface expression of HLA-DR and DQ molecules on dendritic cells with SLE risk genotypes, which increases to over 4-fold after stimulation. Similar analyses of fifteen other SLE risk loci identified 1206 functional variants tightly

linked with disease-associated SNPs and demonstrated that common disease alleles contain multiple causal variants modulating multiple immune system genes.

DOI: [10.7554/eLife.12089.001](https://doi.org/10.7554/eLife.12089.001)

---

## Introduction

Systemic Lupus Erythematosus (SLE) is a complex autoimmune disease resulting from a profound loss of immune tolerance to self-antigens (*Olsen and Karp, 2014; Theofilopoulos, 1995a; 1995b; Fairhurst et al., 2006*). The disease initiates with the production of autoantibodies against a spectrum of self-antigens (typically >10 in SLE patients), focused on nucleic acids and nucleic-acid-associated proteins. Disease pathology begins with the deposition of immune complexes in various target tissues, leading to the activation of inflammatory effector mechanisms that damage critical organ systems. Patients with SLE can present with combinations of symptoms, including skin rashes, oral ulcers, glomerulonephritis, neurologic disorders, severe vasculitis, and a distinct form of arthritis (*Tsokos, 2011*). This extensive heterogeneity in clinical presentation presumably reflects variations in the sites of immune complex deposition and induced inflammation among patients, but also suggests that SLE may be a collection of related diseases, rather than a single pathogenic process. A generalized loss in immune tolerance by the humoral immune system and the aberrant activation of inflammatory effector mechanisms at the sites of immune complex deposition, however, are consistent features of SLE.

Susceptibility to SLE is caused by a combination of genetic and environmental factors (*Fairhurst et al., 2006; Harley et al., 2009; Deng and Tsao, 2010; Rai and Wakeland, 2011*). Current thought postulates that a collection of common risk alleles mediates the development of an autoimmune-prone immune system which, when coupled with poorly-defined environmental triggers, becomes dysregulated, leading to the development of autoantibodies and the initiation of disease pathologies. Genome-wide association analyses (GWAS) have identified more than 50 SLE risk loci to date, indicating that susceptibility is quite polygenic (*Harley et al., 2009; Nath et al., 2008; Harley et al., 2008; Kim et al., 2012; Graham et al., 2006; 2008; 2009; Hom et al., 2008; Gateva et al., 2009; Relle et al., 2015*). A variety of candidate genes have been identified within these risk loci, including: HLA-DR and HLA-DQ class II alleles, IRF5, ITGAM (CD11b), STAT4/STAT1, TNFAIP3, and BLK. The functional effects or 'endophenotypes' that these disease genes contribute to the disease process have not been clearly delineated.

GWAS utilize a dense array of single nucleotide polymorphisms (SNP) to map the positions of risk loci within the human genome to relatively small segments, termed linkage disequilibrium (LD) blocks (typically < 200 Kb in length). Within these LD blocks, recombination is infrequent and polymorphisms form stable combinations or 'haplotypes' that persist within populations for extended periods (*Balding, 2006; de Bakker et al., 2005; Frazer et al., 2007*). Disease associated 'tagging' SNPs are postulated to be imbedded in specific haplotypes that contain the functional variations that impact disease susceptibility. The characteristics of these functional variations and the endophenotypes that they contribute to disease processes are a poorly described aspect of common disease genetics.

Population sequencing studies have identified extensive variations in both the coding and non-coding regions of the human genome (*Abecasis et al., 2010; 2012; Barreiro and Quintana-Murci, 2010; Laval et al., 2010*). The ENCODE consortium has investigated the functional characteristics of non-coding regions in the human genome in detail and have defined a plethora of regulatory elements impacting transcription levels and cell lineage differentiation, including histone associated regions, transcription factor (TF) binding sites, and DNase hypersensitivity clusters (*Gerstein et al., 2012*). A parallel series of investigations by several research groups have used expression quantitative trait locus (eQTL) analysis to identify common polymorphisms that quantitatively impact gene transcription (*Sheffield et al., 2013; Vernet et al., 2012; Dunham et al., 2012; Bernstein et al., 2010; Cookson et al., 2009; Fairfax et al., 2012; 2014; Gilad et al., 2008*). These findings, coupled with data indicating that many disease-tagging SNPs are localized to non-coding regulatory regions (*Maurano et al., 2012*), suggest that the causal variants for common disease risk alleles may impact regulatory processes, rather than protein structure.

**eLife digest** The human immune system defends the body against microbes and other threats. However, if this process goes wrong the immune system can attack the body's own healthy cells, which can lead to serious autoimmune diseases.

Systemic lupus erythematosus (SLE) is an autoimmune disease in which immune cells often attack internal organs – including the kidneys, nervous system and heart. Over the past decade, multiple genes have been linked with an increased risk of SLE. However, it is largely unknown how the sequences of these genes differ between individuals with SLE and healthy individuals, and the precise changes that lead to an increased risk of SLE are also not clear.

Now, Raj, Rai et al. have determined the genetic sequences of over 700 people with SLE and over 500 healthy individuals and looked for differences that influence susceptibility to the disease. The vast majority of differences were discovered in stretches of DNA that regulate the expression of nearby genes, rather than in DNA that encodes the structures of proteins. Notably, extensive differences were found in a region of the human genome that regulates the production of proteins called Human Leukocyte Antigen class II molecules; which are known to play a critical role in activating the immune system. Raj, Rai et al. found that slight changes to the regulatory DNA sequences resulted in an overabundance of these proteins, which led to a hyperactive immune system that is strongly associated with SLE.

Future studies could now ask if the changes to the regulatory DNA sequences highlighted by Raj, Rai et al. increase susceptibility to other autoimmune disorders as well. It may also be possible to use the increased understanding of how the immune system is regulated to develop new ways to minimize the rejection of organ transplants.

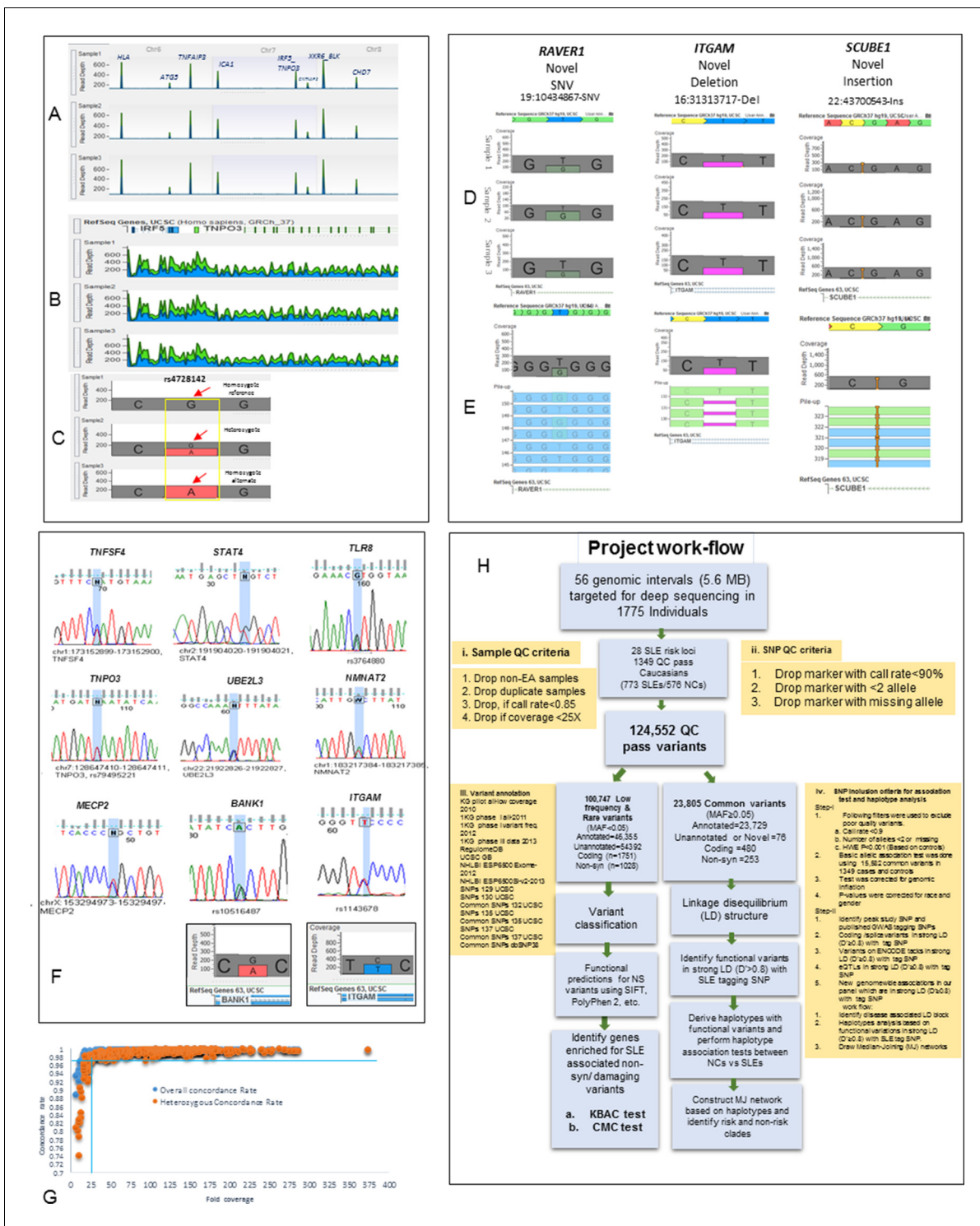
DOI: [10.7554/eLife.12089.002](https://doi.org/10.7554/eLife.12089.002)

Here we describe the targeted, deep sequencing of 28 risk loci for SLE in a population of SLE patients and controls. Our sequencing study identified 124,552 high quality sequence variants contained in these risk loci among 1349 Caucasian cases (773) and controls (576). Detailed analysis of sixteen of these SLE risk loci demonstrate that haplotypes of functional variations in tight LD with SLE- tagging SNPs often impact the expression of multiple genes, resulting in the association of several transcriptional variations with SLE risk haplotypes. Notably, multiple SLE risk haplotypes within the HLA-D region were found to coordinately upregulate HLA-DR, -DQ and a variety of other genes within the antigen processing and presentation pathways for HLA class I and class II molecules. These results reveal a new functional diversification mediated by HLA-D polymorphisms and provide important insights into the molecular mechanisms by which HLA-D and other SLE risk loci potentiate disease.

## Results

### Deep sequencing of SLE risk loci in populations of SLE cases and controls

Targeted genomic sequencing of twenty-eight GWAS-confirmed SLE risk loci was performed using Illumina (Illumina Inc., San Diego, CA) custom enrichment arrays on genomic DNA from 1775 SLE patients and controls (**Supplementary file 1A, Figure 1A–F**). These procedures resulted in >128 fold coverage of the genomic segments containing these SLE risk loci (**Supplementary file 1B, Figure 1G**). Our bioinformatics pipeline (**Figure 1H**) defined 1349 samples of European American (EA) ancestry (**Figure 2A**) that carried 124,552 high quality variants of which 114,487 are single nucleotide variations (SNVs) and 10,065 are insertion/deletions (In/Del). This sequence-based variant database, which identifies an average of one variant every 39 basepairs in the targeted regions, provides a comprehensive assessment of genomic diversity at SLE risk loci in the EA population. The functional properties of these variants were annotated using multiple databases cataloguing the functional properties of human genomic variation (**Figure 2B,C**), including the phase 3 release from the 1000 genome study (**Auton et al., 2015**), the PolyPhen/SIFT (**Adzhubei et al., 2010; Ng and Henikoff, 2003**) coding region database, the ENCODE (**Pazin, 2015**) and RegulomeDB



**Figure 1.** Sequencing quality metrics and work flow pipeline. (A) Depth of sequence reads across chromosomes 6, 7 and 8 for three samples, illustrating enrichment efficiency for targeted regions. (B) Zoom in read depth analysis of *IRF5-TNPO3* gene region (~228 Kb) for three different samples. (C) Genotype calls for a SNP in *IRF5* illustrating read depth across a typical variant position. (D) Examples of data used to genotype a novel SNV in *RAVER1*, a novel deletion in *ITGAM* and a novel insertion in *SCUBE1* gene. (E) The distribution of variant calls in forward and reverse sequencing reads. (F) About 35 SNPs from various targeted genes were confirmed by Sanger sequencing. Sanger sequencing results were further validated by calculating read depths for reference and alternate alleles in heterozygous samples, as shown for *ITGAM* and *BANK1*. (G) This figure compares fold coverage versus SNP concordance rate for a subset of samples that were both sequenced and genotyped with the ImmunoChip. *Figure 1 continued on next page*



Figure 1 continued

v1 SNP array. (H) A diagram of the work flow pipeline for bioinformatics analysis of the sequencing data including quantitative information for the number of variants passing filters at each step.

DOI: [10.7554/eLife.12089.003](https://doi.org/10.7554/eLife.12089.003)

databases (Boyle et al., 2012), and several eQTL databases for immune cell lineages (Fairfax et al., 2014; Raj et al., 2014; Westra et al., 2013). The specific technologies, bioinformatics algorithms, and quality assessments used to generate these data are discussed in Materials and Methods and relevant data are provided in **Supplementary file 1B** and **Figures 1A–H**. Overall, these sequence analyses identified 70,070 previously annotated variations and 54,482 novel or unannotated variations within the EA cohort. Functional annotation defined about 40% of the variants in the dataset as regulatory, based on their inclusion in eQTL datasets or their localization into ENCODE-defined regulatory segments (**Figures 2B,C**).

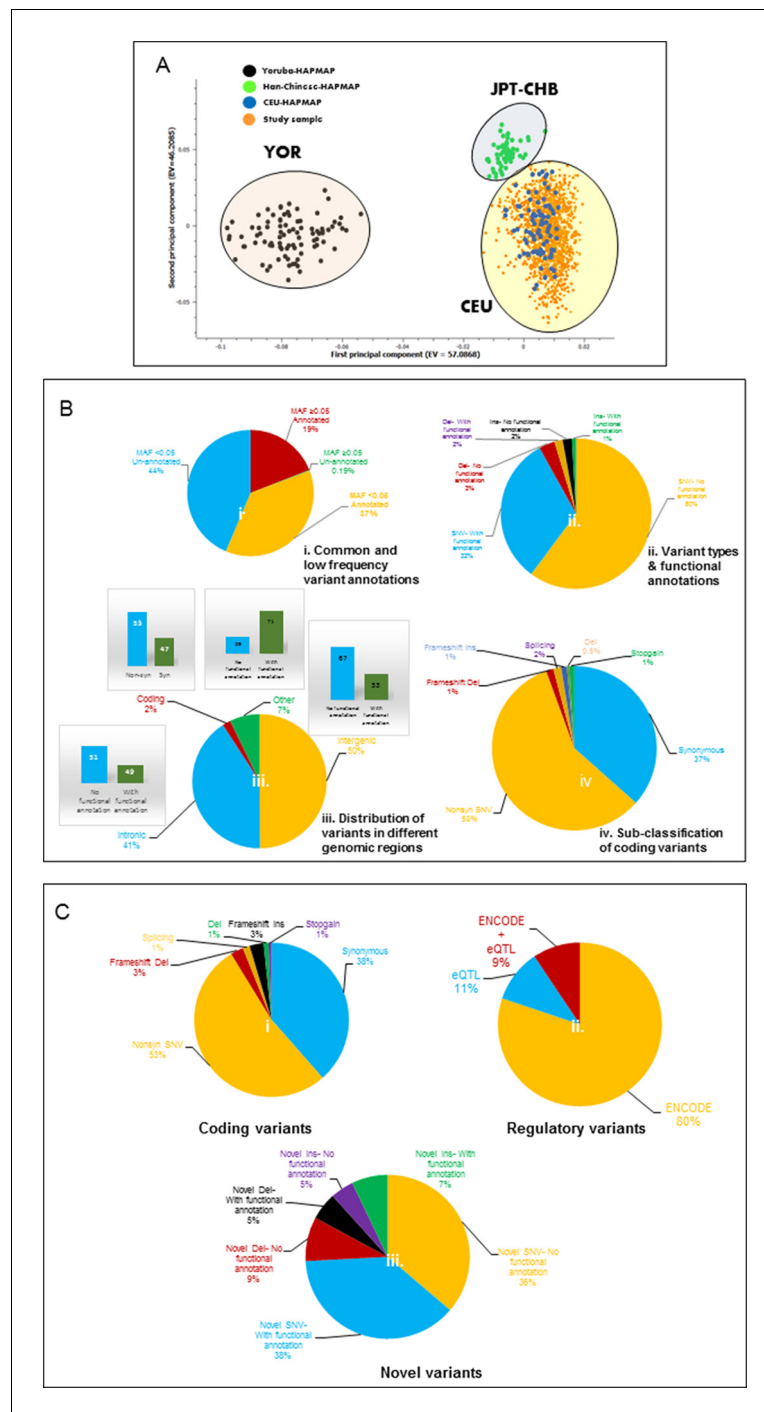
### Association analysis of common variants with SLE

As shown in **Figure 3A**, multiple variants in 26 of the 28 risk loci were strongly associated with susceptibility to SLE, with seven loci reaching genome-wide significance ( $p \leq 5 \times 10^{-8}$ ), ten reaching suggestive significance ( $p \leq 5 \times 10^{-5}$ ), and nine reaching confirmatory significance ( $p \leq 10^{-3}$ ) (tabulated in **Supplementary file 1D**). We also replicated associations previously reported in SLE GWAS for 36 SNPs at ten loci (**Supplementary file 1E**), although the bulk of the strongest associations detected in the sequence dataset were variants that were not previously reported to be associated with SLE. As tabulated in **Supplementary file 1F**, 673 variants in the sequencing data set exhibited similar or stronger associations with disease than published tagging SNPs, and 345 of these were categorized as functional. This is presented in **Figure 3B**, in which functional variants are shown as yellow points, variants with no functional annotations in blue, and previously identified tagging SNPs in red. Zoom in Manhattan plots of *TNFAIP3* and *ITGAM* are also shown. These results show that multiple, new variants had the strongest disease-associations in 27 of the 28 risk loci and that 14 of the peak variants are annotated as functional.

Sixteen risk loci were selected for more detailed analyses, based predominantly on the presence of multiple variations showing strong associations with disease. **Table 1** provides association statistics, identifies the strongest associated variants, and tabulates the coding and non-coding functional variants in tight LD with the peak signal(s) in each locus. As shown, conditional analyses identified four risk loci with multiple, independent signals. This indicates that *NMNAT2-SMG7*, *TNFSF4*, *HLA-D*, and *XKR6* each contained two or more LD blocks with potentially regulatory variants which might be contributing to disease susceptibility independently. In this regard, we attribute regulatory characteristics to these variants based on published studies from the ENCODE consortium and other research groups (see **Supplementary file 1F** for details). Additional studies will be required to confirm these regulatory properties and delineate the precise mechanisms impacting disease-relevant mechanisms. As shown, 1206 functionally annotated variants were in tight LD ( $D' > 0.8$ ) with the 21 peak risk signals and all but 7 of these were non-coding, regulatory variants. These results demonstrate that multiple functional variations are in tight LD with the peak disease associated signal in every risk locus.

### Haplotype analysis of functional variants in tight LD with peak tagging SNPs

The strategy utilized to assess the association of functional variations with disease is outlined in **Figure 1H** (iv) and illustrated for the *STAT4* risk locus in **Figure 4**. As shown in **Figure 4A** and tabulated in **Supplementary file 1B**, targeted sequencing of the 104.2 kb *STAT4* risk locus produced an average of 100.17-fold coverage and identified 2273 high quality variants. The LD structure of this region was assessed using 104 common markers ( $MAF > 0.1$ ). As shown, two distinct LD blocks were identified and the ~68 Kb LD block that encompasses the 3' portion of *STAT4* contained the SLE disease-tagging SNPs. **Figure 4B** plots the disease association of all of the common variants within this LD block and **Figure 4C** demonstrates that conditioning with the strongest SLE tagging SNP (rs12612769) accounts for all of the disease association within *STAT4*. These results indicate that functional variations in tight LD with rs12612769 are responsible for the disease-associated



**Figure 2.** Principal component analysis (PCA) and variant summary. (A) Principal component analysis (PCA), showing clustering of study cohort (orange points) with the CEU (blue points) HAPMAP reference group for Caucasians. (B) (i) Pie chart showing percentages of annotated and unannotated variants in common ( $MAF \geq 0.05$ ) and low frequency ( $MAF < 0.05$ ) categories. (B) (ii) Pie chart showing percentages of potentially functional single nucleotide variants (SNVs) and structural variants (InDels) defined by ENCODE and eQTL data. (B) (iii) Pie chart showing the distribution of variants in various genomic regions and percentage of potential functional variants in each. (B) (iv) Pie chart showing classification of coding variants into various sub-categories. (C) (i) Pie chart showing classification of common frequency coding/splice variants. (C) (ii) Pie chart showing percentages of ENCODE and/or eQTL defined potentially functional common regulatory variants. (C) (iii) Pie chart showing the percentages of un-annotated or novel SNVs and InDels with potentially functional annotations.

DOI: [10.7554/eLife.12089.004](https://doi.org/10.7554/eLife.12089.004)

endophenotypes of the STAT4 locus. **Figure 4D** demonstrates that seven functional variants are in strong LD ( $D' > 0.8$ ) with rs12612769 (strongest tagging SNP in this analysis) and rs7574865 (strongest tagging SNP from literature). **Figure 4E** presents 4 prevalent (frequency  $> 0.05$ ) haplotypes formed by these functional variants, which in sum account for  $> 90\%$  of the chromosomes found among the 1349 EA samples. As shown, HAP2 is strongly associated with susceptibility to SLE ( $6.88E-08$ ) and HAP1 is associated with protection ( $6.00E-04$ ).

**Figure 4F** presents the patterns of sequence divergence that distinguish these haplotypes, utilizing the median neighbor joining (MJ) algorithm (Bandelt et al., 2000). MJ analysis is a phylogenetic algorithm that models sequence-based allelic divergence of haplotypes within species. For MJ diagrams, the spheres (termed nodes) represent individual haplotypes in the network and their size is proportional to their frequency. The pie charts overlaid on each node represent the relative frequency of that haplotype in cases (red) and controls (white). The individual SNPs that distinguish each node are listed along the line that connects them and the length of the line is roughly proportional to the number of SNPs that distinguish the haplotypes. The network is progressive, such that the two nodes at opposite ends of the network are most divergent. In essence, MJ analysis provides a visually informative illustration of the relationships of a set of haplotypes segregating within a population.

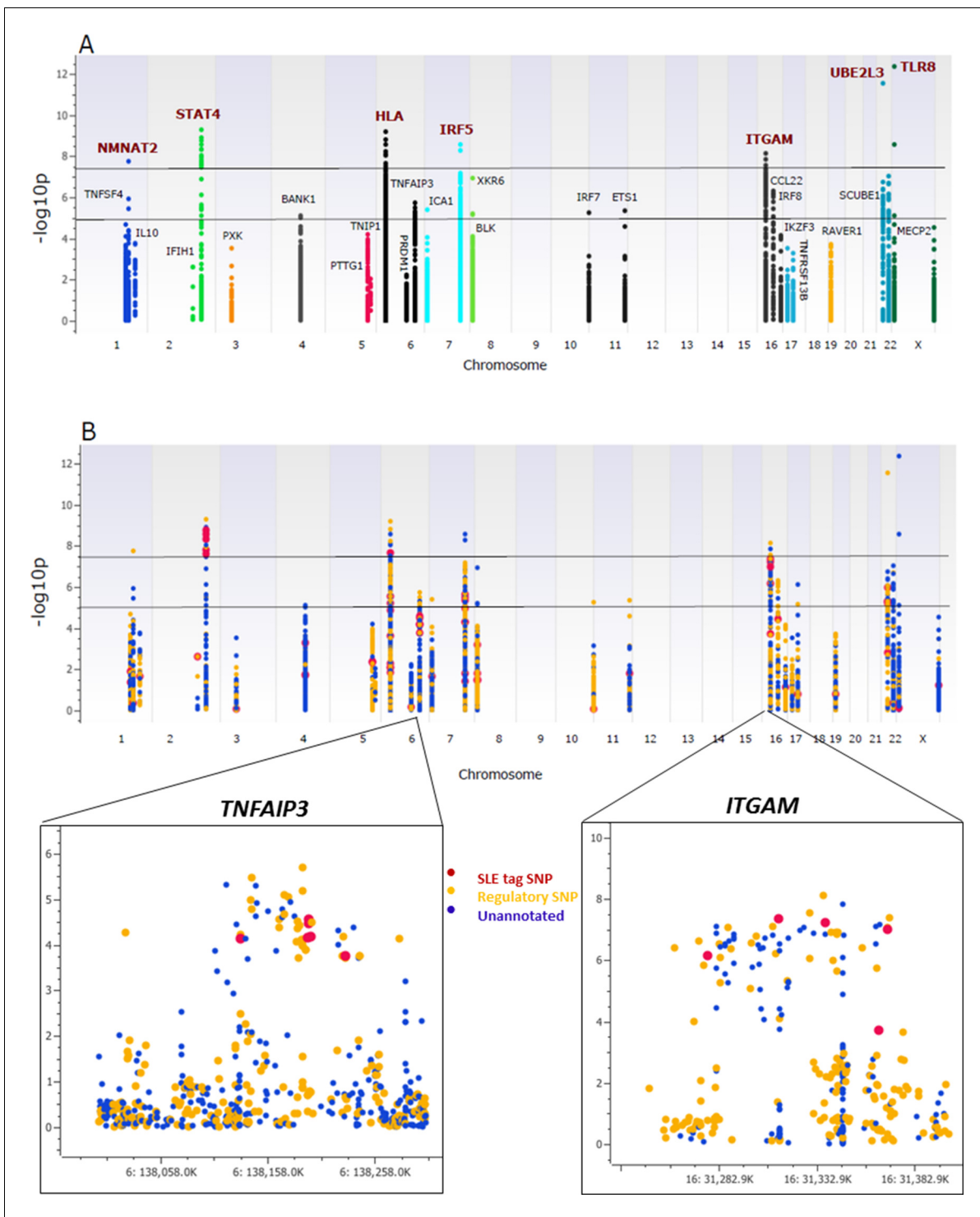
Several features of STAT4 polymorphisms within the EA population are apparent from this analysis. First, HAP1, HAP3, and HAP4 form a clade of protective haplotypes (nodes highlighted in blue), all with decreased frequencies in SLE patients. Further, both the peak signal SNP in this analysis (SNP5) and the peak GWAS SNP from the literature (SNP6) together with three functional variants, SNP2, SNP8, and SNP9, distinguish the disease-associated HAP2 (highlighted in red) from the haplotypes in the protective clade. As listed in **Supplementary file 2**, SNP8 (rs10181656) is located within a binding site for the CCCTC-binding factor (CTCF), which is a chromatin insulator that inhibits transcription and plays a role in defining the borders of transcriptional domains. SNP9 (rs7582694) is located within an ENCODE-defined segment containing transcription binding sites for ESR1 (estrogen response elements) and FOS1. Both of these transcription factors are active in multiple tissues and immune cell lineages and both are annotated by ENCODE with strong effect scores and good regulomeDB scores, suggesting that these variations mediate transcriptional endophenotypes in several cell lineages.

Finally, SNP2 (rs11889341) is the most potent of several eQTL variants within the STAT4 risk locus is very strongly associated with SLE susceptibility ( $p < 4.8 \times 10^{-9}$ ), and impacts the transcription levels of STAT1 and STAT4 (**Supplementary file 2**). As shown in **Figure 4G**, our eQTL dataset for monocyte-derived macrophages (MDM) identifies a significant increase in baseline STAT1 and STAT4 transcription with the T allele of SNP2, which associates this phenotype with susceptibility to SLE. Several other SNPs distinguishing the protective and risk haplotypes were also associated with STAT1 and/or STAT4 transcription levels in published eQTL databases from ex vivo monocytes or peripheral blood (**Supplementary file 2** and **Figure 4H**). These results indicate that the transcription of both STAT1 and STAT4 are impacted by variants in tight LD with the SLE tag variant and that the disease risk allele is associated with increased transcription of both genes in multiple cell types.

## Multiple SLE associations within the HLA-D region

Sequence analysis of the HLA-D region revealed 15129 common variants (1 variant/29.6 bp) distributed throughout the 448 Kb segment analyzed. These variations occur predominantly in non-coding regions, indicating that the entire HLA-D region is diversified. This result is consistent with previous genomic sequencing analyses of HLA-D that defined 4–5 phylogenetic clades with ancient origins for this segment of HLA-D (Raymond et al., 2005). **Figure 5A** presents the LD structure of HLA-D within the EA cohort, based on 8062 common variants ( $MAF > 0.15$ ). Overall, LD in the region is high and the LD structure is very complex, with multiple partial LD associations exhibited between various blocks throughout the region.

More common (15129) than rare (12076) variants were detected in HLA-D, which differs significantly from the roughly 5-fold excess of rare variants that are typically detected throughout the human genome and at other SLE risk loci in our study (Abecasis et al., 2010; 2012; Hu et al., 2014) (**Supplementary file 1B**). Previous studies in many species have demonstrated that much of the extant polymorphisms found in MHC class I and class II genes have ancient origins and persist in populations over evolutionary timespans (McConnell et al., 1988; Lawlor et al., 1988;



**Figure 3.** Association analysis of sequencing variants from 28 SLE risk loci. (A) Manhattan plot of 15582 common variants (MAF>0.05) plotting  $-\log_{10}$  p-value of SLE association (y-axis) versus chromosomal location (x-axis). Horizontal lines mark threshold of significant ( $p=10^{-8}$ ) and suggestive ( $p=10^{-5}$ ) genome-wide significance threshold. (B) Same Manhattan plot using color coding to identify functional variants (yellow), variants with no current functional annotation (blue), and previously identified SLE GWAS tagging SNPs (red). Zoom in picture of Manhattan plot for *TNFAIP3* and *ITGAM* gene is shown.

DOI: 10.7554/eLife.12089.005



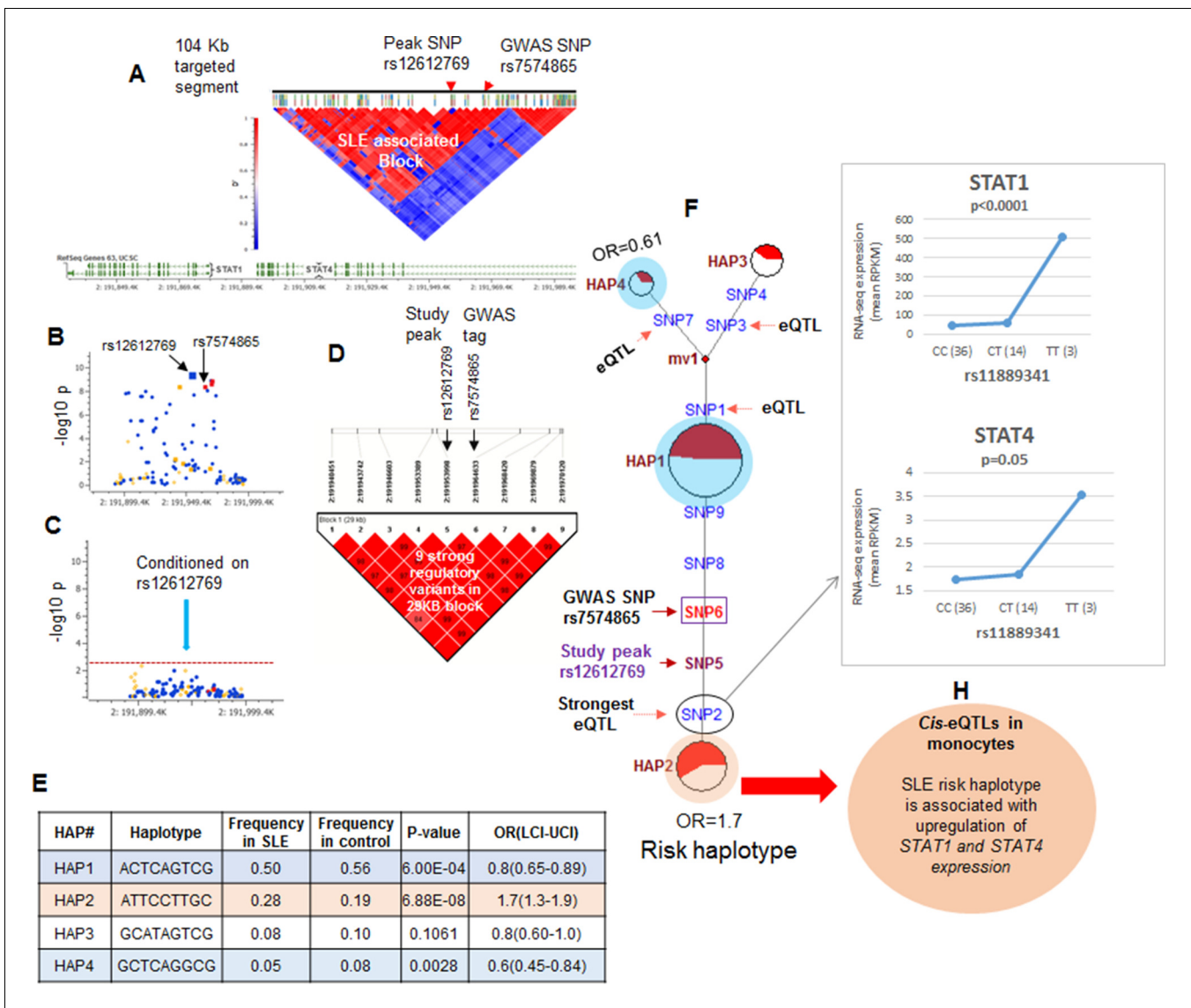
**Table 1.** Characteristics of disease associated variants at sixteen SLE risk loci.

Risk locus	Signal	Peak SNP	Minor allele	Odds ratio (Minor allele)	Allele Freq. (Cases)	Allele Freq. (Controls)	SLE association P-value	SLE associated Annotated variants	Variants in LD with peak SNP (D' >0.8)		
									Total variants	Total potentially functional variants	Total coding variants
STAT4	1	rs12612769	C	1.7	0.29	0.19	5E-10	52	49	9	0
HLA-D	1	rs9271593 (XL9)	C	1.7	0.55	0.42	7E-10	835	530	398	0
	2	rs9274678 (DQB1)	G	2.1	0.24	0.13	6E-09	736	216	69	0
	3	rs36101847 (DRB1)	T	0.5	0.13	0.23	8E-09	760	296	126	0
ITGAM-ITGAX	1	rs41476751	C	1.9	0.25	0.15	8E-09	153	121	62	3
IRF5_TNPO3	1	rs34350562	G	1.8	0.23	0.14	3E-09	245	189	124	0
UBE2L3	1	rs181366	T	1.5	0.27	0.20	2E-07	82	79	55	1
BANK1	1	rs4699260	T	0.7	0.20	0.28	9E-06	267	143	29	2
TNIP1	1	rs62382335	A	1.4	0.14	0.10	6E-05	46	22	16	0
TNFAIP3	1	rs57087937	T	1.9	0.10	0.06	2E-06	69	63	40	1
CCL22-CX3CL1	1	rs223889	T	1.5	0.34	0.27	5E-07	32	25	20	0
RAVER1-ZGLP1	1	rs35186095	T	1.3	0.21	0.17	2E-04	43	24	19	0
ICA1	1	rs74787882	A	0.7	0.06	0.09	2E-03	34	10	6	0
TNFSF4	1	rs1819717	G	0.7	0.29	0.36	2E-05	73	30	14	0
	2	rs4916313	C	1.3	0.39	0.32	2E-04		30	21	0
BLK	1	rs7822109	C	0.8	0.46	0.52	9E-05	97	61	38	0
XKR6	1	rs4840545	A	2.0	0.13	0.07	1E-07	335	51	23	0
	2	rs7000132	C	0.9	0.42	0.46	5E-04		178	118	0
NMNAT2-SMG7	1	rs41272536	G	2.9	0.11	0.05	2E-08	33	8	8	0
	2	rs111487113	A	0.6	0.13	0.18	5E-04		17	5	0
ETS1	1	rs34516251	A	0.8	0.18	0.21	7E-03	18	10	6	0

DOI: 10.7554/eLife.12089.006

**Gyllensten and Erlich, 1989; Edwards et al., 1997; She and Wakeland, 1991).** The preponderance of common variants in this sequence dataset is consistent with an ancient origin of these HLA-D polymorphisms within the human lineage. In addition, the LD structure of these HLA-D variations in our EA cohort is very similar to the LD structure obtained for this segment of HLA-D in the 2504 human genomes in the 1000 genome project (**Figure 5Ai and ii**).

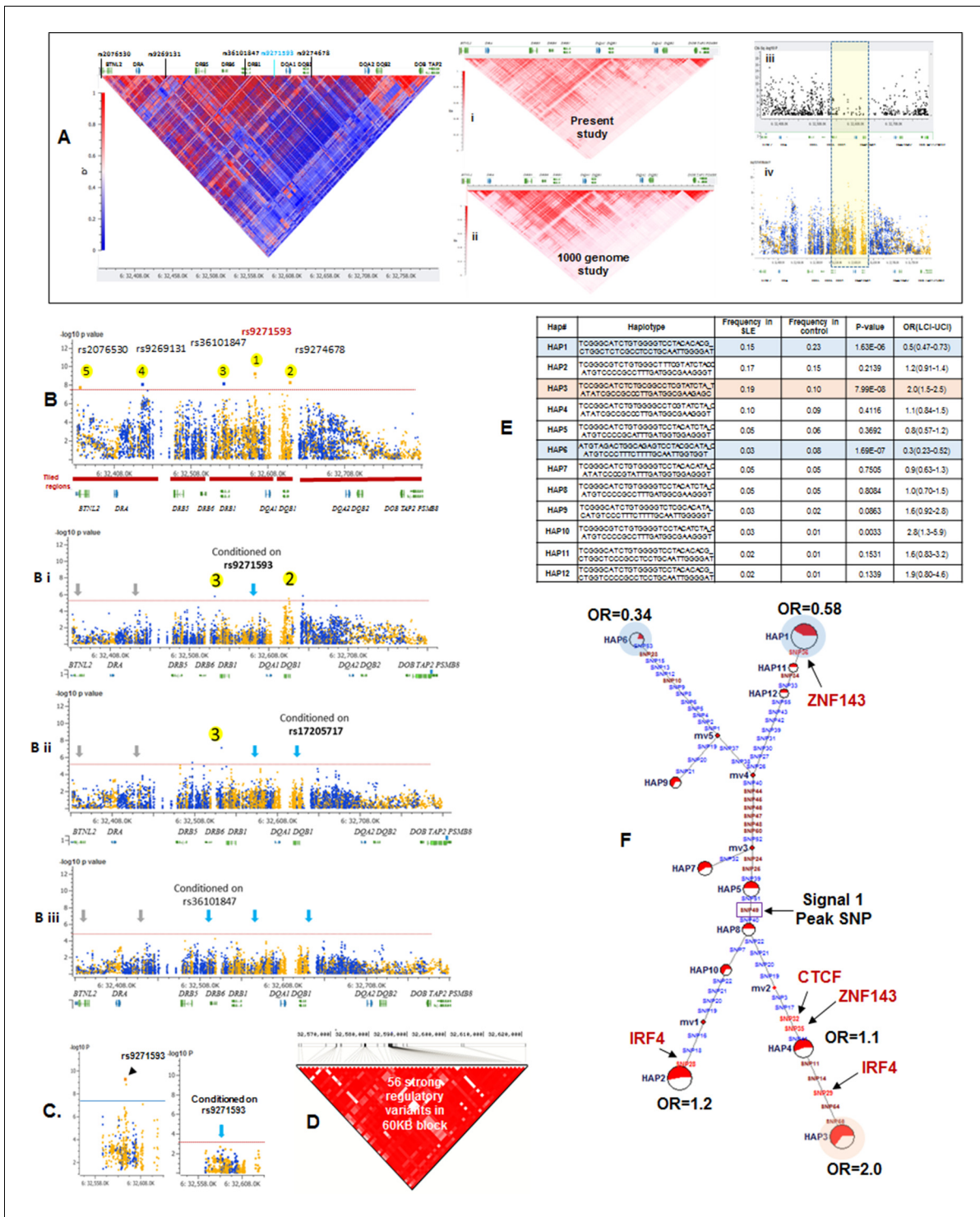
**Figure 5B** plots the association of variants throughout HLA-D with SLE and uses color-coding to distinguish regulatory variants (yellow) from variants with no current functional annotation (blue). As shown, 3786 variations in HLA-D are nominally associated with SLE ( $p < 0.05$ ) and 1797 of these have annotated regulatory functions. Five separate segments within HLA-D contained variants achieving genome-wide significance for association with SLE. The peak SLE association was with rs9271593 ( $p = 6.50E-10$ ), which is one of 687 SLE-associated regulatory variants mapping to the XL9 regulatory component within the intergenic region separating DRB1 and DQA1 (**Majumder et al., 2006; Majumder et al., 2008**). This ~50 kb segment is heavily annotated with ENCODE-defined regulatory sequences controlling chromatin structure and/or the binding of specific transcription factors. As shown in **Figure 5Aiii & iv** the XL9 segment of HLA-D is not adequately covered by SNP typing



**Figure 4.** LD structure, haplotypes and MJ networks analysis at *STAT4* locus. (A) LD structure of *STAT4* sequenced segment is shown above molecular map of the genomic segment showing *STAT1* and *STAT4* exon structure. The locations of GWAS tagging SNPs are shown above LD plot, which was produced with 104 markers (MAF $\geq$ 10%) in 1349 Caucasians. (B) Zoom in Manhattan plot showing SLE association levels of individual sequence variants in *STAT4* LD block containing *STAT4* tagging SNPs. Yellow points indicate functional variants, blue points indicate un-annotated variants and red points identify GWAS and study peak tagging SNPs. (C) Conditional analysis on peak SNP rs12612769 removes all significant associations with SLE within the LD block. (D) LD block based on nine potentially functional SLE associated variants used for haplotype analysis. (E) Derived haplotypes with SLE association results. (F) Median-joining (MJ) network analysis of *STAT4* haplotypes. Spheres (termed nodes) represent the locations of each haplotype (from table in E) within the network and the size of the node is proportional to the overall frequency of that haplotype in the dataset. Each node is overlaid with a pie chart that reflects the frequency of that haplotype in cases (red) versus controls (white). The lines connecting the nodes are labeled with the variants that distinguish the connected nodes and the length is proportional to the number of variants. Haplotypes with significant ( $p < 0.05$ ) association with SLE are highlighted with red (risk) and blue (non-risk). Study peak SNP, SLE GWAS tag SNP and eQTLs are indicated with arrows, boxes and circles within their locations within the network. (G) Presents cis-eQTL effects observed with SNP2 on *STAT1* and *STAT4* in macrophage RNAseq analysis. (H) Similar eQTL effects observed in published eQTL databases in literature.

DOI: 10.7554/eLife.12089.007

arrays such as the Immunochip v.1, possibly accounting for the failure to detect this potent SLE association in previous GWAS analyses. The second strongest SLE association signal was with rs9274678 ( $p = 6.21E-09$ ), which is located in a segment extending 5' from the DQB1 proximal promoter. The third genome-wide SLE signal was with rs36101847 ( $p = 8.33E-09$ ), which has no functional annotation and is located in close proximity to DRB1 exon 2, which encodes the peptide binding segment of



**Figure 5.** LD structure, haplotypes and MJ Network analysis of XL9 region. (A) The LD structure of HLA-D region is shown below a molecular map of the region. The locations of the genes and five genome-wide SLE association signals are marked. Peak association signal is coded blue. (A) (i) The HLA-D LD structure in 1349 Caucasians from present study assayed with -8062 common (MAF>0.15) variants. (A) (ii) The HLA-D LD structure in 2504 samples representing twenty-six cohorts from the world population. Data obtained by analysis of the 1000 Genome project datasets using the same -8062 variants analyzed in A (i). (A) (iii) SNP content of ImmunoChip v.1 across HLA-D region. (A) (iv) High quality common variant calls in this region from targeted sequencing in this study. Highlighted area boxes the XL9 through DQB1 5' segment regulatory region. Yellow points indicate potentially functional variants and blue points indicate un-annotated sequencing variants. (B) Zoom Manhattan plot of all common (MAF>0.05) variants using color coding to identify functional (yellow) and non-annotated (blue) variants. The locations of peak association signals are marked. A molecular map of the *Figure 5 continued on next page*

Figure 5 continued

region and the tiled regions for targeted sequencing are identified at the bottom. Gaps reflect the locations of long stretches of highly repetitive regions that cannot be assembled. (B) (i) The residual association level after conditioning on peak signal 1 in XL9. (B) (ii) Residual association level after conditioning on both signal 1 (XL9) and signal 2 (DQB1 5' segment). (B) (iii) No significant associations remain after conditioning on signal 1 (XL9), signal 2 (DQB1 5' segment), & signal 3 (DRB1). Yellow points identify potentially functional variants and blue points indicate un-annotated variants. (C) Conditional analysis on peak SNP rs9271593 (XL9 signal) showing that all significantly associated variants are in tight LD. (D) A 60KB LD block generated with 56 variants from XL9 region with strong regulatory scores and association with SLE. (E) Twelve haplotypes generated with HAPLOVIEW using the 56 regulatory variants. Frequencies in cases and controls, association statistics, and odds ratios are provided. Protective (blue) and risk (red) haplotypes are highlighted. (F) Median neighbor-joining (MJ) network produced as described in the text. Annotation is the same as presented in legend for **Figure 2**. Variants that disrupt binding sites of CTCF, ZNF143, and IRF4 are labeled.

DOI: [10.7554/eLife.12089.008](https://doi.org/10.7554/eLife.12089.008)

DRB1. The fourth signal is rs9269131 ( $p=9.92E-09$ ), which has no functional annotation and is in close proximity to DRA1. Finally, the fifth signal is rs2076530 ( $p=2.21E-08$ ), which is a regulatory SNP in proximity to BTNL2 that has previously been associated with sarcoidosis and autoimmunity (*Hofmann et al., 2013*).

The independence of these five SLE association signals was assessed by conditional analysis, beginning with the peak SNP (XL9 region, rs9271593) and using a forward stepwise regression method (**Figure 5B**). As shown in **Figure 5Bi**, conditioning with rs9271593 removed the associations of signals 4 and 5 with SLE, consistent with the LD associations revealed in **Figure 5A**. These results indicate that the DRA1 and BTNL2 disease association signals are in strong LD with variants in the XL9 region. However, the DQB1 promoter and DRB1 signals, although significantly diminished, were still significantly associated with disease after removal of the peak XL9 signal, indicating that these signals were somewhat independent. Removal of the XL9 and DQB1 promoter signals left only the DRB1 signal with marginally significant SLE association and removal of all three signals removed all significant association of HLA-D with disease (**Figure 5Bii & iii**). Thus, these conditional analyses identified three independent disease associations, each localized to important regulatory or coding elements within HLA-D. Their basic properties and SLE association characteristics are summarized in **Table 1**.

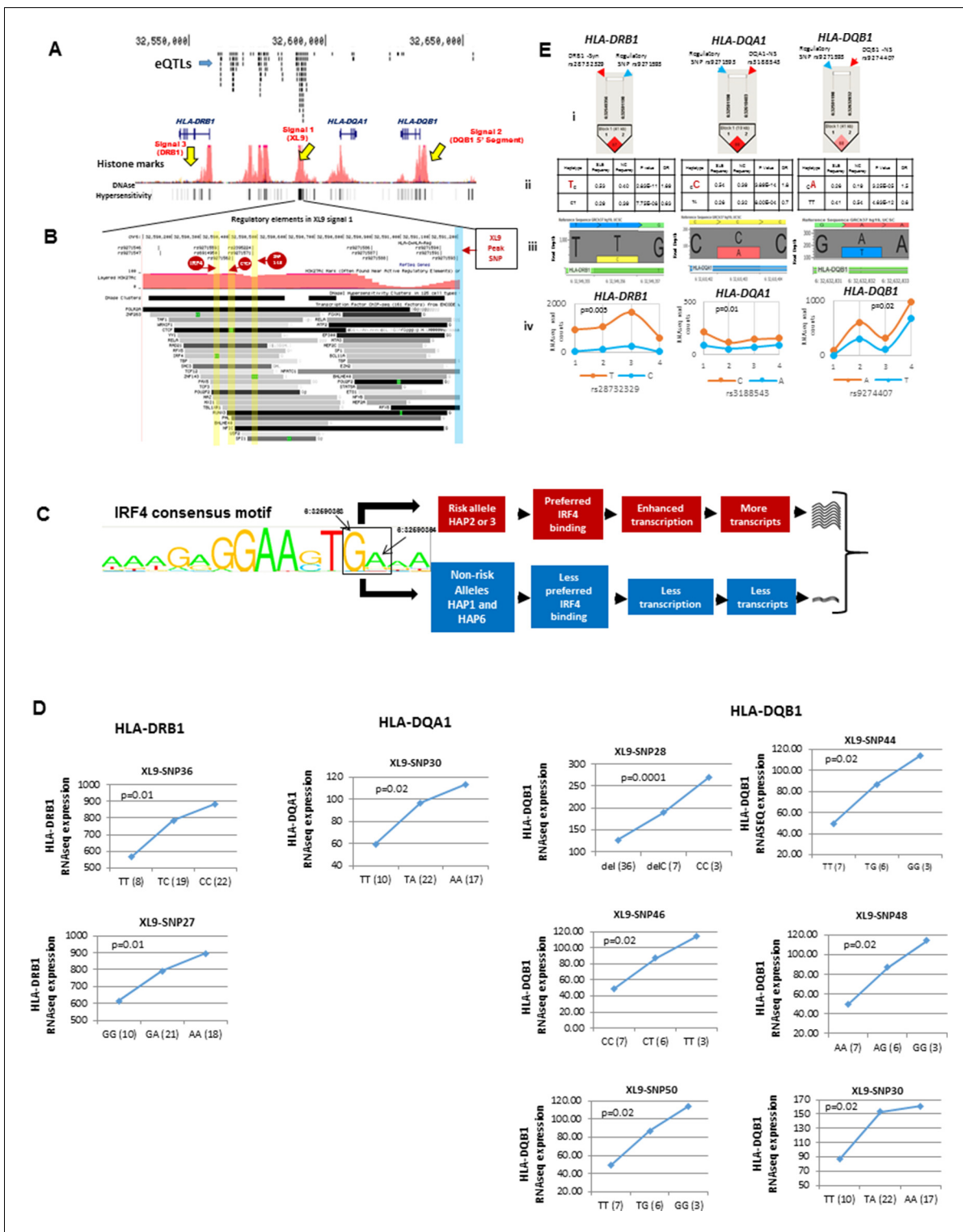
### XL9 polymorphisms mediate quantitative variations in HLA-DR and –DQ transcription

As shown in **Figure 5C**, conditioning on the XL9 signal (rs9271593) completely removed the SLE association of variants within the DRB1 to DQA1 intergenic segment, indicating that rs9271593 tags the XL9 functional haplotype responsible for this disease signal. The content of regulatory variants in strong LD ( $D' > 0.8$ ) with the XL9 signal was very high (**Table 1**) and consequently detailed haplotype analysis was focused on variants with strong ENCODE functional effect scores ( $>500$  for at least one transcription factor binding site), eQTL effects, and associations with SLE. This identified 56 functional variants that formed 12 haplotypes of which HAP3 was strongly associated with SLE susceptibility (OR 2.0,  $p < 7.04 E-08$ ) while HAP1 (OR 0.58,  $p < 1.63 E-06$ ) and HAP6 (OR 0.34,  $p < 1.69 E-07$ ) were protective (**Figure 5D,E, Supplementary file 2**). As shown in **Figure 5F**, MJ analysis found that the protective (shaded in blue) and risk (shaded in red) haplotypes form separate clades at opposite ends of the network, indicating that these two extremes in disease association are also extremes in the divergence of regulatory variations.

**Figure 6A** overlays the locations of the three peak disease signals in HLA-D with the multitude of regulatory elements located within the ~130 Kb segment spanning *HLA-DRB1* through *HLA-DQB1*. This genomic segment contains 5 separate regions with dense arrays of sequence elements that regulate chromatin structure (histone marks, DNase I clusters) and transcription factor binding. As shown in the top tract within **Figure 6A**, more than 150 variations within this small genomic segment are eQTLs that have been shown to impact the transcription of 72 genes in various immune cell lineages (**Supplementary file 2**). Our own RNA-SEQ-based eQTL dataset for MDM associates many of these variations with quantitative variations in the transcription of DRB1, DQA1, and DQB1. All of the MDM eQTL variants impacting HLA-DR and DQ expression are associated with variants in XL9 (**Figure 6D**).

XL9 contains binding sites for multiple factors affecting chromatin structure (CTCF, ZNF143) and has been shown to assemble HLA-DR and HLA-DQ proximal promoters into a transcriptional





**Figure 6.** Chromatin architecture and transcriptional regulation at SLE associated XL9 region. (A) A snap shot of the ~140 Kb DRB1-DQB1 segment that contains three genome-wide association signals for SLE. The locations of HLA class II genes and the peak signals are marked. The locations of some of the more than 750 eQTLs variants mapped into this region are overlaid onto ENCODE defined regulatory elements (Histone marks and DNA hyper sensitivity clusters). (B) A snap shot of a ~1 Kb segment in the center of the XL9 that contains 13 of the 56 strong regulatory variants that constitutes the XL9 haplotype. The positions of the canonical protein binding motifs of CTCF, IRF4 and ZNF143 highlighted in yellow and the peak XL9 SNP highlighted in blue. The locations of about 30 binding sites for transcription factor that are located within this same region and are also impacted genetic variation are also listed. (C) The consensus sequence for IRF4 binding in XL9 is shown with the locations of the two nucleotide variants boxed

*Figure 6 continued on next page*

Figure 6 continued

and marked. The consensus sequence for IRF4 binding (GA) are the alleles present in XL9 risk haplotypes. The alternative alleles for these two nucleotides, which are much less frequent in IRF4 binding motifs, are in protective haplotypes. The red and blue highlighted paths describe the predicted effects of these variations on IRF4-mediated transcription of HLA-DR and HLA-DQ, with risk haplotypes highlighted in red and protective haplotypes highlighted in blue. (D) shows cis eQTL effects observed with SLE associated XL9 region regulatory variants. SNPs were found to impact the expression level of *HLA-DRB1*, *HLA-DQA1* and *HLA-DQB1* gene in monocyte derived macrophages (MDMs). In each plot, x-axis shows three genotypes of a given eQTL SNP and y-axis shows RNAseq expression values in RPKM. SNP numbers correspond to XL9 variants in **Figure 5F**. (E) Part i shows LD between peak regulatory SNP and a coding SNP in *HLA-DRB1*, *DQA1* and *DQB1*. Part ii highlights the SLE associated coding allele sequence and shows the association statistics on peak regulatory and coding SNP haplotype for above three genes. Part iii shows the allelic bias in transcription in *DRB1*, *DQA1* and *DQB1* gene in human macrophages, demonstrated in terms of significantly different number of RNA sequencing reads for SLE risk and non-risk allele. Part iv shows the transcriptional bias between risk and protective alleles for HLA class II genes in four heterozygous human donors for these IRF4 variants.

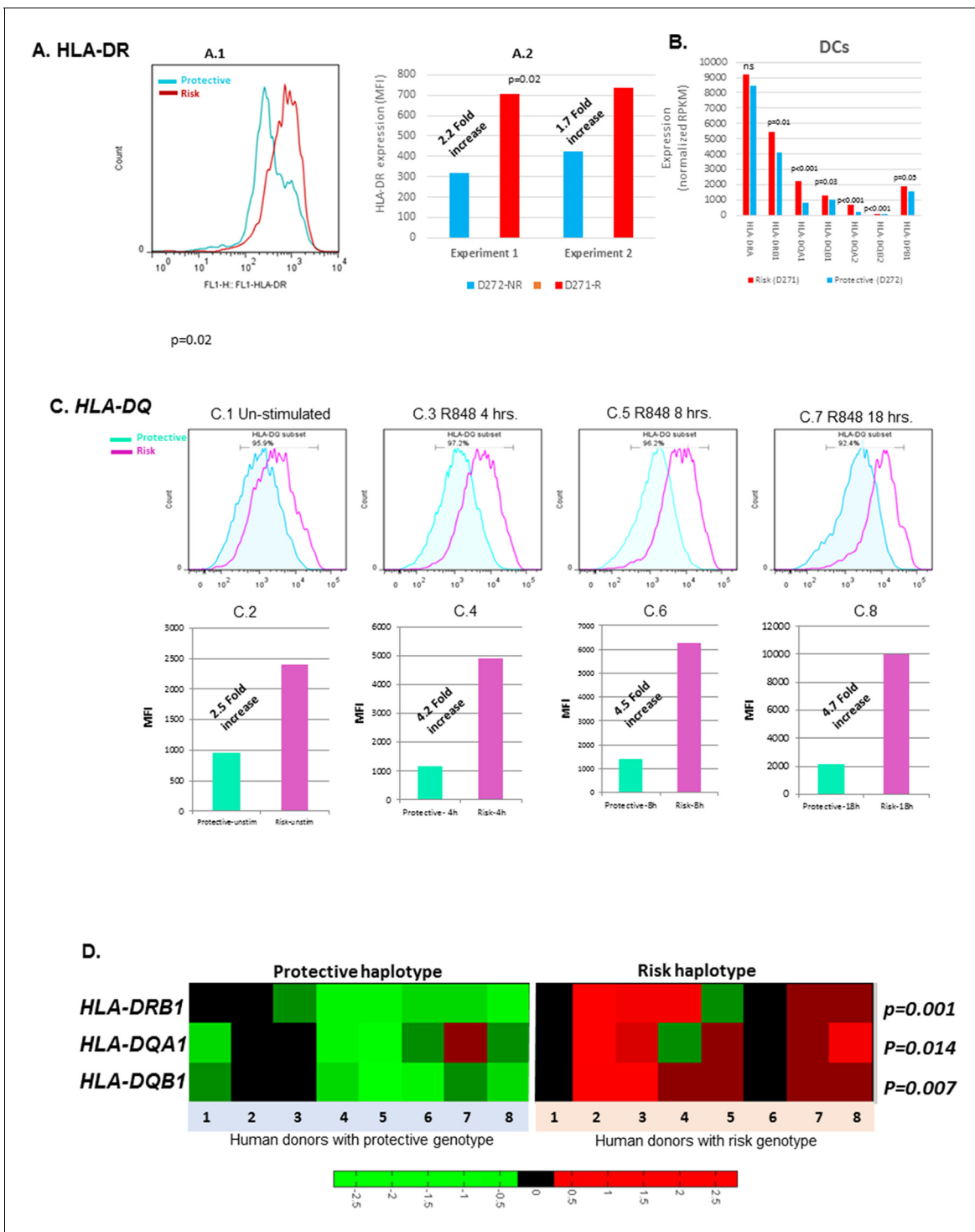
DOI: [10.7554/eLife.12089.009](https://doi.org/10.7554/eLife.12089.009)

complex that facilitates coordinate, tissue-specific transcription (**Bailey et al., 2015; Xi et al., 2007; Liu et al., 2008; Whitfield et al., 2012**). This assembly is mediated by interactions of the transcriptional insulator protein CTCF with cohesion and other chromatin regulatory components such as ZNF143. Current data supports the idea that the XL9 transcriptional complex contains multiple regulatory domains that interact with a variety of transcription factors to control the coordinated expression of HLA-DR and -DQ in lymphoid, myeloid, and thymic epithelial cell lineages. For example, recent studies by Singh and co-workers have demonstrated that the level of transcription of HLA class II molecules in dendritic cells is strongly controlled by the transcription factor IRF4, which is one of several transcription factor binding sites located within XL9 (**Vander Lugt et al., 2014**). Notably, IRF4 up-regulation of MHC class II molecules was shown by these investigators to be strongly associated with disease severity in a murine model of experimental autoimmune encephalomyelitis (EAE).

**Figure 6B** presents a fine map of variations occurring within the regulatory sequence elements in a 1 Kb segment in the center of XL9. This segment contains more than 30 transcription and chromatin configuration factor binding sites. Thirteen of the fifty-six potent regulatory variants in the XL9 haplotypes are located within this small segment. Five of these variants are located in the consensus binding sites for CTCF, ZNF143, or IRF4 and are predicted to impact their binding properties (**Figure 6B**). As shown in the MJ network in **Figure 5F**, four of the five motif variants occur within the SLE risk clade and differ between the protective and risk clades. Interestingly, one of the IRF4 variants occurs in the final link of the risk clade leading to HAP3 (strongest SLE association), while the second IRF4 variant is in the final branch leading to risk HAP2.

**Figure 6C** diagrams the nucleotide changes in the binding motif of IRF4 caused by the two XL9 IRF4 variants, both of which are predicted to strongly impact the binding of IRF4. For both of these variants, alleles carrying the unmodified IRF4 consensus binding sequence are associated with the risk haplotypes, while those with nucleotides that are less common to the consensus motif are present in all of the non-risk haplotypes (**Figure 5F, Figure 6C, and Supplementary file 2**). As diagrammed in **Figure 6C**, this suggests that the transcription of HLA-DR and -DQ should be increased in individuals carrying risk HAP2 and HAP3 due to increased IRF4 binding to XL9 regulatory domains. As shown in **Figure 6D**, the HAP2 and HAP3 alleles of variants within this short segment, including an IRF4 motif variant, are strongly associated with increased transcription of *HLA-DRB1*, *HLA-DQA1*, and *HLA-DQB1* in our eQTL panel of MDM cell cultures and in datasets from the literature (**Supplementary file 2**).

The crucial role of the XL9 region in the chromatin configuration of the HLA-DR and DQ transcriptional complex suggests that XL9 variations should impact the transcription of both DR and DQ in a cis-active, chromosome-specific manner. To test this, we measured allele-specific transcription in four individuals that are heterozygous for XL9 alleles within our MDM eQTL panel. Each of these individuals carries a HAP3 XL9 risk haplotype together with either a HAP1 or HAP6 XL9 protective haplotype. We assessed the allelic bias of transcription for *DRB1*, *DQA1*, and *DQB1* in these heterozygotes utilizing coding region SNPs in tight LD with XL9 regulatory variants. As shown in **Figure 6E**, coding region variants linked with XL9 risk associated regulatory variants are present in a significantly higher proportion of RNA-SEQ reads than coding variants linked to protective alleles. The allelic bias in RNA-SEQ reads for *HLA-DRB1*, *HLA-DQA1*, and *HLA-DQB1* is presented for a representative coding SNP in **Figure 6E** (iii). This bias in SNP read depth was a consistent feature for



**Figure 7.** Cell surface expression of HLA-CLASS II genes. (A.1) Monocyte-derived dendritic cell (MDDC) surface expression of HLA-DR in a culture produced from a homozygote for protective (blue) and homozygote for risk (red) HLA-D haplotypes. This experiment was repeated in same donors. (A.1) shows flow data. (A.2) shows the MFIs from repeated experiments. p-value shown in (A.2) was calculated on mean MFIs from two experiments. (B) shows normalized RNAseq expression on HLA-class II genes in dendritic cells on same donors presented in (A). (C.1–C.8) shows HLA-DQ surface expression on MDDC cultures from a homozygote for protective (blue) and heterozygote for risk (red) HLA-D haplotype. Flow data and respective MFIs are shown on MDDCs at steady state (C.1 and C.2), at 4 hr (C.3 and C.4), 8 hr (C.5 and C.6) and 18 hr (C.7 and C.8) after stimulation with TLR7/8 ligands. (D) heatmap on RNAseq data on lymphoblastoid cell line (LCL) from 1000 genome project compare expression level of HLA-class II genes between individuals homozygous for HLA-D protective and risk haplotype.

DOI: 10.7554/eLife.12089.010

SNP variants throughout exon 2 and exon 3 for all three genes in all four heterozygous individuals [Figure 6E(iv)]. Taken together, these experiments demonstrate that XL9 regulatory variations modulate the level of transcription of HLA-DR and HLA-DQ in a chromosome-specific manner.

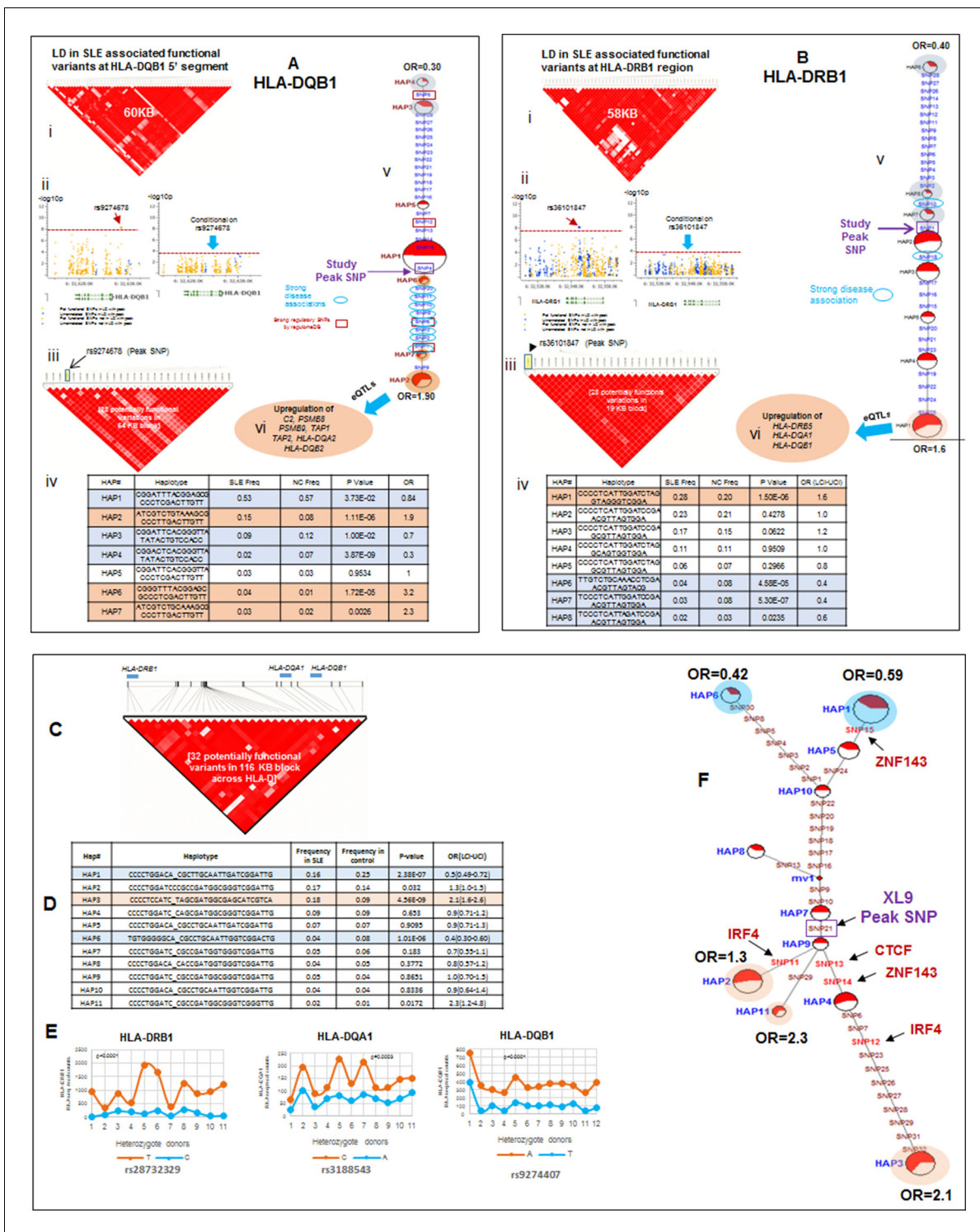
The functional implications of increased transcription of HLA-D by SLE risk associated XL9 alleles are contingent upon their impact on the surface expression levels of HLA-DR and HLA-DQ molecules on immune cell lineages. To test this, quantitative flow cytometry was performed on monocyte-derived dendritic cell cultures (MDDC) derived from the PBMC of individuals with specific XL9 haplotypes. As shown in Figure 7A.1 and A.2, HLA-DR surface expression is roughly 2.5-fold higher on unstimulated MDDC from a HAP3 (risk) homozygote in comparison to a HAP1 (protective) homozygote. As shown in Figure 7A.2, this statistically significant variation in surface expression was fully reproducible. RNA-SEQ analyses of these same MDDC cultures confirmed the increased transcription of HLA-D genes by the XL9 HAP3 donor (Figure 7B). Similarly, the surface expression of HLA-DQ on MDDC from a risk/protective heterozygote is greater than the expression levels of those from a homozygote for a protective haplotype (Figure 7C.1–C.2). This increased surface expression of HLA-DQ is maintained on dendritic cells following activation with the TLR7/8 ligand R848 in a time course over 18 hours, indicating that HLA-D risk haplotypes drive higher levels of HLA class II molecule surface expression during TLR activation and dendritic cell maturation (Figure 7C.3–C.8). Taken together, these findings indicate that variations in the XL9 regulatory region modify chromatin structure and transcription factor binding, leading to a significant increase in the surface expression of HLA class II in the dendritic cell lineage of individuals expressing SLE risk alleles of HLA-D. Finally, the transcription of several genes within the HLA complex are strongly upregulated in lymphoblastoid cell lines from risk versus protective XL9 haplotype homozygotes in the 1000 genome RNA-SEQ lymphoblastoid cell line (Lappalainen et al., 2013) data set (Figure 7D). These results, obtained via an identical analysis of a public dataset, are an independent replicate of our findings of increased expression of important HLA genes in individuals carrying HLA-D haplotypes associated with SLE.

### Regulatory effects and disease associations of composite HLA-D haplotypes

Detailed analyses of the SLE signal within the segment 5' of HLA-DQ (signal 2) revealed tight LD with twenty eight regulatory variants that are distributed through a ~40 KB segment extending 5' from the DQB1 transcription start site. The properties of these regulatory variants are provided in Supplementary file 2 and conditioning and MJ analyses are provided in Figure 8Ai-vi. Interestingly, none of the regulatory variants are correlated with eQTL effects that impact DQB1, although they are associated with transcriptional effects on other genes within the antigen processing pathway of HLA, such as PSMB8, PSMB9, TAP1, TAP2, DQA2, and DQB2. Among the variants in tight LD, seven have strong ( $\geq 900$ ) effect scores from ENCODE and 4 of these are also scored within the 1 or 2 category in the RegulomeDB database, making it likely that they impact transcription factor binding and transcription (Supplementary file 2). ENCODE predicts that these variants will impact the binding of many transcription factors including POLR2A, RELA, BATF, RUNX3, TBP, TAF1, PAX5, RFX5, EP300, and NFIC. The twenty eight variants formed seven haplotypes of which HAP4 was protective (OR 0.3,  $p < 3.88 \times 10^{-9}$ ) and HAP2 conduced risk (OR 1.9,  $p < 1.11 \times 10^{-6}$ ). As shown, two of the strongest regulatory variants and all of the variants showing strong associations with SLE are located on the final MJ branch leading to the risk clade. These results indicate that the most potent regulatory variants identified by ENCODE and associated with the increased transcription of multiple components of the antigen presentation pathway (APP), are all associated with increased risk for SLE.

The DRB1 signal (signal 3) is also associated with twenty eight regulatory variations; however, their predicted functional properties are weaker than those found in the XL9 or DQB1 signals (Supplementary file 2). The DRB1 regulatory variants are distributed in a region extending from the peak SNP through DRB1 and about 10 Kb 5' from the DRB1 start site towards the XL9 regulatory region. As shown in Figure 8Biv, they form 8 haplotypes of which HAP6 is protective (OR 0.4,  $p = 4.58 \times 10^{-5}$ ) and HAP1 is risk (OR 1.6,  $p = 1.50 \times 10^{-6}$ ). The DRB1 regulatory variants do not have strong ENCODE or RegulomeDB scores and do not contain eQTL associations with DRB1 expression. However, associations with DRB5, DQA1, and DQB1 transcription have been reported for variants in the DRB1 regulatory haplotypes. Also, although the DRB1 peak variant is strongly associated with SLE, only 2 of the regulatory variants in tight LD with this peak variant are strongly associated with SLE (Figure 8Bv) and those variants were not included in the MJ network branch proximal to





**Figure 8.** LD structure, haplotypes and MJ network analysis in *HLA-DQB1* and *HLA-DRB1* region. (A) (i) LD structure at *HLA-DQB1* 5' region generated with 68 common (MAF $\geq$ 10%) potentially functional variants in 1349 samples. (A) (ii) Zoom Manhattan plot showing SLE variant association levels and conditional analysis on peak SNP rs9274678. (A) (iii) LD block structure of 28 potentially functional SLE associated SNPs which are used for downstream haplotype analysis. (A) (iv) Haploview generated seven haplotypes from these 28 functional variants. Frequencies in cases and controls and association statistics are provided. Risk (red) and protective (blue) haplotypes are color highlighted. (A) (v) MJ networks analysis to illustrate divergence of risk and protective regulatory haplotypes. (A) (vi) eQTL variations from public databases for variants in strongest risk haplotype. (B) (i) LD structure at *HLA-DRB1* region generated with 66 common (MAF $\geq$ 10%) potentially functional variants in 1349 samples. (B) (ii) Zoom Manhattan plot showing SLE variant association levels and conditional analysis on peak SNP rs36101847. (B) (iii) LD block structure of 28 potentially functional SLE associated SNPs which

Figure 8 continued on next page

Figure 8 continued

are used for downstream haplotype analysis. (B) (iv) Haploview generated eight haplotypes from these 28 functional variants. Frequencies in cases and controls and association statistics are provided. Risk (red) and protective (blue) haplotypes are color highlighted. (B) (v) MJ networks analysis to illustrate divergence of risk and protective regulatory haplotypes. (B) (vi) eQTL variations from public databases for variants in strongest risk haplotype. Panel (C) 116 kb LD block generated with 32 SLE associated potentially functional variations from the three independent association signals in HLA-D region. (D) Haplotype association statistics in cases and controls with risk (red) and protective (blue) haplotypes highlighted. (E) Allelic bias in level of transcription for HLA-class II genes between SLE risk and non-risk alleles in 11 independent heterozygous donors (measured as shown in **Figure 6**). Number of RNA sequencing reads were compared between chromosome carrying risk (orange line) versus non-risk (blue line) allele for each class II gene. (F) MJ network analysis illustrating the relationships of risk and non-risk haplotypes based on 32 functional variations. SLE associated variants sitting exactly within specific protein binding motifs i.e. IRF4, CTCF and ZNF143 are highlighted with arrows.

DOI: [10.7554/eLife.12089.011](https://doi.org/10.7554/eLife.12089.011)

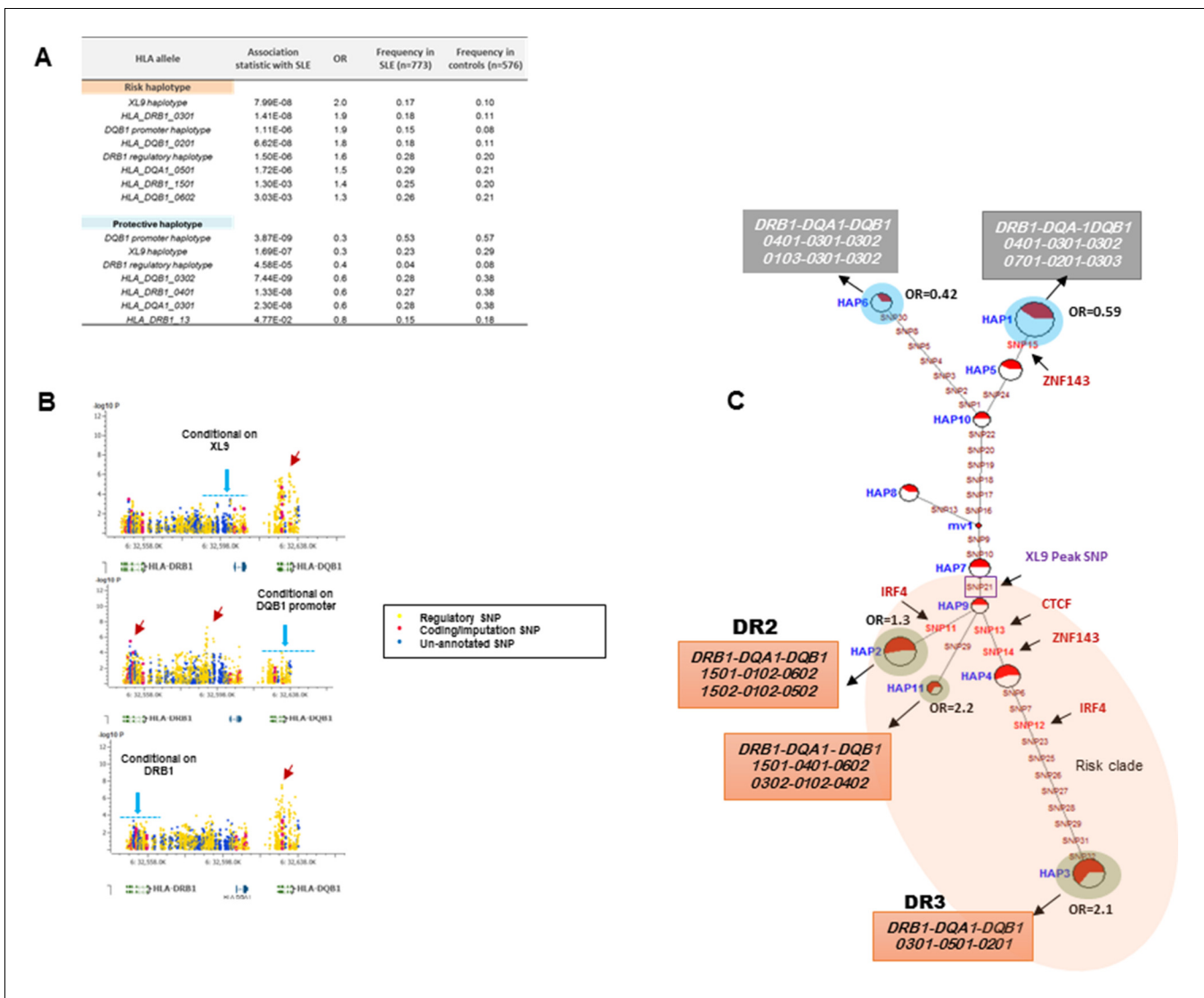
the risk clade. Taken together, these results suggest that these DRB1 regulatory variations may not play a dominant role in the endophenotypes causing the association with SLE.

The SLE associations of regulatory variations spanning the entire HLA-D interval were assessed using composite haplotypes formed with 32 regulatory variants with strong SLE association signals that were derived from the three independent SLE associated signals. As shown in **Figure 8C**, these variants spanned a 116 Kb block containing DRB1, DQA1, and DQB1 and are all in tight LD. The composite analysis formed eleven haplotypes that accounted for more than 90% of all of the chromosomes identified within the EA panel (**Figure 8D**). As shown, HAP1 ( $p=2.38E-07$ , OR = 0.59) and HAP6 ( $p=1.01E-06$ , OR = 0.42) are protective, and HAP3 ( $p=4.56E-09$ , OR = 2.1), HAP2 ( $p=0.032$ , OR = 1.3), and HAP11 ( $p=0.0172$ , OR = 2.3) are risk. MJ analysis revealed a pattern similar to that obtained for XL9 haplotypes (**Figure 5F** versus **Figure 8F**). As shown in **Figure 8E**, an assessment of chromosome-specific transcription in MDM cultures from eleven heterozygotes for the rs9271593 (peak XL9 signal) revealed consistent and highly significant increases in transcription of HLA-DRB1, HLA-DQA1, and HLA-DQB1 for chromosomes carrying risk-associated variants. These results demonstrate that all of the regulatory haplotypes carrying the risk allele of rs9271593, which is the XL9 peak SNP (**Figure 8F**), transcribe higher levels of HLA-D class II genes than protective regulatory haplotypes.

### Comparing HLA class II alleles and HLA-D regulatory haplotypes in SLE susceptibility

The HLA-D sequence data allowed the imputation of standard HLA-DRB1, HLA-DQA1, and HLA-DQB1 class II alleles with four digit accuracy for the EA cohort, using algorithms and strategies described previously (**Morris et al., 2012**). The imputed HLA class II allele designations were used to assess the associations of HLA class II alleles with SLE in our EA cohort and to assess the relationships of these classical HLA-D class II alleles with the defined HLA-D regulatory haplotypes. The table in **Figure 9A** lists the SLE association statistics for the most strongly associated HLA class II alleles and HLA-D regulatory haplotypes in the EA cohort. As shown, HLA-DRB1\_0301 and HLA-DQB1\_0201 were the most strongly associated HLA-D class II alleles in this analysis, which is consistent with several previous studies of HLA-D associations with SLE (**Morris et al., 2012**; **Armstrong et al., 2014**; **Fu et al., 2011**; **Furukawa et al., 2014**; **Kim et al., 2014**; **Morris et al., 2014**; **Fernando et al., 2012**). The disease associations and odds ratios detected for HAP3 of the XL9 signal (**Figure 5F**) and HAP2 of the DQB1 promoter signal are equivalent with these HLA class II alleles. Similarly, as shown in the bottom of **Figure 9A**, both regulatory haplotypes and HLA class II alleles are strongly associated with protection from SLE. Thus, HAP4 of the DQB1 promoter signal, HAP6 of the XL9 signal, and HAP7 of the DRB1 signal all have potent association statistics and odds ratios for decreased frequencies in cases, as do the classic HLA-D class II alleles HLA-DQB1\_0302, HLA-DRB1\_0402, and HLA-DQA1\_0301.

**Figure 9B** presents conditional analyses of the SLE associations contributed by the HLA-D regulatory haplotypes and the imputed HLA class II alleles. The top plot illustrates that XL9 regulatory polymorphisms can completely remove the associations of HLA-DRB1 and HLA-DQA1 imputation variants with SLE, but does not remove the DQB1 imputation or 5' region regulatory haplotype associations. Similarly, conditioning on the DQB1 regulatory haplotype removes the association of HLA-DQB1 imputation SNPs with disease, but has little effect on the imputation or regulatory variations in DRB1, DQA1, or XL9 (**Figure 9B**, middle panel). Finally, conditioning on HLA-DRB1 imputation

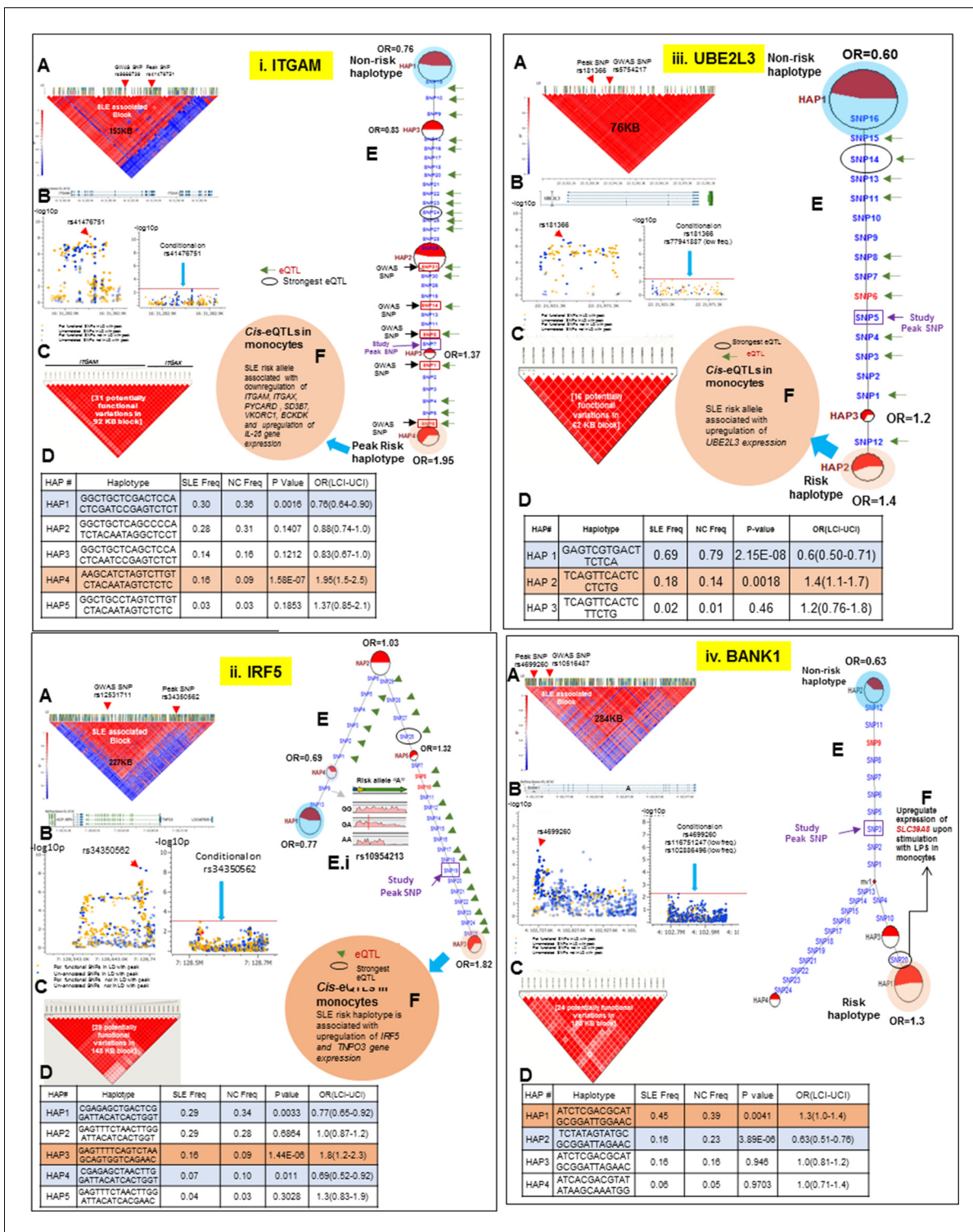


**Figure 9.** HLA-D regulatory haplotypes and classical HLA alleles. (A) SLE association statistics of regulatory and classical HLA alleles in this study. (B) Conditional analysis on peak regulatory signals in XL9, DQB1 and DRB1 regions. (C) Median-joining (MJ) network analysis of 32 regulatory variants spanning HLA-DRB1 to DQB1 region. SLE associated variants sitting directly on canonical binding motif of CTCF, IRF4 and ZNF143 transcription factor are indicated with arrows. The HLA DRB1-DQA1-DQB1 haplotypes associated with each of the risk and protective regulatory haplotypes are presented. DOI: 10.7554/eLife.12089.012

variants removes the association of XL9 regulatory variants and the HLA-DQA1 and HLA-DQB1 imputation SNPs, but does not significantly impact the association of the DQB1 regulatory haplotype with SLE. These analyses indicate that variations in XL9 and HLA-DRB1 class II alleles are in tight LD and represent a combined contribution to SLE, while variations in the segment 5' of DQB1 independently contribute to SLE susceptibility.

Finally, **Figure 9C** presents the MJ network formed by the composite HLA-D regulatory haplotypes (from **Figure 8F**) and overlays the imputed DRB1-DQA1-DQB1 HLA class II alleles present in individuals homozygous for the protective and risk HLA-D regulatory haplotypes. As shown, LD is very strong, but incomplete between the classical HLA class II alleles and the regulatory haplotypes. Notably, all homozygotes for regulatory HAP3 (peak risk) are also homozygous for DRB1\_0301, DQA1\_0501, DQB1\_0201, which is the HLA-D class II haplotype found in the extended DR3 haplotype (Kachru, 1984; Smolen et al., 1987; Hohler and Buschenfelde, 1994; Schur et al., 1990; Niu et al., 2015). Similarly, regulatory risk HAP2 is predominantly associated with the DRB1\_1501, DQA1\_0401, DQB1\_0602 haplotype, which has also been previously associated with susceptibility to





**Figure 10.** LD structure, haplotypes and MJ network analysis of *ITGAM*, *IRF5*, *UBE2L3* and *BANK1*. Panel 10 (i) shows *ITGAM*, Panel 10 (ii) shows *IRF5*, Panel 10 (iii) shows *UBE2L3* and Panel 10 (iv) shows *BANK1* genetic association analysis. (A) LD structure of studied intervals generated with common ( $MAF \geq 10\%$ ) variants in 1349 samples, 221 in case of *ITGAM*, 400 in case of *IRF5*, 84 in case of *UBE2L3* and 430 variants in case of *BANK1*. (B) Zoom Manhattan plot of all common variants in studied region showing SLE association levels and conditional analysis on peak SNP/s. (C) LD block based on potentially functional SLE associated SNPs which are used for downstream haplotype analysis. (D) Haploview generated haplotypes from functional variants. Frequencies in cases and controls and association statistics are provided. Risk (red) and protective (blue) haplotypes are color highlighted. (E) MJ networks analysis to illustrate divergence of risk and protective regulatory haplotypes. Haplotype with significant p value ( $p < 0.05$ ) are highlighted

Figure 10 continued on next page



Figure 10 continued

with red (risk) and blue (non-risk) color. Study peak SNP, previously known SLE GWAS tag SNP and eQTLs are indicated with arrows. (F) eQTL variations from public databases for variants in strongest risk haplotype.

DOI: [10.7554/eLife.12089.013](https://doi.org/10.7554/eLife.12089.013)

SLE. Overall, the DR and DQ alleles that have been associated with SLE in previous studies of EA cohorts are found among the regulatory risk haplotypes and are absent from the protective clade (Morris *et al.*, 2012; 2014; Niu *et al.*, 2015; Ramos *et al.*, 2010). Taken together, these results are consistent with the strong LD within this small genomic segment of HLA and suggest that the regulatory variations and the peptide binding groove polymorphisms are two aspects of HLA-D diversification that are tightly intertwined within allelic lineages of the HLA-D region.

## Regulatory haplotypes in non-HLA risk loci

**Table 2** provides a summary for all sixteen SLE risk loci that have been analyzed in this study. Detailed analyses for the fourteen loci not discussed above are presented in **Figures 10–13**. Several characteristics of the genetic variations that underlie common SLE susceptibility alleles are revealed by these data. First, maximal risk for disease is associated with specific haplotypes typically composed of five or more functional variations that are in strong LD with the peak risk variant. The overwhelming majority of these variants are in regulatory elements (1199 of 1206, **Table 1**) and carry ENCODE scores indicating that they are potent functional polymorphisms. They occur as stable haplotypes within the EA population and are predicted to impact multiple endophenotypes. MJ analysis revealed that the risk and protective/non-risk haplotypes are typically at opposite ends of the networks (15 of 16 risk loci), indicating that significant variations in disease risk are most strongly associated with multiple functional changes. Furthermore, for some risk loci, multiple haplotypes are significantly associated with either risk or protection, but with varying odds ratios, indicating that a spectrum of functional haplotypes with varying disease risk contributions underlie the disease association of individual risk loci.

As tabulated in **Table 2**, peak risk haplotypes have greater odds ratios for SLE susceptibility than individual peak tagging SNPs from the original published GWAS studies and from the targeted association studies performed here. Overall, the peak risk haplotype had a higher odds ratio than the peak GWAS tagging SNP for 13 of 16 loci, resulting in an overall 17% increase in the average odds ratio for the sixteen loci tested. This result is consistent with theoretical predictions of the increase in odds ratio that would be achieved by specifically identifying causative variants in complex disease risk loci, thus supporting the presence of the causal variants of SLE within the identified functional haplotypes (Gusev *et al.*, 2013; Yang *et al.*, 2010).

Thirteen of the SLE risk loci characterized in detail here were identified previously by our group and others and the detailed sequence analyses of these risk loci has confirmed and extended these previous findings. Several of these loci contain long haplotypes, as discussed for *IRF5-TNPO3*, *ITGAM-ITGAX*, *TNFAIP3*, *UBE2L3* and *BANK1*, while *TNFSF4* and *XKR6* each contain two independent association signals in separate LD blocks (**Figures 10–13** and **Tables 1** and **2**). Our analyses found that regulatory haplotypes often contain variants impacting several eQTLs, chromatin structure and transcription regulatory elements. The ENCODE defined regulatory elements for *POLR2A*, *CTCF*, *IRF4*, *RELA*, *STAT5A*, *RFX5*, *RUNX3* were the most common regulatory elements affected by SLE associated variants (**Supplementary file 2**). Finally, three risk loci, *CCL22-CX3CL1* (**Figure 11.iii**), *ZGLP1-RAVER1* (**Figure 11.iv**), and *ICA1* (**Figure 12.i**), which were comparatively less well-studied for SLE association, were detected with strong statistical associations in this EA cohort (**Table 1**, **Supplementary file 2**). Detailed sequence analysis of these loci identified significant associations of these genes with SLE and identified SLE associated haplotypes impacting multiple regulatory components. More results on these loci have been incorporated into the relevant Figure legend for each risk locus.

## Discussion

These analyses provide a comprehensive assessment of the genomic variations associated with SLE disease alleles. We identified 345 regulatory variations impacting gene transcription within these loci

**Table 2.** Summary of SLE association and functional characteristics of peak variants and functional haplotypes for 16 SLE risk loci.

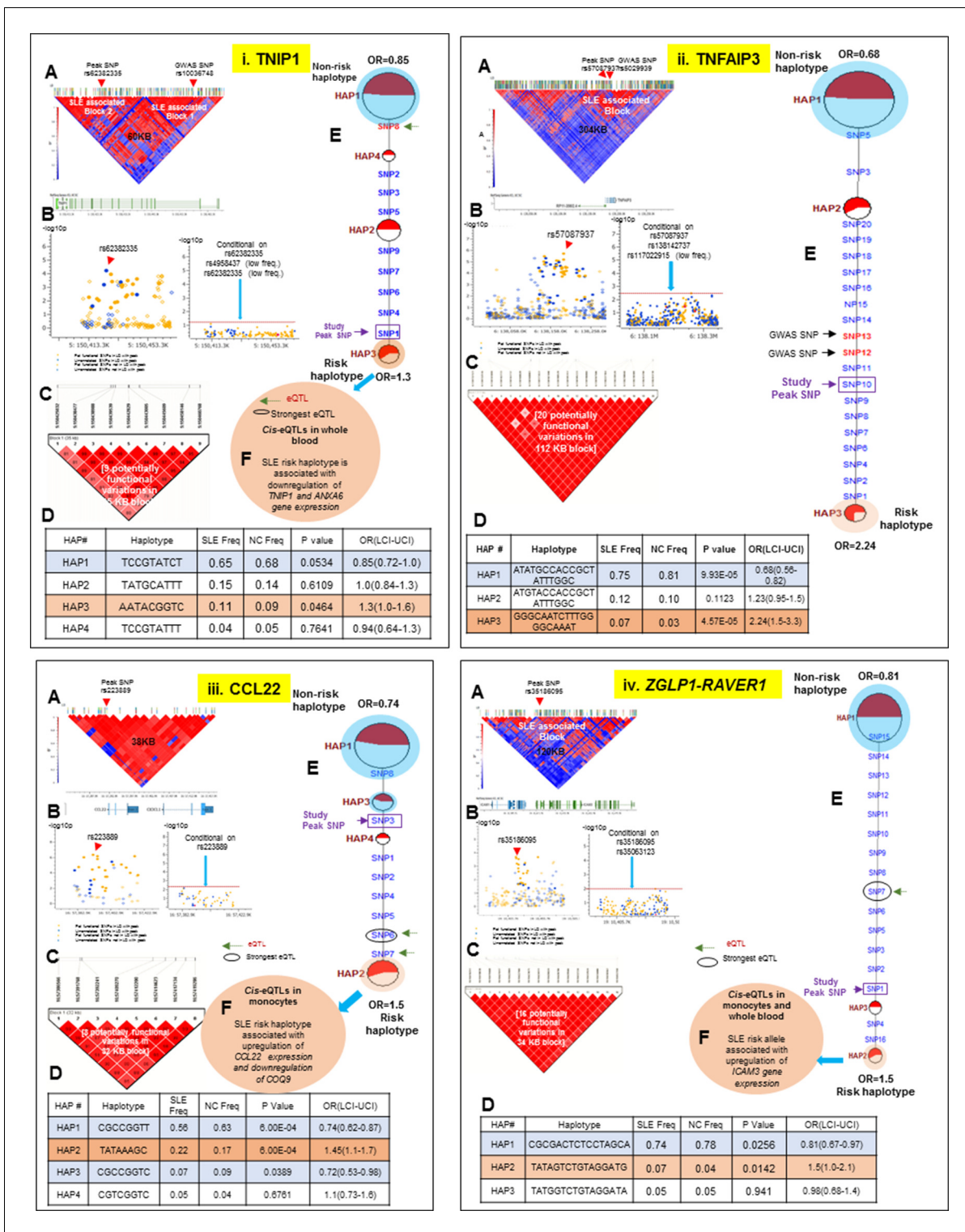
Gene	Known GWAS (tag) SNP	GWAS reference	GWAS (tag) SNP OR	Study peak SNP	Study peak SNP OR	Peak risk-associated functional haplotype	Risk haplotype OR	Increase in OR of haplotype versus GWAS SNP	Increase in OR of haplotype versus study peak SNP	Related Figure in the manuscript	Strongest ENCODE effect cell line/tissue	eQTL data cell type/tissue
STAT4	rs7574865	Lee et al., 2012	1.4	rs12612769	1.7	ATTCCTTGC	1.7	0.3	0	Figure 4	Mammary gland, Epithelial	Monocyte, macrophage
HLA-D	rs1150754	Taylor et al., 2011	1.54	rs9271593 (XL9)	1.6910364	CCCCTCCATC_ TAGCGATGGCG AGCATCGTCA	2.1	0.56	0.4089636	Figure 8C-F	B-lymphocyte, lymphoblastoid	Monocyte
ITGAM-ITGAX	rs9888739	Harley et al., 2008	1.6	rs41476751	1.8696398	AAGCATC TAGCTTT GTCTACAA TAGTCTCTC	1.95	0.35	0.0803602	Figure 10.i	B-lymphocyte, lymphoblastoid	Monocyte, Peripheral blood
IRF5-TNPO3	rs12531711	Chung et al., 2011	1.5	rs34350562	1.7593583	GAGTT TTCAGTCTA AGCAGT GGTGAGAAC	1.8	0.3	0.0406417	Figure 10.ii	Epithelial cell (Lung), B-lymphocyte	Monocyte, macrophage
UBE2L3	rs754217	Chung et al., 2011	1.3	rs181366	1.5217361	TCAGTTTCACTCCTCTG	1.4	0.10	-0.1140361	Figure 10.iii	Epithelial cell (Lung), B-lymphocyte	Monocyte
BANK1	rs10516487	Kozyrev et al., 2008	1.3	rs4699260	1.25	ATCTCGACGCA TCGGGA TTGGAAC	1.3	0	0.05	Figure 10.iv	Hela-S3, Epithelial, Fibroblast	Monocyte
TNIP1	rs10036748	Han et al., 2009; Galimberti et al., 2008	1.2	rs62382335	1.37	AAATACGGTC	1.3	0.12	-0.05	Figure 11.i	B-lymphocyte, lymphoblastoid	Peripheral blood
TNFAIP3	rs5029939	Graham et al., 2008	2.2	rs57087937	1.9092441	GGGCAATCT TTGGGGCAAAT	2.2	0.04	0.3307559	Figure 11.ii	B-lymphocyte, lymphoblastoid, hepatocyte	no data
CCL22-CX3CL1	rs223889	Galimberti et al., 2008	1.4	rs223889	1.45	TATAAAGC	1.5	0.05	0	Figure 11.iii	B-lymphocyte, lymphoblastoid	Monocyte
ZGLIP-RAVER1	rs35186095	Present study	1.3	rs35186095	1.3173789	TATAGTCT GTAGGATG	1.5	0.2	0.1826211	Figure 11.iv	Fibroblast, K-562, Hela-S3, B-lymphocyte	Monocyte
ICA1	rs10156091	Harley et al., 2008	1.3	rs74787882	1.5	GGGT	1.5	0.2	0	Figure 12.i	B-lymphocyte, lymphoblastoid	no data
BLK	rs13277113	Hom et al., 2008	1.3	rs7822109	1.26	ATTTGCCCA	1.3	0	0.04	Figure 12.ii	B-lymphocyte, lymphoblastoid	Monocyte, Peripheral blood
ETS1	rs7932088	Yang et al., 2010	1.2	rs34516251	1.23	GGGCGA	1.4	0.2	0.17	Figure 12.iii	B-lymphocyte, lymphoblastoid, Epithelial	Monocyte

Table 2 continued on next page

Table 2 continued

Gene	Known GWAS (tag) SNP	GWAS (tag) SNP OR	GWAS reference	Study peak SNP	Study peak SNP OR	Peak risk-associated functional haplotype	Risk haplotype OR	Increase in OR of haplotype versus GWAS SNP	Increase in OR of haplotype versus study peak SNP	Related Figure in the manuscript	Strongest ENCODE effect cell line/ tissue	eQTL data cell type/ tissue
TNFSF4	rs2205960	1.3	Han et al., 2009	rs4916313	1.3	TCCATCTTCGA	1.3	0	0	Figure 13.i	Epithelial cell (Lung), Fibroblast, HeLa-S3	no data
NMNAT2-SMG7	rs2022013	1	Cunninghame Graham et al., 2011	rs111487113	1.3	TCACTAAC	1.3	0.3	0	Figure 13.ii	Primary Th1 T cells	no data
XKR6	rs11783247	1.2	Harley et al., 2008	rs7000132	1.2	TGTCGGGGCTT	1.2	0.03	0.03	Figure 13.iii	Neuroblastoma, Mammary gland, Fibroblast	Monocyte

DOI: 10.7554/eLife.12089.014



**Figure 11.** LD structure, haplotypes and MJ network analysis of *TNIP1*, *TNFAIP3*, *CCL22* and *ZGLP1-RAVER1*. Panel 11 (i) shows *TNIP1*, Panel 11 (ii) shows *TNFAIP3*, Panel 11 (iii) shows *CCL22* and Panel 11 (iv) shows *ZGLP1-RAVER1* genetic association analysis. (A) LD structure of studied intervals generated with common ( $MAF \geq 10\%$ ) variants in 1349 samples, 140 in case of *TNIP1*, 356 in case of *TNFAIP3*, 30 in case of *CCL22* and 126 variants in case of *ZGLP1-RAVER1*. (B) Zoom Manhattan plot of all common variants in studied region showing SLE association levels and conditional analysis on peak SNP/s. (C) LD block based on potentially functional SLE associated SNPs which are used for downstream haplotype analysis. (D) Haploview generated haplotypes from functional variants. Frequencies in cases and controls which and association statistics are provided. Risk (red) and protective (blue) haplotypes are color highlighted. (E) MJ networks analysis to illustrate divergence of risk and protective regulatory haplotypes. Haplotype with

*Figure 11 continued on next page*

Figure 11 continued

significant p value ( $p < 0.05$ ) are highlighted with red (risk) and blue (non-risk) color. Study peak SNP, previously known SLE GWAS tag SNP and eQTLs are indicated with arrows. (F) eQTL variations from public databases for variants in strongest risk haplotype.

DOI: [10.7554/eLife.12089.015](https://doi.org/10.7554/eLife.12089.015)

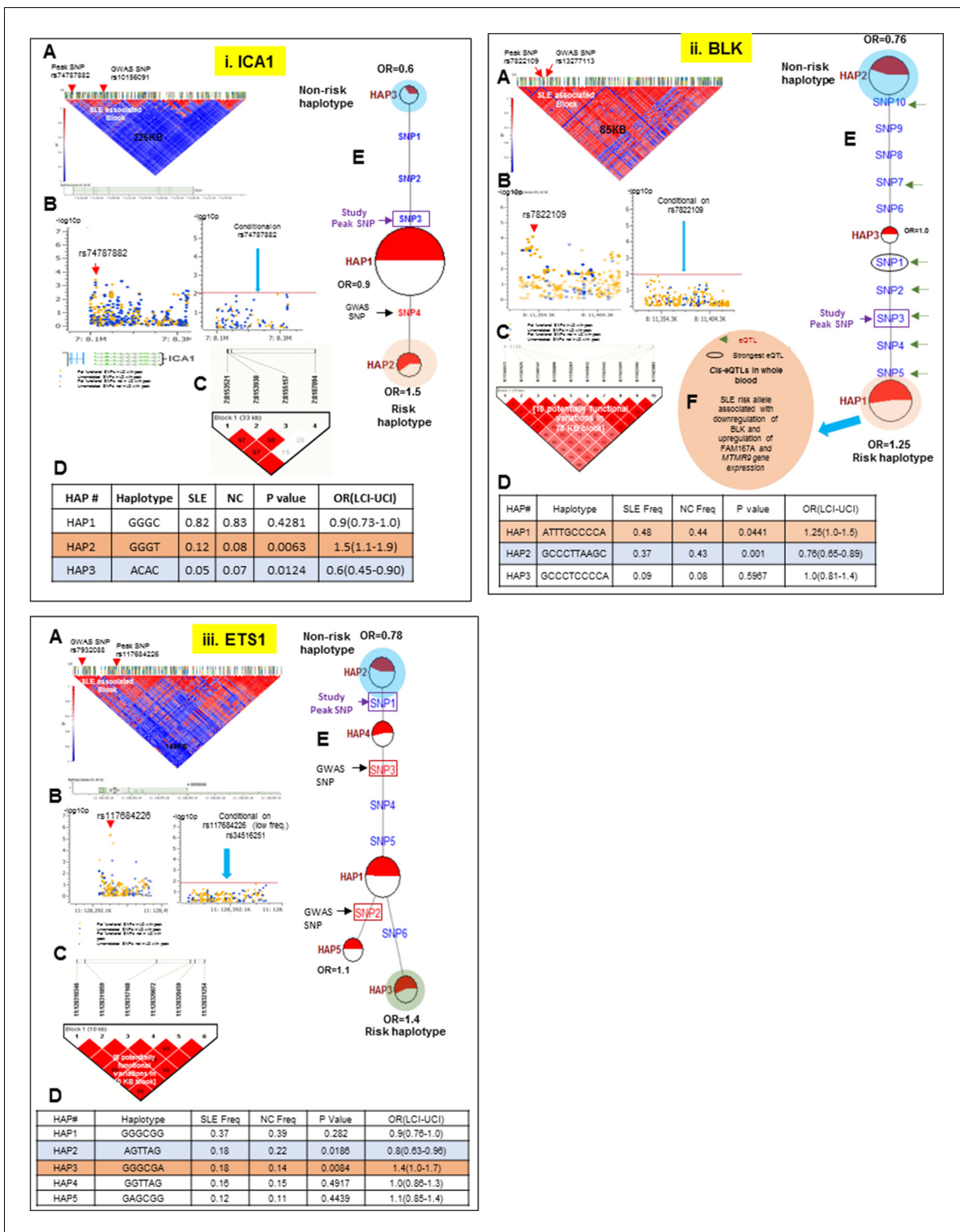
that exhibited stronger disease-associations than previously identified GWAS tagging SNPs (**Figure 3B**). Detailed analyses of the allelic architecture at these loci revealed that SLE disease alleles are haplotypes composed of multiple functional variations and that these variations often modulate several endophenotypes. This architecture is modeled in **Figure 14A**, which depicts the manner in which several functional variants in tight LD with GWAS tagging SNPs mediate transcriptional variations at multiple adjacent genes. As shown, the functional haplotypes identified in our analyses capture all of these causal variants within the LD block, which leads to the identification of a peak risk haplotype with increased disease association. Our analyses identified variants impacting multiple transcriptional changes at nine of the sixteen SLE risk loci, indicating that this level of complexity is prevalent among SLE risk loci. In this regard, we could only utilize our own MDM eQTL database and a few public eQTL databases for immune cell lineages in this analysis (datasets for monocytes, PBMC, and LBL are currently accessible, (Fairfax et al., 2014; Raj et al., 2014; Westra et al., 2013) and it is quite likely that these haplotypes will be found to have additional effects as more datasets become available.

**Figure 14B** presents the odds ratios for disease obtained with GWAS-defined tagging SNPs (Harley et al., 2008; Graham et al., 2008; Adriantoet al., 2011; Taylor et al., 2011) and peak risk haplotypes for the sixteen SLE loci analyzed. As shown, the odds ratio for disease obtained for the peak risk haplotype at each locus was consistently higher than that of the tagging SNP, leading to an average increase of 17% in odds of disease overall (**Table 2** and illustrated in **Figure 14B**). These results support the presence of most or all of the causal variations for disease susceptibility within the identified peak risk haplotypes for each locus. Interestingly, we identified both protective and risk haplotypes with nominal statistical significance ( $p < 0.05$ ) at fourteen of the sixteen risk loci analyzed, suggesting that both types of disease alleles are prevalent and contribute to population risk. MJ analyses consistently found peak risk and protective haplotypes at opposite ends of the network, which suggests that the combined effects of multiple regulatory variations may additively impact disease associations. Consistent with this, HLA-D, STAT4, IRF5, and CCL2-CX3C1 all have multiple haplotypes with different disease associations, suggesting that a spectrum of disease alleles with different impacts on susceptibility may occur at highly variable risk loci. However, some caution is appropriate when interpreting the significance of multiple intermediate risk haplotypes within a network, in that sample numbers for many haplotypes were often small. Consequently, a larger sample of SLE patients will be required for the detailed assessment of the disease risk attributable to all of the prevalent haplotypes at SLE risk loci.

These haplotypes represent stable polymorphisms within the EA population, with six or fewer haplotypes accounting for about 90% of the LD block regulatory sequences segregating in the EA population at individual risk loci (**Table 2** and **Figures 10–13**). We assessed this in more detail by determining the frequencies of the major risk haplotypes for HLA-D, STAT4, IRF5, ITGAM, and UBE2L3 in 2504 individuals derived from 26 global ethnic populations sequenced in the 1000 Genomes project (Abecasis et al., 2010; 2012). As shown in **Figure 14C**, these five risk haplotypes are present at variable frequencies within all of the European, South American, and South Asian populations. However, they are much less frequent in the African populations sampled and, with the exception of UBE2L3, absent from the East Asian populations. This suggests that additional haplotypes will be identified during the analysis of specific ethnic groups as previously shown (Kim-Howard et al., 2014). In addition, although these five SLE risk haplotypes are predominantly found in European populations and populations with significant European admixture, they are also detectable with low frequencies in some African populations. Based on this distribution, it is likely that these haplotypes arose prior to human global colonization and that they were present with divergent frequencies in the ancestral founding populations of modern ethnic groups.

This dataset can also be used to assess the percentage of population disease risk that is attributable to the combined effects of all of these risk loci. As tabulated in **Table 2**, the cumulative risk associated with the GWAS tagging SNPs for all sixteen loci sums to 6.04 fold, while the same value for





**Figure 12.** LD structure, haplotypes and MJ network analysis of *ICA1*, *BLK* and *ETS1*. Panel 12 (i) shows *ICA1*, Panel 12 (ii) shows *BLK* and Panel 12 (iii) shows *ETS1* genetic association analysis. (A) LD structure of studied intervals generated with common (MAF $\geq$ 10%) variants in 1349 samples, 370 in case of *ICA1*, 258 in case of *BLK* and 209 variants in case of *ETS1* (B) Zoom Manhattan plot of all common variants in studied region showing SLE association levels and conditional analysis on peak SNP/s. (C) LD block based on potentially functional SLE associated SNPs which are used for downstream haplotype analysis. (D) Haploview generated haplotypes from functional variants. Frequencies in cases and controls and association statistics are provided. Risk (red) and protective (blue) haplotypes are color highlighted. (E) MJ networks analysis to illustrate divergence of risk and protective regulatory haplotypes. Haplotype with significant p value (p<0.05) are highlighted with red (risk) and blue (non-risk) color. Study peak SNP, Figure 12 continued on next page

Figure 12 continued

previously known SLE GWAS tag SNP and eQTLs are indicated with arrows. (F) eQTL variations from public databases for variants in strongest risk haplotype.

DOI: [10.7554/eLife.12089.016](https://doi.org/10.7554/eLife.12089.016)

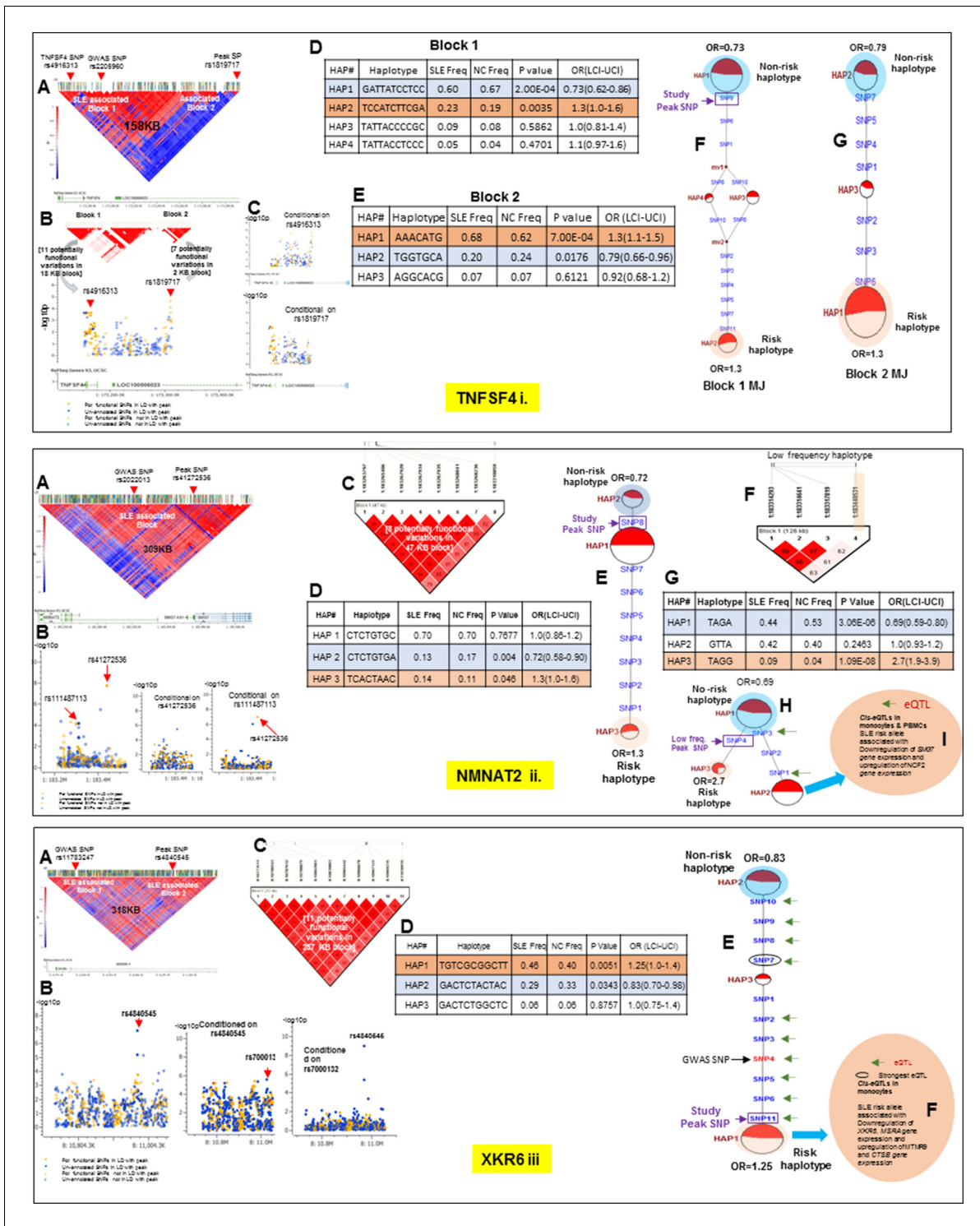
all peak risk haplotypes is 8.8, indicating that improved resolution of disease alleles increases the disease risk and proportion of 'heritability' that is associated with these common disease alleles. Assuming that the contribution of all risk loci for SLE sum to 29 (Alarcon-Segovia et al., 2005; Deapen et al., 1992), then the sum of these sixteen loci would account for about one third of the genetic heritability for SLE. In this regard, a contentious debate concerning the contribution of 'common' (MAF > 0.05) versus 'rare' (MAF << 0.05) disease risk alleles to the overall heritability of common diseases has persisted among investigators in complex phenotype genetics for several years (Raychaudhuri, 2011; Cirulli and Goldstein, 2010; Manolio et al., 2009; Pritchard and Cox, 2002). Although it is clear that rare alleles contribute to disease susceptibility in small subsets of patients (Hunt et al., 2013; Lee-Kirsch et al., 2007; Tang et al., 2014; Mitchell et al., 2002), recent analytical studies have firmly established that common disease alleles are responsible for the bulk of the heritability for autoimmune diseases (Yang et al., 2010; Visscher et al., 2010; Stahl et al., 2012). It is likely that 'missing' heritability predominantly reflects an extensive genetic heterogeneity that underlies many common diseases.

An alternative method to measure the cumulative risk attributable to a specific collection of risk loci is via the calculation of population attributable risk (PAR) (Zheng et al., 2008; Bruzzi et al., 1985; Rockhill et al., 1998a; 1998b; Mezzetti et al., 1998; Natarajan et al., 2007; Claus et al., 1996; Kraft et al., 2009; Pepe et al., 2004). This calculation utilizes the odds ratio for disease and the risk allele population frequency to calculate a weighted risk value for each locus and then combines them to assess their contribution to genetic risk for the population as a whole. As shown in Figure 14B and tabulated in Supplementary file 3B, the peak risk haplotypes at these sixteen risk loci account for 66% of the population attributable risk for SLE within this EA cohort (Supplementary file 3A and 3B). PAR and estimates of 'heritability' differ in that PAR calculations do not assume a specific level of population genetic risk (i.e. 29), but instead simply calculates the proportion of risk that cannot be accounted for by the variables assayed within the population studied, and thus determining the proportion of risk that is attributable to tested factors. The calculated PAR in this analysis indicates that these sixteen loci contribute a significant proportion of disease risk within our population. However, a larger cohort and broader list of risk loci will be essential to estimate genetic risk in all populations and account for a larger proportion of SLE heritability.

Our results provide a system using sequence analyses that can efficiently and accurately identify disease risk alleles within large populations. Further, we define a path forward for the development of useful genetic tools for assessing disease risk. The next phase of genetic analyses of autoimmune disease will involve assessing the functional properties of these disease alleles, sorting out their interactions during disease development, and developing analytical tools for the accurate quantitation of genetic risk for disease in individual genomes (Ray and Hacohen, 2015; Ghodke-Puranik and Niewold, 2015; Lewi et al., 2015; Wang et al., 2015; Mohan and Putterman, 2015).

## HLA-D polymorphisms, antigen presentation pathways, and autoimmune disease

The most intriguing result of our sequence analyses is the discovery of a strong association between SLE susceptibility and HLA-D polymorphisms that regulate HLA class II gene expression. The HLA-D region is consistently a potent susceptibility locus in autoimmunity and significant effort has focused on defining the molecular mechanisms that mediate autoimmunity in the context of specific HLA-D class II alleles (Morris et al., 2012; 2014; Armstrong et al., 2014; Kim et al., 2014; Niu et al., 2015; Graham et al., 2007; Cruz-Tapias et al., 2012; Todd et al., 1987). Multiple genetic studies have identified coding variations in the peptide binding sites of MHC class II molecules as key genetic components of the disease associations, strongly supporting the hypothesis that allelic variations in the antigen presentation process underlie autoimmune disease (Kim et al., 2014; Morris et al., 2014; Fernando et al., 2012; Raychaudhuri et al., 2012). The dominant paradigm has been that the peptide binding regions of disease-associated HLA class II alleles have unique



**Figure 13.** LD structure, haplotypes and MJ network analysis of *TNFSF4*, *NMNAT2* and *XKR6*. Panel 13 (i) shows *TNFSF4*, Panel 13 (ii) shows *NMNAT2* and Panel 13 (iii) shows *XKR6* genetic association analysis. These three interval showed more than one independent LD block associated with SLE in our analysis. (A) LD structure of studied intervals generated with common (MAF ≥ 10%) variants in 1349 samples, 152 in case of *TNFSF4*, 411 in case of *NMNAT2* and 643 variants in case of *XKR6*. In case of *TNFSF4*, (B) shows two SLE associated LD blocks and zoom Manhattan plot of all common variants in studied region. (C) showing SLE association levels and conditional analysis on peak SNP/s. (D and E) Haplotype generated from functional variants in block 1 and block 2, respectively. Frequencies in cases and controls and association statistics are provided. Risk (red) and protective (blue) haplotypes are color highlighted. Similarly, (F and G) shows MJ networks analysis to illustrate divergence of risk and protective regulatory haplotypes from block 1 and block 2, respectively. Haplotype with significant p value (p < 0.05) are highlighted with red (risk) and blue (non-risk) *Figure 13 continued on next page*

Figure 13 continued

color. Study peak SNP, previously known SLE GWAS tag SNP and eQTLs are indicated with arrows. In case of NMNAT2 (13.ii), (B) shows zoom Manhattan plot of all common variants in studied region showing SLE association levels and conditional analysis on peak SNP/s. Panel C: LD block based on potentially functional SLE associated SNPs which are used for downstream haplotype analysis. (D) Haploview generated haplotypes from functional variants. Frequencies in cases and controls and association statistics are provided. Risk (red) and protective (blue) haplotypes are color highlighted. (E) MJ networks analysis to illustrate divergence of risk and protective regulatory haplotypes. (F) LD block based on a low frequency SLE associated variant (G) Low frequency haplotype association analysis (H) MJ networks analysis with low frequency haplotype and (I) eQTL variations from public databases for variants in risk haplotype. In case of XKR6 (13.iii), (B) shows zoom Manhattan plot of all common variants in studied region showing SLE association levels and conditional analysis on peak SNP/s. (C) LD block based on potentially functional SLE associated SNPs which are used for downstream haplotype analysis. (D) Haploview generated haplotypes from functional variants. Frequencies in cases and controls and association statistics are provided. Risk (red) and protective (blue) haplotypes are color highlighted. (E) MJ networks analysis to illustrate divergence of risk and protective regulatory haplotypes. Haplotype with significant p value ( $p < 0.05$ ) are highlighted with red (risk) and blue (non-risk) color. Study peak SNP, previously known SLE GWAS tag SNP and eQTLs are indicated with arrows. (F) eQTL variations from public databases for variants in strongest risk haplotype.

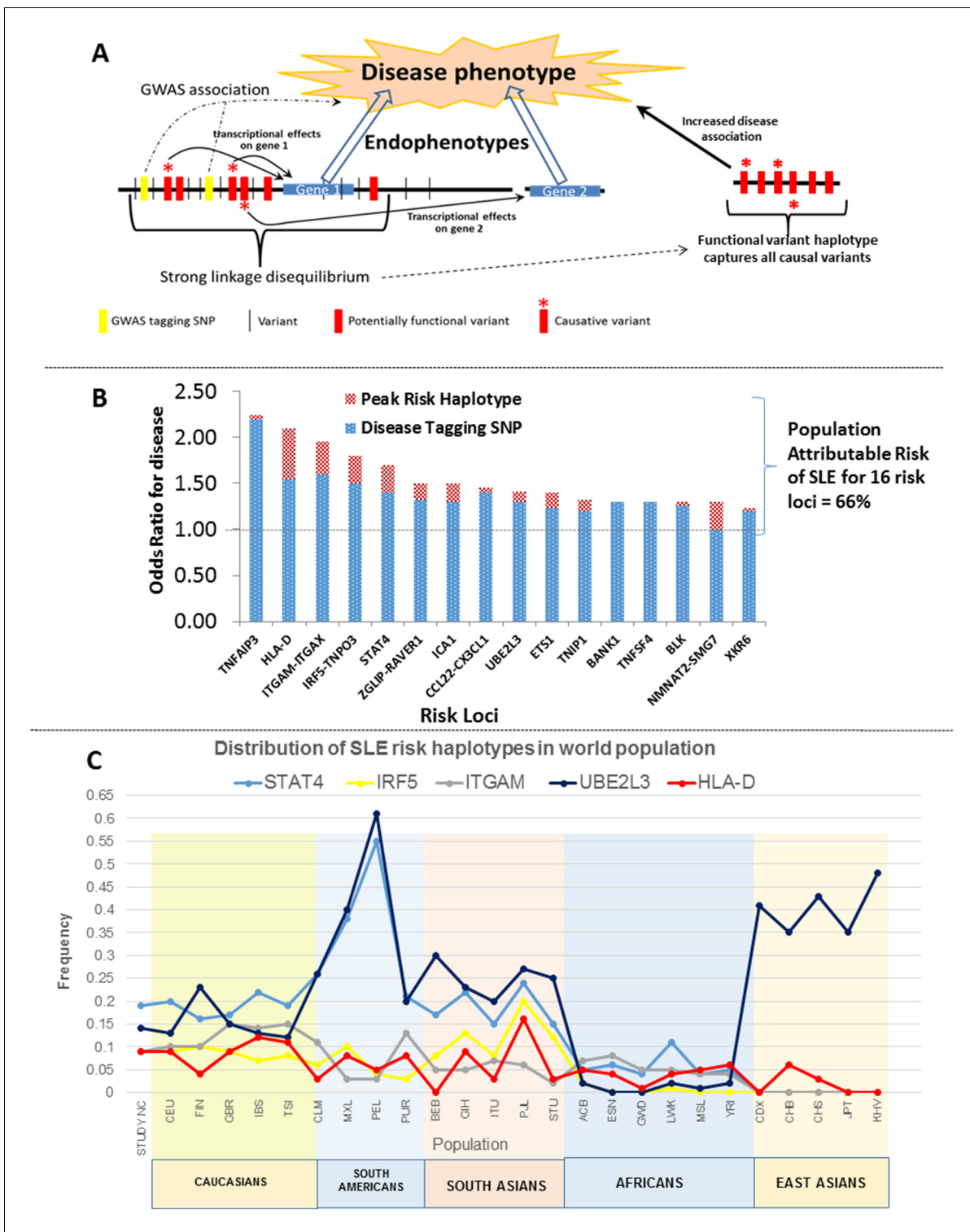
DOI: 10.7554/eLife.12089.017

peptide binding properties that present a novel spectrum of self-peptides or modified self-peptides in a manner capable of eliciting autoimmunity. Solid data supporting this mechanism have been developed by decades of experiments, notably for insulin peptides in autoimmune diabetes (Unanue, 2014). However, many studies have found that multiple self-antigens are recognized by T cells clones isolated from the earliest stages of disease development, suggesting that HLA-D associated autoimmunity is initiated against multiple self-antigens by a heterogeneous T cell response. Notably, SLE patients have a profound breach in immune tolerance and typically produce autoantibodies binding more than ten different self-antigens, with the diversity of autoantigens recognized increasing as individuals approach disease diagnosis (Olsen and Karp, 2014; Arbuckle et al., 2003; Li et al., 2005). Further, multiple HLA class II DR and DQ alleles are associated with SLE susceptibility, which indicates that HLA class II alleles with highly divergent peptide binding properties are capable of promoting disease development. In this regard, classic studies of the association of DR2 and DR3 with susceptibility to SLE have shown that DR2/DR3 heterozygotes are more strongly associated with disease susceptibility than the individual haplotypes (Graham et al., 2007). Taken together, these data suggest that SLE is associated with an extensive array of divergent HLA class II alleles that would be predicted to present a diverse array of self-peptides.

Our data indicate that all of the HLA-DR and -DQ alleles that are strongly associated with susceptibility to SLE are in strong LD with XL9 regulatory haplotypes that increase HLA class II gene transcription. Boss and co-workers have shown that XL9 contains CTCF elements that interact with cohesion molecules and other chromatin factors to assemble a transcriptional complex that brings multiple HLA class II promoters into close proximity with an array of transcription factor binding sites (Majumder et al., 2006; 2008). The XL9 model in Figure 15A.i, which is adapted from a model presented by Majumder et al. (2008), illustrates the key role of chromatin configuration in the coordinated transcription of HLA class II genes. Consistent with the chromatin structure effects of XL9, our data indicate that transcriptional variations are chromosome specific in HLA-D heterozygotes, with polymorphisms in the XL9 regulatory haplotype modulating transcription of DR and DQ genes in a cis-specific fashion. This indicates that the level of transcription dictated by XL9 will be specific for the adjacent HLA-DR or DQ allele, thus making expression levels an additional allele-specific facet of HLA-D class II molecules. Classic studies have demonstrated allele-specific variations in the expression levels of MHC class II molecules in murine MHC heterozygotes and shown that these levels strongly impacted the stimulation of antigen-specific T cells in autoimmune disease models (Ridgway et al., 1998). Our data suggest that these early studies revealed an important facet of MHC diversity that strongly impacts the development of autoimmunity.

XL9 contains the peak association signal with SLE in our sequencing dataset, indicating that these regulatory haplotypes play an important role in the development of SLE autoimmunity. We suspect that the failure to detect XL9 associations in previous GWAS reflects the low frequency of many of the associated variants and the paucity of SNPs in this region on the commonly utilized SNP typing arrays (see Figure 5.iii & iv). Our data demonstrate that individuals homozygous for XL9 HAP3 (risk) variations have more than two-fold higher surface expression for HLA-DR and DQ molecules at baseline which increases to 4-fold after stimulation with TLR-ligand on monocyte-derived dendritic cells





**Figure 14.** Model of allelic architecture for functional variations in common disease risk loci. (A) A working model of the architecture of the variations within common disease risk loci. Disease associated tagging SNPs associate an LD block with a disease phenotype. Within this LD block, multiple variations are in tight LD, including nonfunctional, functional, and causal variants. Causal variants potentiate the disease phenotype by modulating endophenotypes. In this model, causal variants impact two adjacent genes, one of which is not located within the LD block, both of which contribute endophenotypes towards disease. Haplotype and MJ analysis using functional variants in tight LD with original tagging SNP define haplotypes that contain all of the causal variants. The peak risk haplotype defines a disease allele with increased disease association in comparison to the original GWAS tagging SNP. (B) A plot of all of the odds ratios attributable to the GWAS tagging SNP (blue bars) versus the peak risk haplotype (additional red Figure 14 continued on next page



Figure 14 continued

bar) for each of the sixteen risk loci analyzed in detail. A consistent gain in odds ratio for SLE was obtained with regulatory haplotypes that averaged 17% in the present study. (C) Frequency of *STAT4*, *IRF5-TNPO3*, *ITGAM-ITGAX*, *UBE2L3* and *HLA-D* SLE risk haplotypes among our own study and 26 ethnic populations characterized in the 1000 Genomes project. The x-axis of the graph shows population groups and y axis show frequency of haplotypes.

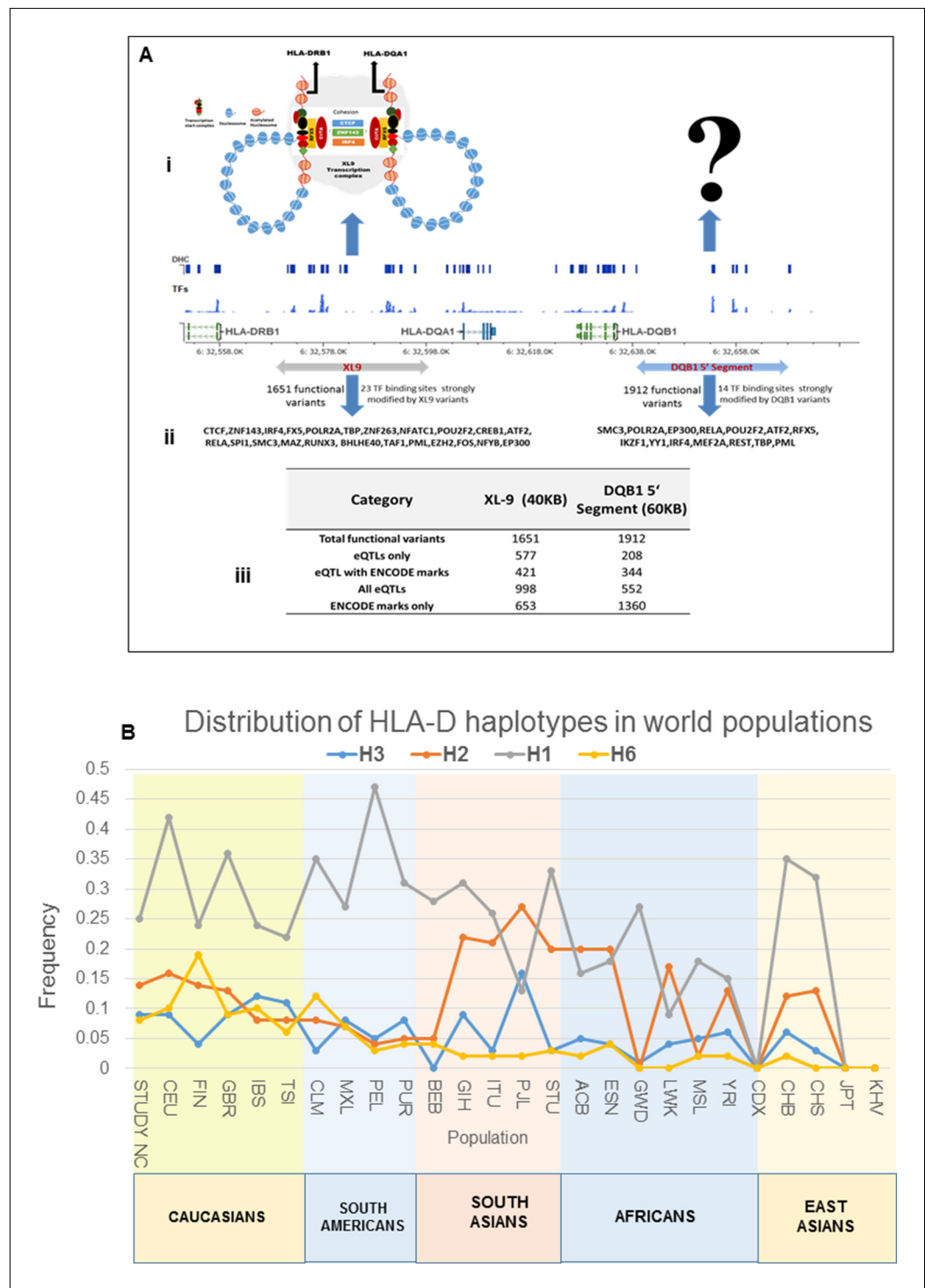
DOI: [10.7554/eLife.12089.018](https://doi.org/10.7554/eLife.12089.018)

than HAP1 (protective) homozygotes (**Figure 7**). This increase is maintained during the maturation of these dendritic cells via TLR7/8 stimulation, thus supporting the functional significance of these transcriptional variations to immune mechanisms known to impact immune response activation. Potent polymorphisms in the binding motif of the IRF4 transcription factor are in the final MJ branch to the SLE-associated XL9 HAP2 and HAP3 haplotypes and it is likely that these two polymorphisms are causal and contribute significantly to this expression change. Recent studies by Vander Lugt et al (**Vander Lugt et al., 2014**) have demonstrated that IRF4 is a key component of the transcriptional regulation of HLA class II molecules in dendritic cells and that upregulation of MHC class II molecules strongly promotes susceptibility to autoimmunity in an animal model. Based on these data, we hypothesize that the XL9-mediated increase in surface expression of HLA-DR and DQ in dendritic cells is predominantly responsible for the association of XL9 regulatory haplotypes with susceptibility to SLE.

The extensive diversification of HLA-D regulatory elements has implications well beyond the association of these polymorphisms with SLE. HLA class II molecules are expressed on a variety of immune cell lineages, including monocytes, macrophages, dendritic cells, B cells, activated T cells and thymic epithelial cells. Expression levels are tightly controlled by a variety of transcription factors unique to these cell lineages and their expression impacts a variety of functional processes in the immune system (**Cresswell, 1994; Krawczyk et al., 2004; Steimle et al., 1994; Reith et al., 2005**). For example, increased surface expression of class II molecules is an essential event in the maturation of dendritic cells, in that higher surface expression is crucial to the increased capacity of mature dendritic cells to effectively present antigens to naïve T cells (**Cella et al., 1997a; 1997b; Banchereau and Steinman, 1998; Pierre, 1997**). As tabulated in **Figure 15A.ii**, twenty-three transcription factor binding sites in XL9 and fourteen in the DQB1 5' segment are strongly modified by the extensive variations present within these genomic segments. Overall, we identified a total of 1651 functional variants in XL9 and 1912 functional variants in the DQB1 5' segment, indicating that the regulatory characteristics of these transcriptional complexes will be highly diversified among HLA-D haplotypes. This level of polymorphism is readily comparable to that observed in the codons for the peptide binding regions of the HLA class II molecules. Interestingly, as shown in **Figure 15B**, both the protective and risk haplotypes that we identified for the entire HLA-D region are present with varying frequencies throughout the global population. These findings are consistent with the characteristics of other highly polymorphic regions of HLA, including the HLA-D class II coding alleles, and indicate that these regulatory haplotypes have evolved together with the HLA class II coding regions over long periods. Given all of these characteristics, we propose that these regulatory variations represent an essential and highly selected characteristic of the diversification of HLA-D that strongly impacts a variety of immune functions.

These regulatory HLA-D haplotypes probably have divergent effects on the expression levels of HLA class II molecules in different cell lineages. That is, XL9 HAP3 clearly upregulates HLA class II expression in the myeloid lineage, however, that may not be true for HAP3 B cells, T cells, or thymic epithelial cells. In addition, XL9 variants impact the function of a variety of transcription factors whose activity is modified by innate system activation signals in specific cell lineages, indicating that modulations of HLA class II expression levels by activation signals may also be differentially associated with individual haplotypes. Taken together, these features indicate that the regulatory polymorphisms identified in this study will have multifaceted effects on the adaptive immune response and probably result in a significant diversification in the functioning of HLA class II antigen presentation among HLA haplotypes.

Finally, a variety of eQTL studies have identified the HLA complex as a 'master' regulatory complex that impacts the expression of genes throughout the genome (**Fehrmann et al., 2011; 2012**). An analysis of the available eQTL databases identified a total of 72 genes whose transcription



**Figure 15.** Model of chromatin architecture and transcription regulatory elements in XL9 and DQB1 segments. (A) (i-ii) A model showing the XL9 transcription complex and three important proteins (CTCF, IRF4 and ZNF143) which may be impacted by SLE associated genetic variants hitting canonical motifs in XL9 region (Adapted from *Majumdar et al., 2008*). The chromatin structure of the regulatory complex produced in the DQB1 5' segment is hypothetical and currently unknown. A chromosomal map of HLA-DRB1 through HLA-DQB1 region showing ENCODE defined regulatory marks, eQTLs and most strongly impacted transcription factors by XL9 and the DQB1 5' segment is shown below these models. The transcription factor binding sites impacted by functional variations within these regions are shown below the molecular map. (A) (iii) A table listing the numbers of and characteristics

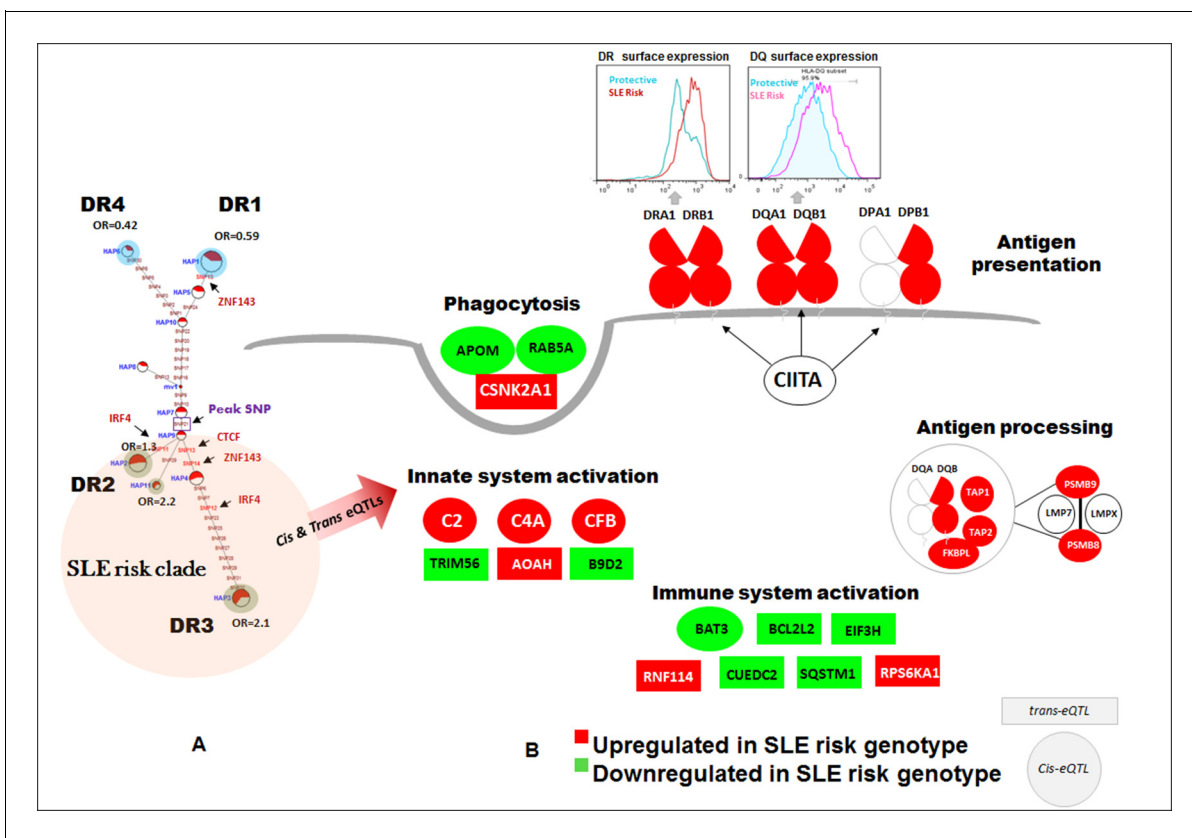
*Figure 15 continued on next page*

Figure 15 continued

of functional variants in these two regulatory regions of HLA-D. (B): Global distribution of the major risk and protective haplotypes from the composite HLA-D region analysis.

DOI: 10.7554/eLife.12089.019

levels are reported to be modulated by variations in the HLA-D region. Many of these eQTL targets are encoded in other segments of the HLA complex, as well as 35 that are located on other chromosomes. A variety of immune system genes are included in this list and **Figure 16** illustrates the pattern of up and down expression that distinguishes the HAP3 risk haplotype from the HAP1 and HAP6 protective haplotypes. As shown, essentially all of the HLA class II molecules and a variety of gene products involved in antigen processing, peptide loading, and surface expression are up-regulated by the HAP3 risk haplotype. This result illustrates the extensive functional variations that are associated with a single HLA-D risk haplotype within the EA population. Whether these eQTLs reflect the formation of remarkably intricate transcriptional complexes, or (more likely) very strong LD throughout the HLA complex remains to be determined. However, these results indicate that regulatory polymorphisms in HLA-D affect a plethora of immune system genes that are involved in various pathways of the adaptive immune system. It is likely that these regulatory variations are an integral element of the functional diversification of HLA and that they will ultimately be found to modulate functions throughout the immune system.



**Figure 16.** SLE risk haplotype upregulates the antigen presentation pathway (APP). All of the composite HLA-D haplotypes within the risk clade (highlighted in red) contain eQTL variants reported to impact 72 genes in the publicly available eQTL datasets utilized in this study. The patterns of increased or decreased transcription associated with all of these haplotypes is modeled on the left, with red indicating increased expression and green indicating decreased expression relative to the protective haplotypes shaded in blue. All of the HLA-DR, HLA-DQ, and HLA-DP class II molecules, along with a variety of gene products involved in the APP pathway are upregulated in all SLE risk haplotypes. A variety of other genes in the immune system, including some with known associations with SLE susceptibility (C2, C4A) are also modulated.

DOI: 10.7554/eLife.12089.020

## Materials and methods

### Targeted sequencing of SLE risk loci

Genomic sequencing libraries were prepared from 1775 SLE patient or control samples contributed by 5 collaborating sites in the U.S.A and Europe (**Supplementary file 1A**). All subjects gave their written informed consent and research protocols and methods employed were approved by the UT Southwestern Institutional Review Board. More than 50% of SLE cases and all of the control samples used in the present study were new recruitments and have not been used in any previous association or GWAS on SLE. Target enrichment and deep sequencing was carried out in the UT Southwestern Medical Center IIMT Genomics Core. 1 µg picogreen measured genomic DNA was sonicated using Covaris S220 platform to generate 300–400 bp genomic fragments. The sequencing libraries were generated using TruSeq (Illumina) or KAPA Biosystem library preparation kits (KK8232). Each sample was ligated with custom designed Illumina-compatible adaptors with unique 6 base barcodes following the kit manufacturer's protocol. The custom target enrichment array (Illumina Inc. San Diego, CA, [www.illumina.com](http://www.illumina.com)) was designed to capture the complete genome sequence of 28 confirmed or potential SLE risk loci (**Supplementary file 1B**). The Illumina custom enrichment system theoretically captured sequence information for ~99.94% of the ~4.4 Mbs of genome targeted in these risk loci. The enriched libraries were sequenced using a paired-end 100 bp protocol to produce 1–2 Gb of high quality data per sample.

### Sequence alignment and variant calling

Sequence reads were demultiplexed and each sample's reads were aligned to the human genome (HG19) using BWA-MEM, with base quality recalibration and local realignment performed with the Genome Analysis Toolkit (GATKv2) (*Li et al., 2009; McKenna et al., 2010; DePristo et al., 2011*). As illustrated in **Figure 1A**, target enrichment was highly specific and efficient, typically resulting in >70% of reads on target and resulting in >128X average coverage for the 28 risk loci analyzed (**Supplementary file 1B**). **Figure 1B** illustrates that coverage within the targeted segments was comprehensive, with relatively uniform read depth throughout the non-repetitive regions. In general, continuous sequences could be derived for >85% of the targeted intervals in assembled sequences, with only extended regions (i.e. >1 Kb) of highly repetitive sequences failing to assemble.

Variant analysis with the GATK Haplotype caller identified a total of 215880 variations in the targeted regions using analytic technologies and thresholds analogous to those of the 1000 Genomes Project (**Supplementary file 1B**) (*Abecasis et al., 2010; 2012*).

### Defining high quality variants in the EA population

As listed in **Figure 1G,H**, we used additional criteria to filter this dataset to create an ethnically-matched, case-control cohort with uniform coverage and high quality variant calls. **Figure 1C** illustrates the read depth and balanced allele representation of variants that passed all filters and **Figure 2A** presents principal component analysis of sample ethnicity in comparison to standardized populations from the HAPMAP dataset (*International HapMap et al., 2003*). We started with 1775 samples, all of which were sequenced with targeted array. Of these, 88 samples had missing case/control status information and 249 were PCA outliers as they did not cluster with HapMap CEU reference population in principal component analysis. Of the remaining 1438 PCA pass samples, 11 and 5 were excluded due poor call rate (<85%) and being duplicate, respectively. Furthermore, 73 samples were excluded due to poor sequencing fold coverage (<25x, n=54) and significant p value ( $p > 0.001$ ) of HWE in controls (n=19). Thus, 1349 samples which included 773 SLEs and 576 normal controls passed all quality criteria and were used for genetic association analysis. Application of these filters defined 1349 samples of EA ancestry and identified 124,552 high quality variants, of which 114,487 are single nucleotide variations (SNVs) and 10,065 are insertion/deletion (In/Del) polymorphisms. Data supporting the accuracy of these variant calls is provided in **Figure 1F**, which presents Sanger sequence validation of several key variants. In addition, the concordance of sequence-based SNV calls versus SNV calls with the Illumina ImmunoChip was >99.8% for samples with >25X average coverage (**Figure 1G**), which supports the overall accuracy of variant detection throughout the genomic segments assayed. This sequence-based variant database, which identifies an average of one variant every 39 basepairs through 4.4 Mb of genomic DNA in 28 validated SLE

risk loci, provides a comprehensive assessment of genomic diversity at SLE risk loci in the EA population. Raw sequencing data (FASTQ files) for all targeted intervals in 1349 individuals is available on request ([www.utsouthwestern.edu/labs/wakeland/about/contact.html](http://www.utsouthwestern.edu/labs/wakeland/about/contact.html)).

### Variant annotation

Variants were annotated using multiple databases cataloguing the functional properties of human genomic variation, including the recent phase 3 release from the 1000 genome study, the PolyPhen/SIFT coding region database, the most recent release of the ENCODE database, and several recent eQTL databases for immune cell lineages. The outcome of these analyses for all variants in the final dataset is summarized in **Figure 2B,C**. Our sequence analyses identified 70070 previously annotated variations and 54482 unannotated or novel variations. All but 76 of the novel variations were low frequency or rare ( $MAF < 0.05$ ) within the EA cohort, which is consistent with expectations (**Supplementary file 1C**). Functional annotation determined that about 40% of the variants in the dataset were potentially functional, most of which were categorized as regulatory based on their localization into ENCODE-defined regulatory segments or inclusion in eQTL datasets (**Figure 2B.i-iv**). **Figures 2C.i-iii** shows summary statistics on just common coding, regulatory and novel variants.

### ImmunoChip genotyping and Sanger sequencing

A subset ( $n=536$ ) of study samples was also genotyped with the ImmunoChipv1, an Illumina Infinium genotyping chip which contains 196524 genomewide markers. SNP concordance analysis was done between sequencing and immunoChip genotypes in order to validate the quality of sequencing calls. Raw image files from immunoChip array were imported into Genome Studio (GS version 1.9.4) and SNPs were called. The genotype outputs from Genome Studio were then imported into SNP & Variation suite (SVS version 7.6.8 win64) for further quality control (QC) and downstream association analysis. In addition, about 35 SLE associated SNPs from various targeted genes were also confirmed by Sanger sequencing method.

### Quality control criteria

Quality control filters were applied to both samples and markers. All the duplicates and those with call rate  $< 85\%$  were excluded from the analysis. Principal component analysis (PCA) was done to remove any population stratification. About 2902 markers ( $MAF \geq 0.05$ ) present in both Omni1 HapMap data set and our sequencing data were used for PCA (**Figure 2A**). PCA clusters were further confirmed by doing another PCA based on ~26000 markers present on Omni1 HapMap data set and ImmunoChip v1.0 on subset samples ( $n=408$ ). Samples that did not cluster with HapMap CEU reference population were excluded after visual inspection of plots. SNP concordance was done between sequencing and immunoChip data on subset of samples ( $n=408$ ), which showed a genotype concordance rate  $\geq 98\%$  at  $\geq 25x$  fold coverage. Therefore, samples with fold coverage  $< 25x$  were excluded from downstream analysis. For genetic association tests, autosomal markers showing HWE  $p < 0.001$  in controls were excluded.

### Genetic association tests and haplotype analysis

Of the 23,805 common ( $MAF \geq 0.05$ ) markers detected in 28 targeted loci, 15,582 quality pass variants were used for genetic association tests. A basic allelic association test was performed with 1349 PCA confirmed and age matched European cases ( $n=773$ ) and controls ( $n=576$ ). The association test was controlled for genomic inflation using Golden Helix scripts where we first determined uncorrected genomic inflation factor  $\lambda$  value which was 3.0. We further corrected data for batch effects and stratification with PCA using numeric association and regression analysis in Golden Helix. Finally, we corrected association results using this inflation value. This removed observed genomic inflation in association results. Chi-square  $p$  values were further corrected for gender bias, using the covariate regression module in SVS, Golden Helix software. A total of 5561 markers across 28 loci showed significant ( $p < 0.05$ ) association with SLE. The LD structure and haplotype analysis was performed in SVS, Golden Helix and using Haploview v4.2. Regulatory haplotypes were generated on potentially functional variants with strong LD ( $\geq 0.8$ ) to the study peak and/or previously known SLE tagging SNPs. An allele was defined as potentially functional if it is a coding variation i.e. non-



synonymous, synonymous, UTR or splice variant, and or an ENCODE's histone mark, transcription factor binding site, DNase I hypersensitivity clusters or an expressed quantitative trait loci (eQTL).

### Population attributable risk (PAR) calculation

PAR has been used in genome-wide association studies (Ziegler and studies, 2009; Bonnelykke et al., 2013; Wang et al., 2010). It combines information on risk allele frequencies and genotypic relative risks to estimate the excess fraction of cases that would not occur if no one in the population carried the risk allele. The case/control design of present study provided authors an opportunity to apply odds ratio (OR) and PAR to calculate the relative risk of SLE disease. As it has already been shown and present study also confirms that SLE is a polygenic disease, assessment of cumulative disease risk by calculating joint PAR from all the loci is a reasonable approach. First we performed conditional analysis on peak SNP from all the 16 loci analyzed in the present study. We observed residual SLE association after conditioning on each locus (Supplementary file 3A), which suggests significant and cumulative contribution of each locus to SLE risk within this population. Then, we used 16 SLE risk haplotypes to assess the percentage of population disease risk that is attributable to the combined effects of all of these risk loci (Supplementary file 3B). PAR was calculated using methods applied in other complex diseases (Zheng et al., 2008).

### Production of monocyte-derived macrophages and dendritic cells

Human peripheral blood mononuclear cells (PBMCs) were enriched by density gradient centrifugation of peripheral blood from healthy human donors through a Ficoll-Hypaque gradient. Monocytes were isolated from PBMC either by negative selection using an EasySep Human Monocyte Enrichment Kit (STEMCELL Technologies), or by cell culture dish adherence as plastic adherence method. For monocyte isolation by adherence, PBMCs were plated in tissue culture treated dishes and incubated for 2 hours at 37°C in a humidified CO<sub>2</sub> incubator. Non-adherent cells were discarded by washing three times. For the generation of monocyte-derived dendritic cells (MDDCs) or human monocyte-derived macrophages (MDMs), monocytes were cultured in RPMI-1640 with 10% FBS, 2 mM L-glutamine, 10 mM HEPES, 1 mM sodium pyruvate, 100 U/ml penicillin, 100 µg/ml streptomycin supplemented with 100 ng/ml GM-CSF and 50 ng/ml IL-4 or 50 ng/ml M-CSF, respectively. The culture media which contained fresh GM-CSF and IL-4 or M-CSF were replaced every 2 days. MDDCs and MDMs were harvested on day 7. MDDCs or MDMs were seeded in 6 or 12 well plates at a density of  $1 \times 10^6$  or  $5 \times 10^5$  cells/ well, respectively, treated and incubated with 10 µg/ml R848 for 18 hr. R848 (InvivoGen, tlr-r848) is an imidazoquinoline compound with potent anti-viral activity. This low synthetic molecule activates immune cells via the TLR7/TLR8 MyD88-dependent signaling pathway (Hemmi et al., 2002)

### Flow cytometry

MDDCs were incubated with anti-HLA-DR FITC (clone G46-6) for 20 min on ice and washed three times with PBS. MDDCs were acquired on a FACS Calibur (BD Biosciences) and data analyzed using FlowJo software. For HLA-DQ staining, MDDC's were stimulated with 1 µg/ml R848 for 4, 8 and 18 hr. Cells were harvested and washed twice with PBS. Staining was done with HLA-DQ-FITC (Clone: Tu169) in PBS for 30 min on ice followed by two washes with PBS and ran on BD FACSCALIBUR. Data was analyzed using FlowJO software.

### RNA-Seq data production and analysis

RNA was extracted using TRIZOL (Life Technologies) and RNeasy Mini Kit (QIAGEN) according to the manufacturer's protocol. RNA quantity and purity was assessed on a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific), and integrity was measured on an Agilent Bioanalyzer 2100 (Agilent Technologies). RNA-seq libraries were prepared with the Illumina TruSeq RNA Sample Preparation kit (Illumina) according to the manufacturer's protocol. Libraries were validated on an Agilent Bioanalyzer 2100. Six RNAseq libraries were sequenced on a SE50 (single end 50 base pair) HiSeq 2500 lane, which yielded an average of about  $30 \times 10^6$  reads/sample. We used CLC Genomics Workbench 7 for bioinformatics and statistical analysis of the sequencing data. This approach used by CLC Genomics Workbench is based on method developed by Mortazavi et al. (Mortazavi et al., 2008). Human Genome GRCh37 was used as reference sequence. The reference has 33615 genes

and 30842 transcripts. All uniquely mapping reads to the genes were counted. Alignment with mismatch cost of '2', Insertion cost '3' Deletion cost of '3' was used. The maximum number of hits for a read was set to 1 meaning that only reads those maps uniquely were considered. The steady state expression of various genes was calculated in terms of RPKM values. For eQTL analysis RPKM values were normalized as described previously (Dozmorov and Lefkovits, 2009; Dozmorov et al., 2011) as well as for population stratification or batch effect and cis-eQTL results were corrected for gender and ethnicity.

## Public databases used

We have accessed multiple public databases to validate and functionally annotate sequencing variants identified in the present study. We used DNA and RNA sequencing data-based variants from the 1000 Genome Project samples (<http://www.1000genomes.org/>); and downloaded DNA sequencing data from the phase III dataset (<ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>) for haplotype analysis of 2504 genomic samples from the global human population. Similarly, FASTQ files of RNA sequencing data (<http://www.geuvadis.org/web/geuvadis/RNAseq-project>) of lymphoblastoid cell lines derived from 369 Europeans were downloaded and used for HLA class II expression analysis. Our analysis of this data is shown in **Figure 7D** as a heatmap. The ENCODE database ([www.encodeproject.org](http://www.encodeproject.org)) was used to annotate variants for transcription factors binding motif, DNase hypersensitivity cluster and histone marks. Similarly, the RegulomeDB database (<http://regulomedb.org/>) was used to annotate potentially regulatory variations. Finally, UCSC genome browser (<https://genome.ucsc.edu/>) was used to generate custom tracks on sequencing variants and for the general visualization of study data.

## Source data text

### ITGAM-ITGAX

We replicated association with previously reported GWAS SNPs in *ITGAM* [Integrin, Alpha M (Complement Component 3 Receptor 3 Subunit)] and *ITGAX* [Integrin, Alpha X (Complement Component 3 Receptor 4 Subunit)] gene and identified new common variants showing strong association with SLE in our analysis (**Supplementary file 2**). The peak associated SNP rs41476751 [OR (LCI-UCI) = 1.87 (1.5–2.2),  $p=7.84E-09$ ] was mapped to the same LD block where previous GWAS SNP was tagged (**Figure 10.i**). Conditioning on peak SNP removed all the disease association (**Figure 10.i B**). We found 30 potentially functional variations including four known SLE GWAS SNPs in strong LD with peak (**Supplementary file 2**). Haplotypes were then derived on these 31 potentially functional variations and genetic association analysis was done in cases vs. controls (**Figure 10.i D**). Association results showed that haplotype 4 (HAP4) which include all previously reported GWAS SLE alleles also carry multiple potentially functional variations (eQTLs) and pose strongest SLE risk [OR (LCI-UCI) = 1.95 (1.5–2.5),  $p=1.58E-07$ ]. Median-joining network analysis illustrated accumulation of multiple potentially functional alleles in the risk haplotype (**Figure 10.iE**). Analysis of published eQTLs data showed that risk haplotype carries multiple regulatory variations which cumulatively contribute to the down-regulation of *ITGAM*, *ITGAX* and *PYCARD* genes (**Figure 10.iF**). Interestingly, we observed that OR of strongest haplotype (HAP4) was significantly increased as compared to known GWAS tag (rs9888739) as well as study peak SNP (**Table 2**).

### IRF5-TNPO3

We replicated previous SLE associations at interferon regulatory factor 5 (*IRF5*) and transportin 3 (*TNPO3*) gene region and also identified many new potentially functional variants showing strong association with disease (**Supplementary file 2**). Multiple association signals at this locus were mapped to *IRF5*, *TNPO3* and *TP1P2* gene region with peak SNP rs34350562 [OR (LCI-UCI) = 1.76 (1.4–2.1),  $p=2.86E-09$ ] mapped near *TNPO3* gene (**Figure 10.iiA**). **Figure 11.iiB-C** shows Manhattan plot of SLE association and conditional analysis on peak SNP, respectively. Total of 29 SLE associated potentially functional variants were identified in the associated LD block which also included four previously known GWAS tags (10.iiC). Haplotypes were generated based on all 29 variants and haplotype association analysis was done. These results showed that haplotype 3 (HAP3) which carries previously reported GWAS SLE SNPs and multiple potential functional alleles identified in present sequencing study pose the greatest risk for SLE [OR (LCI-UCI) = 1.8 (1.2–2.3),  $p=1.44E-06$ ]

(**Figure 10.iiD**). Median-joining network analysis showed that risk haplotype HAP3 differs from the non-risk haplotype (HAP1) by multiple functional variations which regulates transcription of local genes (**Figure 10.iiE**). Further analysis of eQTL SNPs showed that SLE risk alleles were associated with upregulation of *IRF5* and *TNPO3* expression in monocytes, PBMCs and MDMs (**Figure 10.iiF**). In addition, a 3'UTR truncation of *IRF5* gene rs10954213 (SNP9) was also a part of strongest risk haplotype. Our analysis shows that there are multiple potential functional variants present at this locus which contributes to the SLE susceptibility.

### UBE2L3

*UBE2L3* (Ubiquitin-conjugating enzyme E2L3) gene is a known SLE susceptibility gene. We observed strong association signal at this locus and identified multiple regulatory variants (**Supplementary file 2**). The peak association was observed with SNP rs181366 [OR (LCI-UCI) = 1.5 (2.0–3.1),  $p=1.98E-07$ ] in same LD block where previously reported GWAS tag was located (**Figure 10.iiiA**). We replicated previous SLE associations at this locus and identified many new associations. In addition to *UBE2L3*, strong association signal were also mapped to *SCUBE1* (Signal Peptide, CUB Domain, EGF-Like 1) gene with SNP rs4647815 [OR (LCI-UCI) = 1.7 (1.2–2.3),  $p=5.46E-06$ ] and *YDJC* (YdjC Homolog (Bacterial) gene with SNP rs2298429 [OR (LCI-UCI) = 1.5 (1.2–1.8),  $p=9.26E-07$ ]. Manhattan plot of association and conditioning analysis are shown in **Figure 10.iiiB**. There was low frequency variant which was not in strong LD with peak common variant. We identified 15 potentially functional variations in strong LD with peak SNP including previously known GWAS tags for SLE. Haplotypes were derived based on these 16 markers and haplotype association analysis was done (**Figure 10.iiiC**). As shown in **Figure 10.iiiD**, haplotype 2 (HAP2) which carries GWAS SLE allele and multiple potentially functional alleles was the greatest risk haplotype [OR (LCI-UCI) = 1.41 (1.1–1.7),  $p=1.80E-03$ ]. Median-joining network analysis illustrated that risk haplotype (HAP2) differs from non-risk haplotype (HAP1) by 16 putatively functional changes (**Figure 10.iiiE**). Further, SNPs with published eQTL effects shows that SLE associated risk haplotype associate with upregulation of *UBE2L3* expression in peripheral blood data (**Figure 10.iiiF**). Comparison of tag SLE SNP versus haplotype suggest that present regulatory risk haplotype accounts for increased disease risk (OR=1.4) than known GWAS SNP rs5754217 (OR=1.3) alone (**Table 2**).

### BANK1

We observed peak association with an intronic SNP rs4699260 [OR (LCI-UCI) = 1.5(1.3–1.8),  $p=8.54E-06$ ] in *BANK1* (B-cell scaffold protein with ankyrin repeats 1) gene. Previously reported SLE associations were replicated and some new variants were identified in same LD block (**Figure 10.ivA**). Association signal was in the range of suggestive genome-wide significance (**Figure 10.ivB**). We also observed few low frequency variants in the same block. **Figure 10.ivB** also shows conditional analysis based on peak common variants and two low frequency variants. Annotation of SLE associated variants for possible potentially functional effects revealed 23 variants in strong LD with peak associated SNP as well as previously known GWAS tags (**Figure 10.ivC**). So, we performed haplotype analysis on all 24 potentially functional variations (**Supplementary file 2**). Haplotype association analysis showed that haplotype 1 (HAP1) is the strongest risk haplotype [OR (LCI-UCI) = 1.3 (1.0–1.4),  $p=0.004$ ] (**Figure 10.ivD**), which differs from non-risk haplotype by at least 13 potentially functional variations (**Figure 10.ivE**). An interesting eQTL (SNP 20, rs17208914) was observed in risk haplotype which associate with the upregulation of *SLC39A8* gene expression upon stimulation by LPS in monocyte.

### TNIP1

Multiple strong association signals were observed with *TNIP1* (TNFAIP3-interacting protein 1) gene (**Supplementary file 2**). The peak associated SNP in our analysis was rs62382335 [OR (LCI-UCI) = 1.37 (1.0–1.8),  $p=6.42E-05$ ] located in a different block than previously known GWAS tag (**Figure 11.iA**). Still, both were in strong LD with each other due to long LD block. We observed a modest association signal at this locus (**Figure 11.iB**). Conditional analysis showed that in addition to peak common variant, there were few low frequency variants associated with SLE (**Figure 11.iB**). We generated haplotypes on 9 potentially functional variations including peak and GWAS tag SNP (**Figure 11.iC**). The haplotype analysis showed that haplotype 3 (HAP3) is the strongest risk haplotype

[OR (LCI-UCI) = 1.30 (1.0–1.6),  $p=0.04$ ] (**Figure 11.iD**). Median-joining network analysis showed that risk haplotype (HAP3) differs from non-risk haplotype (HAP1) by 9 potentially functional changes (**Figure 11.iE**), which include an eQTL associated with downregulation of *TNIP1* and *ANXA6* (Annexin A6) gene expression in PBMCs (**Figure 11.iF**).

### TNFAIP3

Multiple strong association signals were observed in *TNFAIP3* (tumor necrosis factor, alpha-induced protein 3) gene in present study. The peak associated SNP rs57087937 [OR (LCI-UCI) = 1.9 (1.4–2.5),  $p=2.05E-06$ ] is a ENCODE defined regulatory variant with strong potential to impact binding of several transcription factors in intron 1 region of *TNFAIP3* gene. We also replicated previously reported association with rs5029939 [OR (LCI-UCI) = 2.0 (1.5–2.9),  $p=2.86E-05$ ] which was located in the same LD block (**Figure 11.iiA**). In addition, some new potentially regulatory variants were observed in our analysis (**Supplementary file 2**). **Figure 11.iiB** shows Manhattan plot of SLE association and conditional analysis, respectively. We identified 20 potentially functional variations including peak SNP and known GWAS alleles (SNP rs5029930 (OR=1.5,  $p=7.74355E-05$ ), rs7750604 (OR=1.5,  $p=8.37827E-05$ ), rs719149 (OR=1.5,  $p=0.0001$ ) and rs719150 (OR=1.5,  $p=0.0001$ ) in strong LD ( $D' \geq 0.8$ ) (**Figure 11.iiC**). Next, haplotypes were derived on these 20 markers and association analysis was done. Haplotype association results showed that haplotype 3 (HAP3) which also carries previously known SLE alleles confer strongest risk [OR (LCI-UCI) = 2.2 (1.4–2.9),  $p=4.57E-05$ ] to SLE (**Figure 11.iiD**). Median-joining network analysis shows that risk haplotype (HAP3) differs from non-risk haplotype (HAP1) by 20 potentially functional variations (**Figure 11.iiE**).

### CCL22-CX3CL1

Strong association signals were observed in *CCL22* [Chemokine (C-C Motif) Ligand 22(CCL22)] and *CX3CL1* [Chemokine (C-X3-C Motif) Ligand 1] genes (**Supplementary file 2**). The peak association was observed with SNP rs223889 [OR (LCI-UCI) = 1.5(1.2–1.7),  $p=4.93E-07$ ] in *CCL22* gene (**Figure 11.iiiA**). SLE association statistics is shown in Manhattan plot (**Figure 11.iiiB**). All the SLE association was gone after conditioning on peak SNP. Peak SNP is an eQTL associated with downregulation of *COQ9* gene expression in peripheral blood (**Westra et al., 2013**). Seven potentially functional variations were identified in strong LD with peak signal (**Figure 11.iiiC**). Four common haplotypes were identified based on 8 potentially functional variations. The haplotype association test results showed that HAP2 [OR (LCI-UCI) = 1.5 (1.1–1.7),  $p=6.00E-04$ ] confer strongest SLE risk as compared to HAP1 which is protective [OR (LCI-UCI) = 0.74 (0.62–0.87),  $p=6.00E-04$ ] (**Figure 11.iiiD**). Median-joining network analysis showed that risk haplotype (HAP2) differs from non-risk haplotype (HAP1) by 8 potentially functional variations (**Figure 11.iiiE**). eQTL analysis showed that strongest risk haplotype is associated with upregulation of *CCL22* and downregulation of *COQ9* gene expression in monocytes (**Figure 11.iiiF**).

### ZGLP1-RAVER1

We observed multiple SLE association signals at this locus which were mapped to various local genes including *ZGLP1* (zinc finger, GATA-like protein 1), *FDX1L* (Ferredoxin 1-like), *RAVER1* (Homo sapiens ribonucleoprotein, PTB-binding 1), *ICAM1* (Intercellular adhesion molecule 1), and *TYK2* (Tyrosine kinase 2) genes (**Supplementary file 2**). Peak SNP rs35186095 [OR (LCI-UCI) = 1.32 (1.0–1.6),  $p=2.10E-04$ ] was sitting in the middle of a big LD block (**Figure 11.ivA**), mapped to intron 9 of *ZGLP1* gene. Manhattan plot illustrate the SLE association statistics as shown in **Figure 11.ivB**. The conditioning on first and second peak variant removed all the observed association. We identified 16 potentially functional variations in SLE associated block and derived haplotypes on these (**Figure 11.ivC**). Haplotype association analysis showed that haplotype 2 (HAP2) is the strongest risk haplotype [OR (LCI-UCI) = 1.5 (1.0–2.1),  $p=0.01$ ] (**Figure 11.ivD**). Median-joining network illustrates number of variants that differs between risk (HAP2) and non-risk (HAP1) haplotypes (**Figure 11.ivE**). The eQTL data from literature suggests that SLE risk haplotype is associated with upregulated expression of *ICAM3* in human monocytes and PBMCs (**Figure 11.ivF**).



## ICA1

We observed modest association of *ICA1* (Islet Cell Autoantigen 1, 69kDa) gene with SLE in our study. The peak association was observed with SNP rs74787882 [OR (LCI-UCI) = 0.67 (0.49–0.92),  $p=1.61E-03$ ] which was mapped to the previously associated LD block (**Figure 12.iA**). We replicated previous association with SNP rs10156091 [OR (LCI-UCI) = 1.5 (1.1–1.9),  $p=0.02$ ]. Manhattan plot shown in **Figure 12.iB** shows the strength of SLE association at this locus. All the association was gone after conditioning on peak variant. Haplotypes were derived based on four potentially functional variants in the SLE associated LD block (**Figure 12.iC**). Haplotype association test results showed that haplotype 2 (HAP2) is the strongest risk haplotype in *ICA1* (**Figure 12.iD**). Median-Joining network shows that HAP2 is differed from non-risk haplotype HAP3 by four potentially functional variants (**Figure 12.iE**). Associated data is provided in **Supplementary file 2**.

## BLK

The SLE association signal at *BLK* (B lymphoid tyrosine kinase) gene was moderate in our analysis. However, we did replicate previously reported associations at this locus. SLE associated peak SNP rs7822109 [OR (LCI-UCI) = 0.79 (0.68–0.92),  $p=8.57E-05$ ] is a potentially functional (ENCODE, eQTL effects) variant. It is in strong LD ( $D' \geq 0.8$ ) with previously reported SLE GWAS SNPs (**Figure 12.iiA**). Manhattan plot in **Figure 12.iiB** shows SLE association strength in present study and conditional analysis. We derived haplotypes based on 10 potentially functional variations in this LD block and performed haplotype association test (**Figure 12.iiC**). Results shows that haplotype 1 [OR (LCI-UCI) = 1.25 (1.0–1.5),  $p=0.04$ ] pose strongest risk for SLE (**Figure 12.iiD**). Several eQTLs cumulated into SLE risk haplotype 1 associate with down regulation of *BLK* and upregulation of *FAM167A* and *MTMR9* genes in peripheral blood (**Figure 12.iiE-F**). Associated data is provided in **Supplementary file 2**.

## ETS1

Peak SLE association at *ETS1* (V-ETS avian erythroblastosis virus E26 oncogene homolog 1) gene was observed with a low frequency SNP rs117684226 [OR (LCI-UCI) = 0.47 (0.35–0.63),  $p=4.87E-06$ ] located in the previously implicated LD block (**Figure 12.iiiA**). In addition, rs34516251 was the strongest common SLE risk alleles located in the same block. Manhattan plot in **Figure 12.iiiB** shows the SLE association statistics. Conditional analysis revealed that peak low frequency variant and strongest common allele accounts for all the observed association at this locus in the present study (**Figure 12.iiiB**). We identified 6 potentially functional variants in SLE associated LD block and derived haplotypes on them (**Figure 12.iiiC**). Haplotype analysis showed that haplotype 3 (HAP3) is the greatest risk haplotype [OR (LCI-UCI) = 1.4 (1.0–1.7),  $p=0.008$ ] (**Figure 12.iiiD**), which differs from non-risk haplotype HAP2 by 5 variants as illustrated in Median-Joining network (**Figure 12.iiiE**). Associated data is provided in **Supplementary file 2**.

## TNFSF4

We observed two independent association signal in *TNFSF4* (Tumor necrosis factor (ligand) superfamily member 4) gene region (**Figure 13.iA**). First signal was SNP rs4916313 [1.34 [OR (LCI-UCI) = 1.34 (1.1–1.5),  $p=0.0002$ ] in previously associated LD block (block 1). This SNP is a strong regulatory variant defined by ENCODE data. In block 1, we identified 11 potentially functional variations in strong LD with peak SNP. Manhattan plot shows the strength of SLE association in block1 and block2 (**Figure 13.iB**). Conditioning analysis on peak variants in each block showed that both the signals are independent of each other (**Figure 13.iC**). Haplotypes association analysis in block1 showed that haplotype 2 (HAP2) was the greatest risk haplotype [OR (LCI-UCI) = 1.33 (1.1–1.6),  $p=0.003$ ] (**Figure 13.iD**), which differs from HAP1 by 11 variations (**Figure 13.iF**). In block 2, where SNP rs1819717 was the strongest association [OR (LCI-UCI) = 1.36 (1.1–1.6),  $p=2.29E-05$ ] was mapped to an uncharacterized gene LOC100506023. Haplotype analysis based on seven potentially functional SNPs showed that HAP1 which carries multiple potentially functional variants pose strongest risk for SLE from this region (**Figure 13.iE**). Median-joining network analysis shows distribution of various risk alleles between risk and non-risk haplotypes from block 2 (**Figure 13.iG**).

Associated data is provided in **Supplementary file 2**.

## NMNAT2

We observed a modest association of polymorphisms in *NMNAT2* (Nicotinamide nucleotide adenylyl transferase 2) gene with SLE in present study (**Supplementary file 2**). The peak associated SNP was though a low frequency variant rs41272536 [OR (LCI-UCI) = 2.9 (2.0–4.0),  $p=1.87E-08$ ] but reached genome wide level of significance. It was pretty much in LD with previously reported tag SNP (**Figure 13.iiA**). This SNP is a ENCODE defined strong regulatory variant which can impact binding of several transcription factors, enhancers and insulators. Another strongly associated common variant was SNP rs111487113. Conditional analysis on peak and strongest common alleles suggested that two effects are independently associated with SLE (**Figure 13.iiB**). We identified 7 potentially functional variations in strong LD with the strongest common variant and derived haplotypes (**Figure 13.iiC**). Haplotype association analysis showed that haplotype 3 (HAP3) was the strongest risk haplotype [OR (LCI-UCI) = 1.3 (1.0–1.6),  $p=0.04$ ] (**Figure 13.iiD-E**). We also derived haplotypes based on three potential functional variants in modest LD with low frequency peak variant and found that HAP3 which carry low frequency risk allele have strong SLE risk (OR=2.7) for SLE (**Figure 13.iiF-H**). Also, we found that eQTL variations in low frequency risk haplotype were associated with down regulation of *SMG7* (Homo sapiens smg-7 homolog, nonsense mediated mRNA decay factor) and upregulation of *NCF2* (Neutrophil cytosolic factor 2) gene expression in monocytes and PBMCs (**Figure 13.iiI**).

## XKR6

We observed more than one association signals at *XKR6* (XK, Kell blood group complex subunit related family, member 6) gene (**Supplementary file 2**). The peak associated SNP was rs4840545 [OR (LCI-UCI) = 1.95 (1.5–2.5),  $p=1.27E-07$ ] was a low frequency marker mapped to a different LD block than previously reported GWAS tag (**Figure 13.iiiA**). SNP rs7000132 was the strongest common allele associated with SLE at this locus but it was not in complete LD with peak low frequency variant as shown by the conditional analysis (**Figure 13.iiiB**). We identified 10 potentially functional variations in strong LD with rs7000132 including previously reported SLE associated rs11783247 SNP and derived haplotypes (**Figure 13.iiiC**). Haplotype association test showed that HAP1 was the greatest risk haplotype associated with SLE [OR (LCI-UCI) = 1.25 (1.0–1.4),  $p=0.005$ ] (**Figure 13.iiiD**). Median-Joining network analysis illustrated that risk haplotype 1 differs from non-risk haplotype 2 by several eQTL SNPs (**Figure 13.iiiE**), which are associated with down regulation of *XKR6*, *MSRA* gene expression and upregulation of *MTMR9* and *CTSB* gene expression in monocytes (**Figure 13.iiiF**).

## Statistical analyses

We used SNP & Variation suite (SVS) of Golden Helix (version 7.6.8 win64, Golden Helix, Inc., Bozeman, MT, [www.goldenhelix.com](http://www.goldenhelix.com)) for genetic analysis. SNP conditioning analysis was done using regression module of SVS. Haploview v.2 software was used for visualization of LD plots and haplotype analysis (Barrett et al., 2005). GraphPad Prism 6.0 software was used for statistical analysis and graphics. Correlations between continuous variables were determined using Pearson's  $r$  in GraphPad Prism 6.0. Discontinuous variables were compared by Fisher's exact test. P values  $<0.05$  were considered significant.

## Acknowledgements

We thank the many SLE patients and control participants whose sample contributions were essential for these studies. We also thank all of the personnel in the IIMT Genomics Core at UT Southwestern Medical Center for their excellent technical support and participation. These studies were supported by multiple grants from the NIH, the Alliance for Lupus Research, and the Walter M. and Helen D. Bader Center for Research on Arthritis and Autoimmune Diseases.

## Additional information

### Funding

Funder	Grant reference number	Author
Alliance for Lupus Research		Prithvi Raj Ran Song Shaheen Khan
Walter M. and Helen D. Bader Center		Edward K Wakeland
NIH Office of the Director	AR055503	David R Karp Quan Zhen Li Patrick M Gaffney Edward K Wakeland
NIH Office of the Director	AI045196	Edward K Wakeland
NIH Office of the Director	AR058959	Graham B Wiley Jennifer A Kelly

The funding from above listed agencies supported sample collection, data generation, data analysis, etc., and man power for the present study

### Author contributions

PR, EKW, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article, Contributed unpublished essential data or reagents; ER, MM, GBW, JAK, CC, COJ, Acquisition of data, Analysis and interpretation of data, Contributed unpublished essential data or reagents; RS, SK, KV, BZ, ID, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; BEW, Analysis and interpretation of data, Drafting or revising the article, Contributed unpublished essential data or reagents; CA, CL, FC-J, BRL, NJO, CKG, CAW, JBH, SKN, JAJ, BPT, QZL, Conception and design, Acquisition of data, Analysis and interpretation of data; CP, Conception and design, Analysis and interpretation of data, Drafting or revising the article; DRK, PMG, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article

### Author ORCIDs

Edward K Wakeland,  <http://orcid.org/0000-0002-7107-0992>

### Ethics

Human subjects: All the study subjects gave their written informed consent for the study. All the research protocols and methods employed were approved by UT Southwestern Institutional Review Board.

## Additional files

### Supplementary files

- Supplementary file 1. (A) SLE patients and controls analyzed in this study (B) Genomic intervals of SLE risk loci targeted for sequencing (C) Characteristics of unannotated/novel common variants ( $MAF \geq 0.05$ ) detected in this study (D) Peak association signal detected for each of the 28 SLE risk loci (E) Association status of previously published GWAS tagging SNPs (F) Sequencing variants that are strongly associated with SLE.

DOI: [10.7554/eLife.12089.021](https://doi.org/10.7554/eLife.12089.021)

- Supplementary file 2. Summary of functional properties of all variants in tight LD with disease tagging SNPs used for haplotype analysis.

DOI: [10.7554/eLife.12089.022](https://doi.org/10.7554/eLife.12089.022)

- Supplementary file 3. (A) Conditional analysis on SLE associated 16 peak SNPs.(B) Calculation of joint PAR on 16 SLE risk loci.

DOI: [10.7554/eLife.12089.023](https://doi.org/10.7554/eLife.12089.023)

**Major datasets**

The following previously published datasets were used:

<b>Author(s)</b>	<b>Year</b>	<b>Dataset title</b>	<b>Dataset URL</b>	<b>Database, license, and accessibility information</b>
Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, Jostins L, Plant K, Andrews R, McGee C, Knight JC	2014	Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression	<a href="http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4064786/">http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4064786/</a>	Publicly available at the NCBI Gene Expression Omnibus (Accession no: PMC4064786).
Raj T, Rothamel K, Mostafavi S, Ye C, Lee MN, Replogle JM, Feng T, Lee M, Asinowski N, Frohlich et al.	2014	Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. <i>Science</i>	<a href="http://classic.sciencemag.org/content/344/6183/519.long">http://classic.sciencemag.org/content/344/6183/519.long</a>	<i>Science</i> , 2014. 344 (6183): p. 519-23
Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, Zhernakova A, Zhernakova DV, Veldink JH, Van den Berg LH, Karjalainen J, Withoff S, Uitterlinden AG, Hofman A, Rivadeneira F, 't Hoen PAC, Reinmaa E, Fischer K, Nelis M, Milani L, Melzer D, Ferrucci L, Singleton AB, Hernandez DG, Nalls MA, Homuth G, Nauck M, Radke D, Völker U, Perola M, Salomaa V, Brody J, Suchy-Dacey A, Gharib SA, Enquobahrie DA, Lumley T, Montgomery GW, Makino S, Prokisch H, Herder C, Roden M, Grallert H, Meitinger T, Strauch K, Li Y, Jansen RC, Visscher PM, Knight JC, Psaty BM, Ripatti S, Teumer A, Frayling TM, Metspalu A, van Meurs JBJ, Franke L	2013	Systematic identification of trans eQTLs as putative drivers of known disease associations	<a href="http://www.nature.com/ng/journal/v45/n10/full/ng.2756.html">http://www.nature.com/ng/journal/v45/n10/full/ng.2756.html</a>	<i>Nat Genet</i> , 2013. 45 (10): p. 1238-43
The 1000 Genomes Project Consortium	2015	A global reference for human genetic variation	<a href="http://www.nature.com/nature/journal/v526/n7571/full/nature15393.html">http://www.nature.com/nature/journal/v526/n7571/full/nature15393.html</a>	<i>Nature</i> , 2015. 526 (7571): p. 68-74
Lappalainen et al.	2013	Transcriptome and genome sequencing uncovers functional variation in humans	<a href="http://www.geuvadis.org/web/geuvadis/RNA-seq-project">http://www.geuvadis.org/web/geuvadis/RNA-seq-project</a>	<i>Nature</i> , 2013. 501 (7468): p. 506-511

Pazin MJ	2015	Using the ENCODE Resource for Functional Annotation of Genetic Variants.	<a href="https://www.encodeproject.org/">https://www.encodeproject.org/</a>	Cold Spring Harb Protoc, 2015. 2015(6): p. 522-36
Boyle AP1, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M	2012	Annotation of functional variation in personal genomes using RegulomeDB	<a href="http://regulomedb.org/">http://regulomedb.org/</a>	Genome Res, 2012. 22(9): p. 1790-7.

## References

- Abecasis GR**, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA, Consortium G, 1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. *Nature* **467**:1061–1073. doi: [10.1038/nature09534](https://doi.org/10.1038/nature09534)
- Abecasis GR**, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA, Consortium G, 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**:56–65. doi: [10.1038/nature11632](https://doi.org/10.1038/nature11632)
- Adrianto I**, Wen F, Templeton A, Wiley G, King JB, Lessard CJ, Bates JS, Hu Y, Kelly JA, Kaufman KM, Guthridge JM, Alarcón-Riquelme ME, BIOLUPUS and GENLES Networks, Anaya JM, Bae SC, Bang SY, Boackle SA, Brown EE, Petri MA, Gallant C, Ramsey-Goldman R, Reveille JD, Vila LM, Criswell LA, Edberg JC, Freedman BI, Gregersen PK, Gilkeson GS, Jacob CO, James JA, Kamen DL, Kimberly RP, Martin J, Merrill JT, Niewold TB, Park SY, Pons-Estel BA, Scofield RH, Stevens AM, Tsao BP, Vyse TJ, Langefeld CD, Harley JB, Moser KL, Webb CF, Humphrey MB, Montgomery CG, Gaffney PM. 2011. Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. *Nature Genetics* **43**:253–258. doi: [10.1038/ng.766](https://doi.org/10.1038/ng.766)
- Adzhubei IA**, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nature Methods* **7**:248–249. doi: [10.1038/nmeth0410-248](https://doi.org/10.1038/nmeth0410-248)
- Alarcón-Segovia D**, Alarcón-Riquelme ME, Cardiel MH, Caeiro F, Massardo L, Villa AR, Pons-Estel BA, on behalf of the Grupo Latinoamericano de Estudio del Lupus Eritematoso (GLADEL). 2005. Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. *Arthritis & Rheumatism* **52**:1138–1147. doi: [10.1002/art.20999](https://doi.org/10.1002/art.20999)
- Arbuckle MR**, McClain MT, Rubertone MV, Scofield RH, Dennis GJ, James JA, Harley JB. 2003. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *New England Journal of Medicine* **349**:1526–1533. doi: [10.1056/NEJMoa021933](https://doi.org/10.1056/NEJMoa021933)
- Armstrong DL**, Zidovetzki R, Alarcón-Riquelme ME, Tsao BP, Criswell LA, Kimberly RP, Harley JB, Sivils KL, Vyse TJ, Gaffney PM, Langefeld CD, Jacob CO. 2014. GWAS identifies novel SLE susceptibility genes and explains the association of the HLA region. *Genes and Immunity* **15**:347–354. doi: [10.1038/gene.2014.23](https://doi.org/10.1038/gene.2014.23)
- Auton A**, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. 1000 Genomes Project Consortium. The 1000 Genomes Project C. 2015. A global reference for human genetic variation. *Nature* **526**:68–74. doi: [10.1038/nature15393](https://doi.org/10.1038/nature15393)
- Bailey SD**, Zhang X, Desai K, Aid M, Corradin O, Cowper-Sal-lari R, Akhtar-Zaidi B, Scacheri PC, Haibe-Kains B, Lupien M. 2015. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. *Nature Communications* **2**:6186. doi: [10.1038/ncomms7186](https://doi.org/10.1038/ncomms7186)
- Balding DJ**. 2006. A tutorial on statistical methods for population association studies. *Nature Reviews Genetics* **7**:781–791. doi: [10.1038/nrg1916](https://doi.org/10.1038/nrg1916)
- Banchereau J**, Steinman RM. 1998. Dendritic cells and the control of immunity. *Nature* **392**:245–252. doi: [10.1038/32588](https://doi.org/10.1038/32588)
- Bandelt HJ**, Macaulay V, Richards M. 2000. Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Molecular Phylogenetics and Evolution* **16**:8–28. doi: [10.1006/mpev.2000.0792](https://doi.org/10.1006/mpev.2000.0792)
- Barreiro LB**, Quintana-Murci L. 2010. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nature Reviews Genetics* **11**:17–30. doi: [10.1038/nrg2698](https://doi.org/10.1038/nrg2698)
- Bernstein BE**, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS, Thomson JA. 2010. The NIH roadmap epigenomics mapping consortium. *Nature Biotechnology* **28**:1045–1048. doi: [10.1038/nbt1010-1045](https://doi.org/10.1038/nbt1010-1045)
- Boyle AP**, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. 2012. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research* **22**:1790–1797. doi: [10.1101/gr.137323.112](https://doi.org/10.1101/gr.137323.112)
- Bruzzi P**, Green SB, Byar DP, Brinton LA, Schairer C. 1985. Estimating the population attributable risk for multiple risk factors using case-control data. *American Journal of Epidemiology* **122**:904–914.
- Bønnelykke K**, Matheson MC, Pers TH, Granell R, Strachan DP, Alves AC, Linneberg A, Curtin JA, Warrington NM, Standl M, Kerkhof M, Jonsdottir I, Bukvic BK, Kaakinen M, Sleimann P, Thorleifsson G, Thorsteinsdottir U, Schramm K, Baltic S, Kreiner-Møller E, Simpson A, St Pourcain B, Coin L, Hui J, Walters EH, Tiesler CM, Duffy DL, Jones G, Ring SM, McArdle WL, Price L, Robertson CF, Pekkanen J, Tang CS, Thiering E, Montgomery



- GW, Hartikainen AL, Dharmage SC, Husemoen LL, Herder C, Kemp JP, Elliot P, James A, Waldenberger M, Abramson MJ, Fairfax BP, Knight JC, Gupta R, Thompson PJ, Holt P, Sly P, Hirschhorn JN, Blekic M, Weidinger S, Hakonarsson H, Stefansson K, Heinrich J, Postma DS, Custovic A, Pennell CE, Jarvelin MR, Koppelman GH, Timpson N, Ferreira MA, Bisgaard H, Henderson AJ, Australian Asthma Genetics Consortium (AAGC), EARly Genetics and Lifecourse Epidemiology (EAGLE) Consortium. 2013. Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. *Nature Genetics* **45**:902–906. doi: [10.1038/ng.2694](https://doi.org/10.1038/ng.2694)
- Cella M**, Engering A, Pinet V, Pieters J, Lanzavecchia A. 1997a. Inflammatory stimuli induce accumulation of MHC class II complexes on dendritic cells. *Nature* **388**:782–787. doi: [10.1038/42030](https://doi.org/10.1038/42030)
- Cella M**, Sallusto F, Lanzavecchia A. 1997b. Origin, maturation and antigen presenting function of dendritic cells. *Current Opinion in Immunology* **9**:10–16. doi: [10.1016/S0952-7915\(97\)80153-7](https://doi.org/10.1016/S0952-7915(97)80153-7)
- Chung SA**, Taylor KE, Graham RR, Nititham J, Lee AT, Ortmann WA, Jacob CO, Alarcón-Riquelme ME, Tsao BP, Harley JB, Gaffney PM, Moser KL, Petri M, Demirci FY, Kamboh MI, Manzi S, Gregersen PK, Langefeld CD, Behrens TW, Criswell LA. SLEGEN. 2011. Differential Genetic Associations for Systemic Lupus Erythematosus Based on Anti-dsDNA Autoantibody Production. *PLoS Genetics* **7**:e1001323. doi: [10.1371/journal.pgen.1001323](https://doi.org/10.1371/journal.pgen.1001323)
- Cirulli ET**, Goldstein DB. 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics* **11**:415–425. doi: [10.1038/nrg2779](https://doi.org/10.1038/nrg2779)
- Claus EB**, Schildkraut JM, Thompson WD, Risch NJ. 1996. The genetic attributable risk of breast and ovarian cancer. *Cancer* **77**:2318–2324. doi: [10.1002/\(SICI\)1097-0142\(19960601\)77:11<2318::AID-CNCR21>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-0142(19960601)77:11<2318::AID-CNCR21>3.0.CO;2-Z)
- Cookson W**, Liang L, Abecasis G, Moffatt M, Lathrop M. 2009. Mapping complex disease traits with global gene expression. *Nature Reviews Genetics* **10**:184–194. doi: [10.1038/nrg2537](https://doi.org/10.1038/nrg2537)
- Cresswell P**. 1994. Assembly, transport, and function of MHC class II molecules. *Annual Review of Immunology* **12**:259–291. doi: [10.1146/annurev.iy.12.040194.001355](https://doi.org/10.1146/annurev.iy.12.040194.001355)
- Cruz-Tapias P**, Pérez-Fernández OM, Rojas-Villarraga A, Rodríguez-Rodríguez A, Arango M-T, Anaya J-M. 2012. Shared HLA class II in six autoimmune diseases in latin america: a meta-analysis. *Autoimmune Diseases* **2012**:1–10. doi: [10.1155/2012/569728](https://doi.org/10.1155/2012/569728)
- Cunningham Graham DS**, Morris DL, Bhangale TR, Criswell LA, Syvänen AC, Rönnblom L, Behrens TW, Graham RR, Vyse TJ. 2011. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. *PLoS Genetics* **7**:e1002341. doi: [10.1371/journal.pgen.1002341](https://doi.org/10.1371/journal.pgen.1002341)
- de Bakker PIW**, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. 2005. Efficiency and power in genetic association studies. *Nature Genetics* **37**:1217–1223. doi: [10.1038/ng1669](https://doi.org/10.1038/ng1669)
- Deapen D**, Escalante A, Weinrib L, Horwitz D, Bachman B, Roy-Burman P, Walker A, Mack TM. 1992. A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis and Rheumatism* **35**:311–318.
- Deng Y**, Tsao BP. 2010. Genetic susceptibility to systemic lupus erythematosus in the genomic era. *Nature Reviews Rheumatology* **6**:683–692. doi: [10.1038/nrrheum.2010.176](https://doi.org/10.1038/nrrheum.2010.176)
- DePristo MA**, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**:491–498. doi: [10.1038/ng.806](https://doi.org/10.1038/ng.806)
- Dozmorov I**, Lefkowitz I. 2009. Internal standard-based analysis of microarray data. part 1: analysis of differential gene expressions. *Nucleic Acids Research* **37**:6323–6339. doi: [10.1093/nar/gkp706](https://doi.org/10.1093/nar/gkp706)
- Dozmorov IM**, Jarvis J, Saban R, Benbrook DM, Wakeland E, Aksentijevich I, Ryan J, Chiorazzi N, Guthridge JM, Drewe E, Tighe PJ, Centola M, Lefkowitz I. 2011. Internal standard-based analysis of microarray data—analysis of functional associations between HVE-genes. *Nucleic Acids Research* **39**:7881–7899. doi: [10.1093/nar/gkr503](https://doi.org/10.1093/nar/gkr503)
- Dunham I**, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee B-K, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shores N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Dunham I, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S, Kellis M, Khatun J, Kheradpour P, Kundaje A, Lassmann T, Li Q, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SCJ, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HS, Yip KY, Zhuang J, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M, Pazin MJ, Lowdon RF, Dillon LAL, Adams LB, Kelly CJ, Zhang J, Wexler JR, Green ED, Good PJ, Feingold EA, Bernstein BE, Birney E, Crawford GE, Dekker J, Elnitski L, Farnham PJ, Gerstein M, Giddings MC, Gingeras TR, Green ED, Guigó R, Hardison RC, Hubbard TJ, Kellis M, Kent WJ, Lieb JD, Margulies EH, Myers RM, Snyder M, Stamatoyannopoulos JA, Tenenbaum SA, Weng Z, White KP, Wold B, Khatun J, Yu Y, Wrobel J, Risk BA, Gunawardena HP, Kuiper HC, Maier CW, Xie L, Chen X, Giddings MC, Bernstein BE, Epstein CB, Shores N, Ernst J, Kheradpour P, Mikkelsen TS, Gillespie S, Goren A, Ram O, Zhang X, Wang L, Issner R, Coyne MJ, Durham T, Ku M, Truong T, Ward LD, Altshuler RC, Eaton ML, Kellis M, Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Batut P, Bell I, Bell K, Chakraborty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Dutttagupta R, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena HP, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Li G, Luo OJ, Park E, Preall JB, Presaud K, Ribeca P, Risk BA, Robyr D, Ruan X, Sammeth M, Sandhu KS, Schaeffer L, See L-H, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Hayashizaki Y, Harrow J, Gerstein M, Hubbard TJ, Reymond A, Antonarakis SE, Hannon GJ, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR, Rosenbloom KR, Sloan CA, Learned K, Malladi VS, Wong MC, Barber GP, Cline MS, Dreszer TR, Heitner SG, Karolchik D, Kent

- WJ, Kirkup VM, Meyer LR, Long JC, Maddren M, Raney BJ, Furey TS, Song L, Grasfeder LL, Giresi PG, Lee B-K, Battenhouse A, Sheffield NC, Simon JM, Showers KA, Safi A, London D, Bhinge AA, Shestak C, Schaner MR, Ki Kim S, Zhang ZZ, Mieczkowski PA, Mieczkowska JO, Liu Z, McDaniel RM, Ni Y, Rashid NU, Kim MJ, Adar S, Zhang Z, Wang T, Winter D, Keefe D, Birney E, Iyer VR, Lieb JD, Crawford GE, Li G, Sandhu KS, Zheng M, Wang P, Luo OJ, Shahab A, Fullwood MJ, Ruan X, Ruan Y, Myers RM, Pauli F, Williams BA, Gertz J, Marinov GK, Reddy TE, Vielmetter J, Partridge E, Trout D, Varley KE, Gasper C, Bansal A, Pepke S, Jain P, Amrhein H, Bowling KM, Anaya M, Cross MK, King B, Muratet MA, Antoshechkin I, Newberry KM, McCue K, Nesmith AS, Fisher-Aylor KI, Pusey B, DeSalvo G, Parker SL, Balasubramanian S, Davis NS, Meadows SK, Eggleston T, Gunter C, Newberry JS, Levy SE, Absher DM, Mortazavi A, Wong WH, Wold B, Blow MJ, Visel A, Pennachio LA, Elnitski L, Margulies EH, Parker SCJ, Petrykowska HM, Abyzov A, Aken B, Barrell D, Barson G, Berry A, Bignell A, Boychenko V, Bussotti G, Chrast J, Davidson C, Derrien T, Despacio-Reyes G, Diekhans M, Ezkurdia I, Frankish A, Gilbert J, Gonzalez JM, Griffiths E, Harte R, Hendrix DA, Howald C, Hunt T, Jungreis I, Kay M, Khurana E, Kokocinski F, Leng J, Lin MF, Loveland J, Lu Z, Manthavadi D, Mariotti M, Mudge J, Mukherjee G, Notredame C, Pei B, Rodriguez JM, Saunders G, Sboner A, Searle S, Sisu C, Snow C, Steward C, Tanzer A, Tapanari E, Tress ML, van Baren MJ, Walters N, Washietl S, Wilming L, Zadissa A, Zhang Z, Brent M, Haussler D, Kellis M, Valencia A, Gerstein M, Reymond A, Guigó R, Harrow J, Hubbard TJ, Landt SG, Fritze S, Abyzov A, Addleman N, Alexander RP, Auerbach RK, Balasubramanian S, Bettinger K, Bhardwaj N, Boyle AP, Cao AR, Cayting P, Charos A, Cheng Y, Cheng C, Eastman C, Eastman C, Euskirchen G, Fleming JD, Grubert F, Habegger L, Hariharan M, Harmanci A, Iyengar S, Jin VX, Karczewski KJ, Kasowski M, Lacroute P, Lam H, Lamarre-Vincent N, Leng J, Lian J, Lindahl-Allen M, Min R, Miotto B, Monahan H, Moqtaderi Z, Mu XJ, O'Geen H, Ouyang Z, Patacsil D, Pei B, Raha D, Ramirez L, Reed B, Rozowsky J, Sboner A, Shi M, Sisu C, Slifer T, Witt H, Wu L, Xu X, Yan K-K, Yang X, Yip KY, Zhang Z, Struhl K, Weissman SM, Gerstein M, Farnham PJ, Snyder M, Tenenbaum SA, Penalva LO, Doyle F, Karmakar S, Landt SG, Bhanvadia RR, Choudhury A, Domanus M, Ma L, Moran J, Patacsil D, Slifer T, Victorsen A, Yang X, Snyder M, White KP, Auer T, Centanin L, Eichenlaub M, Gruhl F, Heermann S, Hoeckendorf B, Inoue D, Kellner T, Kirchmaier S, Mueller C, Reinhardt R, Schertel L, Schneider S, Sinn R, Wittbrodt B, Wittbrodt J, Weng Z, Whitfield TW, Wang J, Collins PJ, Aldred SF, Trinklein ND, Partridge EC, Myers RM, Dekker J, Jain G, Lajoie BR, Sanyal A, Balasundaram G, Bates DL, Byron R, Canfield TK, Diegel MJ, Dunn D, Ebersol AK, Frum T, Garg K, Gist E, Hansen RS, Boatman L, Haugen E, Humbert R, Jain G, Johnson AK, Johnson EM, Kutayavin TV, Lajoie BR, Lee K, Lotakis D, Maurano MT, Neph SJ, Neri FV, Nguyen ED, Qu H, Reynolds AP, Roach V, Rynes E, Sabo P, Sanchez ME, Sandstrom RS, Sanyal A, Shafer AO, Stergachis AB, Thomas S, Thurman RE, Vernot B, Vierstra J, Vong S, Wang H, Weaver MA, Yan Y, Zhang M, Akey JM, Bender M, Dorschner MO, Groudine M, MacCoss MJ, Navas P, Stamatoyannopoulos G, Kaul R, Dekker J, Stamatoyannopoulos JA, Dunham I, Beal K, Brazma A, Flicek P, Herrero J, Johnson N, Keefe D, Lukk M, Luscombe NM, Sobral D, Vaquerizas JM, Wilder SP, Batzoglou S, Sidow A, Hussami N, Kyriazopoulou-Panagiotopoulou S, Libbrecht MW, Schaub MA, Kundaje A, Hardison RC, Miller W, Giardine B, Harris RS, Wu W, Bickel PJ, Banfai B, Boley NP, Brown JB, Huang H, Li Q, Li JJ, Noble WS, Billes JA, Buske OJ, Hoffman MM, Sahu AD, Kharchenko PV, Park PJ, Baker D, Taylor J, Weng Z, Iyer S, Dong X, Greven M, Lin X, Wang J, Xi HS, Zhuang J, Gerstein M, Alexander RP, Balasubramanian S, Cheng C, Harmanci A, Lochovsky L, Min R, Mu XJ, Rozowsky J, Yan K-K, Yip KY, Birney E. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**:57–74. doi: [10.1038/nature11247](https://doi.org/10.1038/nature11247)
- Edwards SV**, Chesnut K, Satta Y, Wakeland EK. 1997. Ancestral polymorphism of mhc class II genes in mice: implications for balancing selection and the mammalian molecular clock. *Genetics* **146**:655–668.
- Fairfax BP**, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, Jostins L, Plant K, Andrews R, McGee C, Knight JC. 2014. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**:1246949. doi: [10.1126/science.1246949](https://doi.org/10.1126/science.1246949)
- Fairfax BP**, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, Ellis P, Langford C, Vannberg FO, Knight JC. 2012. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nature Genetics* **44**:502–510. doi: [10.1038/ng.2205](https://doi.org/10.1038/ng.2205)
- Fairhurst AM**, Wandstrat AE, Wakeland EK. 2006. Systemic lupus erythematosus: multiple immunological phenotypes in a complex genetic disease. *Advances in Immunology* **92**:1–69. doi: [10.1016/S0065-2776\(06\)92001-X](https://doi.org/10.1016/S0065-2776(06)92001-X)
- Fehrmann RSN**, Jansen RC, Veldink JH, Westra H-J, Arends D, Bonder MJ, Fu J, Deelen P, Groen HJM, Smolonska A, Weersma RK, Hofstra RMW, Buurman WA, Rensen S, Wolfs MGM, Platteel M, Zhernakova A, Elbers CC, Festen EM, Trynka G, Hofker MH, Saris CGJ, Ophoff RA, van den Berg LH, van Heel DA, Wijmenga C, te Meerman GJ, Franke L. 2011. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genetics* **7**:e1002197. doi: [10.1371/journal.pgen.1002197](https://doi.org/10.1371/journal.pgen.1002197)
- Fernando MM**, Freudenberg J, Lee A, Morris DL, Boteva L, Rhodes B, Gonzalez-Escribano MF, Lopez-Nevot MA, Navarra SV, Gregersen PK, Martin J, Vyse TJ. 2012. Transancestral mapping of the MHC region in systemic lupus erythematosus identifies new independent and interacting loci at MSH5, HLA-DPB1 and HLA-g. *Annals of the Rheumatic Diseases* **71**:777–784. doi: [10.1136/annrheumdis-2011-200808](https://doi.org/10.1136/annrheumdis-2011-200808)
- Frazer KA**, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Wayne MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB,

- Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J, International HapMap Consortium. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**:851–861. doi: [10.1038/nature06258](https://doi.org/10.1038/nature06258)
- Fu SM, Deshmukh US, Gaskin F. 2011. Pathogenesis of systemic lupus erythematosus revisited 2011: end organ resistance to damage, autoantibody initiation and diversification, and HLA-DR. *Journal of Autoimmunity* **37**: 104–112. doi: [10.1016/j.jaut.2011.05.004](https://doi.org/10.1016/j.jaut.2011.05.004)
- Furukawa H, Kawasaki A, Oka S, Ito I, Shimada K, Sugii S, Hashimoto A, Komiya A, Fukui N, Kondo Y, Ito S, Hayashi T, Matsumoto I, Kusaoi M, Amano H, Nagai T, Hirohata S, Setoguchi K, Kono H, Okamoto A, Chiba N, Suematsu E, Katayama M, Migita K, Suda A, Ohno S, Hashimoto H, Takasaki Y, Sumida T, Nagaoka S, Tsuchiya N, Tohma S. 2014. Human leukocyte antigens and systemic lupus erythematosus: a protective role for the HLA-DR6 alleles DRB1\*13:02 and \*14:03. *PLoS ONE* **9**:e87792. doi: [10.1371/journal.pone.0087792](https://doi.org/10.1371/journal.pone.0087792)
- Galimberti D, Scalabrini D, Fenoglio C, De Riz M, Comi C, Venturelli E, Cortini F, Piola M, Leone M, Dianzani U, D'Alfonso S, Monaco F, Bresolin N, Scarpini E. 2008. Gender-specific influence of the chromosome 16 chemokine gene cluster on the susceptibility to Multiple Sclerosis. *Journal of the Neurological Sciences* **267**: 86–90. doi: [10.1016/j.jns.2007.10.001](https://doi.org/10.1016/j.jns.2007.10.001)
- Gatava V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, Ortmann W, Kosoy R, Ferreira RC, Nordmark G, Gunnarsson I, Svenungsson E, Padyukov L, Sturfelt G, Jönsen A, Bengtsson AA, Rantapää-Dahlqvist S, Baechler EC, Brown EE, Alarcón GS, Edberg JC, Ramsey-Goldman R, McGwin G, Reveille JD, Vilá LM, Kimberly RP, Manzi S, Petri MA, Lee A, Gregersen PK, Seldin MF, Rönblom L, Criswell LA, Syvänen A-C, Behrens TW, Graham RR. 2009. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nature Genetics* **41**:1228–1233. doi: [10.1038/ng.468](https://doi.org/10.1038/ng.468)
- Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addleman N, Bhardwaj N, Boyle AP, Cayting P, Charos A, Chen DZ, Cheng Y, Clarke D, Eastman C, Euskirchen G, Fietze S, Fu Y, Gertz J, Grubert F, Harman A, Jain P, Kasowski M, Lacroute P, Leng J, Lian J, Monahan H, O'Geen H, Ouyang Z, Partridge EC, Patasil D, Pauli F, Raha D, Ramirez L, Reddy TE, Reed B, Shi M, Slifer T, Wang J, Wu L, Yang X, Yip KY, Zilberman-Schapira G, Batzoglou S, Sidow A, Farnham PJ, Myers RM, Weissman SM, Snyder M. 2012. Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**:91–100. doi: [10.1038/nature11245](https://doi.org/10.1038/nature11245)
- Ghodke-Puranik Y, Niewold TB. 2015. Immunogenetics of systemic lupus erythematosus: a comprehensive review. *Journal of Autoimmunity* **64**:125–136. doi: [10.1016/j.jaut.2015.08.004](https://doi.org/10.1016/j.jaut.2015.08.004)
- Gilad Y, Rifkin SA, Pritchard JK. 2008. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends in Genetics* **24**:408–415. doi: [10.1016/j.tig.2008.06.001](https://doi.org/10.1016/j.tig.2008.06.001)
- Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, Burt NP, Guiducci C, Parkin M, Gates C, Plenge RM, Behrens TW, Wither JE, Rioux JD, Fortin PR, Cunninghame Graham D, Wong AK, Vyse TJ, Daly MJ, Altshuler D, Moser KL, Gaffney PM. 2008. Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nature Genetics* **40**:1059–1061. doi: [10.1038/ng.200](https://doi.org/10.1038/ng.200)
- Graham RR, Hom G, Ortmann W, Behrens TW. 2009. Review of recent genome-wide association scans in lupus. *Journal of Internal Medicine* **265**:680–688. doi: [10.1111/j.1365-2796.2009.02096.x](https://doi.org/10.1111/j.1365-2796.2009.02096.x)
- Graham RR, Kozyrev SV, Baechler EC, Reddy MVPL, Plenge RM, Bauer JW, Ortmann WA, Koeuth T, Escibano MFG, Pons-Estel B, Petri M, Daly M, Gregersen PK, Martín J, Altshuler D, Behrens TW, Alarcón-Riquelme ME, Collaborative Groups the Argentine and Spanish. 2006. A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nature Genetics* **38**:550–555. doi: [10.1038/ng1782](https://doi.org/10.1038/ng1782)
- Graham RR, Ortmann W, Rodine P, Espe K, Langefeld C, Lange E, Williams A, Beck S, Kyogoku C, Moser K, Gaffney P, Gregersen PK, Criswell LA, Harley JB, Behrens TW. 2007. Specific combinations of HLA-DR2 and DR3 class II haplotypes contribute graded risk for disease susceptibility and autoantibodies in human SLE. *European Journal of Human Genetics* **15**:823–830. doi: [10.1038/sj.ejhg.5201827](https://doi.org/10.1038/sj.ejhg.5201827)
- Gusev A, Bhatia G, Zaitlen N, Vilhjalmsón BJ, Diogo D, Stahl EA, Gregersen PK, Worthington J, Klareskog L, Raychaudhuri S, Plenge RM, Pasaniuc B, Price AL. 2013. Quantifying missing heritability at known GWAS loci. *PLoS Genetics* **9**:e1003993. doi: [10.1371/journal.pgen.1003993](https://doi.org/10.1371/journal.pgen.1003993)

- Gyllenstein UB**, Erlich HA. 1989. Ancient roots for polymorphism at the HLA-DQ alpha locus in primates. *Proceedings of the National Academy of Sciences of the United States of America* **86**:9986–9990. doi: [10.1073/pnas.86.24.9986](https://doi.org/10.1073/pnas.86.24.9986)
- Han J-W**, Zheng H-F, Cui Y, Sun L-D, Ye D-Q, Hu Z, Xu J-H, Cai Z-M, Huang W, Zhao G-P, Xie H-F, Fang H, Lu Q-J, Xu J-H, Li X-P, Pan Y-F, Deng D-Q, Zeng F-Q, Ye Z-Z, Zhang X-Y, Wang Q-W, Hao F, Ma L, Zuo X-B, Zhou F-S, Du W-H, Cheng Y-L, Yang J-Q, Shen S-K, Li J, Sheng Y-J, Zuo X-X, Zhu W-F, Gao F, Zhang P-L, Guo Q, Li B, Gao M, Xiao F-L, Quan C, Zhang C, Zhang Z, Zhu K-J, Li Y, Hu D-Y, Lu W-S, Huang J-L, Liu S-X, Li H, Ren Y-Q, Wang Z-X, Yang C-J, Wang P-G, Zhou W-M, Lv Y-M, Zhang A-P, Zhang S-Q, Lin D, Li Y, Low HQ, Shen M, Zhai Z-F, Wang Y, Zhang F-Y, Yang S, Liu J-J, Zhang X-J. 2009. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nature Genetics* **41**:1234–1237. doi: [10.1038/ng.472](https://doi.org/10.1038/ng.472)
- Harley ITW**, Kaufman KM, Langefeld CD, Harley JB, Kelly JA. 2009. Genetic susceptibility to SLE: new insights from fine mapping and genome-wide association studies. *Nature Reviews Genetics* **10**:285–290. doi: [10.1038/nrg2571](https://doi.org/10.1038/nrg2571)
- Harley JB**, Alarcón-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, Vyse TJ, Langefeld CD, Nath SK, Guthridge JM, Cobb BL, Mirel DB, Marion MC, Williams AH, Divers J, Wang W, Frank SG, Namjou B, Gabriel SB, Lee AT, Gregersen PK, Behrens TW, Taylor KE, Fernando M, Zidovetzki R, Gaffney PM, Edberg JC, Rioux JD, Ojwang JO, James JA, Merrill JT, Gilkeson GS, Seldin MF, Yin H, Baechler EC, Li QZ, Wakeland EK, Bruner GR, Kaufman KM, Kelly JA, International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN). 2008. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PTK, KIAA1542 and other loci. *Nature Genetics* **40**:204–210. doi: [10.1038/ng.81](https://doi.org/10.1038/ng.81)
- Hemmi H**, Kaisho T, Takeuchi O, Sato S, Sanjo H, Hoshino K, Horiuchi T, Tomizawa H, Takeda K, Akira S. 2002. Small anti-viral compounds activate immune cells via the TLR7 MyD88-dependent signaling pathway. *Nature Immunology* **3**:196–200. doi: [10.1038/ni758](https://doi.org/10.1038/ni758)
- Hofmann S**, Fischer A, Nothnagel M, Jacobs G, Schmid B, Wittig M, Franke A, Gaede KI, Schürmann M, Petrek M, Mrazek F, Pabst S, Grohé C, Grunewald J, Ronninger M, Eklund A, Rosenstiel P, Höhne K, Zissel G, Müller-Quernheim J, Schreiber S. 2013. Genome-wide association analysis reveals 12q13.3–q14.1 as new risk locus for sarcoidosis. *European Respiratory Journal* **41**:888–900. doi: [10.1183/09031936.00033812](https://doi.org/10.1183/09031936.00033812)
- Hom G**, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, Lee AT, Chung SA, Ferreira RC, Pant PVK, Ballinger DG, Kosoy R, Demirci FY, Kamboh MI, Kao AH, Tian C, Gunnarsson I, Bengtsson AA, Rantapää-Dahlqvist S, Petri M, Manzi S, Seldin MF, Rönnblom L, Syvänen A-C, Criswell LA, Gregersen PK, Behrens TW. 2008. Association of systemic lupus erythematosus with C8orf13–BLK and ITGAM–ITGAX. *New England Journal of Medicine* **358**:900–909. doi: [10.1056/NEJMoa0707865](https://doi.org/10.1056/NEJMoa0707865)
- Hu X**, Zhang B, Liu W, Paciga S, He W, Lanz TA, Kleiman R, Dougherty B, Hall SK, McIntosh AM, Lawrie SM, Power A, John SL, Blackwood D, St Clair D, Brandon NJ. 2014. A survey of rare coding variants in candidate genes in schizophrenia by deep sequencing. *Molecular Psychiatry* **19**:858–859. doi: [10.1038/mp.2013.131](https://doi.org/10.1038/mp.2013.131)
- Hunt KA**, Mistry V, Bockett NA, Ahmad T, Ban M, Barker JN, Barrett JC, Blackburn H, Brand O, Burren O, Capon F, Compston A, Gough SCL, Jostins L, Kong Y, Lee JC, Lek M, MacArthur DG, Mansfield JC, Mathew CG, Mein CA, Mirza M, Nutland S, Onengut-Gumuscu S, Papouli E, Parkes M, Rich SS, Sawcer S, Satsangi J, Simmonds MJ, Trembath RC, Walker NM, Wozniak E, Todd JA, Simpson MA, Plagnol V, van Heel DA. 2013. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. *Nature* **498**:232–235. doi: [10.1038/nature12170](https://doi.org/10.1038/nature12170)
- Höhler T**, Büschenfelde KH. 1994. Systemic lupus erythematosus. *The New England Journal of Medicine* **331**:1235.
- International HapMap Consortium** Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, Ch’ang L-Y, Huang W, Liu B, Shen Y, Tam PK-H, Tsui L-C, Waye MMY, Wong JT-F, Zeng C, Zhang Q, Chee MS, Galver LM, Kruglyak S, Murray SS, Oliphant AR, Montpetit A, Hudson TJ, Chagnon F, Ferretti V, Leboeuf M, Phillips MS, Verner A, Kwok P-Y, Duan S, Lind DL, Miller RD, Rice JP, Saccone NL, Taillon-Miller P, Xiao M, Nakamura Y, Sekine A, Sorimachi K, Tanaka T, Tanaka Y, Tsunoda T, Yoshino E, Bentley DR, Deloukas P, Hunt S, Powell D, Altshuler D, Gabriel SB, Zhang H, Zeng C, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Aniagwu T, Marshall PA, Matthew O, Nkwojima C, Royal CDM, Leppert MF, Dixon M, Stein LD, Cunningham F, Kanani A, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Donnelly P, Marchini J, McVean GAT, Myers SR, Cardon LR, Abecasis GR, Morris A, Weir BS, Mullikin JC, Sherry ST, Feolo M, Altshuler D, Daly MJ, Schaffner SF, Qiu R, Kent A, Dunston GM, Kato K, Niikawa N, Knoppers BM, Foster MW, Clayton EW, Wang VO, Watkin J, Gibbs RA, Belmont JW, Sodergren E, Weinstock GM, Wilson RK, Fulton LL, Rogers J, Birren BW, Han H, Wang H, Godbout M, Wallenburg JC, L’Archevêque P, Bellemare G, Todani K, Fujita T, Tanaka S, Holden AL, Lai EH, Collins FS, Brooks LD, McEwen JE, Guyer MS, Jordan E, Peterson JL, Spiegel J, Sung LM, Zacharia LF, Kennedy K, Dunn MG, Seabrook R, Shillito M, Skene B, Stewart JG, Valle (chair) DL, Clayton (co-chair) EW, Jorde (co-chair) LB, Belmont JW, Chakravarti A, Cho MK, Duster T, Foster MW, Jasperse M, Knoppers BM, Kwok P-Y, Licinio J, Long JC, Marshall PA, Ossorio PN, Wang VO, Rotimi CN, Royal CDM, Spallone P, Terry SF, Lander (chair) ES, Lai (co-chair) EH, Nickerson (co-chair) DA, Abecasis GR, Altshuler D, Bentley DR, Boehnke M, Cardon LR, Daly MJ, Deloukas P, Douglas JA, Gabriel SB, Hudson RR, Hudson TJ, Kruglyak L, Kwok P-Y, Nakamura Y, Nussbaum RL, Royal CDM, Schaffner SF, Sherry ST, Stein LD, Tanaka T, International HapMap C. 2003. The international HapMap project. *Nature* **426**:789–796. doi: [10.1038/nature02168](https://doi.org/10.1038/nature02168)



- Kachru RB.** 1984. A significant increase of HLA-DR3 and DR2 in systemic lupus erythematosus among blacks. *The Journal of Rheumatology* **11**:471–474.
- Kim K,** Bang S-Y, Lee H-S, Okada Y, Han B, Saw W-Y, Teo Y-Y, Bae S-C. 2014. The HLA-DRβ1 amino acid positions 11–13–26 explain the majority of SLE–MHC associations. *Nature Communications* **5**:5902. doi: [10.1038/ncomms6902](https://doi.org/10.1038/ncomms6902)
- Kim K,** Brown EE, Choi CB, Alarcón-Riquelme ME, Kelly JA, Glenn SB, Ojwang JO, Adler A, Lee HS, Boackle SA, Criswell LA, Alarcón GS, Edberg JC, Stevens AM, Jacob CO, Gilkeson GS, Kamen DL, Tsao BP, Anaya JM, Guthridge JM, Nath SK, Richardson B, Sawalha AH, Kang YM, Shim SC, Suh CH, Lee SK, Kim CS, Merrill JT, Petri M, Ramsey-Goldman R, Vilá LM, Niewold TB, Martin J, Pons-Estel BA, Vyse TJ, Freedman BI, Moser KL, Gaffney PM, Williams A, Comeau M, Reveille JD, James JA, Scofield RH, Langefeld CD, Kaufman KM, Harley JB, Kang C, Kimberly RP, Bae SC, BIOLUPUS, GENLES. 2012. Variation in the ICAM1-ICAM4-ICAM5 locus is associated with systemic lupus erythematosus susceptibility in multiple ancestries. *Annals of the Rheumatic Diseases* **71**:1809–1814. doi: [10.1136/annrheumdis-2011-201110](https://doi.org/10.1136/annrheumdis-2011-201110)
- Kim-Howard X,** Sun C, Molineros JE, Maiti AK, Chandru H, Adler A, Wiley GB, Kaufman KM, Kottyan L, Guthridge JM, Rasmussen A, Kelly J, Sanchez E, Raj P, Li Q-Z, Bang S-Y, Lee H-S, Kim T-H, Kang YM, Suh C-H, Chung WT, Park Y-B, Choe J-Y, Shim SC, Lee S-S, Han B-G, Olsen NJ, Karp DR, Moser K, Pons-Estel BA, Wakeland EK, James JA, Harley JB, Bae S-C, Gaffney PM, Alarcon-Riquelme M, Acevedo E, Acevedo E, La Torre IG-D, Maradiaga-Cecena MA, Cardiel MH, Esquivel-Valerio JA, Rodriguez-Amado J, Moctezuma JF, Miranda P, Perandones C, Aires B, Castel C, Laborde HA, Alba P, Musuruana J, Goecke A, Foster C, Orozco L, Baca V, Looger LL, Nath SK. 2014. Allelic heterogeneity in NCF2 associated with systemic lupus erythematosus (SLE) susceptibility across four ethnic populations. *Human Molecular Genetics* **23**:1656–1668. doi: [10.1093/hmg/ddt532](https://doi.org/10.1093/hmg/ddt532)
- Kozyrev SV,** Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, Sanchez E, Gunnarsson I, Svenungsson E, Sturfelt G, Jönsen A, Truedsson L, Pons-Estel BA, Witte T, D'Alfonso S, Barizzone N, Barizzone N, Danieli MG, Gutierrez C, Suarez A, Junker P, Lastrup H, González-Escribano MF, Martin J, Abderrahim H, Alarcón-Riquelme ME. 2008. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nature Genetics* **40**:211–216. doi: [10.1038/ng.79](https://doi.org/10.1038/ng.79)
- Kraft P,** Wacholder S, Cornelis MC, Hu FB, Hayes RB, Thomas G, Hoover R, Hunter DJ, Chanock S. 2009. Beyond odds ratios — communicating disease risk based on genetic profiles. *Nature Reviews Genetics* **10**:264–269. doi: [10.1038/nrg2516](https://doi.org/10.1038/nrg2516)
- Krawczyk M,** Peyraud N, Rybtsova N, Masternak K, Bucher P, Barras E, Reith W. 2004. Long distance control of MHC class II expression by multiple distal enhancers regulated by regulatory factor x complex and CIITA. *The Journal of Immunology* **173**:6200–6210. doi: [10.4049/jimmunol.173.10.6200](https://doi.org/10.4049/jimmunol.173.10.6200)
- Lappalainen T,** Sammeth M, Friedländer MR, 't Hoen PAC, Monlong J, Rivas MA, González-Porta M, Kurbatova N, Griebel T, Ferreira PG, Barann M, Wieland T, Greger L, van Iterson M, Almlöf J, Ribeca P, Pulyakhina I, Esser D, Giger T, Tikhonov A, Sultan M, Bertier G, MacArthur DG, Lek M, Lizano E, Buermans HPJ, Padioleau I, Schwarzmayr T, Karlberg O, Ongen H, Kilpinen H, Beltran S, Gut M, Kahlem K, Amstislavskiy V, Stegle O, Pirinen M, Montgomery SB, Donnelly P, McCarthy MI, Flícek P, Strom TM, The Geuvadis Consortium, Lehrach H, Schreiber S, Sudbrak R, Carracedo Ángel, Antonarakis SE, Häslér R, Syvänen A-C, van Ommen G-J, Brazma A, Meitinger T, Rosenstiel P, Guigó R, Gut IG, Estivill X, Dermitzakis ET. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**:506–511. doi: [10.1038/nature12531](https://doi.org/10.1038/nature12531)
- Laval G,** Patin E, Barreiro LB, Quintana-Murci L. 2010. Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. *PLoS ONE* **5**:e10284. doi: [10.1371/journal.pone.0010284](https://doi.org/10.1371/journal.pone.0010284)
- Lawlor DA,** Ward FE, Ennis PD, Jackson AP, Parham P. 1988. HLA-a and b polymorphisms predate the divergence of humans and chimpanzees. *Nature* **335**:268–271. doi: [10.1038/335268a0](https://doi.org/10.1038/335268a0)
- Lee YH,** Bae S-C, Choi SJ, Ji JD, Song GG. 2012. Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis. *Molecular Biology Reports* **39**:10627–10635. doi: [10.1007/s11033-012-1952-x](https://doi.org/10.1007/s11033-012-1952-x)
- Lee-Kirsch MA,** Gong M, Chowdhury D, Senenko L, Engel K, Lee Y-A, de Silva U, Bailey SL, Witte T, Vyse TJ, Kere J, Pfeiffer C, Harvey S, Wong A, Koskenmies S, Hummel O, Rohde K, Schmidt RE, Dominiczak AF, Gahr M, Hollis T, Perrino FW, Lieberman J, Hübner N. 2007. Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 are associated with systemic lupus erythematosus. *Nature Genetics* **39**:1065–1067. doi: [10.1038/ng2091](https://doi.org/10.1038/ng2091)
- Lewis M,** Vyse S, Shields A, Boeltz S, Gordon P, Spector T, Lehner P, Walczak H, Vyse T. 2015. Effect of UBE2L3 genotype on regulation of the linear ubiquitin chain assembly complex in systemic lupus erythematosus. *The Lancet* **385**:S9. doi: [10.1016/S0140-6736\(15\)60324-5](https://doi.org/10.1016/S0140-6736(15)60324-5)
- Li H,** Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 1000 Genome Project Data Processing Subgroup. 2009. The sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352)
- Li QZ,** Zhen QL, Xie C, Wu T, Mackay M, Aranow C, Putterman C, Mohan C. 2005. Identification of autoantibody clusters that best predict lupus disease activity using glomerular proteome arrays. *The Journal of Clinical Investigation* **115**:3428–3439. doi: [10.1172/JCI23587](https://doi.org/10.1172/JCI23587)
- Liu X,** Yu X, Zack DJ, Zhu H, Qian J. 2008. TiGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics* **9**:271. doi: [10.1186/1471-2105-9-271](https://doi.org/10.1186/1471-2105-9-271)



- Majumder P**, Gomez JA, Boss JM. 2006. The human major histocompatibility complex class II HLA-DRB1 and HLA-DQA1 genes are separated by a CTCF-binding enhancer-blocking element. *Journal of Biological Chemistry* **281**:18435–18443. doi: [10.1074/jbc.M601298200](https://doi.org/10.1074/jbc.M601298200)
- Majumder P**, Gomez JA, Chadwick BP, Boss JM. 2008. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *Journal of Experimental Medicine* **205**:785–798. doi: [10.1084/jem.20071843](https://doi.org/10.1084/jem.20071843)
- Manolio TA**, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TFC, McCarroll SA, Visscher PM. 2009. Finding the missing heritability of complex diseases. *Nature* **461**:747–753. doi: [10.1038/nature08494](https://doi.org/10.1038/nature08494)
- Maurano MT**, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutayin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**:1190–1195. doi: [10.1126/science.1222794](https://doi.org/10.1126/science.1222794)
- McConnell TJ**, Talbot WS, McIndoe RA, Wakeland EK. 1988. The origin of MHC class II gene polymorphism within the genus mus. *Nature* **332**:651–654. doi: [10.1038/332651a0](https://doi.org/10.1038/332651a0)
- McKenna A**, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**:1297–1303. doi: [10.1101/gr.107524.110](https://doi.org/10.1101/gr.107524.110)
- Mezzetti M**, La Vecchia C, Decarli A, Boyle P, Talamini R, Franceschi S. 1998. Population attributable risk for breast cancer: diet, nutrition, and physical exercise. *Journal of the National Cancer Institute* **90**:389–394.
- Mitchell DA**, Pickering MC, Warren J, Fossati-Jimack L, Cortes-Hernandez J, Cook HT, Botto M, Walport MJ. 2002. C1q deficiency and autoimmunity: the effects of genetic background on disease expression. *The Journal of Immunology* **168**:2538–2543. doi: [10.4049/jimmunol.168.5.2538](https://doi.org/10.4049/jimmunol.168.5.2538)
- Mohan C**, Putterman C. 2015. Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis. *Nature Reviews Nephrology* **11**:329–341. doi: [10.1038/nrneph.2015.33](https://doi.org/10.1038/nrneph.2015.33)
- Morris DL**, Fernando MMA, Taylor KE, Chung SA, Nititham J, Alarcón-Riquelme ME, Barcellos LF, Behrens TW, Cotsapas C, Gaffney PM, Graham RR, Pons-Estel BA, Gregersen PK, Harley JB, Hauser SL, Hom G, Langefeld CD, Noble JA, Rioux JD, Seldin MF, Harley JB, Alarcón-Riquelme ME, Criswell LA, Gaffney PM, Jacob CO, Kimberly RP, Sivits KLM, Tsao BP, Vyse TJ, Langefeld CD, Vyse TJ, Criswell LA. 2014. MHC associations with clinical and autoantibody manifestations in european SLE. *Genes and Immunity* **15**:210–217. doi: [10.1038/gene.2014.6](https://doi.org/10.1038/gene.2014.6)
- Morris DL**, Taylor KE, Fernando MM, Nititham J, Alarcón-Riquelme ME, Barcellos LF, Behrens TW, Cotsapas C, Gaffney PM, Graham RR, Pons-Estel BA, Gregersen PK, Harley JB, Hauser SL, Hom G, Langefeld CD, Noble JA, Rioux JD, Seldin MF, Criswell LA, Vyse TJ, International MHC and Autoimmunity Genetics Network, Systemic Lupus Erythematosus Genetics Consortium. 2012. 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* **91**:778–793. doi: [10.1016/j.ajhg.2012.08.026](https://doi.org/10.1016/j.ajhg.2012.08.026)
- Mortazavi A**, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nature Methods* **5**:621–628. doi: [10.1038/nmeth.1226](https://doi.org/10.1038/nmeth.1226)
- Natarajan S**, Lipsitz SR, Rimm E. 2007. A simple method of determining confidence intervals for population attributable risk from complex surveys. *Statistics in Medicine* **26**:3229–3239. doi: [10.1002/sim.2779](https://doi.org/10.1002/sim.2779)
- Nath SK**, Han S, Kim-Howard X, Kelly JA, Viswanathan P, Gilkeson GS, Chen W, Zhu C, McEver RP, Kimberly RP, Alarcón-Riquelme ME, Vyse TJ, Li Q-Z, Wakeland EK, Merrill JT, James JA, Kaufman KM, Guthridge JM, Harley JB. 2008. A nonsynonymous functional variant in integrin- $\alpha$ M (encoded by ITGAM) is associated with systemic lupus erythematosus. *Nature Genetics* **40**:152–154. doi: [10.1038/ng.71](https://doi.org/10.1038/ng.71)
- Ng PC**, Henikoff S. 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research* **31**:3812–3814. doi: [10.1093/nar/gkg509](https://doi.org/10.1093/nar/gkg509)
- Niu Z**, Zhang P, Tong Y. 2015. Value of HLA-DR genotype in systemic lupus erythematosus and lupus nephritis: a meta-analysis. *International Journal of Rheumatic Diseases* **18**:17–28. doi: [10.1111/1756-185X.12528](https://doi.org/10.1111/1756-185X.12528)
- Olsen NJ**, Karp DR. 2014. Autoantibodies and SLE—the threshold for disease. *Nature Reviews Rheumatology* **10**:181–186. doi: [10.1038/nrrheum.2013.184](https://doi.org/10.1038/nrrheum.2013.184)
- Pazin MJ**. 2015. Using the ENCODE resource for functional annotation of genetic variants. *Cold Spring Harbor Protocols* **2015**:pdb.top084988–536. doi: [10.1101/pdb.top084988](https://doi.org/10.1101/pdb.top084988)
- Pepe MS**, Janes H, Longton G, Leisenring W, Newcomb P. 2004. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology* **159**:882–890. doi: [10.1093/aje/kwh101](https://doi.org/10.1093/aje/kwh101)
- Pierre P**, Turley SJ, Gatti E, Hull M, Meltzer J, Mirza A, Inaba K, Steinman RM, Mellman I. 1997. Developmental regulation of MHC class II transport in mouse dendritic cells. *Nature* **388**:787–792. doi: [10.1038/42039](https://doi.org/10.1038/42039)
- Pritchard JK**, Cox NJ. 2002. The allelic architecture of human disease genes: common disease-common variant... or not? *Human Molecular Genetics* **11**:2417–2423. doi: [10.1093/hmg/11.20.2417](https://doi.org/10.1093/hmg/11.20.2417)
- Rai E**, Wakeland EK. 2011. Genetic predisposition to autoimmunity – what have we learned? *Seminars in Immunology* **23**:67–83. doi: [10.1016/j.smim.2011.01.015](https://doi.org/10.1016/j.smim.2011.01.015)
- Raj T**, Rothamel K, Mostafavi S, Ye C, Lee MN, Replogle JM, Feng T, Lee M, Asinovski N, Frohlich I, Imboywa S, Von Korff A, Okada Y, Patsopoulos NA, Davis S, McCabe C, Paik H.-i., Srivastava GP, Raychaudhuri S, Hafler DA, Koller D, Regev A, Hacohen N, Mathis D, Benoist C, Stranger BE, De Jager PL. 2014. Polarization of the

- effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**:519–523. doi: [10.1126/science.1249547](https://doi.org/10.1126/science.1249547)
- Ramos PS, Brown EE, Kimberly RP, Langefeld CD. 2010. Genetic factors predisposing to systemic lupus erythematosus and lupus nephritis. *Seminars in Nephrology* **30**:164–176. doi: [10.1016/j.semnephrol.2010.01.007](https://doi.org/10.1016/j.semnephrol.2010.01.007)
- Ray JP, Hacohen N. 2015. Impact of autoimmune risk alleles on the immune system. *Genome Medicine* **7**:57. doi: [10.1186/s13073-015-0182-y](https://doi.org/10.1186/s13073-015-0182-y)
- Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee H-S, Jia X, Alfredsson L, Padyukov L, Klareskog L, Worthington J, Siminovitch KA, Bae S-C, Plenge RM, Gregersen PK, de Bakker PIW. 2012. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nature Genetics* **44**:291–296. doi: [10.1038/ng.1076](https://doi.org/10.1038/ng.1076)
- Raychaudhuri S. 2011. Mapping rare and common causal alleles for complex human diseases. *Cell* **147**:57–69. doi: [10.1016/j.cell.2011.09.011](https://doi.org/10.1016/j.cell.2011.09.011)
- Raymond CK, Kas A, Paddock M, Qiu R, Zhou Y, Subramanian S, Chang J, Palmieri A, Haugen E, Kaul R, Olson MV. 2005. Ancient haplotypes of the HLA class II region. *Genome Research* **15**:1250–1257. doi: [10.1101/gr.3554305](https://doi.org/10.1101/gr.3554305)
- Reith W, LeibundGut-Landmann S, Waldburger JM. 2005. Regulation of MHC class II gene expression by the class II transactivator. *Nature Reviews Immunology* **5**: 793–806. doi: [10.1038/nri1708](https://doi.org/10.1038/nri1708)
- Relle M, Weinmann-Menke J, Scorletti E, Cavagna L, Schwarting A. 2015. Genetics and novel aspects of therapies in systemic lupus erythematosus. *Autoimmunity Reviews* **14**:1005–1018. doi: [10.1016/j.autrev.2015.07.003](https://doi.org/10.1016/j.autrev.2015.07.003)
- Ridgway WM, Ito H, Fasso M, Yu C, Garrison Fathman C. 1998. Analysis of the role of variation of major histocompatibility complex class II expression on nonobese diabetic (NOD) peripheral T cell response. *Journal of Experimental Medicine* **188**:2267–2275. doi: [10.1084/jem.188.12.2267](https://doi.org/10.1084/jem.188.12.2267)
- Rockhill B, Newman B, Weinberg C. 1998a. Use and misuse of population attributable fractions. *American Journal of Public Health* **88**:15–19. doi: [10.2105/AJPH.88.1.15](https://doi.org/10.2105/AJPH.88.1.15)
- Rockhill B, Weinberg CR, Newman B. 1998b. Population attributable fraction estimation for established breast cancer risk factors: considering the issues of high prevalence and unmodifiability. *American Journal of Epidemiology* **147**:826–833. doi: [10.1093/oxfordjournals.aje.a009535](https://doi.org/10.1093/oxfordjournals.aje.a009535)
- Schur PH, Marcus-Bagley D, Awdeh Z, Yunis EJ, Alper CA. 1990. The effect of ethnicity on major histocompatibility complex complement allotypes and extended haplotypes in patients with systemic lupus erythematosus. *Arthritis & Rheumatism* **33**:985–992. doi: [10.1002/art.1780330710](https://doi.org/10.1002/art.1780330710)
- She J-X, Wakeland EK. 1991. Molecular and Genetic Mechanisms Involved in the Generation of Mhc Diversity. *Molecular Evolution of the Major Histocompatibility Complex*. Berlin, Heidelberg: Springer Berlin Heidelberg; 139–154. doi: [10.1007/978-3-642-84622-9\\_11](https://doi.org/10.1007/978-3-642-84622-9_11)
- Sheffield NC, Thurman RE, Song L, Safi A, Stamatoyannopoulos JA, Lenhard B, Crawford GE, Furey TS. 2013. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Research* **23**:777–788. doi: [10.1101/gr.152140.112](https://doi.org/10.1101/gr.152140.112)
- Smolen JS, Klippel JH, Penner E, Reichlin M, Steinberg AD, Chused TM, Scherak O, Graninger W, Hartter E, Zielinski CC. 1987. HLA-DR antigens in systemic lupus erythematosus: association with specificity of autoantibody responses to nuclear antigens. *Annals of the Rheumatic Diseases* **46**:457–462. doi: [10.1136/ard.46.6.457](https://doi.org/10.1136/ard.46.6.457)
- Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, Voight BF, Kraft P, Chen R, Kallberg HJ, Kurreeman FA, Kathiresan S, Wijmenga C, Gregersen PK, Alfredsson L, Siminovitch KA, Worthington J, de Bakker PI, Raychaudhuri S, Plenge RM, Diabetes Genetics Replication and Meta-analysis Consortium, Myocardial Infarction Genetics Consortium. 2012. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nature Genetics* **44**:483–489. doi: [10.1038/ng.2232](https://doi.org/10.1038/ng.2232)
- Steimle V, Siegrist C, Mottet A, Lisowska-Groszpiette B, Mach B. 1994. Regulation of MHC class II expression by interferon-gamma mediated by the transactivator gene CIITA. *Science* **265**:106–109. doi: [10.1126/science.8016643](https://doi.org/10.1126/science.8016643)
- Tang H, Jin X, Li Y, Jiang H, Tang X, Yang X, Cheng H, Qiu Y, Chen G, Mei J, Zhou F, Wu R, Zuo X, Zhang Y, Zheng X, Cai Q, Yin X, Quan C, Shao H, Cui Y, Tian F, Zhao X, Liu H, Xiao F, Xu F, Han J, Shi D, Zhang A, Zhou C, Li Q, Fan X, Lin L, Tian H, Wang Z, Fu H, Wang F, Yang B, Huang S, Liang B, Xie X, Ren Y, Gu Q, Wen G, Sun Y, Wu X, Dang L, Xia M, Shan J, Li T, Yang L, Zhang X, Li Y, He C, Xu A, Wei L, Zhao X, Gao X, Xu J, Zhang F, Zhang J, Li Y, Sun L, Liu J, Chen R, Yang S, Wang J, Zhang X. 2014. A large-scale screen for coding variants predisposing to psoriasis. *Nature Genetics* **46**:45–50. doi: [10.1038/ng.2827](https://doi.org/10.1038/ng.2827)
- Taylor KE, Chung SA, Graham RR, Ortmann WA, Lee AT, Langefeld CD, Jacob CO, Kamboh MI, Alarcón-Riquelme ME, Tsao BP, Moser KL, Gaffney PM, Harley JB, Petri M, Manzi S, Gregersen PK, Behrens TW, Criswell LA. 2011. Risk alleles for systemic lupus erythematosus in a large case-control collection and associations with clinical subphenotypes. *PLoS Genetics* **7**:e1001311. doi: [10.1371/journal.pgen.1001311](https://doi.org/10.1371/journal.pgen.1001311)
- Theofilopoulos AN. 1995a. The basis of autoimmunity: part I mechanisms of aberrant self-recognition. *Immunology Today* **16**:90–98. doi: [10.1016/0167-5699\(95\)80095-6](https://doi.org/10.1016/0167-5699(95)80095-6)
- Theofilopoulos AN. 1995b. The basis of autoimmunity: part II genetic predisposition. *Immunology Today* **16**: 150–159. doi: [10.1016/0167-5699\(95\)80133-2](https://doi.org/10.1016/0167-5699(95)80133-2)
- Todd JA, Bell JI, McDevitt HO. 1987. HLA-DQ $\beta$  gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* **329**:599–604. doi: [10.1038/329599a0](https://doi.org/10.1038/329599a0)

- Tsokos GC.** 2011. Systemic lupus erythematosus. *New England Journal of Medicine* **365**:2110–2121. doi: [10.1056/NEJMra1100359](https://doi.org/10.1056/NEJMra1100359)
- Unanue ER.** 2014. Antigen presentation in the autoimmune diabetes of the NOD mouse. *Annual Review of Immunology* **32**:579–608. doi: [10.1146/annurev-immunol-032712-095941](https://doi.org/10.1146/annurev-immunol-032712-095941)
- Vander Lugt B, Khan AA, Hackney JA, Agrawal S, Lesch J, Zhou M, Lee WP, Park S, Xu M, DeVoss J, Spooner CJ, Chalouni C, Delamarre L, Mellman I, Singh H.** 2014. Transcriptional programming of dendritic cells for enhanced MHC class II antigen presentation. *Nature Immunology* **15**:161–167. doi: [10.1038/ni.2795](https://doi.org/10.1038/ni.2795)
- Vernot B, Stergachis AB, Maurano MT, Vierstra J, Neph S, Thurman RE, Stamatoiyannopoulos JA, Akey JM.** 2012. Personal and population genomics of human regulatory variation. *Genome Research* **22**:1689–1697. doi: [10.1101/gr.134890.111](https://doi.org/10.1101/gr.134890.111)
- Visscher PM, Yang J, Goddard ME.** 2010. A commentary on ‘common SNPs explain a large proportion of the heritability for human height’ by yang et al. (2010). *Twin Research and Human Genetics* **13**:517–524. doi: [10.1375/twin.13.6.517](https://doi.org/10.1375/twin.13.6.517)
- Wang K, Dickson SP, Stolle CA, Krantz ID, Goldstein DB, Hakonarson H.** 2010. Interpretation of association signals and identification of causal variants from genome-wide association studies. *The American Journal of Human Genetics* **86**:730–742. doi: [10.1016/j.ajhg.2010.04.003](https://doi.org/10.1016/j.ajhg.2010.04.003)
- Wang S, Wiley GB, Kelly JA, Gaffney PM.** 2015. Disease mechanisms in rheumatology-tools and pathways: defining functional genetic variants in autoimmune diseases. *Arthritis & Rheumatology* **67**:1–10. doi: [10.1002/art.38800](https://doi.org/10.1002/art.38800)
- Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, Zernakova A, Zernakova DV, Veldink JH, Van den Berg LH, Karjalainen J, Withoff S, Uitterlinden AG, Hofman A, Rivadeneira F, 't Hoen PAC, Reinmaa E, Fischer K, Nelis M, Milani L, Melzer D, Ferrucci L, Singleton AB, Hernandez DG, Nalls MA, Homuth G, Nauck M, Radke D, Völker U, Perola M, Salomaa V, Brody J, Suchy-Dacey A, Gharib SA, Enquobahrie DA, Lumley T, Montgomery GW, Makino S, Prokisch H, Herder C, Roden M, Grallert H, Meitinger T, Strauch K, Li Y, Jansen RC, Visscher PM, Knight JC, Psaty BM, Ripatti S, Teumer A, Frayling TM, Metspalu A, van Meurs JBJ, Franke L.** 2013. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature Genetics* **45**:1238–1243. doi: [10.1038/ng.2756](https://doi.org/10.1038/ng.2756)
- Whitfield TW, Wang J, Collins PJ, Partridge EC, Aldred S, Trinklein ND, Myers RM, Weng Z.** 2012. Functional analysis of transcription factor binding sites in human promoters. *Genome Biology* **13**:R50–90. doi: [10.1186/gb-2012-13-9-r50](https://doi.org/10.1186/gb-2012-13-9-r50)
- Xi H, Shulha HP, Lin JM, Vales TR, Fu Y, Bodine DM, McKay RDG, Chenoweth JG, Tesar PJ, Furey TS, Ren B, Weng Z, Crawford GE.** 2007. Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genetics* **3**:e136. doi: [10.1371/journal.pgen.0030136](https://doi.org/10.1371/journal.pgen.0030136)
- Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM.** 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**:565–569. doi: [10.1038/ng.608](https://doi.org/10.1038/ng.608)
- Zheng SL, Sun J, Wiklund F, Smith S, Stattin P, Li G, Adami H-O, Hsu F-C, Zhu Y, Bälter K, Kader AK, Turner AR, Liu W, Bleecker ER, Meyers DA, Duggan D, Carpten JD, Chang B-L, Isaacs WB, Xu J, Grönberg H.** 2008. Cumulative association of five genetic variants with prostate cancer. *New England Journal of Medicine* **358**: 910–919. doi: [10.1056/NEJMoa075819](https://doi.org/10.1056/NEJMoa075819)
- Ziegler A.** Studies Genome-wide association. 2009. Genome-wide association studies: quality control and population-based measures. *Genetic Epidemiology* **33**:S45–S50. doi: [10.1002/gepi.20472](https://doi.org/10.1002/gepi.20472)