**Supplementary Information**

**Positive selection affects the expression of systemic lupus erythematosus associated loci in human populations**

Victoria Oberreiter1, Tobias Goellner1, David L. Morris2, Helmut Schaschl1,\*

**Affiliations**

1*Department of Evolutionary Anthropology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria*

*Department of Medical and Molecular Genetics, Faculty of Life Sciences & Medicine, King's College London School of Medicine, Great Maze Pond, London SE1 9RT, United Kingdom*

**Supplementary Table 1.** SNP allele frequencies (1000 Genomes) and genomic position of the allele under positive selection at the different SLE risk loci.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Beneficial allele/ alternate allele** | **Position** | **Locus** | **Allele frequencies** | | | |
| **AFR** | **EUR** | **SAS** | **EAS** |
| rs4655706-G/A | 67855152 | *IL12RB2* | 0.20 | 0.58 | 0.56 | 0.34 |
| rs10176621-T/C | 192035043 | *STAT4* | 0.13 | 0.74 | 0.76 | 0.58 |
| rs4681674-G/A;  rs4681834-T/G | 58273815  58263151 | *ABHD6*  *ABHD6* | 0.08  0.11 | 0.30  0.58 | 0.16  0.65 | 0.01  0.78 |
| rs583911-G/A | 159710390 | *IL12A* | 0.16 | 0.47 | 0.43 | 0.70 |
| rs4479588-G/T;  rs9682946-A/G | 188899235  188792843 | *TPRG1*  *TPRG1* | 0.05  0.14 | 0.43  0.93 | 0.29  0.86 | 0.40  0.91 |
| rs9307257-G/A | 102396292 | *BANK1* | 0.87 | 0.54 | 0.56 | 0.76 |
| rs7708392-G/C | 150457485 | *TNIP1* | 0.15 | 0.74 | 0.60 | 0.27 |
| rs11755393-A/G | 74099826 | *UHRF1BP1* | 0.23 | 0.65 | 0.71 | 0.54 |
| rs2428793-C/G;  rs6460087-A/G | 74099826  74065589 | *GTF2I*  *GTF2I* | 0.47  0.20 | 0.07  0.67 | 0.02  0.68 | 0.04  0.88 |
| rs1015877-C/G | 75296212 | *HIP1* | 0.20 | 0.34 | 0.29 | 0.32 |
| rs10086521-T/C | 10783769 | *XKR6* | 0.03 | 0.35 | 0.37 | 0.77 |
| rs4841563-T/G | 11425081 | *BLK* | 0.05 | 0.32 | 0.47 | 0.92 |
| rs75346988-G/A | 5097324 | *JAK2* | 0.82 | 1.0 | 1.0 | 1.0 |
| rs2732552-T/C | 35084592 | *CD44* | 0.69 | 0.45 | 0.28 | 0.25 |
| rs6539014-A/G | 102277782 | *DRAM1* | 0.02 | 0.43 | 0.33 | 0.09 |
| rs10774625-A/G | 111910215 | *ATXN2* | 0.02 | 0.48 | 0.07 | 0.0 |
| rs1274641-G/A | 68668994 | *RAD51B* | 0.68 | 0.82 | 0.83 | 0.97 |
| rs7197422-C/G | 11121071 | *CLEC16A* | 0.30 | 0.59 | 0.57 | 0.66 |
| rs140490-G/T | 21921686 | *UBE2L3* | 0.50 | 0.81 | 0.57 | 0.49 |
| rs61616683-T/C | 39755773 | *SYNGR1* | 0.04 | 0.23 | 0.35 | 0.78 |

**Supplementary Table 2.** Pairwise *FST*–values of SNPs under positive selection at SLE risk loci for different human populations (Pop): AFR – African population (LWK), EUR – European population (GBR), SAS – South Asian population (BEB), EAS – East Asian population (CHB).

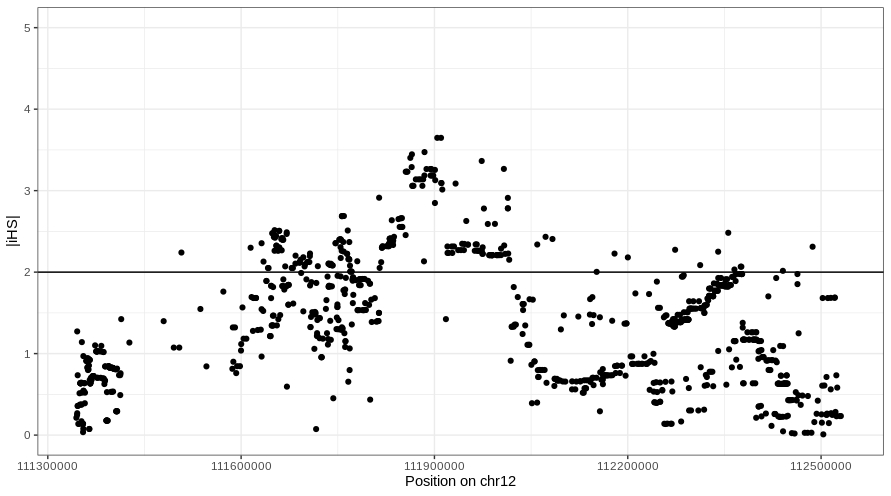
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | **Pairwise** ***FST*** | | | | |
| **Risk region** | **Risk gene\_locus** | **Positively selected SNP** | **Pop** | **AFR** | **EUR** | **SAS** | **EAS** |
| 1p31.3 | *IL12RB2* | rs4655706 | AFR | - |  |  |  |
| EUR | 0.273 | - |  |  |
| SAS | 0.211 | 0.002 | - |  |
| EAS | 0.071 | 0.081 | 0.040 | - |
| 2q32.3 | *STAT4* | rs10176621 | AFR | - |  |  |  |
| EUR | 0.572 | - |  |  |
| SAS | 0.577 | 0 | - |  |
| EAS | 0.426 | 0.03 | 0.03 | - |
| 3p14.3 | *ABHD6* | rs4681674  rs4681834 | AFR | - |  |  |  |
| EUR | 0.195 | - |  |  |
| SAS | 0 | 0.152 | - |  |
| EAS | 0.087 | 0.380 | 0.121 | - |
| AFR | - |  |  |  |
| EUR | 0.354 | - |  |  |
| SAS | 0.478 | 0.0172 | - |  |
| EAS | 0.593 | 0.089 | 0.020 | - |
| 3q25.33 | *IL12A* | rs583911 | AFR | - |  |  |  |
| EUR | 0.189 | - |  |  |
| SAS | 0.166 | 0 | - |  |
| EAS | 0.459 | 0.110 | 0.131 | - |
| 3q28 | *TPRG1* | rs4479588  rs9682946 | AFR | - |  |  |  |
| EUR | 0.320 | - |  |  |
| SAS | 0.138 | 0.064 | - |  |
| EAS | 0.313 | 0 | 0.065 | - |
| AFR | - |  |  |  |
| EUR | 0.865 | - |  |  |
| SAS | 0.799 | 0.015 | - |  |
| EAS | 0.879 | 0 | 0.028 | - |
| 4q24 | *BANK1* | rs9307257 | AFR | - |  |  |  |
| EUR | 0.235 | - |  |  |
| SAS | 0.219 | 0 | - |  |
| EAS | 0.041 | 0.093 | 0.08 | - |
| 5q33.1 | *TNIP1* | rs7708392 | AFR | - |  |  |  |
| EUR | 0.451 | - |  |  |
| SAS | 0.232 | 0.070 | - |  |
| EAS | 0.026 | 0.312 | 0.111 | - |
| 6p21.31 | *UHRF1BP1* | rs11755393 | AFR | - |  |  |  |
| EUR | 0.419 | - |  |  |
| SAS | 0.461 | 0 | - |  |
| EAS | 0.252 | 0.037 | 0.059 | - |
| 7q11.23 | *GTF2I* | rs2428793  rs6460087 | AFR | - |  |  |  |
| EUR | 0.242 | - |  |  |
| SAS | 0.393 | 0.071 | - |  |
| EAS | 0.385 | 0.047 | 0.002 | - |
| AFR | - |  |  |  |
| EUR | 0.227 | - |  |  |
| SAS | 0.317 | 0.008 | - |  |
| EAS | 0.595 | 0.203 | 0.126 | - |
| *HIP1* | rs1015877 | AFR | - |  |  |  |
| EUR | 0 | - |  |  |
| SAS | 0.041 | 0 | - |  |
| EAS | 0.119 | 0.003 | 0.020 | - |
| 8p23.1 | *XKR6* | rs10086521 | AFR | - |  |  |  |
| EUR | 0.322 | - |  |  |
| SAS | 0.455 | 0.023 | - |  |
| EAS | 0.743 | 0.296 | 0.168 | - |
| *BLK* | rs4841563 | AFR | - |  |  |  |
| EUR | 0.164 | - |  |  |
| SAS | 0.387 | 0.077 | - |  |
| EAS | 0.851 | 0.591 | 0.373 | - |
| 9p24 | *JAK2* | rs75346988 | AFR | - |  |  |  |
| EUR | 0.226 | - |  |  |
| SAS | 0.221 | nan | - |  |
| EAS | 0.237 | nan | nan | - |
| 11p13 | *CD44* | rs2732552 | AFR | - |  |  |  |
| EUR | 0.164 | - |  |  |
| SAS | 0.274 | 0.016 | - |  |
| EAS | 0.365 | 0.062 | 0.008 | - |
| 12q23.2 | *DRAM1* | rs7137913  rs6539014 | AFR | - |  |  |  |
| EUR | 0.537 | - |  |  |
| SAS | 0.277 | 0.106 | - |  |
| EAS | 0.156 | 0.216 | 0.020 | - |
| AFR | - |  |  |  |
| EUR | 0.453 | - |  |  |
| SAS | 0.276 | 0.058 | - |  |
| EAS | 0.057 | 0.308 | 0.126 | - |
| 12q24.12 | *SH2B3\_ATXN2* | rs10774625 | AFR | - |  |  |  |
| EUR | 0.519 | - |  |  |
| SAS | 0.070 | 0.375 | - |  |
| EAS | 0.005 | 0.504 | 0.045 | - |
| 14q24.1 | *RAD51B* | rs1274641 | AFR | - |  |  |  |
| EUR | 0.062 | - |  |  |
| SAS | 0.040 | 0 | - |  |
| EAS | 0.253 | 0.100 | 0.127 | - |
| 16p13 | *CLEC16A* | rs8049882 | AFR | - |  |  |  |
| EUR | 0.164 | - |  |  |
| SAS | 0.174 | 0 | - |  |
| EAS | 0.281 | 0.020 | 0.015 | - |
| 22q11.21 | *UBE2L3* | rs140490 | AFR | - |  |  |  |
| EUR | 0.225 | - |  |  |
| SAS | 0.003 | 0.163 | - |  |
| EAS | 0.013 | 0.133 | 0 | - |
| 22q13.1 | *SYNGR1* | rs61616683 | AFR | - |  |  |  |
| EUR | 0.179 | - |  |  |
| SAS | 0.382 | 0.073 | - |  |
| EAS | 0.738 | 0.457 | 0.234 | - |

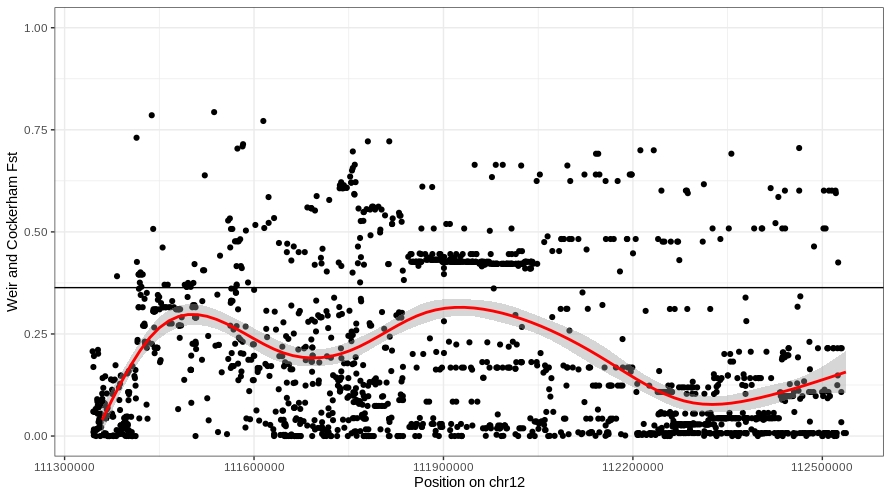
nan = not a number

**Supplementary Table 3.** The locations of SLR risk loci under positive selection in the Reactome Pathway Browser (<https://reactome.org/>); timing of selection (in kya) and potential target gene of positive selection are also given. N.d. = no data.

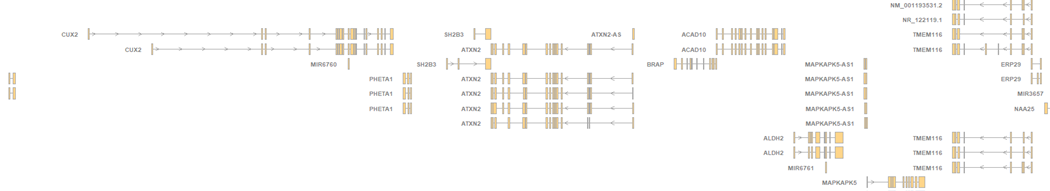
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SNP under positive selection** | **GTEx functions** | **Timing of selection** | **Potential target of positive selection** | **Reactome - locations in the**  **PathwayBrowser** |
| rs4655706 | eQTL | 9 - 11 | *IL12RB2* | immune system |
| rs10176621 | eQTL | 19 - 21 | *STAT4* | immune system |
| rs4681674; rs4681834 | eQTL;  eQTL | 7 - 10  9 - 13 | *ABHD6* | hemostasis, signal transduction |
| rs583911 | eQTL | 13 - 19 | *IL12A* | immune system |
| rs4479588; rs9682946 | -  - | 10 - 16  14 - 19 | *-* | n.d |
| rs9307257 | eQTL | 5 - 9 | *FLJ20021* | n.d |
| rs7708392 | eQTL | 15 - 20 | *TNIP1* | metabolism of proteins |
| rs11755393 | eQTL | 12 - 15 | *UHRF1BP1* | n.d |
| rs2428793; rs6460087 | -  eQTL | 15 - 21  14 - 19 | *GTF2I* | n.d |
| rs1015877 | eQTL | 20 - 25 | *HIP1* | vesicle-mediated transport |
| rs10086521 | eQTL | 24 - 31 | *BLK* | n.d |
| rs4841563 | eQTL | 28 - 35 | *BLK* | gene expression |
| rs75346988 |  | 8 - 9 | *-* | immune system |
| rs2732552 | eQTL | 10 - 14 | *CD44* | extracellular matrix organization, hemostasis, immune system, metabolism |
| rs6539014 | eQTL | 7 - 11 | *DRAM1* | n.d |
| rs10774625 | eQTL | 2 - 3 | *ALDH2* | metabolism |
| rs1274641 | eQTL | 30 - 35 | *RDH12* | DNA repair, hemostasis |
| rs7197422 | eQTL | 35 - 45 | *CLEC16A* | n.d |
| rs140490 | eQTL | 23 - 28 | *UBE2L3* | immune system; metabolism of proteins |
| rs61616683 | eQTL | 23 - 33 | *SYNGR1* | immune system |

**Supplementary Figure 1.** Scatter plots with (A) |iHS| and (B) *FST*values across the genomic region *SH2B3–ATXN2* (chr12q24.12). (A) y-axis of the plot shows the normalized iHS values for each SNP; the x-axis shows the genomic region on chr12. The black line indicates statistical significance (*p* < 0.05) level for |iHS| > 2.0. (B) y-axis of the plot shows the pairwise *FST* values for each SNP (GBR vs LWK); the y-axis shows the genomic region; the black line indicates statistical significance level (*p* < 0.05) for FST values ≥ 0.32; red line indicates the mean *FST* across the locus.

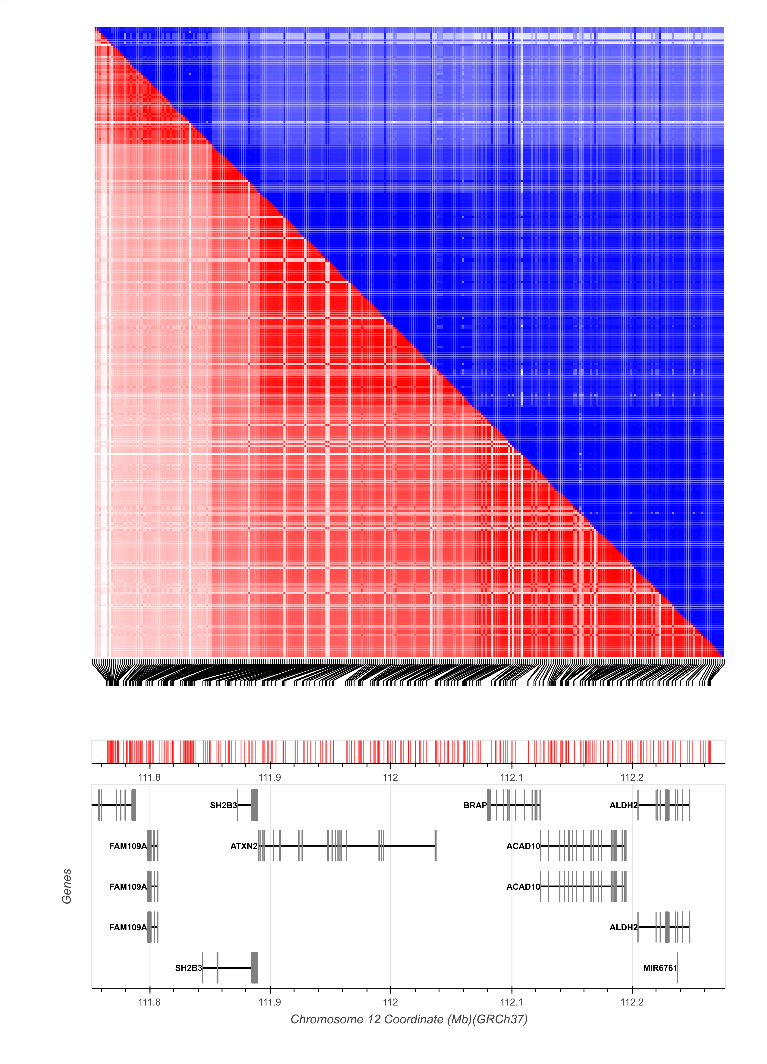
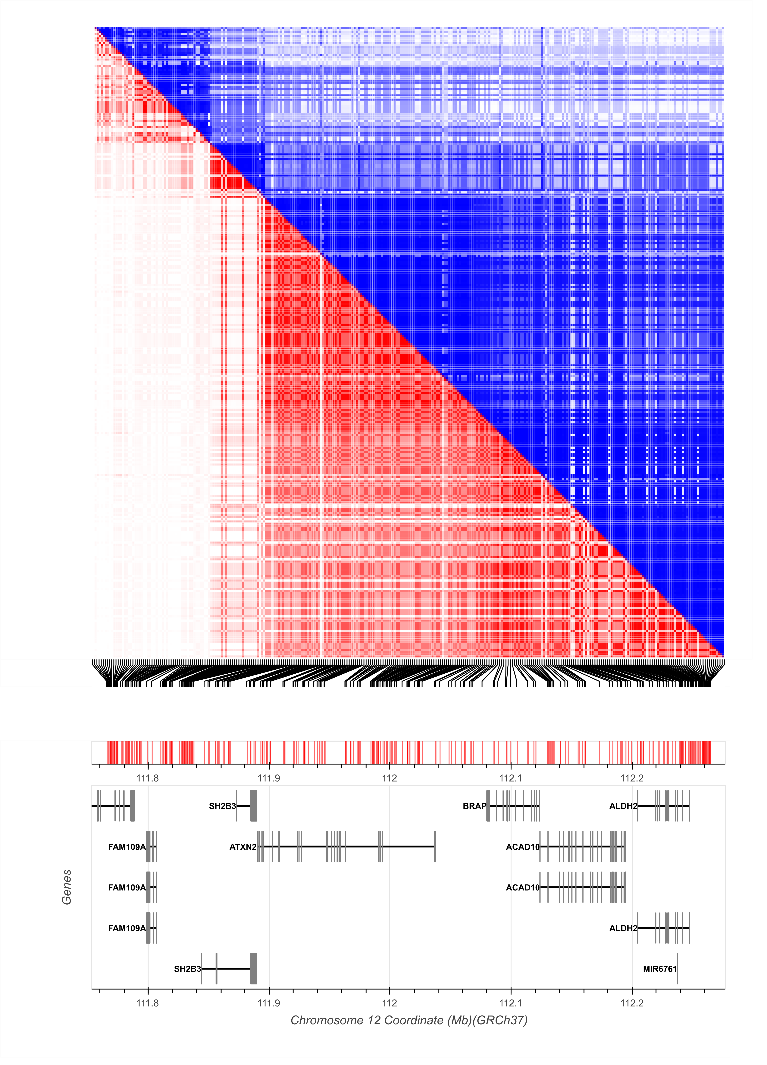
****(A)

****

(B)



**Supplementary Figure 2**. Linkage disequilibrium (LD) plot of the SLE risk region *SH2B3*-*ATXN2* at the chromosome region 12q24.12 for the populations GBR and CHB (for population details see Materials and Methods).



CHB

GBR

**Supplementary Table 4.** Genome wide association studies (GWASs) with reported susceptibility loci to SLE. Studies were retrieved using NHGRI–EBI GWAS catalog (<https://www.ebi.ac.uk/gwas/>).

|  |  |  |  |
| --- | --- | --- | --- |
| **PUBMEDID** | **STUDY** | **Studied population** | **REPORTED GENE(S)** |
| 18204098 | Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. | European | HLA-DQA1, STAT4, TNPO3, IRF5, ITGAM, ITGAX, C8orf13, BLK |
| 18204446 | Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. | European | HLA region, ITGAM, TNPO3, IRF5, KIAA1542, PXK |
| 18204447 | Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. | European | BANK1, TNFAIP3, STAT4, BLK |
| 19165918 | Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. | European | TNFAIP3, STAT4, BLK |
| 19357697 | Association of BANK1 and TNFSF4 with systemic lupus erythematosus in Hong Kong Chinese | Hong Kong Chinese | BANK1, TNFSF4 |
| 19838193 | Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. | Han Chinese | STAT4, TNFSF4, ETS1, BLK, IKZF1, IRF5, TNFAIP3, NR, UBE2L3, HIC2, RASGRP3, HLA-DRB1, PRDM1, ATG5, LRRC18, WDFY4, SLC15A4, TNIP1, HIP1, ZNF689, PRR14 |
| 19838195 | A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. | European | HLA-DRB1, STAT4, IRF5, ITGAM, BLK, TNIP1, TNFAIP3, PTPN22, PRDM1, JAZF1, PHRF1, TNFSF4, UHRF1BP1, IL10 |
| 20169177 | Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. | Han Chinese | WDFY4, ETS1 |
| 20662065 | Identification of candidate loci at 6p21 and 21q22 in a genome-wide association study of cardiac manifestations of neonatal lupus. | European | NFKBIL1, LTA, AIF1, TNF, LTB |
| 21044949 | ELF1 is associated with systemic lupus erythematosus in Asian populations. | Han Chinese | ELF1 |
| 21408207 | Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. | European | TNXB, HLA-DRB3, STAT4, IRF5, ITGAM, HLA-DQA2, IRF5 |
| 22291604 | A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus erythematosus in Japanese. | Japanese | AFF1 |
| 23053960 | Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis. | European | HLA-DQA2, TNPO3, STAT4, STAT4, BLK, IRF5 |
| 23273568 | Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as associated with systemic lupus erythematosus in Asians. | Han Chinese | HLA-DRB1, STAT4, GPR19, DGUOK, TET3, CD80, TNFAIP3, BLK, UHRF1BP1, TNFSF4, CDKN1B, DRAM1, ARID5B, RTKN2, IRF5, WDFY4, CREBL2, ETS1 |
| 23740937 | A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. |  | TNPO3, IRF5, STAT4, PXK, KIAA0319L, ITGAM, IRF8, TNFAIP3, BLK, JAZF1, TNIP1, CSK |
| 24163247 | Allelic heterogeneity in NCF2 associated with systemic lupus erythematosus (SLE) susceptibility across four ethnic populations | European, African-American, Hispanic, Korean | NCF2 |
| 24871463 | GWAS identifies novel SLE susceptibility genes and explains the association of the HLA region. | European | C2, HLA-DQA2, FAM98B, TYRO3, EDEM3, TNPO3, RNF114, ITGAM, IRF5, HLA-DQA1, STAT4, ZNF184, SPATA8 |
| 25338677 | Genetic analysis of the pathogenic molecular sub-phenotype interferon-alpha identifies multiple novel loci involved in systemic lupus erythematosus. | European | PRKG1, A1CF, CSTF2T |
| 25620976 | Lupus risk variants in the PXK locus alter B-cell receptor internalization | European, African American, Asian, Hispanic | PXK |
| 25862617 | Genome-wide search followed by replication reveals genetic interaction of CD80 and ALOX5AP associated with systemic lupus erythematosus in Asian populations. | Han Chinese | CD80 |
| 26301688 | Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. | European | LPHN2, TNM3, ANKRD30A, IL23R, PTPN22, ATG16L1, DAG1, PTGER4, ZNF365, INS, SMAD3, NOD2, NKX2-3, ANKRD55, IL12B, LRRK2, IL5, SUOX, SBK1, ADCY7, IL2RA, CD40LG, TNFSF15, ZMIZ1, IL21, CARD9, FUT2, PSMG1 |
| 26316170 | Identification of a New Susceptibility Locus for Systemic Lupus Erythematosus on Chromosome 12 in Individuals of European Ancestry. | European | HLA-DQA1, HLA-DQB1, TNXB |
| 26502338 | Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. | European | MHC class III, STAT4, IRF5, TNPO3, TNXB, HLA-DQB1, DPCR, MUC21, SMG7, NCF2, ITGAM, IRF5, TNIP1, TNFAIP3, PTPN22, MIR146A, CD44, BLK, UBE2L3, YDJC, DHCR7, NADSYN1, TNFSF4, IRF8, BANK1, CIITA, SOCS1, CLEC16A, TCF7, SKP1, IRAK1, MECP2, ARHGAP4, NAA10, RENBP, HCFC1, TMEM187, MIR718, CSK, SCAMP5, PPCDC, ABHD6, PXK, PDHB, PRDM1, ATG5, IKZF1, TYK2, IKZF2, SLC15A4, UBE2L3, IRF7, HLA-DQ1, FCGR2A, SLC44A4, FCGR2A, FCGR2B, FCGR3B, TNFSF4, LOC100506023, WDFY4, IFIH1, IL12A, JAZF1, SPRED2, ARID5B, ETS1, FLI1, CXorf21, CSK, RAD51B, IRF7, LRRC56, LMNTD2, RASSF7, MIR210HG, PHRF1, CDHR5, LYST, PLD2, SH2B3, ATXN2, IL10, IKZF3, UHRF1BP1, HER-2, C17orf37, GRB7, ZNFN1A3, ZBPB2, GSDMB, IKZF3, ERBB2, CIITA, SOCS1 |
| 26606652 | Genome-Wide Association Study in an Amerindian Ancestry Population Reveals Novel Systemic Lupus Erythematosus Risk Loci and the Role of European Admixture. | Native American | TNPO3, IRF5, IRF5, ITGAM, IRAK1, HLA-DQA1, STAT4, HLA-DQB1, HLA-DRB1, TNIP1, LOC100506023, USMG5, BC040734, HLA-B, NCF2 |
| 26663301 | Identification of a Systemic Lupus Erythematosus Risk Locus Spanning ATG16L2, FCHSD2, and P2RY2 in Koreans. | Korean | STAT4, IKZF1, C7orf72, PERP, TNFAIP3, PRDX6, TNFSF4, HIP1, LOC392787, IRF5, ETS1, FCHSD2, P2RY2, ATG16L2, BLK, TMEM187, MECP2, IRAK1, WDFY4, TNFSF4, PRDX6 |
| 26980576 | Response to Intravenous Cyclophosphamide Treatment for Lupus Nephritis Associated with Polymorphisms in the FCGR2B-FCRLA Locus. | Korean | FCGR2B, FCRLA, FCGR3B, RPL31P11 |
| 27193031 | A combined large-scale meta-analysis identifies COG6 as a novel shared risk locus for rheumatoid arthritis and systemic lupus erythematosus. | European | PTPN22, TNPO3, IRF5, TNFAIP3, COG6, BLK, UBE2L3, ICAM3, TYK2 |
| 27399966 | Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. | Han Chinese | MHC class III, STAT4, SMG7, NCF2, ITGAM, IRF5, TNIP1, BLK, JAK2, IRF8, TNFAIP3, LBH, TNFSF4, MIR146A, LPP, TPRG1-AS1, GTF2IRD1-GTF2I, UBE2L3, PTPN22, TYK2, PTPRC, CD45, ATXN1, PRDM1, ATG5, BACH2, IKBKE, CD44, FCGR2A, WDFY4, IRF7, UHRF1BP1, TCF7, SKP1, BANK1, RNASEH2C, CSK, IKZF2, ZFP90, TET3 |
| 29494758 | Genome-wide association study identifies three novel susceptibility loci for systemic lupus erythematosus in Han Chinese. | Han Chinese | KIT, TRAPPC11, GPR78 |
| 29848360 | Genome-wide association study meta-analysis identifies five new loci for systemic lupus erythematosus. | European | GRB2, SMYD3, LAT2, ARHGAP27, ST8SIA4 |
| 29967481 | Genetic variants in systemic lupus erythematosus susceptibility loci, XKR6 and GLT1D1 are associated with childhood-onset SLE in a Korean cohort. | Korean | XKR6, GLT1D1 |

**Supplementary Table 5.** 1000 Genomes populations and genetic ancestries included in the study (<ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/20130502/>).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Number of individuals** | **Population code** | **Population**  **description** | **Super population** | **Super population code** |
| 99 | LWK | Luhya in Webuye, Kenya | Africa | AFR |
| 113 | GWD | Gambians from The Gambia | Africa | AFR |
| 99 | ESN | Esan in Nigeria | Africa | AFR |
| 91 | GBR | British in England and Scotland | Europe | EUR |
| 99 | FIN | Finnish in Finland | Europe | EUR |
| 107 | TSI | Toscani in Italia | Europe | EUR |
| 86 | BEB | Bengali from Bangladesh | South Asian | SAS |
| 102 | ITU | Indian Telugu from the UK | South Asian | SAS |
| 96 | PJL | Punjabi from Lahore, Pakistan | South Asian | SAS |
| 103 | CHB | Han Chinese in Beijing, China | East Asian | EAS |
| 104 | JPT | Japanese in Tokyo, Japan | East Asian | EAS |
| 99 | KHV | Kinhin Ho Chi Minh City, Vietnam | East Asian | EAS |