A multilayered post-GWAS analysis pipeline defines functional variants and target genes for systemic lupus erythematosus (SLE)

Mehdi Fazel-Najafabadi[1*], Loren L. Looger[2,3*], Harikrishna Reddy-Rallabandi[1], Swapan K. Nath[1†]

[1]Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104, USA
[2]Department of Neurosciences, University of California, San Diego, San Diego, CA 92121, USA
[3]Howard Hughes Medical Institute, University of California, San Diego, San Diego, CA 92121, USA

*Contributed equally

†To whom correspondence should be addressed:
Swapan K. Nath, PhD
Oklahoma Medical Research Foundation
825 NE 13th Street
Oklahoma City, OK 73104
Email: Swapan-Nath@omrf.org

Competing Interest Statement
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Funding information
This work was supported by Grant # AR060366 from National Institutes of Health to SKN.

Ethics statements
Studies involving human subjects
This study involves various genomic data downloaded from public databases (1000Genomes data, various chromatin marks from ENCODE, and Summary GWAS), and cell lines bought from ATCC. This project was approved by Institutional Review Board, OMRF.

Inclusion of identifiable human data
No potentially identifiable human images or data is presented in this study.

Abstract

Objectives: Systemic lupus erythematosus (SLE), an autoimmune disease with incompletely understood etiology, has a strong genetic component. Although genome-wide association studies (GWAS) have revealed multiple SLE susceptibility loci and associated single nucleotide polymorphisms (SNPs), the precise causal variants, target genes, cell types, tissues, and mechanisms of action remain largely unknown.

Methods: Here, we report a comprehensive post-GWAS analysis using extensive annotation, molecular modeling, and integrative functional genomic and epigenomic analyses to optimize fine-mapping. We compile and cross-reference immune cell-specific expression quantitative trait loci (*cis*- and *trans*-eQTLs) with promoter capture Hi-C, allele-specific chromatin accessibility, and massively parallel reporter assay data to define predisposing variants and target genes. To assess our predictions, we experimentally validate a locus using CRISPR/Cas9 genome editing, qPCR, and Western blot.

Results: Anchoring on 452 index SNPs, we selected 9,931 high-linkage disequilibrium (r^2>0.8) SNPs and defined 182 independent non-HLA SLE loci. 3,746 SNPs from 143 loci were identified as regulating 564 unique genes. Target genes are enriched in lupus-related tissues and associated with other autoimmune diseases. Of these, 329 SNPs (106 loci) showed significant allele-specific chromatin accessibility or enhancer activity, indicating regulatory potential. Using CRISPR/Cas9, we validated rs57668933 as a functional variant regulating multiple targets, including SLE risk gene *ELF1*, in B-cells.

Conclusion: We demonstrate and validate post-GWAS strategies for utilizing multi-dimensional data to prioritize likely causal variants with cognate gene targets underlying SLE pathogenesis. Our results provide a catalog of significantly SLE-associated SNPs and loci, target genes, and likely biochemical mechanisms, to guide experimental characterization.

Introduction

Systemic lupus erythematosus (SLE, lupus) is a complex autoimmune disease with substantial genetic underpinnings, *e.g.*, strong familial aggregation[1-3], large twin concordance (monozygotic>dizygotic)[1,4], and high sibling recurrence risk ratio (λ_s~30)[5]. >50 candidate gene studies and genome-wide association studies (GWAS) have identified >100 SLE risk loci (*p*-value<5×10^{-8}), across multiple ethnicities[6-11]. However, these loci explain only ~30% of SLE heritability (h^2)[7,8,12].

In addition to incomplete knowledge of precise risk loci and alleles underlying GWAS peaks, it is not generally understood how such alleles mechanistically contribute to disease. For a given locus, GWAS often reports a sole ("index") single nucleotide polymorphism (SNP), which may or may not itself be functional, but is likely in linkage disequilibrium (LD) with disease-predisposing SNPs[13,14]. As in other complex diseases, >90% of reported SLE index SNPs are non-coding (intronic and intergenic). A major challenge in the post-GWAS era is to precisely identify predisposing coding and non-coding SNPs and their associated target genes, and to determine the molecular mechanisms underlying disease risk.

Accurate association determination remains a nontrivial challenge in clinical genomics and genome informatics. Generally, post-GWAS analyses combine multiple GWAS signals using LD structure and epigenetics[15-17]. Additional data sources, such as multiple independent, consistent annotations, greatly assists prioritization of likely functional SNPs[18-20]. SNPs can modulate transcription factor (TF) binding and chromatin structure, altering gene regulation. Indeed, *cis*- and *trans*-expression quantitative trait locus (eQTL) analyses frequently link disease-associated alleles to specific gene/isoform expression[21]. Together, annotating open, active chromatin from DNase I hypersensitivity and Assay for Transposase-Accessible Chromatin (ATAC-seq) peaks[22], alongside individual genomic regulatory elements (promoters, enhancers, silencers, *etc*.) using histone marks, chromatin modifiers, and transcription factors, and by *in silico* bioinformatics[23,24], yields a powerful framework for testing GWAS hypotheses.

Multiple databases (*e.g.*, ENCODE, RoadMap) integrate histone mark data from common cell lines to create consensus regulatory region annotation[25-27]. Moreover, combining ATAC-seq and

gene expression data yields chromatin accessibility QTL (caQTLs) to identify SNPs with allele-specific effects (*e.g.*, allelic imbalance)[28,29]. Genomic regulatory elements communicate with one another and target genes through complex three-dimensional chromatin interactions (topologically associating domains, TADs). Various chromatin conformation capture (3C) technologies[30,31] annotate TADs and other chromatin features. Such interactions, particularly direct enhancer-promoter interactions[32], underlie long-range enhancer activity[33,34], and can aid GWAS interpretation.

Finally, recent methods like Massively Parallel Reporter Assays (MPRAs) simultaneously screen thousands of SNPs for transcriptional enhancer activity, providing information about SNP genomic context and allele behavior[35].

We compiled all available data on the above features and cross-referenced them to predict locus/SNP functionality. Genomic regions can assume different activities in different cell types; we matched datasets taken from the same cell type. Further, when possible, annotations were taken from data derived solely from immune cells[23,32,36-39]. This ensures consistency and identifies cell types relevant to pathogenesis[40]. This approach has revealed target genes and cells associated with rheumatoid arthritis[41] and breast cancer[42], among others.

For protein-coding SNPs, these techniques are applied, as many exons contain TF-binding sites and/or promoters/enhancers[43]; we also consider effects on protein structure and function.

Using an integrated approach, we collate and reassess published SLE-GWAS association signals along with high-LD SNPs, incorporating diverse data on underlying genomic features. For each locus, we define potential functional variants and their cognate target genes in immune cells. We prioritize functional SNPs and assign associated target genes and modulated biochemical pathways. As proof of concept, we used CRISPR-Cas9 genome editing, qPCR, and protein blotting to validate allelic effects of a candidate SNP on SLE risk gene *ELF1*.

Methods

Study design

Our workflow and study design are shown in **Figure 1**. 1) From qualified studies, we collated all reported and replicated index and correlated SNPs to define statistically-independent SLE susceptibility loci. 2) We predicted SNP effects in statistically-independent loci and annotated them in regulatory tiers. 3) We performed molecular modeling on coding SNPs. 4) We leveraged cell type-specific *cis*- and –*trans*-expression quantitative trait loci (eQTLs) and promoter-capture Hi-C (PCHiC) data to define "enhancer" and "promoter" SNPs and target genes. 5) We estimated locus overrepresentation in molecular pathways and gene ontology categories and identified cell type-specific SNP enrichment in epigenetic features. 6) We used chromatin accessibility QTLs (caQTLs) and massively parallel reporter assay (MPRA) data to identify allele-specific effects. 7) We experimentally validated a functional variant using CRISPR/Cas9-based activation/silencing in B-cells.

Collating variants

We performed an exhaustive literature search from SLE association studies (to September 2021), both GWAS and candidate-gene studies with sample sizes >2,000, selecting genome-wide significant ($P<5x10^{-8}$) index SNPs (**Table 1**). Given numerous SNPs at each locus, strong linkage disequilibrium (LD) between them, and sparse genotyping coverage, index SNPs likely don't mechanistically underlie GWAS signals. To identify likely "causal" SNPs underlying association signals, we first expanded each locus to its LD region. Loci were considered independent if separated by >250 kb and in very low LD ($r^2<0.2$). The HLA region (hg19, chr6: 28,477,797-33,448,354) was condensed into a single locus and removed.

We extracted all population-relevant LD SNPs from major populations (EUR=European, EAS=East Asian, SAS=South Asian, AFR=African, and AMR=Mixed American) in 1000Genomes Project Phase III around each index SNP using NIH LDlink[44], using only high-LD SNPs ($r^2>0.8$).

Regulatory region annotation

PredictSNP2[45], a consensus classifier combining five prediction tools supplemented by experimental annotation from eight databases, initiated SNP annotation, supplemented by GENCODE GRCh37.36[46,47], Haploreg[48], Ensembl GRCh37.87 (https://ftp.ensembl.org/pub/grch37), GenoSTAN[27], silencerDB[26], eQTLgen[49], ImmuNexUT[50], and EpiMap[25].

For enhancer annotation, we used GenoSTAN (127 cell types and tissues from ENCODE and Roadmap[27]) and EpiMap[25] (3,030 aggregated and uniformly-reprocessed references from ENCODE, Roadmap, and Genomics of Gene Regulation). We imputed 14,952 datasets across 859 biosamples and 18 chromatin marks/assays. We used GenoSTAN to predict promoter and enhancer activity and silencerDB (33,060 experimentally-validated and 5,045,547 predicted silencers from machine learning) for silencers. For allele-specific enhancers, we used Massively Parallel Reporter Assays (MPRAs) on >3,000 SNPs with both alleles present[51]. We also used chromatin accessibility quantitative trait loci (caQTLs)[29,36] to further fine-map and annotate (allele-specific) SNP regulatory aspects.

Target genes

We used two methods to determine SNP targets. First, SNPs were annotated with *cis*- and/or *trans*-expression QTLs (eQTLs) and splicing QTLs (sQTLs) using multiple databases. We extracted immune cell-specific expression data from GTEx V8[36], Blood eQTL[39], DICE[37], ImmuNexUT[50], and eQTLgen[49]. SNP-gene pairs (GTEx QTLs with p<1e-4 and/or "is_best" flag, or false-discovery rate (FDR) < 0.05 for other sources) were recorded. We obtained blood sQTLs from GTEx V8. For *trans*-eQTLs, we overlapped SNPs with blood cell-based curated *trans*-eQTLs from the eQTLgen Consortium[49] (31,684 individuals from 37 cohorts). We set the *trans*-eQTL FDR threshold stringently (1e-5). To identify SNPs interacting with enhancers and promoters through chromatin interactions, we overlapped associated SNPs within anchors of chromatin interactions in immune cells with available promoter capture Hi-C (PCHiC)[23,32] data. Given that some interactions between enhancers and targets are distal, we annotated colocation of candidate SNPs within PCHiC loops. To identify candidate promoter-SNP pairs, we annotated each SNP with nearby PCHiC loops from immune cells.

SNP/geneset enrichment analysis

Gene targets of functional SNPs were tested for enrichment in Gene Ontology (GO) categories, pathway membership, and disease association. Enrichment analysis was carried out using FUMA[17] and epiCOLOC[52] on different genesets identified through target-type annotations.

Transcription factor binding

Binding sites were annotated from UCSC Genome Browser GRCh37/hg19 JASPAR core 22.

Protein models

Protein models were taken from AlphaFold2 and illustrated with PyMOL.

CRISPR-based functional validation

We used CRISPR/Cas9 activation/silencing (CRISPRa/i) to bring activating or silencing domains to rs57668933. Briefly, single-guide RNA (sgRNA)/Cas9-RNP complex was prepared at room temperature in Cas9 buffer. RNP complex was transfected into NA18535 LCL cells with Neon Electroporation and allowed to express for 72 hours.

qPCR

qRT-PCR was performed on WT, CRISPRa, and CRISPRi cells, as described elsewhere[53]. RNA was isolated from WT and CRISPRa/i cells using an RNA Mini kit (Zymo Research). RNA was synthesized using iScript Reverse Transcription Supermix cDNA synthesis kit (Bio-Rad). We measured *ELF1* expression and analyzed results for significance using Prism V.7 (GraphPad).

Results

Defining independent SLE candidate loci

Overall, we identified 452 reported genome-wide significant ($P<5x10^{-8}$) non-HLA index SNPs from 76 different GWAS and candidate gene studies (**Supplementary Table 1**). Most index SNPs derived from East Asian and European ancestry (**Supplementary Figure 1**). Most (242, 53.5%) index SNPs lay within 145 genes; 210 (46.5%) were intergenic (**Figure 1, Supplementary Table 1**).

We then collected SNPs in high (r^2>0.8) linkage disequilibrium (LD) with index SNPs, finding 9,479 – totaling 9,931 SNPs for our study. We binned these into 182 statistically-independent loci (**Supplementary Table 1&2, Supplementary Figure 1c**); with median locus size of 57.7 kb [range 314 bp – 1.15 Mb]. Of the 182 loci, 89 contained single index SNPs; the rest had 2-14 (**Supplementary Table 7**). Total linked SNPs per locus ranged from 1-1,148 (median 26; **Supplementary Table 7**). Correlated SNPs per index SNP ranged from 1-146 (median 24). Fifteen loci had single index SNPs and no LD-SNPs; conversely, LOC_180 had two index SNPs and 1,146 LD-SNPs. The physical distance between index SNPs and LD-SNPs varied from 1 bp to 499 kb (median 14 kb).

The 182 independent loci contain 426 genes. Of our 9,931 total SNPs, 47.0% are intronic, 0.9% synonymous, 0.9% missense, and 51.2% intergenic (**Figure 1**, **Supplementary Table 2**). We annotated all 9,931 SNPs for *cis*- and *trans*-regulatory effects. The 89 missense SNPs potentially alter both gene product and expression level and were additionally characterized by molecular modeling.

Annotation pipeline

To annotate and prioritize these 9,931 SNPs at 182 loci/426 genes, we established this pipeline: 1) collate eQTLs, 2) collate PCHiC, 3) combine these initially classify SNPs, 4) add histone mark and MPRA data, 5) refine GWAS peaks with caQTLs, 6) experimentally test prioritized SNPs. caQTLs appear much narrower than many other GWAS signals[29]; however, of immune cells, they are currently only available for B-cells. As such, we placed them late in our pipeline, so that the initial prioritization covers all cell types. As caQTL data becomes more widely available for other cell types, placing this step earlier in the pipeline could narrow GWAS peaks sooner.

eQTLs

We first annotated all SNPs with *cis*- and *trans*-eQTLs and associated target genes, using only immune cell-specific data. Most SNPs (9,052) have ≥1 significant *cis*-eQTL [range 0-31; 856 SNPs have single *cis*-eQTLs and 5,539 have ≤5 *cis*-eQTLs] (**Supplementary Table 2**). *cis*-eQTL targets are enriched in immune-related genes, with many being known SLE risk loci. In

LOC_13, rs17849501 (Neutrophil cytosol factor 2, *NCF2*) is an eQTL of several genes in multiple immune cell types. We experimentally demonstrated strong, allele-dependent enhancer activity of this SNP[54]. In LOC_66, rs2431697 (intergenic) affects expression of multiple genes across cell types. This SNP has been experimentally shown to physically associate with the promoter of miRNA-146a (10.1038/s41467-020-20460-1), a potent immune regulator[55] and SLE biomarker[56]. In LOC_76, SLE risk SNP rs2230926 (Tumor necrosis factor, alpha-induced protein 3, *TNFAIP3*) greatly increases neutrophil extracellular traps and citrullinated epitopes in SLE patients[57]. In LOC_83, rs13239597 (intergenic) is an experimentally-validated allele-specific enhancer of Interferon regulatory factor 5 (*IRF5*)[58], a key SLE risk gene.

For *trans*-eQTLs (having target genes >1 Mb or on another chromosome), we identified 75 SNPs from 22 loci targeting 272 unique genes (range 1-149 per locus; 20 out of 22 *trans*-eQTLs had 1-11 target genes) with FDR <1e-5 (**Supplementary Tables 8, 9**). Among them, 13 target genes were distal, and 259 on different chromosomes. Among 75 *trans*-eQTL SNPs, 73 were also identified as *cis*-eQTLs. At LOC_121 (*SH2B3*, *ATXN2*), all 8 *trans*-eQTL SNPs showed >100 target genes, demonstrating substantial interactions across the genome. Consistent with this, *SH2B3* (*a.k.a.* lymphocyte adaptor protein, *LNK*) links numerous immune signaling pathways to inflammation[59] and is a major immune regulator. LOC_79 (*IKZF1*) had one *trans*-eQTL (rs4917014) with 50 target genes. Many target genes were themselves immune-related and often SLE-associated. For example, rs1990760, a regulatory coding SNP at *IFIH1* (LOC_31), is defined for lupus susceptibility[60]. This SNP is also a *trans*-eQTL targeting nine genes (*MX1, IFI44L/IFI44, HERC5, IFIT1/IFI6, OAS3/OAS2, HERC6*) significantly enriched in immune response network and type I/II interferon signaling genes. Interestingly, seven (*MX1, IFI44L/IFI44, HERC5, IFIT1/IFI6, OAS3*) and four (*MX1, IFI44L/IFI44, HERC5)* target genes were also targeted by LOC_97 and LOC_99, respectively (**Supplementary Table 8**), suggesting that coregulation of core genes further amplifies *trans*-effects in an omnigenic model[61].

Chromatin interactions

We performed an independent analysis using PCHiC data on immune cell lines[23,32]. The 6,198 SNPs had ≥1 PCHiC connection (762 SNPs had 1; 3,322 SNPs had ≤5; maximum 93).

Combining eQTL and PCHiC datasets, our SNPs target 3,504 unique genes (**Supplementary Tables 2, 4**).

SNP categorization

Concordance between eQTL and PCHiC annotation suggests that a given SNP has a strong regulatory role; thus, we based SNP tiers on this intersection (**Figures 1-2, Supplementary Table 4**). Tier1 includes SNPs annotated by both methods with non-zero target gene overlap. These SNPs (3,746 from 143 loci) have strong evidence of controlling expression of specific target genes. 1,906 SNPs (17 loci; Tier2) were annotated by both methods but targeted different genes. Tiers 3a and 3b (546 SNPs, 11 loci, 3,400 SNPs, 6 loci) showed either PCHiC or eQTL activity, respectively, but not both. Finally, 333 (Tier4) exhibited neither activity. Of 9,052 *cis*-eQTL SNPs, 3,746, 1,906, and 3,400 were categorized as Tier1, Tier2, and Tier3b, respectively. Of 75 *trans*-eQTL SNPs, 50, 11, and 14 were Tier1, Tier2, and Tier3b, respectively.

Regulatory elements

Linked SNPs were closely associated with transcriptional regulatory regions annotated by GenoSTAN and other databases; 4,332 (43.6%) lie in annotated promoter, enhancer, and/or silencer regions (**Supplementary Table 2**). Of 9,059 eQTL SNPs, 3,457 (38.1%) lie in enhancers, 625 (6.9%) in promoters, 485 (5.4%) in both, and 670 (7.3%) in silencers. We observed median 13 transcriptional element-associated SNPs per locus (4 loci had no such SNPs; LOC_71 had 360). The bulk were Tier1/Tier2 SNPs, indicating a relationship between transcriptional regulatory elements and eQTL/PCHiC activity. Enhancer SNPs that are also eQTL SNPs had a median distance of 47.2 kb to their target genes' transcription start sites (TSS); for Tier1 SNPs, this distance was 45.0 kb. Enhancer SNPs that are also PCHiC SNPs had a median distance of 214.5 kb to their target genes' TSS; for Tier1 SNPs, this distance was 193.3 kb (**Supplementary Table 1, 4)**. Tier1 SNPs are substantially closer to their target genes than other tiers, consistent with strong regulatory effects.

Of all regulatory element-associated SNPs, 117 (from 32 loci) were Tier1 SNPs with 1-4 common target genes, leading to 58 unique genes targeted in both eQTLs and PCHiC. Of enhancer SNPs, 93 (26 loci) were Tier1, together targeting 44 unique genes (**Supplementary**

Table 2). These SNPs, which are in annotated enhancers, are involved in chromatin interactions, and transcriptionally regulate specific target genes, represent highly prioritized candidates and are given further attention below.

Massively parallel reporter assays

As an independent measure of SNP effects on transcription, we mined massively parallel reporter assay (MPRA) datasets, which characterize enhancers in high-throughput[62]. We examined MPRA data from B-cells (GM12878)[51]. A total of 2,614 SNPs appeared in this dataset, and 42 out of 51 significant ones showed allele-specific expression (ASE; FDR<0.01; **Supplementary Tables 12, 13**). MPRA-ASE SNPs were overwhelmingly non-coding: 50 intergenic, 46 intronic, 1 synonymous, 1 missense.

Deleteriousness scores

We annotated SNPs with pre-computed deleteriousness scores (predictSNP2, CADD, GWAVA)[45]. Of exonic SNPs (177 in 61 loci, 91 unique protein-coding genes; 89 missense from 43 loci, 57 unique genes), the algorithms identified 11, 26, and 37 deleterious SNPs, respectively. For missense SNPs, 9, 17, and 37, respectively, were labeled deleterious. For non-coding SNPs, 516 (55% intronic, 45% intergenic) were identified as deleterious by at least one algorithm (**Supplementary Table 3**).

Chromatin accessibility

We next annotated our SNPs according to two measures of chromatin accessibility: DNase hypersensitivity and ATAC-seq (whole blood). As before, Tier1 had by far the largest signals, followed by Tier2 and 3a (**Supplementary Figure 2**). Tiers 3b and 4 showed essentially zero enrichment.

caQTL SNPs

To identify SNPs with allele-specific chromatin accessibility, we searched a caQTL database from lymphoblastoid (B-cell) cell lines (LCLs) from ten ethnicities[29]. caQTL peaks are quite narrow[63]; however, the method is new and of immune cells, has only yet been applied to LCLs. Thus, although the technique dramatically reduces SNP numbers, these results are specific to B-

cells. SLE, of course, manifests through numerous cell types; this analysis is only a subset of associated SNPs. As caQTLs are determined in more cell types, this analysis can be extended.

Of our SNPs (covering 100 loci), 295 appeared as caQTLs in ≥1 ancestry. Among our 182 loci, 100 had ≥1 caQTL SNP (range 1-16); 73 loci had ≥1 Tier1 caQTL SNP. All but one caQTL SNP were also eQTL SNPs. Of 295 caQTL SNPs, 194 are Tier1, 46 Tier2, 6 Tier3a, 48 Tier3b, and 1 Tier4 (**Supplementary Table 11**). Thus, caQTL SNPs are heavily enriched in high-tier SNPs, illustrating that SNP-driven changes in chromatin accessibility strongly contribute to downstream expression and chromatin interaction phenotypes. Of 295 caQTL SNPs, 235 (79.7%) lie in enhancers, 91 (30.8%) in promoters, 63 (21.4%) in both, and 19 (6.4%) in silencers. This is consistent with eQTL and MPRA data, although caQTL SNPs are substantially more enriched with enhancer SNPs.

Transcription factor binding

Next, we independently annotated transcription factor (TF) binding sites using epiCOLOC[52]. Tier1 SNPs showed by far the most TFs (89) with binding site enrichment (**Supplementary Figure 3**), with Tier2 next. Tier3a showed small enrichment, and Tiers 3b and 4 were negligible. TFs highly represented in Tier1/Tier2 SNPs include Brachyury/TBXT, TCF4, MYB, and NFKB1–all critical immune-linked proteins. In fact, these TFs have clear involvement in SLE pathogenesis[64-69]. Altogether, TFBS enrichment strongly correlates with eQTL/PCHiC activity, and enriched TFs were immune-linked and SLE-associated.

Tissue expression

Next, for our collected loci, we tabulated expression in diverse tissues using FUMA GENE2FUNC. Tier1 loci target genes were significantly enriched (FDR <0.001) in whole blood and lymphocytes (**Supplementary Figure 4, Supplementary Table 5**). As before, lower tiers were much less enriched in these tissues and demonstrated less tissue enrichment overall.

Disease and pathway association

We searched disease GWAS association catalogs; Tier1 loci target genes were significantly overrepresented in 154 out of 310 traits/diseases. SLE, rheumatoid arthritis (RA), and

inflammatory bowel disease (IBD) were particularly enriched in GWAS hitting these loci. Lower tiers were much less linked to GWAS hits. Similarly, Tier1 loci target genes were highly enriched among KEGG pathways (36 out of 68) and gene ontology (GO) classifications (464 out of 1,374), whereas lower tiers were not. Tier1 loci-associated pathways included immune system regulation, cytokine production, phosphorus metabolism, and regulation of protein modification and interferon signaling. Further studies are required to flesh out exact pathways and mechanisms by which these highly associated SNPs contribute to dyshomeostasis and SLE progression; these results will prioritize avenues for experimental investigation.

Missense SNPs

We highlight several missense SNPs predicted to dramatically disrupt protein function. rs78555129 mutates a universally conserved arginine in adipolin (CTRP12/FAM132A/C1QTNF12) to cysteine, perturbing protein folding and presumably interactions with its (currently unknown) receptor (**Supplementary Figure 5a**). Adipolin is an anti-inflammatory adipokine implicated in diabetes, arthritis, and obesity[70]. In B-cell scaffold protein with ankyrin repeats 1 (BANK1), rs10516487 destabilizes the protein (**Supplementary Figure 5b**), likely interfering with its interactions with TRAF6 and MyD88 in innate immune signaling[71]. rs201802880 in Neutrophil Cytosolic Factor 1 (NCF1/p47phox) mutates a universally conserved residue (**Supplementary Figure 5c**), leading to protein destabilization. NCF1 is a subunit of NADPH oxidase, which produces reactive oxygen in the phagosome, critical for immune responses[72]. rs2230926 in TNFα-Induced Protein 3 (TNFAIP3) mutates a universally conserved residue important for protein stability (**Supplementary Figure 5d**). TNFAIP3 is indispensable to TNF signaling and immune activation and is an SLE risk gene[73].

CRISPR-based validation of rs57668933

To validate our approach, we used CRISPRa and CRISPRi targeting the rs57668933 locus for activation and inhibition, respectively (**Figure 3e**). Both activation domains doubled *ELF1* expression levels, and both suppressor domains halved it. Thus, rs57668933 is indeed within a strong enhancer, consistent with our other data.

Discussion

We have established a state-of-the-art SNP and locus analysis pipeline for assimilating data regarding gene expression, chromatin accessibility and interactions, histone marks, transcription factor binding, tissue expression, and disease association. Our analysis pipeline dramatically reduces large sets of associated SNPs to several likely causal SNPs for experimental validation. This pipeline will be useful for diverse genetic association studies.

After carefully gathering all high-quality SLE GWAS and candidate gene studies up to September 2021 and their high-LD SNPs from 1000Genomes Project Phase3, we defined 182 statistically-independent, non-HLA loci totaling ~10,000 SNPs. Our analysis first focused on SNPs with effects on gene expression; unsurprisingly, these SNPs were overwhelmingly non-coding in nature, and very often localized to enhancer regions. We also found many missense SLE-associated SNPs. These SNPs had high deleteriousness scores from other pipelines; in fact, 30% of the most deleterious SNPs were missense, compared to 1.7% of all SNPs. This dramatic enrichment supports their involvement; it should be noted, though, that CADD and other programs generally view missense SNPs as more deleterious. In further support, our molecular modeling showed that many missense SNPs adversely affect protein structure and/or function.

Intriguingly, we found several examples of SNPs encompassing both effects: they were simultaneously missense SNPs with adverse predicted effects on protein function, and enhancer SNPs affecting expression of multiple other genes. This is perhaps unsurprising, as exonic transcription factor binding sites and enhancers are common[43]. For instance, we again found rs1143679, mutating a key residue of integrin alpha M (ITGAM) and disrupting multiple transcription factor binding sites, dramatically weakening enhancer activity[74].

Only 45 loci contained missense SNPs; most SNPs were non-coding. Our pipeline prioritized SLE risk SNPs and tiered them according to target gene expression (eQTL) and chromatin interactions (PCHiC). We obtained 3,746 Tier1 SNPs, where the two independent experiments identified common regulated genes. Of these, 1,913 are also enhancer-SNPs. Overall, 100 loci had ≥1 caQTL SNP (total 295), and 22 loci had ≥1 allele-specific enhancer SNP (total 42). Together, these constitute 106 out of 182 total SLE loci (329 total SNPs) flagged by ≥3 independent experimental methods regarding gene regulation: eQTL-chromatin interaction-

chromatin accessibility (B-cells) or eQTL-chromatin interaction-enhancer histone marks (**Table 1, Supplementary Table 2**). Adding MPRA data (GM12878 cells), yielded a final set of six loci (6 SNPs; Table 2) flagged by all available experimental methods with highly significant changes in eQTLs, chromatin accessibility, and target gene expression. These SNPs are predicted to be highly associated with SLE, with effects manifested through enhancer-driven alteration of target gene expression, and mediated at least somewhat through B-cells.

We examined these SNPs in detail. rs57668933 (intron of lymphoid cell transcription factor E74-like factor 1, *ELF1*), at LOC_125 (Chr 13), controls *ELF1* expression (**Figure 3a-b**). The protective T allele correlates with higher *ELF1* expression in T-cells, B-cells, and monocytes in healthy controls (**Supplementary Figure 7**) and shows high allele-specific chromatin accessibility (**Figure 3c**) and enhancer activity (**Figure 3d**). *ELF1* has been previously reported as an SLE risk gene (lead SNP rs7329174[75])–we show that rs57668933 is instead the likely causal SNP, with the risk allele yielding lower chromatin accessibility and *ELF1* expression. ELF1 represses FcRγ expression[76]; SLE patients' T-cells express essentially no ELF1 but high levels of FcRγ, which activates immune reactivity and promotes nephritis[77]. In B-cells, ELF1 regulates antibody heavy chain production[78]. This SNP disrupts universally conserved binding sites for the tumor suppressors p63 and p73 (**Supplementary Figure 6a**), both with strong immune contributions[79].

All six SNPs show much experimental evidence linking them to SLE. Most disrupt highly conserved binding sites of critical immune transcription factors (**Supplementary Table 14, Supplementary Figure 3**). Target genes and disrupted transcription factors are known autoimmune risk genes, implicated in multiple diseases (**Supplementary Table 5**). Many selected SNPs are far from index SNPs and do not appear in the literature, highlighting the pipeline's ability to localize signals in large GWAS peaks.

Beyond the six most highly selected SNPs (Table2), our Tier1 hits and associated targets were very strongly enriched in immune-related genes. High-tier SNPs were also greatly enriched in SNPs flagged as deleterious by other methods. Overall, putative risk loci and target genes were overwhelmingly enriched in immune genes, with many being known risk for SLE, rheumatoid

arthritis, systemic sclerosis, Crohn's disease, Sjögren's syndrome, primary biliary cholangitis, and particularly inflammatory bowel disease. We experimentally validated a high-priority SNP with CRISPR/Cas9 gene activation/silencing, confirming that this site does indeed have dramatic enhancer activity, likely underlying SLE association. This experimental support for SNPs and loci prioritized by our analysis supports its utility in selecting likely underlying SNPs from GWAS peaks.

Our study has two major limitations: 1) Sparse MPRA data means that some loci might have unflagged causal variants, 2) Sparse caQTL data limits the strongest conclusions to B-cells at this time. Generation of more MPRA data and caQTL data in other cell types will refine existing loci, add further ones, and extend applicability to more diseases.

In conclusion, we demonstrate and validate strategies, and provide a pipeline that will be useful for diverse post-GWAS studies. The SLE dataset we generated will serve as a roadmap for future studies verifying causality and establishing underlying biochemical mechanisms.

Table 1. Summary of all SLE loci.

Locus	CHR	Boundaries	SNP	Index/LD	Nearest Gene
LOC_1	1	1156655 - 1191870	rs6697886	1/104	C1QTNF12
LOC_2	1	8431607 - 8505058	rs301807	1/10	RERE
LOC_3	1	24518206 - 24519920		1/1	IFNLR0
LOC_4	1	38258007 - 38379018	rs28469609	1/43	MTF1
LOC_5	1	67787691 - 67891029	rs11209064	2/55	IL12RB2
LOC_6	1	114303808 - 114377568		2/0	PTPN22
LOC_7	1	117040622 - 117104215	rs10924104	2/40	CD58; NAP1L4P1
LOC_8	1	157108159 - 157119915	rs116785379	1/2	ETV3
LOC_9	1	157486336 - 157538786	rs34273689	1/90	FCRL5
LOC_10	1	161469054 - 161596283		3/30	FCGR2A; FCGR2C
LOC_11	1	173177392 - 173376184	rs6664517	13/157	AL645568.1
LOC_12	1	174396030 - 174923045	rs72717613	1/420	RABGAP1L
LOC_13	1	183225237 - 183591098	rs10911363	7/39	NCF2
LOC_14	1	184636486 - 184723135		1/39	EDEM3
LOC_15	1	192513661 - 192544795	rs2984920	1/34	AL390957.1
LOC_16	1	198543027 - 198670469		2/116	PTPRC
LOC_17	1	201977073 - 201986311		1/0	ELF3
LOC_18	1	206642539 - 206647450		2/7	IKBKE
LOC_19	1	206939904 - 206955041	rs3024493	1/3	IL10; IL19
LOC_20	1	235890096 - 236041129	rs4660117	1/27	LYST
LOC_21	1	246434447 - 246444082		1/3	SMYD3
LOC_22	2	7573079 - 7584668		1/24	AC013460.1
LOC_23	2	30442402 - 30492116	rs906866	3/48	LBH
LOC_24	2	33701890 - 33702203	rs13385731	2/0	RASGRP3
LOC_25	2	61040651 - 61173382		1/11	LINC01185
LOC_26	2	65559027 - 65667272	rs1876518	3/58	SPRED2
LOC_27	2	74200833 - 74219948		3/24	TET3
LOC_28	2	111868604 - 111940585	rs12613243	1/20	BCL2L11
LOC_29	2	136555659 - 136761853	rs2278682	3/75	LCT; MCM6
LOC_30	2	144013184 - 144028568	rs10153706	1/28	ARHGAP15
LOC_31	2	163025929 - 163211491		5/244	FAP; IFIH1
LOC_32	2	191399581 - 191434502		1/0	AC108047.1
LOC_33	2	191900449 - 191973034	rs7574865	8/38	STAT4
LOC_34	2	198492316 - 198954774	rs13034353	2/216	PLCL1
LOC_35	2	204690355 - 204738919		1/52	CTLA3
LOC_36	2	213585035 - 213593970		1/3	AC093865.1
LOC_37	2	213862922 - 213890232		1/9	IKZF2
LOC_38	3	28068394 - 28079260	rs1813375	1/15	LINC01967
LOC_39	3	58261741 - 58473899		4/56	PXK
LOC_40	3	72200387 - 72256927	rs7637844	1/0	LINC00870
LOC_41	3	119111870 - 119272391	rs9877891	7/22	TIMMDC1; CD80
LOC_42	3	159625393 - 159748367	rs2936303	3/62	IL12A-AS1
LOC_43	3	169476991 - 169528523	rs3821383	3/47	LRRC34
LOC_44	3	188451078 - 188472383	rs1568669	1/11	LPP
LOC_45	4	953193 - 983809	rs11248061	3/16	DGKQ
LOC_46	4	2540146 - 2760732	rs4690053	2/126	FAM193A
LOC_47	4	8558199 - 8568191		1/11	GPR78
LOC_48	4	40301264 - 40308368	rs13136820	1/4	LINC02265
LOC_49	4	55547533 - 55553801		1/10	KIT
LOC_50	4	79626160 - 79679733		1/31	AC112253.1
LOC_51	4	84141253 - 84161920	rs4693592	1/51	AC114781.2
LOC_52	4	87888054 - 87976055	rs340643	2/50	AFF1
LOC_53	4	102712542 - 102762581	rs6811141	6/70	BANK1
LOC_54	4	108968701 - 109090112		1/0	LEF1
LOC_55	4	123073009 - 123551032		2/76	ADAD1; IL21
LOC_56	4	184603297 - 184618470		1/5	TRAPPC11
LOC_57	5	1282319 - 1286516		2/6	TERT
LOC_58	5	35850149 - 35916174		1/10	IL7R; CAPSL
LOC_59	5	100084878 - 100291657	rs10060686	3/216	ST8SIA4
LOC_60	5	127733961 - 127853142		1/15	FBN2
LOC_61	5	130665788 - 131259361		1/4	FNIP1
LOC_62	5	131812897 - 131835395	rs61175929	1/60	IRF1
LOC_63	5	133418739 - 133433641		3/21	AC008608.1
LOC_64	5	150386395 - 150462638	rs10036748	5/26	GPX3; TNIP1
LOC_65	5	158883027 - 158944457		1/6	LINC01845
LOC_66	5	159879978 - 159887336	rs2431697	2/0	MIR3142HG
LOC_67	6	238790 - 259719		2/24	AL035696.1
LOC_68	6	16299343 - 16761722		1/0	ATXN1
LOC_69	6	25184408 - 26339131	rs17598658	4/93	CARMIL1; H2BC6
LOC_70	6	27498217 - 27665920	rs10807029	1/20	CD83P1
LOC_71	6	34549107 - 35356143	rs6934662	8/794	PPARD
LOC_72	6	36695519 - 36722789	rs236469	1/26	CPNE5
LOC_73	6	90936894 - 91002494	rs614120	1/13	BACH2
LOC_74	6	106564236 - 106598933		3/15	ATG5
LOC_75	6	116690849 - 116694120		2/0	DSE
LOC_76	6	137959235 - 138243739	rs200820567	9/49	TNFAIP3
LOC_77	6	154562302 - 154579861	rs2141289	1/15	AL357075.4
LOC_78	7	28142088 - 28209953	rs702814	3/15	JAZF1
LOC_79	7	50227828 - 50348043	rs876039	6/26	IKZF1
LOC_80	7	67014434 - 67084823		1/36	MTATP6P21
LOC_81	7	73434106 - 74193642		5/37	GTF2IRD1
LOC_82	7	75167934 - 75209951		6/21	HIP1
LOC_83	7	128563721 - 128764737	rs3778752	14/114	IRF5; TNPO3
LOC_84	8	8088230 - 8155475	rs2945248	2/44	ALG1L13P
LOC_85	8	8622877 - 8649881	rs2428	1/30	MFHAS1
LOC_86	8	10712945 - 10802146	rs6985109	5/55	XKR6
LOC_87	8	11270993 - 11402063	rs67934857	13/84	AF131216.5; BLK
LOC_88	8	42128820 - 42189978		1/0	IKBKB
LOC_89	8	56835673 - 57044066	rs189658553	2/198	LYN; RPS20
LOC_90	8	71017438 - 71330166	rs71517442	2/95	NCOA2
LOC_91	8	72891748 - 72913114	rs9298192	1/14	MSC-AS1
LOC_92	8	79555186 - 79657666	rs3808619	2/65	ZC2HC1A; IL7
LOC_93	8	128192981 - 128197856	rs2456452	1/11	CASC19
LOC_94	8	129324232 - 129465024		2/50	LINC00824
LOC_95	9	4981602 - 4984530		1/1	JAK2

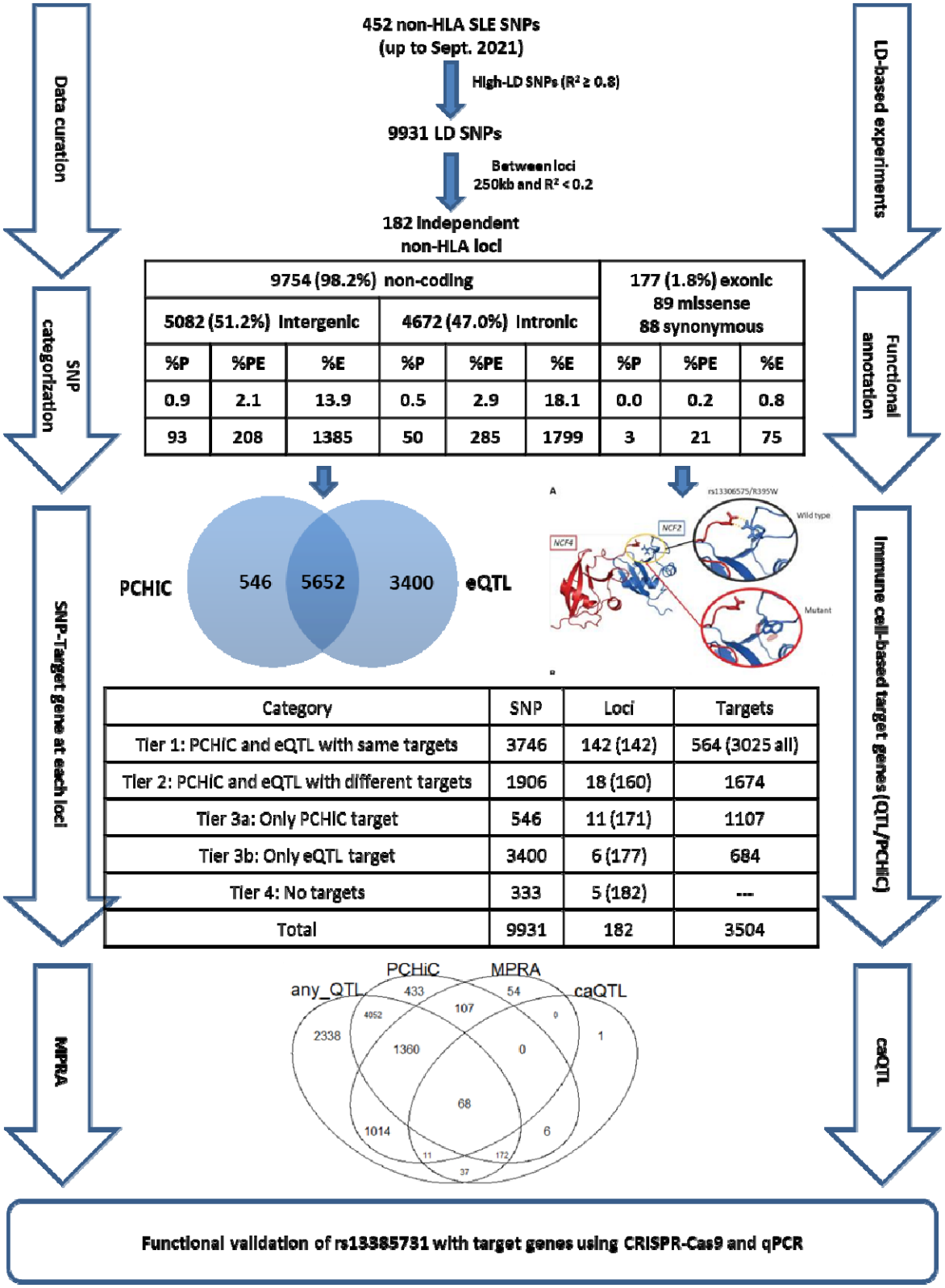
Locus	CHR	Boundaries	SNP	Index/LD	Nearest Gene
LOC_96	9	21171267 - 21320324	rs10757201	2/147	IFNA22P
LOC_97	9	102337143 - 102605963	rs1405209	2/64	NR4A3
LOC_98	10	5894714 - 5914581		1/12	ANKRD16
LOC_99	10	50014917 - 50122181	rs7086101	5/121	WDFY4
LOC_100	10	63785089 - 63825807	rs56140430	2/21	ARID5B
LOC_101	10	64399617 - 64443139	rs2393909	2/24	AC067752.1
LOC_102	10	73466709 - 73506129	rs3802712	2/34	CDH23
LOC_103	10	104973061 - 105175131		1/25	NT5C2; INA
LOC_104	10	105671683 - 105700775		1/5	STN1
LOC_105	10	112633671 - 112799757	rs73343848	1/27	BBIP1
LOC_106	11	551235 - 635569	rs59115876	5/87	IRF7; CDHR5
LOC_107	11	3875757 - 4114440		1/0	STIM1
LOC_108	11	18303597 - 18362382		1/1	HPS5; GTF2H1
LOC_109	11	35070068 - 35123574	rs2785201	5/51	PDHX
LOC_110	11	65378028 - 65564926	rs10791824	4/31	AP5B1; OVOL1
LOC_111	11	68814887 - 68869034	rs7942690	2/35	TPCN2
LOC_112	11	71132868 - 71225082	rs11606611	1/66	NADSYN1
LOC_113	11	72499768 - 72895102		3/6	FCHSD2
LOC_114	11	118480115 - 118735476	rs2508573	4/42	DDX6
LOC_115	11	128297318 - 128504173	rs12576753	6/17	ETS1
LOC_116	12	4134873 - 4152163		1/7	AC084375.1
LOC_117	12	12760658 - 12874462	rs12811932	5/44	CDKN1B
LOC_118	12	43130547 - 43200941		1/9	LINC02450
LOC_119	12	102271358 - 102405908		1/0	DRAM1
LOC_120	12	103912112 - 103965115		1/0	AC084364.3
LOC_121	12	111826477 - 112059557		4/8	ATXN2
LOC_122	12	121099302 - 121378566	rs904628	2/141	CABP1
LOC_123	12	129276658 - 129307699	rs35907548	7/71	SLC15A4
LOC_124	12	133038182 - 133042182		1/0	AC079031.2
LOC_125	13	41529773 - 41588832	rs57668933	2/15	ELF1
LOC_126	13	50143361 - 50192528		1/20	RCBTB1
LOC_127	13	100084039 - 100104407	rs749114	1/13	TM9SF2
LOC_128	14	35831811 - 35832666		1/1	AL133163.2
LOC_129	14	68728425 - 68760141	rs3784099	2/14	RAD51B
LOC_130	14	88370343 - 88383035	rs28626750	1/13	GALC
LOC_131	14	103238582 - 103290221	rs12880641	1/62	TRAF3
LOC_132	14	105386039 - 105416010	rs2819426	3/51	PLD4; AHNAK2
LOC_133	15	38728250 - 38927386	rs7173565	4/10	RASGRP1
LOC_134	15	75079474 - 75392795	rs34180494	3/23	CSK; SCAMP5
LOC_135	15	77824646 - 77830430	rs1317320	1/19	AC046168.1
LOC_136	15	97595545 - 97626101		1/9	AC055873.1
LOC_137	15	101529012 - 101550214		1/1	LRRK1
LOC_138	16	11038360 - 11291722	rs2041670	7/82	CLEC16A
LOC_139	16	23871457 - 23901376		2/1	PRKCB
LOC_140	16	30584430 - 30827205	rs3812999	2/88	PRR14; RNF40
LOC_141	16	31260235 - 31369803	rs4632147	7/100	ITGAM; ITGAX
LOC_142	16	50068422 - 50139799		1/67	HEATR3
LOC_143	16	57352124 - 57403500	rs9921681	3/13	CCL22

Locus	CHR	Boundaries	SNP	Index/LD	Nearest Gene
LOC_144	16	58247523 - 58268561	rs2731741	1/70	CCDC113
LOC_145	16	68551277 - 68663156	rs28537207	3/197	ZFP90; RNU4-36P
LOC_146	16	79739978 - 79755446		1/25	MAFTRR
LOC_147	16	85966683 - 86020039	rs8052690	6/42	AC092723.4
LOC_148	16	87390630 - 87443734	rs10431963	1/26	MAP1LC3B
LOC_149	17	4706123 - 4712617		1/4	PLD2
LOC_150	17	7208373 - 7240391		2/8	ACAP1
LOC_151	17	16839901 - 16845467		2/1	TNFRSF13B
LOC_152	17	37885383 - 38088150	rs34758895	7/245	MIEN1; IKZF3
LOC_153	17	43422855 - 43457886		1/5	RNA5SP443
LOC_154	17	47448102 - 47554350	rs2671655	1/0	AC091180.5
LOC_155	17	73304710 - 73417662	rs8072449	3/158	GRB2
LOC_156	17	76372972 - 76393736		1/5	PGS1
LOC_157	18	67518031 - 67562657	rs1788103	3/36	CD226
LOC_158	18	77377925 - 77386912	rs118075465	1/7	AC068473.4
LOC_159	19	936297 - 952429	rs2238580	1/16	ARID3A
LOC_160	19	2131148 - 2208859	rs2864419	1/49	DOT1L
LOC_161	19	6689065 - 6699330		1/31	C3
LOC_162	19	10392638 - 10481532	rs2569693	5/20	TYK2
LOC_163	19	16438661 - 16443718	rs11086029	1/9	KLF2
LOC_164	19	18383794 - 18637194	rs28375303	3/100	IQCN; SSBP4
LOC_165	19	33035097 - 33106621		2/21	PDCD5
LOC_166	19	49788205 - 49918814	rs7257053	2/45	SLC6A16; TEAD2
LOC_167	19	50162909 - 50182697		1/2	IRF3
LOC_168	19	52021247 - 52127053		2/30	SIGLEC6
LOC_169	19	55730976 - 55739813		2/15	TMEM86B
LOC_170	20	1507507 - 1558508		1/27	AL049634.1
LOC_171	20	44730245 - 44749251		1/11	CD40
LOC_172	20	48429020 - 48605930	rs117447227	1/92	RNF114
LOC_173	22	18648861 - 18654105		1/19	USP18
LOC_174	22	21798351 - 21985094	rs1034329	7/110	UBE2L3; YDJC
LOC_175	22	39739187 - 39756650	rs2069235	2/9	SYNGR1
LOC_176	22	40291139 - 40317126		1/13	GRAP2
LOC_177	X	12839152 - 12907658		3/3	PRPS2
LOC_178	X	30572729 - 30577846		1/5	CXorf21
LOC_179	X	53081414 - 53111428		1/37	GPR173
LOC_180	X	56295245 - 57406814	rs5913948	2/1146	KLF8; NBDY
LOC_181	X	149663590 - 149673253		1/5	MAMLD1
LOC_182	X	153189819 - 153378375	rs3027878	11/95	IRAK1; MECP2

Table 2. Six most significant SNPs and their target genes.

rsID	CHR	Risk/Non-Risk	Closest Gene	Common target genes
rs13385731	2	T/C	RASGRP3	RASGRP3, FAM98A
rs2936303	3	G/A	IL12A-AS1	IL12A, TRIM59
rs10036748	5	T/C	TNIP1	TNIP1, ANXA6, GPX3
rs2431697	5	T/C	MIR3142HG	PTTG1, SLU7
rs57668933	13	C/T	ELF1	ELF1
rs2069235	22	A/G	SYNGR1	PDGFB, MGAT3, RPL3

Figure 1. Study framework and summary. Tier 1: PCHiC and QTL with at least one same target gene, Tier 2: PCHiC and QTL with different targets, Tier 3a: Only PCHiC target, Tier 3b: Only



QTL target, Tier 4: No targets

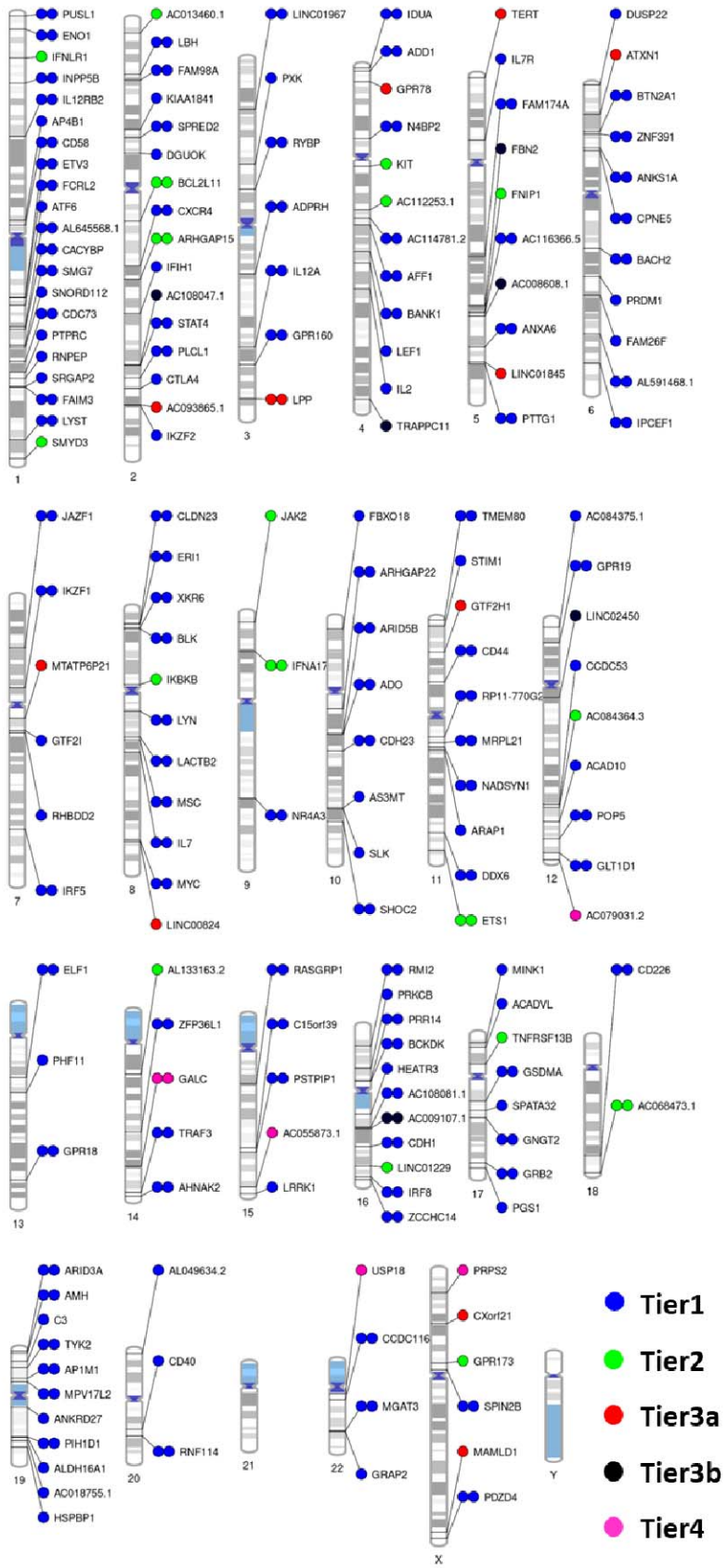
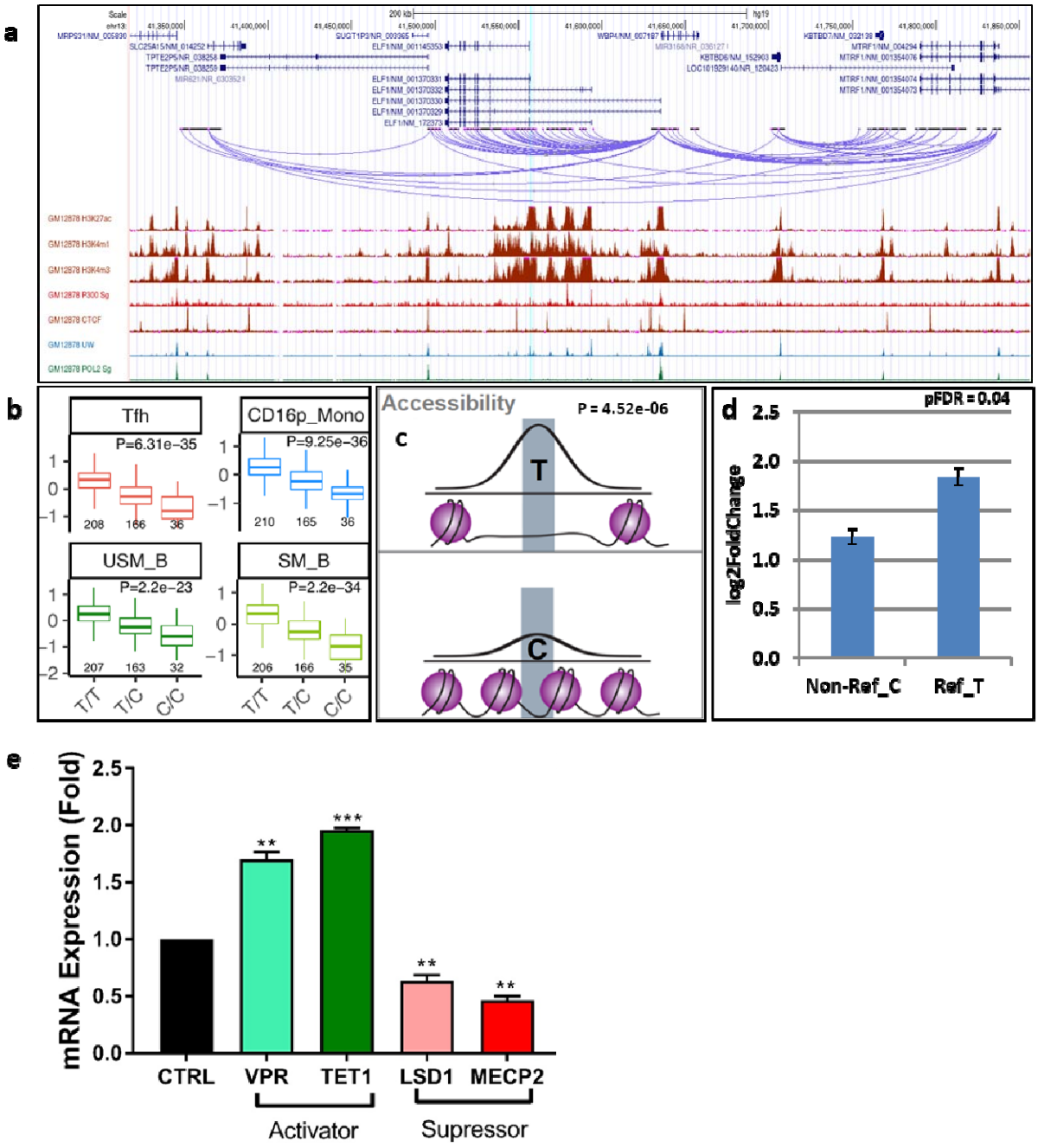


Figure 2. Distribution of 182 non-HLA SLE loci across the human genome. Loci colored by highest SNP tier. Tier1 names are common target genes from both eQTL and PCHiC data; other tiers are named by closest positional gene. Loci with double dot have ≥1 significant experimentally validated (caQTL or MPRA) allele-specific SNP. Single dots mean no experimentally validated SNPs are yet known.

Figure 3. Analysis of LOC_125. This locus has two index SNPs and 15 high-LD SNPs. Of these, rs57668933 is a caQTL-SNP and has allele-specific enhancer activity. a) Visualized connections of rs57668933 and neighboring regions based on PCHiC, alongside various histone marks. b) The SNP lies in *ELF1*, a significant eQTL gene. Significant genotype-specific gene expression in follicular T-helper cells (Tfh), CD16^{+} monocytes, and unstimulated and stimulated B-cells; c) chromatin accessibility of alleles. Reference allele T has higher chromatin accessibility. d) MPRA data shows T has higher enhancer activity. e) CRISPR-dCas9-based targeting of rs57668933-containing region by two activators (dCas9-VPR and dCas9-TET1) and two suppressors (dCas9-MECP2, dCas9-LSD1) and ELF1 (target gene) mRNA expression.



REFERENCES

1 Alarcon-Segovia, D. *et al.* Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. *Arthritis Rheum* **52**, 1138-1147 (2005). https://doi.org:10.1002/art.20999

2 Kuo, C. F. *et al.* Familial Aggregation of Systemic Lupus Erythematosus and Coaggregation of Autoimmune Diseases in Affected Families. *JAMA Intern Med* **175**, 1518-1526 (2015). https://doi.org:10.1001/jamainternmed.2015.3528

3 Sestak, A. L., Shaver, T. S., Moser, K. L., Neas, B. R. & Harley, J. B. Familial aggregation of lupus and autoimmunity in an unusual multiplex pedigree. *J Rheumatol* **26**, 1495-1499 (1999).

4 Deapen, D. *et al.* A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis Rheum* **35**, 311-318 (1992).

5 Lawrence, J. S., Martins, C. L. & Drake, G. L. A family survey of lupus erythematosus. 1. Heritability. *J Rheumatol* **14**, 913-921 (1987).

6 Bentham, J. *et al.* Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet* **47**, 1457-1464 (2015). https://doi.org:10.1038/ng.3434

7 Molineros, J. E. *et al.* Confirmation of five novel susceptibility loci for systemic lupus erythematosus (SLE) and integrated network analysis of 82 SLE susceptibility loci. *Hum Mol Genet* **26**, 1205-1216 (2017). https://doi.org:10.1093/hmg/ddx026

8 Morris, D. L. *et al.* Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. *Nat Genet* (2016). https://doi.org:10.1038/ng.3603

9 Sun, C. *et al.* High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. *Nat Genet* (2016). https://doi.org:10.1038/ng.3496

10 Wang, Y. F. *et al.* Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. *Nat Commun* **12**, 772 (2021). https://doi.org:10.1038/s41467-021-21049-y

11 Yin, X. *et al.* Meta-analysis of 208370 East Asians identifies 113 susceptibility loci for systemic lupus erythematosus. *Ann Rheum Dis* (2020). https://doi.org:10.1136/annrheumdis-2020-219209

12 Langefeld, C. D. *et al.* Transancestral mapping and genetic load in systemic lupus erythematosus. *Nat Commun* **8**, 16021 (2017). https://doi.org:10.1038/ncomms16021

13 Flister, M. J. *et al.* Identifying multiple causative genes at a single GWAS locus. *Genome Res* **23**, 1996-2002 (2013). https://doi.org:10.1101/gr.160283.113. Epub 2013 Sep 4.

14 MacArthur, D. G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469-476 (2014). https://doi.org:10.1038/nature13127

15 Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *Am J Hum Genet* **102**, 717-730 (2018). https://doi.org:10.1016/j.ajhg.2018.04.002

16 Gotoda, T. From Association to Function in the Post-GWAS Era. *J Atheroscler Thromb* **22**, 442-444 (2015). https://doi.org:10.5551/jat.ED017. Epub 2015 Apr 30.

17 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017). https://doi.org:10.1038/s41467-017-01261-5.
18 Cano-Gamez, E. & Trynka, G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front Genet* **11**, 424 (2020). https://doi.org:10.3389/fgene.2020.00424
19 Simovski, B. *et al.* Coloc-stats: a unified web interface to perform colocalization analysis of genomic features. *Nucleic Acids Research* **46**, W186-W193 (2018). https://doi.org:10.1093/nar/gky474
20 Trynka, G. *et al.* Disentangling the Effects of Colocalizing Genomic Annotations to Functionally Prioritize Non-coding Variants within Complex-Trait Loci. *Am J Hum Genet* **97**, 139-152 (2015). https://doi.org:10.1016/j.ajhg.2015.05.016.
21 Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet* **99**, 1245-1260 (2016). https://doi.org:10.1016/j.ajhg.2016.10.003. Epub 2016 Nov 17.
22 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013). https://doi.org:10.1038/nmeth.2688. Epub 2013 Oct 6.
23 Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369-1384 e1319 (2016). https://doi.org:10.1016/j.cell.2016.09.037
24 Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876-880 (2005). https://doi.org:10.1038/nature03877. Epub 2005 Jun 29.
25 Adsera, C. B., Park, Y. P., Meuleman, W. & Kellis, M. Integrative analysis of 10,000 epigenomic maps across 800 samples for regulatory genomics and disease dissection. *bioRxiv* (2019).
26 Doni Jayavelu, N., Jajodia, A., Mishra, A. & Hawkins, R. D. Candidate silencer elements for the human and mouse genomes. *Nat Commun* **11**, 1061 (2020). https://doi.org:10.1038/s41467-020-14853-5
27 Zacher, B. *et al.* Accurate Promoter and Enhancer Identification in 127 ENCODE and Roadmap Epigenomics Cell Types and Tissues by GenoSTAN. *PLoS One* **12**, e0169249 (2017). https://doi.org:10.1371/journal.pone.0169249
28 Broekema, R. V., Bakker, O. B. & Jonkers, I. H. A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open Biol* **10**, 190221 (2020). https://doi.org:10.1098/rsob.190221
29 Tehranchi, A. *et al.* Fine-mapping cis-regulatory variants in diverse human populations. *Elife* **8** (2019). https://doi.org:10.7554/eLife.39595
30 Cavalli, G. & Misteli, T. Functional implications of genome topology. *Nat Struct Mol Biol* **20**, 290-299 (2013). https://doi.org:10.1038/nsmb.2474.
31 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014). https://doi.org:10.1016/j.cell.2014.11.021. Epub 2014 Dec 11.
32 Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet* (2015). https://doi.org:10.1038/ng.3286

33 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012). https://doi.org:10.1038/nature11082.
34 Symmons, O. *et al.* Functional and topological characteristics of mammalian regulatory domains. *Genome Res* **24**, 390-400 (2014). https://doi.org:10.1101/gr.163519.113. Epub 2014 Jan 7.
35 Inoue, F. & Ahituv, N. Decoding enhancers using massively parallel reporter assays. *Genomics* **106**, 159-164 (2015). https://doi.org:10.1016/j.ygeno.2015.06.005
36 The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318 (2020). https://doi.org:10.1126/science.aaz1776
37 Chandra, V. *et al.* Promoter-interacting expression quantitative trait loci are enriched for functional genetic variants. *Nature Genetics* **53**, 110-119 (2021). https://doi.org:10.1038/s41588-020-00745-3
38 Munz, M. *et al.* Qtlizer: comprehensive QTL annotation of GWAS results. *Scientific Reports* **10**, 20417 (2020). https://doi.org:10.1038/s41598-020-75770-7
39 Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-1243 (2013). https://doi.org:10.1038/ng.2756. Epub 2013 Sep 8.
40 Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-1195 (2012). https://doi.org:10.1126/science.1222794
41 Fonseka, C. Y., Rao, D. A. & Raychaudhuri, S. Leveraging blood and tissue CD4+ T cell heterogeneity at the single cell level to identify mechanisms of disease in rheumatoid arthritis. *Current opinion in immunology* **49**, 27-36 (2017). https://doi.org:10.1016/j.coi.2017.08.005
42 Baxter, J. S. *et al.* Capture Hi-C identifies putative target genes at 33 breast cancer risk loci. *Nat Commun* **9**, 1028 (2018). https://doi.org:10.1038/s41467-018-03411-9.
43 Stergachis, A. B. *et al.* Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* **342**, 1367-1372 (2013). https://doi.org:10.1126/science.1243490
44 Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555-3557 (2015). https://doi.org:10.1093/bioinformatics/btv402. Epub 2015 Jul 2.
45 Bendl, J. *et al.* PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions. *PLoS computational biology* **12**, e1004962 (2016). https://doi.org:10.1371/journal.pcbi.1004962
46 Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research* **47**, D766-D773 (2019). https://doi.org:10.1093/nar/gky955
47 GENCODE. *Human Release 36 (mapped to GRCh37) [Internet]*, <https://www.gencodegenes.org/human/release_36lift37.html> (
48 Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **40**, D930-934 (2012). https://doi.org:10.1093/nar/gkr917. Epub 2011 Nov 7.

49 Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* **53**, 1300-1310 (2021). https://doi.org:10.1038/s41588-021-00913-z
50 Ota, M. *et al.* Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell* **184**, 3006-3021.e3017 (2021). https://doi.org:10.1016/j.cell.2021.03.056
51 Lu, X. *et al.* Global discovery of lupus genetic risk variant allelic enhancer activity. *Nat Commun* **12**, 1611 (2021). https://doi.org:10.1038/s41467-021-21854-5
52 Zhou, Y., Sun, Y., Huang, D. & Li, M. J. epiCOLOC: Integrating Large-Scale and Context-Dependent Epigenomics Features for Comprehensive Colocalization Analysis. *Frontiers in Genetics* **11** (2020). https://doi.org:10.3389/fgene.2020.00053
53 Singh B, M. G. P. Z. X. F.-N. M. B. S. C. S. C. T. C. O. Y. C. K. H. K. Y. G. Lupus susceptibility region containing CDKN1B rs34330 mechanistically influences expression and function of multiple target genes, also linked to proliferation and apoptosis. *Arthritis Rheumatol* (2021).
54 Kim-Howard, X. *et al.* Allelic heterogeneity in NCF2 associated with systemic lupus erythematosus (SLE) susceptibility across four ethnic populations. *Hum Mol Genet* **23**, 1656-1668 (2014). https://doi.org:10.1093/hmg/ddt532
55 Rusca, N. & Monticelli, S. MiR-146a in Immunity and Disease. *Mol Biol Int* **2011**, 437301 (2011). https://doi.org:10.4061/2011/437301
56 Wang, G. *et al.* Serum and urinary cell-free MiR-146a and MiR-155 in patients with systemic lupus erythematosus. *J Rheumatol* **37**, 2516-2522 (2010). https://doi.org:10.3899/jrheum.100308
57 Odqvist, L. *et al.* Genetic variations in A20 DUB domain provide a genetic link to citrullination and neutrophil extracellular traps in systemic lupus erythematosus. *Ann Rheum Dis* **78**, 1363-1370 (2019). https://doi.org:10.1136/annrheumdis-2019-215434
58 Thynn, H. N. *et al.* An Allele-Specific Functional SNP Associated with Two Systemic Autoimmune Diseases Modulates IRF5 Expression by Long-Range Chromatin Loop Formation. *J Invest Dermatol* **140**, 348-360.e311 (2020). https://doi.org:10.1016/j.jid.2019.06.147
59 Devallière, J. & Charreau, B. The adaptor Lnk (SH2B3): an emerging regulator in vascular cells and a link between immune and inflammatory signaling. *Biochem Pharmacol* **82**, 1391-1402 (2011). https://doi.org:10.1016/j.bcp.2011.06.023
60 Molineros, J. E. *et al.* Admixture Mapping in Lupus Identifies Multiple Functional Variants within IFIH1 Associated with Apoptosis, Inflammation, and Autoantibody Production. *PLOS Genetics* **9**, e1003222 (2013). https://doi.org:10.1371/journal.pgen.1003222
61 Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022-1034.e1026 (2019). https://doi.org:10.1016/j.cell.2019.04.014
62 Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nature Biotechnology* **30**, 271-277 (2012). https://doi.org:10.1038/nbt.2137
63 Kumasaka, N., Knights, A. J. & Gaffney, D. J. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nature Genetics* **48**, 206-213 (2016). https://doi.org:10.1038/ng.3467

64 Deng, Y. *et al.* Expression characteristics of interferon-stimulated genes and possible regulatory mechanisms in lupus patients using transcriptomics analyses. *EBioMedicine* **70**, 103477 (2021). https://doi.org:10.1016/j.ebiom.2021.103477
65 Gao, M., Wang, C. H., Sima, X. & Han, X. M. NFKB1 -94 insertion/deletion ATTG polymorphism contributes to risk of systemic lupus erythematosus. *DNA Cell Biol* **31**, 611-615 (2012). https://doi.org:10.1089/dna.2011.1389
66 Kitajima, T. *et al.* Histological detection of c-myb and c-myc proto-oncogene expression in infiltrating cells in cutaneous lupus erythematosus-like lesions of MRL/l mice by in situ hybridization. *Clinical Immunology and Immunopathology* **62**, 119-123 (1992). https://doi.org:https://doi.org/10.1016/0090-1229(92)90031-I
67 Nascimento, D. Q. *et al.* Expression of the miR-9-5p, miR-125b-5p and its target gene NFKB1 and TRAF6 in childhood-onset systemic lupus erythematosus (cSLE). *Autoimmunity* **55**, 515-519 (2022). https://doi.org:10.1080/08916934.2022.2128781
68 Sisirak, V. *et al.* Genetic evidence for the role of plasmacytoid dendritic cells in systemic lupus erythematosus. *J Exp Med* **211**, 1969-1976 (2014). https://doi.org:10.1084/jem.20132522
69 Teng, L. *et al.* TGF-β-mediated upregulation of brachyury contributes to constitutively up-regulated type I collagen expression in skin fibroblasts: possible role in systemic sclerosis. *Eur J Dermatol* **29**, 359-365 (2019). https://doi.org:10.1684/ejd.2019.3607
70 Enomoto, T. *et al.* Adipolin/C1qdc2/CTRP12 Protein Functions as an Adipokine That Improves Glucose Metabolism *. *Journal of Biological Chemistry* **286**, 34552-34558 (2011). https://doi.org:10.1074/jbc.M111.277319
71 Georg, I., Díaz-Barreiro, A., Morell, M., Pey, A. L. & Alarcón-Riquelme, M. E. BANK1 interacts with TRAF6 and MyD88 in innate immune signaling in B cells. *Cellular & Molecular Immunology* **17**, 954-965 (2020). https://doi.org:10.1038/s41423-019-0254-9
72 Olsson, L. M. *et al.* A single nucleotide polymorphism in the NCF1 gene leading to reduced oxidative burst is associated with systemic lupus erythematosus. *Annals of the Rheumatic Diseases* **76**, 1607-1613 (2017). https://doi.org:10.1136/annrheumdis-2017-211287
73 Graham, R. R. *et al.* Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet* **40**, 1059-1061 (2008). https://doi.org:10.1038/ng.200
74 Maiti, A. K. *et al.* Combined protein- and nucleic acid-level effects of rs1143679 (R77H), a lupus-predisposing variant within ITGAM. *Hum Mol Genet* **23**, 4161-4176 (2014). https://doi.org:10.1093/hmg/ddu106. Epub 2014 Mar 7.
75 Lessard, C. J. *et al.* Identification of IRF8, TMEM39A, and IKZF3-ZPBP2 as susceptibility loci for systemic lupus erythematosus in a large-scale multiracial replication study. *Am J Hum Genet* **90**, 648-660 (2012). https://doi.org:10.1016/j.ajhg.2012.02.023
76 Juang, Y. T. *et al.* Elf-1 binds to GGAA elements on the FcRgamma promoter and represses its expression. *J Immunol* **179**, 4884-4889 (2007). https://doi.org:10.4049/jimmunol.179.7.4884
77 Bergtold, A., Gavhane, A., D'Agati, V., Madaio, M. & Clynes, R. FcR-bearing myeloid cells are responsible for triggering murine lupus nephritis. *J Immunol* **177**, 7287-7295 (2006). https://doi.org:10.4049/jimmunol.177.10.7287

78 Akbarali, Y., Oettgen, P., Boltax, J. & Libermann, T. A. ELF-1 Interacts with and Transactivates the IgH Enhancer π Site *. *Journal of Biological Chemistry* **271**, 26007-26012 (1996). https://doi.org:10.1074/jbc.271.42.26007

79 Yang, A., Kaghad, M., Caput, D. & McKeon, F. On the shoulders of giants: p63, p73 and the rise of p53. *Trends Genet* **18**, 90-95 (2002). https://doi.org:10.1016/s0168-9525(02)02595-7