A novel regulatory element upstream of TNFAIP3 promoter may be influenced by SLE risk variants

Authors: Ajay Nair*1,2 and Archana S Rao2

1 Division of Genomics and Data Sciences, Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA.
2 School of Basic and Applied Sciences, Dayananda Sagar University, Bengaluru, India

*Corresponding Author: Ajay Nair
email: ajay.nair14@me.com

Abstract
Almost all of Genome Wide Association Studies (GWAS) hits come from non-coding DNA elements. Data from chromatin interaction analyses suggest a long-range interaction with a putative enhancer upstream of TNFAIP3. Disrupting the enhancer may impair TNFAIP3 expression and enhance SLE risk. Two variants, rs10499197 and rs58905141 carried on the SLE risk haplotype are situated near this enhancer and could affect its function. Cloning the regulatory region surrounding rs10499197 into an expression plasmid containing a CRISPR-Cas9 backbone, and then performing a genome editing assay, we found that the variant is located near an enhancer. And any changes to the SNP region might impair enhancer and its ability to regulate TNFAIP3 expression.

Introduction
Completion of the human genome sequence has enabled unparalleled exploration of the human genome. Availability of dense SNP panels with high-throughput, cost-effective genotyping and sequencing platforms, aid in performing unbiased high-resolution genome-wide surveys needed to identify DNA segments that harbor complex disease causal variants.

Understanding SLE, Why it matters. Systemic Lupus Erythematosus (SLE) is an autoimmune disease characterized by dysregulated interferon response and loss of self-tolerance to cellular
antigens. Autoantibody production leads to immune complex formation resulting in wide scale inflammation and multiple organ failure. SLE afflicts women at a rate nine times that of men (Chen et al., 2017; Deng and Tsao, 2017). The precise etiology of SLE is complex and poorly defined, requiring the interplay between inherited common risk alleles, genetic predisposition and environmental triggers.

**TNFAIP3, a key SLE-risk locus.** Currently, there are more than 90 robust SLE susceptibility loci (Mohan and Putterman, 2015; Chen et al., 2017; Deng and Tsao, 2017). Testing association of 311,238 genotyped SNPs in 431 unrelated SLE cases and 2155 controls, three regions met a specified threshold for genome-wide significance: the previously defined HLA, IRF5 regions and a novel SLE locus at chromosome 6q23 within the TNFAIP3 gene (Dieude P. et al, 2010).

*Tumor necrosis factor alpha-inducible protein 3 (TNFAIP3)* is a prominent risk locus identified in many autoimmune diseases including SLE. **TNFAIP3** encodes the ubiquitin-editing enzyme, A20 that negatively regulates Nuclear Factor kappa-B (NF-κB) signaling to attenuate pro-inflammatory immune responses and protect against the damaging effects of chronic inflammation (Han et al., 2009; Cai et al., 2010).

Human GWAS of autoimmune diseases have identified single nucleotide polymorphisms (SNPs) on the **TNFAIP3** locus that encodes A20. However, due to strong linkage disequilibrium, GWAS cannot distinguish causal from benign SNPs. Our group previously identified several potential causal SNPs within the **TNFAIP3** gene locus associated with SLE. In this study, we aim to briefly evaluate the function of a novel regulatory element located upstream of the **TNFAIP3** promoter.

**Materials and Methods**

**Antibodies and cell lines**

Jurkat, Human embryonic kidney cells 293 (HEK293T) and THP-1 were purchased from ATCC. Epstein Barr Virus (EBV)-transformed B cell lines were obtained from the Lupus
Family Registry and Repository (OMRF) with IRB approval. EBV cell lines were selected using genotype data corresponding to the rs10499197 variant. Genotypes were verified by Sanger sequencing. Cell lines were maintained in RPMI 1640 medium supplemented with 10% FBS, 1X penicillin-streptomycin antibiotic mixture (Atlanta Biologicals Inc., GA), 2mM L-glutamine (Lonza™ BioWhittaker™, Walkersville, MD), and 55μM β-mercaptoethanol. The following antibodies were used in the study: anti-GAPDH (Cell Signaling Inc., Danvers, MA); anti-A20 (GeneTex Inc., Atlanta, GA); anti-Histone H3 (acetyl K27) (Abcam, Cambridge, MA); normal Rabbit IgG (EMD Millipore, Burlington, MA). All stock laboratory chemicals were from Sigma Aldrich (St. Louis, MO).

**Western Blotting**

Two million EBV-transformed B-cells were pelleted and washed in cold PBS and lysed with RIPA lysis buffer (25 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA, pH 8, 1% Triton X-100, 0.1% SDS, 0.5% sodium deoxycholate) containing proteases (EMD Millipore, Burlington, MA) and Halt phosphatase inhibitors (Pierce, Rockford, IL). Cells were lysed for 15 minutes on ice followed by syringe lysis with a 27g needle then cleared by centrifugation at max speed for 20 minutes. Protein concentrations were determined by Qubit assay (Life Technologies, Carlsbad, CA). Proteins were denatured in 2X SDS loading buffer by heating to 100 °C for 5-10 min. Proteins were eluted by SDS-PAGE and analyzed by immunoblotting as indicated.

**HiChIP Assay**

A modified version of a previously published HiChIP method was used (Mumbach et al., 2016). Briefly, EBV-transformed B cells were crosslinked with 1% formaldehyde for 10 min at room temperature, then intact nuclei were isolated and DNA digested using MboI restriction enzyme (New England BioLabs, Inc., Ipswich, MA) for 4 hours at 37 °C. The ends of the DNA fragments were filled and labeled with dCTP, dGTP, dTTP, and biotin-dATP. Proximity
ligation was performed at room temperature overnight, and then the ligated chromatin was fragmented using a Covaris S2 sonicator (Woburn, MA) and previously described parameters (Mumbach et al., 2016). The fragmented crosslinked protein/DNA complex was immunoprecipitated using an anti-H3K27acetylated antibody, and then purified by immunomagnetic beads (Magna ChIP™ Protein A+G Magnetic Beads, EMD Millipore Corp, USA). The number of cells in each sample was used to determine the amount of MboI enzyme, immunoprecipitating antibody, and immunomagnetic beads used (Mumbach et al. Nat Methods 2016). DNA was eluted from the beads by incubating at 65 °C for 4 hours with 5% Proteinase K and purified using spin column (DNA Clean & Concentrator™, Zymo Research, Irvine, CA). Purified DNA concentration was quantified by Qubit High Sensitivity Assay Kit (Thermo Fisher Scientific). The yields were used to determine the tagmentation enzyme amount and PCR cycle number. Library construction required a minimum of 2 ng DNA. Biotin-labeled DNA fragments (Thermo Fisher Scientific) were further immunoprecipitated by Streptavidin M-280 Dynabeads (Thermo Fisher Scientific). Finally, the HiChIP library was generated on the streptavidin beads using Nextera DNA Library Prep Kit (Illumina Inc., USA). The final PCR product was purified by two-sided size selection using Ampure XP beads (Beckman Coulter, Brea, CA) to capture DNA fragments between 300 bp and 700 bp. Library quality assessment and sequencing: Several quality controls were performed on the HiChIP library. First, the Qubit High Sensitivity Assay Kit was used to quantify DNA concentration. Secondly, the integrity and quality of the library was assessed by Agilent TapeStation 2200 bioanalyzer (Santa Clara, CA). A library was considered good if it had a clear DNA peak at 300-700 bp. Several HiChIP libraries were pooled together and quantified using the KAPA Biosystem library quantification qPCR kit (KAPA Biosystems, Wilmington, MA). The qPCR was performed on a Roche LightCycler 480 system using primers specific to the HiSeq flowcell. Using the template size determined from the Qubit, the pools were diluted to 2 nM.
and re-quantified using qPCR to confirm the starting concentration. The pooled samples were
denatured for 5 min with 0.1N NaOH, and further diluted to a loading concentration of 1.8 pM.
Libraries were sequenced on the Illumina NextSeq 500 sequencer on a 150-cycle paired-end
Mid flowcell. Analysis pipeline: HiChIP raw reads (fastq files) were aligned to the hg19 human
reference genome using HiC-Pro [Servant, et al., 2015]. Aligned data were processed and
analyzed through hichipper pipeline [hichipper MS]. We used MAC2 for anchor calling in the
pipeline based on ChIP-enriched regions. Loops were derived from the linked paired-end reads
that overlap with anchors. A good quality sample has a minimum of 15% intrachromosomal
interactions.

CRISPR-Cas9 Genome Editing

A CRISPR-Cas9 system from reagent design, construction, validation and cell line expansion
was developed using the established protocol (Ran F.A. et al., 2013). Custom guide RNA
(gRNA) pairs for a 500bp sequence around the rs10499197 (T/G) SNP region, as well as
genotyping primers, were designed in silico via the CRISPR Design Tool (http://tools.genome-
engineering.org). gRNA guide sequences were cloned into pSpCas9(BB)-2A-GFP (PX458)
expression plasmid bearing both gRNA scaffold backbone (BB) and Cas9, pSpCas9(BB).
Completed and sequence-verified pSpCas9 (gRNA) plasmids and optional repair templates for
facilitating HDR were then transfected into HEK293T cells and assayed for their ability to
mediate targeted cleavage. Finally, transfected cells were clonally expanded to derive isogenic
cell lines with defined mutations. Chimeric DNA generated was extracted and cleaned using a
QIAGEN DNA purification kit (QIAGEN Inc., Valencia, CA). DNA was then amplified by
PCR and subjected to Sanger sequencing to decode the clones.

Results
Epigenomic analyses revealed a putative enhancer upstream of the \textit{TNFAIP3} promoter.

Dynamic organization of chromatin in three-dimensional space, specifically enhancer-gene promoter looping, plays a crucial role in regulating gene transcription. To investigate the three-dimensional organization of the TNFAIP3 locus, we performed highly integrative chromatin immunoprecipitation (HiChIP) analysis on chromatin obtained from Epstein Barr virus-transformed B cells (EBV) using antibodies against the epigenetic mark of an active enhancer, acetylation of Histone 3 lysine 27 (H3K27ac). We used hichipper and DNAlandscaper (Lareau and Aryee, 2018) to visualize, in two dimensions, several long-range enhancer-promotor interactions in and around the \textit{TNFAIP3} locus (\textbf{Figure 1}). Notably, our H3K27ac HiChIP data revealed a prominent long-range looping event between the \textit{TNFAIP3} promoter and a region ~55 kb upstream that is enriched with epigenetic marks of an enhancer (H3K4me1, H3K27ac) and clusters of transcription factor binding sites defined by ENCODE ChIP-seq data. The putative enhancer also overlapped SLE risk variants rs10499197 and rs58905141 located on the most 5' end of the SLE-associated ~109 kb risk haplotype (\textbf{Figure 1}). Collectively these results suggest that the putative enhancer located ~55 kb upstream of \textit{TNFAIP3} may play an important role in modulating \textit{TNFAIP3} expression during SLE pathogenesis.

\textit{TNFAIP3} expression is regulated by upstream DNA element.

To further analyze the functional potential of the putative enhancer ~ 55kb upstream of \textit{TNFAIP3}, we generated CRISPR-Cas9 constructs to modify the region using the rs10499197 variant as a landmark. We used two sets of small RNAs (guide (g)-RNA) to enable Cas9 mediated genome editing at the target site in HEK293T cells: gRNA#1 was located 31 bp upstream of rs10499197, and gRNA#10 was located 4 bp downstream of rs10499197 (\textbf{Supplementary Figure}). Several isogenic clones were Sanger sequenced (\textbf{Supplementary Table}), but only three cell clones exhibited varying patterns of CRISPR editing near the target
site. Clone#10-14 (from gRNA#10) had a 22-nucleotide deletion around rs10499197 on one of the alleles and 62-nucleotide deletion around rs10499197 on the other (Figure 2a). Clone#10-9 had a 11-nucleotide deletion around rs10499197 on one allele and a 9-nucleotide deletion downstream of the target site on the other allele (Figure 2a). Clone#1-6 (from gRNA#1) had a homozygous 5-nucleotide deletion upstream of rs10499197 (Figure 2a). Western blotting for A20 protein in whole-cell lysates from the CRISPR clones revealed that the homozygous 5-nucleotide deletion upstream of the rs10499197 significantly reduced \textit{TNFAIP3} expression (Figure 2b). Our CRISPR technology provided an orthologus approach to support the ENCODE data by demonstrating that the ~ 55 kb region upstream of the \textit{TNFAIP3} promoter functions as a likely enhancer necessary to sustain \textit{TNFAIP3} expression. The data also suggests that the rs10499197 may be a causal variant that disrupts \textit{TNFAIP3} gene expression by disrupting the enhancer.

\textbf{Discussion}

In this study, we describe the functional characterization of the rs10499197 variant that is associated with human SLE in the region of \textit{TNFAIP3} on chromosome 6q23 for which previous genetic and bioinformatics analyses suggest there are likely to be causal variants. A key objective of human genetics is the identification and characterization of variants associated with complex diseases. Although extensive platforms such as GWAS help us identify several variants associated with complex genetic diseases only a small fraction are causal. Due to Linkage Disequilibrium (LD) in the human genome, GWAS is usually unable to differentiate causal variants from non-causal variants on the same haplotype. Therefore, separating causal variants from non-causal variants requires a combination of genetic (finemapping, resequencing, imputation) and bioinformatic (variant annotation, modeling building) techniques. The end result, as a consequence, is a list of variants that methodically
characterized to understand the underlying biological function in relation to complex diseases.

The current work offers a functional explanation for the genetic association between SLE and the alleles of the variant rs10499197. Thereby making rs10499197 a potential SLE risk haplotype. The \textit{TNFAIP3} locus exhibits complex genetic architecture. A study looking at the genetic associations of SLE and \textit{TNFAIP3} identified three key risk haplotypes around the \textit{TNFAIP3} gene body (Plenge R.M. et al., 2007): 185kb haplotype upstream of \textit{TNFAIP3}, 109kb haplotype that spans the \textit{TNFAIP3} coding region and a 249kb haplotype downstream of \textit{TNFAIP3}. Our work focuses on a non-coding DNA element 55kb upstream of the \textit{TNFAIP3} promoter. This upstream DNA element harbors SLE variants believed to affect \textit{TNFAIP3} and in turn SLE pathophysiology.

\textbf{Conclusions}

Our future work would include, Chromatin conformation capture assays (3C) that would determine how the upstream DNA element physically interacts with the \textit{TNFAIP3} promoter. An understanding of which will help us establish a novel mechanism for \textit{TNFAIP3} transcriptional regulation.

\textbf{References}


Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 39, 1477-1482.


**Manuscript figures**

![Figure A](image1.png)

**Figure 1. Location of SNP in putative enhancer region.** Two risk variants locate in putative enhancer region. A. The relative locations of three most strong DNA loopings mediated by Pol2 in EBV cell lines. B, C. A zoom in view of two upstream interaction regions. The ENCODE Integrated Regulation super-tracks contain: Chromatin State Segmentation, H3K4Me1 Marks, H3K27Ac Marks, and transcription factor ChIP-seq data. SLE-associated risk variants (rs58905141, rs10499197) overlap with H3K4Me1 and H3K27Ac Marks and are very close to strong ChIP-seq signals (the darkness of the segment is proportional to the signal strength). The ChIA-PET signal ~20kb upstream of the TNFAIP3 promoter did not overlap with any variants present on the risk haplotype.
Figure 2. Evaluation of isogenic TNFAIP3 clones following genome editing. (A) Illustration showing edited upstream putative enhancer region around rs10499197 across three CRISPR-HEK293T clones: #10-14, #10-9, and #1-6. WT represents wildtype cell. The clones either have a homozygous deletion or a heterozygous deletion. (B) A Western blot demonstrating the difference in A20 expression across all the clones when compared to the wildtype. Cell lysate from Jurkat T cells were used as a positive control. GAPDH represents the loading control.
**Supplementary Fig. CRISPR-CAS9 guide RNA design.** (A) Location of putative enhancer upstream of *TNFAIP3* identified by the risk variant rs10499197 in the UCSC genome browser. The top track shows location of the SNPs, the middle track shows the H3K27ac marks that are
indicative of an enhancer region, the bottom track shows the ENCODE defined transcription factor binding tracks (B) Sequence around rs10499197 in the upstream DNA element. (C) location of designed guide RNA, PAM sequence around rs10499197 (T/G).

<table>
<thead>
<tr>
<th>Clone</th>
<th>PCR primer</th>
<th>Heterozygous or Homozygous</th>
<th>Sequence readout</th>
<th>Summary of editing</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1-1</td>
<td>F/R</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1-4</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1-5</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1-6</td>
<td>F</td>
<td>Homozygous</td>
<td>TC/-----AAG</td>
<td>5-nt del</td>
</tr>
<tr>
<td>#1-7</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1-8</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1-9</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1-10</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1-11</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1-12</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1-13</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#10-6</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#10-7</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#10-8</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#10-9</td>
<td>F</td>
<td>Heterozygous</td>
<td>ACTTA-*10/-TGGTAG……; AAT/-----GTAAGT---TGATAT……</td>
<td>11-nt del/ 15-del+ 6 inst (miss SNP T on one parental chromosome)</td>
</tr>
<tr>
<td>#10-10</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#10-11</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#10-12</td>
<td>F</td>
<td>no editing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#10-13</td>
<td>F</td>
<td>no editing</td>
<td>TTAGC-*8/-*14TTGAT; TTAGC-*8/-*54ACAGA……</td>
<td>22-nt del, 62-nt del</td>
</tr>
<tr>
<td>#10-14</td>
<td>F &amp; R</td>
<td>Heterozygous</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Table. PCR screening of Clones. Table columns show (from left to right): isogenic clones; PCR primers used to sanger sequence the clones; editing in one allele or both; sequence readout of the edit.