BMC Medical Genomics

RESEARCH Open Access

A systematic strategy for identifying causal single nucleotide polymorphisms and their target genes on Juvenile arthritis risk haplotypes

Kaiyu Jiang[1], Tao Liu[2], Susan Kales[3], Ryan Tewhey[3], Dongkyeong Kim[4], Yungki Park[4,5] and James N. Jarvis[1,5,6*]

Abstract

Background Although genome-wide association studies (GWAS) have identified multiple regions conferring genetic risk for juvenile idiopathic arthritis (JIA), we are still faced with the task of identifying the single nucleotide polymorphisms (SNPs) on the disease haplotypes that exert the biological effects that confer risk. Until we identify the risk-driving variants, identifying the genes influenced by these variants, and therefore translating genetic information to improved clinical care, will remain an insurmountable task. We used a function-based approach for identifying causal variant candidates and the target genes on JIA risk haplotypes.

Methods We used a massively parallel reporter assay (MPRA) in myeloid K562 cells to query the effects of 5,226 SNPs in non-coding regions on JIA risk haplotypes for their ability to alter gene expression when compared to the common allele. The assay relies on 180 bp oligonucleotide reporters ("oligos") in which the allele of interest is flanked by its cognate genomic sequence. Barcodes were added randomly by PCR to each oligo to achieve > 20 barcodes per oligo to provide a quantitative read-out of gene expression for each allele. Assays were performed in both unstimulated K562 cells and cells stimulated overnight with interferon gamma (IFNg). As proof of concept, we then used CRISPRi to demonstrate the feasibility of identifying the genes regulated by enhancers harboring expression-altering SNPs.

Results We identified 553 expression-altering SNPs in unstimulated K562 cells and an additional 490 in cells stimulated with IFNg. We further filtered the SNPs to identify those plausibly situated within functional chromatin, using open chromatin and H3K27ac ChIPseq peaks in unstimulated cells and open chromatin plus H3K4me1 in stimulated cells. These procedures yielded 42 unique SNPs (total = 84) for each set. Using CRISPRi, we demonstrated that enhancers harboring MPRA-screened variants in the *TRAF1* and *LNPEP/ERAP2* loci regulated multiple genes, suggesting complex influences of disease-driving variants.

Conclusion Using MPRA and CRISPRi, JIA risk haplotypes can be queried to identify plausible candidates for disease-driving variants. Once these candidate variants are identified, target genes can be identified using CRISPRi informed by the 3D chromatin structures that encompass the risk haplotypes.

*Correspondence:
James N. Jarvis
jamesjar@washington.edu

Full list of author information is available at the end of the article

Keywords Causal variant, CRISPRi, Enhancers, Genetics, Haplotype, Juvenile arthritis

Introduction

Juvenile idiopathic arthritis (JIA) is a term used to describe a group of childhood illnesses characterized by chronic inflammation and hypertrophy of synovial membranes. Although more rare than the adult disease it resembles, i.e., rheumatoid arthritis (RA), JIA is one of the most common chronic disease conditions in children [1, 2]. Like RA, JIA has long been recognized as a complex genetic trait, in which multiple genetic loci contribute to disease risk [3]. Although the contribution of any single genetic locus is small, genetic influences on JIA are still quite strong. For example, in a study using the Utah Population Database [4], Sampath et al. found that the relative risk for JIA in siblings was nearly 12-fold that of the broader population (11.6; confidence intervals 4.9–27.5; $p<3\times10^{-8}$), and that for first cousins was nearly 6-fold (5.8; confidence intervals 2.5–13.8; $p<6\times10^{-5}$).

To date more than 30 different risk loci have been shown to contribute to JIA [5–7]; the statistical associations are particularly strong for those published in the Hinks Immunochip study [6] and the more recent McIntosh meta-analysis [7]. However, the field still faces several challenges in elucidating the mechanism(s) through which genetic variants confer risk, the most formidable of which is to identify the variants that exert the relevant biological effects that contribute to risk. The standard approach – fine mapping, re-sequencing, imputation, bioinformatic annotation, and laboratory testing – is inherently laborious and low throughput due to linkage disequilibrium (LD), which renders causal variants statistically indistinguishable from neutral variants on risk haplotypes. Not knowing the causal variants complicates the task of identifying the target genes, i.e., those genes whose function or expression levels are influenced by the risk-driving (causal) SNPs. Until the actual causal variants are known, it will be impossible to clarify the mechanisms through which genetic variants alter normal function and contribute to disease risk.

It is becoming increasingly clear that genetic risk for multiple complex traits is more likely exerted through alterations in genomic regulatory functions that influence the efficiency of transcription rather than through the alterations in the coding sequences of disease-relevant genes [8, 9]. We have published several papers that support this concept for JIA [10–12]. For example, we have shown that the JIA risk loci are highly enriched, compared to genome background, for H3K27me1 and H3K27ac chromatin immunoprecipitation-sequencing (ChIPseq) peaks, epigenetic features typically associated with enhancer function [10]. Furthermore, we have shown that, in vitro, risk variants on the JIA risk haplotypes alter the efficiency of enhancers in the *IL2RA* and *IL6R* risk loci [12]. However, the locus-by-locus approach that we and others have used is time-consuming and inefficient. What the field requires is a rapid method for screening thousands of variants for their effects on gene expression in a single assay [13].

The purpose of this study was to Identify single nucleotide polymorphisms on JIA risk haplotypes that show intrinsic ability to alter gene expression in a massively parallel reporter assay (MPRA). We demonstrate the efficacy of this unbiased approach and its capacity for uncovering previously unsuspected mechanisms of genetic risk. Furthermore, we demonstrate how MPRA information and three dimensional chromatin data can then be used to identify the likely target genes, i.e., the genes regulated by the enhancers harboring the variants identified on MPRA. This systematic approach provides an efficient and effective method for resolving the most vexing aspects of autoimmune disease genetics.

Methods

The general workflow for the MPRA is shown in Fig. 1.

We provide further detail in the following sections.

Selection of genetic variants

We queried SNPs within LD blocks where there is already-established risk for JIA. We used the SNPs in the regions identified by Hinks and Hersh [5, 6], including 1,016 new SNPs that we identified within the JIA-associated LD blocks using deep whole genome sequencing [14]. We used a cut-off of $r^2=0.80$ to choose those SNPs in strong LD with the index SNPs. This procedure identified 7,312 candidates to test.

Oligonucleotide library preparation

The oligonucleotide library was prepared following previously published methods [13]. In brief, oligos were synthesized (Agilent Technologies) as 230 bp sequences containing 200 bp of genomics sequence and 15 bp of adapter sequence at either end (5'ACTGGCCGCTTGA CG [200 bp oligo] CACTGCGGCTCCTGC3'). Unique 20 bp barcodes were added by PCR along with additional constant sequence for subsequent incorporation into a backbone vector by gibson assembly. The oligo library was expanded by electroporation into E.coli and the resulting plasmid library was sequenced by Illumina 2×150 bp chemistry to acquire barcode/oligo pairings. The library underwent restriction digest and green fluorescence protein (GFP) with a minimal TATA promoter was inserted by gibson assembly resulting in the 230 bp oligo sequence positioned directly upstream

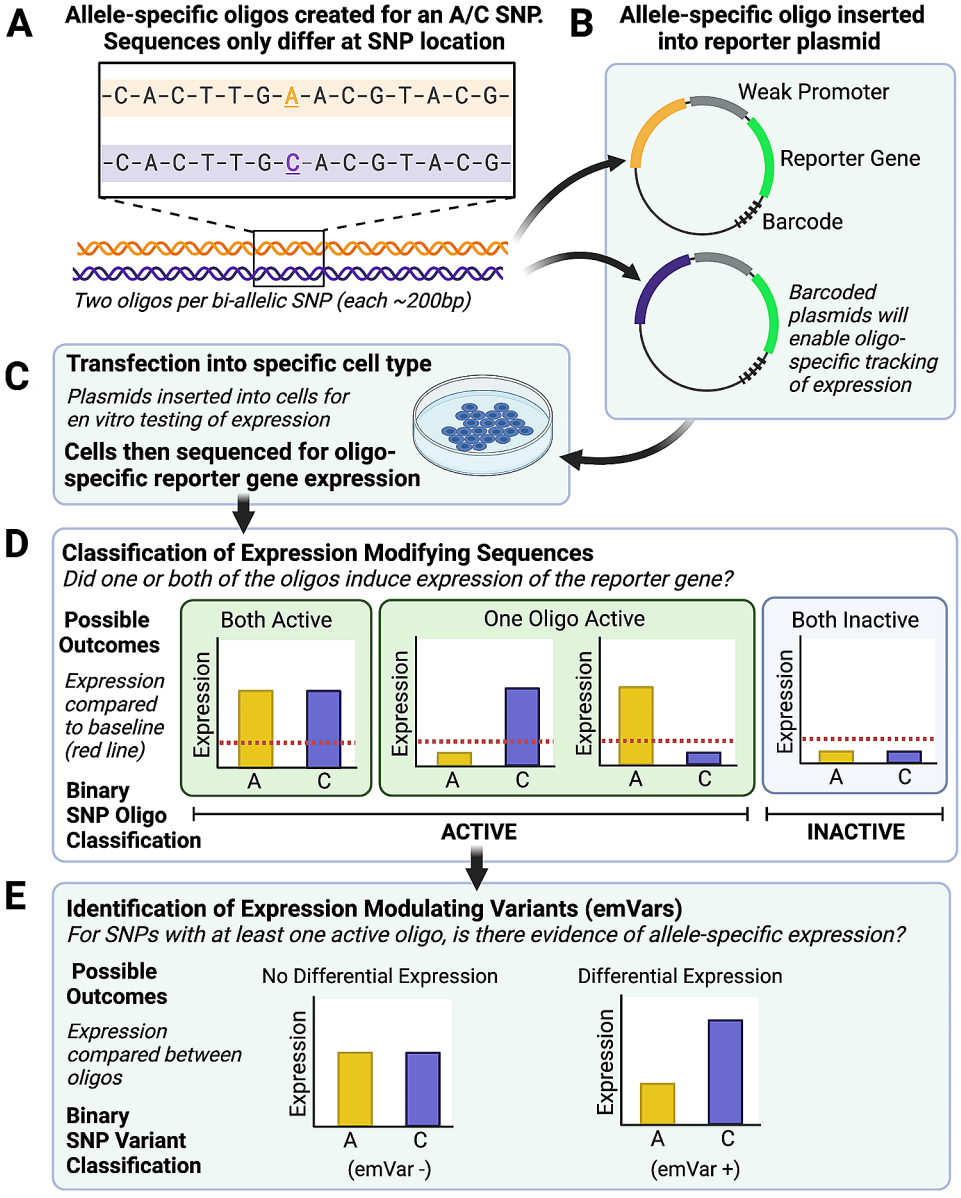

Fig. 1 Summary of work flow for MPRA. Once candidate SNPs are identified, they are cloned into the center of 200 bp oligonucleotides ("oligos") with the cognate flanking sequences and a weak upstream minimal promoter along with a downstream reporter (green fluorescence protein-GFP). After quality control measures to assure that each SNP is represented by a sufficient number of bar codes, oligo libraries are transfected into cells. After incubation with cells, GFP RNA is collected and sequenced. Bar codes allow for the quantification of allele-specific expression levels

of the promoter and the 20 bp barcode falling in the 3′ UTR of GFP. After expansion within E.coli the final massively parallel reporter assay (MPRA) plasmid library was sequenced by Illumina 1×30 bp chemistry to acquire a baseline representation of each oligo within the library.

Transfection of K562 cells

Previous work from our group has shown that JIA risk loci are highly enriched for both H3K4me1 and H3K27ac peaks in both neutrophils and CD4+T cells. Furthermore, for many of the regions queried, H3K4me1/H3k27ac-marked regions were identical in both cell types [10, 11], suggesting that these enhancers regulate common hematopoietic cell functions. This idea is corroborated by ontology analyses of genes within the topologically associated domains (TADs) that encompass the JIA risk haplotypes [12]. Thus, we elected to perform our first-step screening using K562 cells, a myeloid cell line that is easy to transfect and has been used previously with this assay [13]. K562 cells were obtained from the American Tissue Type Collection (ATTC). Cell line authentication was performed using short tandem repeat (STR) loci. STR markers are polymorphic DNA loci that contain repeated nucleotide sequences, and the number

of repeats varies for each individual. Combinations of repeats were used to match cell line to their original reported profile as provided by ATTC.

Cells were grown in RPMI medium (Life Technologies) supplemented with 10% FBS (Life Technologies) maintaining a cell density of 0.5-8×10^5 cells per mL at 37 °C and 5% CO_2. Libraries were electroporated into K562 cells in 100ul volumes using the Amaxa system (Program X-001, Nucleofactor kit V, Lonza). We performed 6 independent replicates with each replicate consisting of ∼2×10^8 cells. In a separate set of experiments, K562 cells were treated with or without IFNγ (250 ng/ml), a biologically relevant dose of a ligand known to contribute to JIA pathobiology [15, 16], 24 h after transfection. The cells were collected 48 h post transfection by centrifugation and washed three times with PBS. The cell pellets were stored at -80 °C.

Preparation of GFP RNA and RNAseq

Total RNA was extracted from cells using Rneasy Midi kit (Qiagen) following the manufacturer's protocol, including the on-column Dnase digestion. A second Dnase treatment was performed on the purified RNA using 5 µL of Turbo Dnase (Life Technologies) in 300 µL of total volume for 1 h at 37 °C. The digestion was stopped with the addition of 3 µL 10% SDS and 30 µL of 0.5 M EDTA followed by a 5 min incubation at 70 °C. The total reaction was then used for pulldown of GFP mRNA. The Dnase digested RNA (300 ul) with 300 µL of 20X SSC (LifeTechnologies), 600 µL of Formamide (Life Technologies) and 10 µL of 10 µm biotin-labeled GFP probes (CCTCGATG TTGTGGCGGGTCTTGAAGTTCACCTTG/3BioTEG; CCAGGATGTTGCCGTCCTCCTTGAAGTCGATGC CC/3BioTEG; CGCCGTAGGTGAAGGTGGTCACGA GGGTGGGCCAG/3BioTEG) was incubated for 2.5 h at 65 °C. Biotin probes were captured using 125 µL of prewashed Streptavidin beads (RNAse clean C1 beads, Life Technologies). The hybridized RNA/probe bead mixture was agitated on a rotator at room temperature for 20 min. Beads were captured by magnet and washed once with 1x SSC and twice with 0.1x SSC. Elution of RNA was performed by the addition of 25 µL water and heating of the water/bead mixture for 2 min at 70 °C followed by immediate collection of eluent on a magnet. A second elution was performed by incubating the beads with an additional 25 µL of water 2 min at 80 °C. A final Dnase treatment was performed in 50 µL total volume using 1 µL of Turbo Dnase incubated for 4 h at 37 °C followed by inactivation with 1 µL of 10% SDS and purification using RNA Ampure XP beads (Beckman Coulter).

First-strand cDNA was synthesized from half of the Dnase-treated GFP mRNA with SuperScript III and a primer specific to the 3' UTR (CCGACTAGCTTGGCC GC) using the manufacturer's recommended protocol.

To minimize amplification bias during the creation of cDNA tag sequencing libraries, samples were amplified by qPCR to estimate relative concentrations of GFP cDNA using 2 µL of sample in a 20 µL PCR reaction containing 10 µL Q5 NEBNext master mix, 2 µL SYBR green I diluted 1:10,000 (Life Technologies) and 0.5 µm of TruSeq_Universal_Adapter (AATGATACGGCGACCA CCGAGATCTACACTCTTTCCCTACACGACGCTCT TCCGATCT) and MPRA_Illumina_GFP_F primers (GT GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCG CCCTGAGCAAAGACC). Samples were amplified with the following conditions: 95 °C for 20 s, 40 cycles (95 °C for 20 s, 65 °C for 20 s, 72 °C for 30 s), 72 °C for 2 min.

To add Illumina sequencing adapters, cDNA samples and 4 mpra: gfp plasmid controls were diluted to match the replicate with the lowest concentration and 20 µL of normalized sample was amplified using the reaction conditions from the qPCR scaled to 50 ul. Amplified cDNA was purified using MinElute PCR Kit (Qiagen) and eluted in 30 µL of EB. Individual sequencing barcodes were added to each sample by amplifying the 20 µL elution in a 50 µL Q5 NEBNext reaction with 0.5 µm of TruSeq_ Universal_Adapter primer and a reverse primer containing a unique 8 bp index (Illumina_Multiplex) for sample demultiplexing post-sequencing. Samples were amplified at 95 °C for 20 s, 6 cycles (95 °C for 20 s, 64 °C for 30 s, 72 °C for 30 s), 72 °C for 2 min. Indexed libraries were purified using MinElute PCR Kit (Qiagen) and pooled according to molar estimates from Agilent TapeStation quantifications. Samples were sequenced using 1×30 bp chemistry on an Illumina HiSeq through the Jackson laboratory in Maine.

Data analysis: identification of SNPs with significant influences on gene expression

We followed the same analysis used previously [13] on our MPRA data from K562 cells. For each of the replicates, we extracted the first 20bps from the sequenced reads and used them to assign the reads to either reference or alternative allele for variants based on our barcode library. We then generated a count-table for each variant, including the names of variants, as well as the counts of the reference and alternative alleles for each replicate. After merging replicates, we generated a master count-table for each of the variant alleles for each of the replicate of DNA (plasmid library), RNA expression in K562 cells, and RNA expression in K562 cells with IFNγ treatment. We used a customized R script based on that described by Tewhey [13] to process the count-table and identify the DNA elements, defined as a region on chromosome that is investigated in MPRA, with active regulatory activities and the variants that can alter the regulatory activities. We used DESeq2 [17], to normalize the counts and fit the dispersion using 'local fit'. The

distribution of the log2 fold changes between RNA and DNA were investigated. Since we found the distribution was not centered at 0, we adjusted the size factors for DNA and RNA samples according to the offset, then performed the normalization again. The oligos showing differential RNA expression relative to the plasmid DNA were then identified by the nbinomWaldTest function in DESeq2 (a Wald Test for the coefficients in a negative binomial generalized linear model) and applying a threshold of 0.01 for the False Discovery Rate and a minimum fold change of 1.5. For DNA elements that displayed significant regulatory activity, we applied a t-test on the log2-transformed RNA/plasmid ratios for each paired replicate (e.g. alternative alleles of K562 replicate 1 vs. reference alleles of K562 replicate 1) to test whether the reference and alternate allele had a similar activity. The *p*-values from t-test were adjusted through the Benjamini-Hochberg process into FDR values and we applied a cutoff of FDR 0.01 to call those variants showing altered regulatory activities between alleles.

Raw data from the MPRA sequencing has been deposited in the National Library of Medicine's Sequence Read Archive (SRA), BioProject accession number PRJNA818294.

Development of K562 cells that express dCas-KRAB

We developed K562 cells expressing a deactivated for of the Cas9 RNA-guided endonuclease (dCas) fused to the Kruppel-associated box protein (KRAB), which deacetylates histones and thus is capable of altering chromatin accessibility and repressing enhancer function. The inducible dCas9-KRAB construct was generated as described previously [18–20]. The dCas9-KRAB were co-transfected with SB100X into K562. The dCas9-KRAB was integrated into genome of K562 cells by SB100X, a hypersensitive transposon [21]. Cells were selected using blasticidin for 10 days. These cells express dCas9-KRAB in a doxycycline-dependent manner (Supplementary Fig. 1).

Identification of genes regulated by enhancers harboring SNPs that screen positive on MPRA

Because most enhancers regulate genes within the same chromatin loop or *topologically associated domain* (TAD) [22], we used publicly available HiC data [12] to identify the likely target genes regulated by the enhancers harboring the SNPs identified by MPRA. As proof-of-concept, we chose to test genes from 2 loci. We chose *TRAF1*, which is of special interest because of its specific association with a polyarticular disease course [23], and *ERAP2/LNPEP*, which are downstream effectors of interferon responses [24], which are known to be enhanced in JIA [15].

Single guide RNAs (sgRNAs) were designed by using a web service from IDT Integrated DNA Technologies, Inc. (www.idtdna.com/site/order/designtool/index/CRISPR_CUSTOM). We designed four gRNAs for each enhancer (Fig. 2). The sgRNA sequence for the HS2 enhancer (HS Cr4, gaaggttacacagaaccaga) was from [25]. The gRNA expression vectors were generated as described previously [18–20]. The sequence information of all constructs was verified by Sanger sequencing.

Next, a pool of gRNAs (4 gRNAs) was transfected into the dCas9-KRAB K562 cells using hypBase plasmid [26], which inserts the gRNA into the genome of the dCas9-KRAB-expressing K562 cells. Cells were selected with puromycin for 10 days to get stably expressed gRNAs- dCas9-KRAB K562 cells. To induce dCas9-KRAB expression, the cells were treated with 1 ug/ml of doxycycline for 48 h before RNA isolation.

RT-qPCR

Total RNA was purified using RNeasy Plus mini kit (Qiagen), and cDNA was synthesized by with iScript™ cDNA synthesis kit according to the directions of the manufacturer (Bio-Rad, Hercules, CA, USA). Quantitative PCR was performed using Power SYBR™ Green PCR Master Mix kit on a StepOne Plus real time PCR system (Applied Biosystems, Foster City, CA, USA). The relative expression level of a gene was normalized by that of GAPDH. The primer sequences were, PHF19: forward- TGGACA GATGGCCTGTACTA and reverse- CCTTCCATAGGA CCCAGTATTT; C5: forward- CTCCTCAGGCCATGT TCATTTA and reverse- TTGTCCTCCAGGCAATTG TT; LNPE: forward- TGAGCAATACACCGCTTTATC A, reverse- GTGCTCATCTTCACACTCTCAG; ERAP2: forward- GACCTCTTCTGCTTCCGATAAA, reverse- GCCAAATATCATTCCACCATTCC; CAST: forward- T GACCGGTCTGAATGTAAAGAG, reverse- TATACTA CACATGGAGGTCCGA; HBE1: forward- TCACTAGC AAGCTCTCAGGC, reverse- AACAACGAGGAGTCT GCCC.

Results

Quality control results

We first sought to determine whether there were significant differences in the number of reads from each oligo for each of the three conditions: *Plasmid alone*, *K562 cells*, and *K562* cells incubated with IFN gamma (*K562 IFN)*. Sequencing reads were scanned for all barcodes, then mapped and counted for each matched oligo. The *Plasmid* condition provides the control of MPRA read counts. The raw counts of reads mapped for each oligo are shown in the heatmap shown in Supplementary Fig. 2. (each row represents a separate oligo; values are log2 counts).

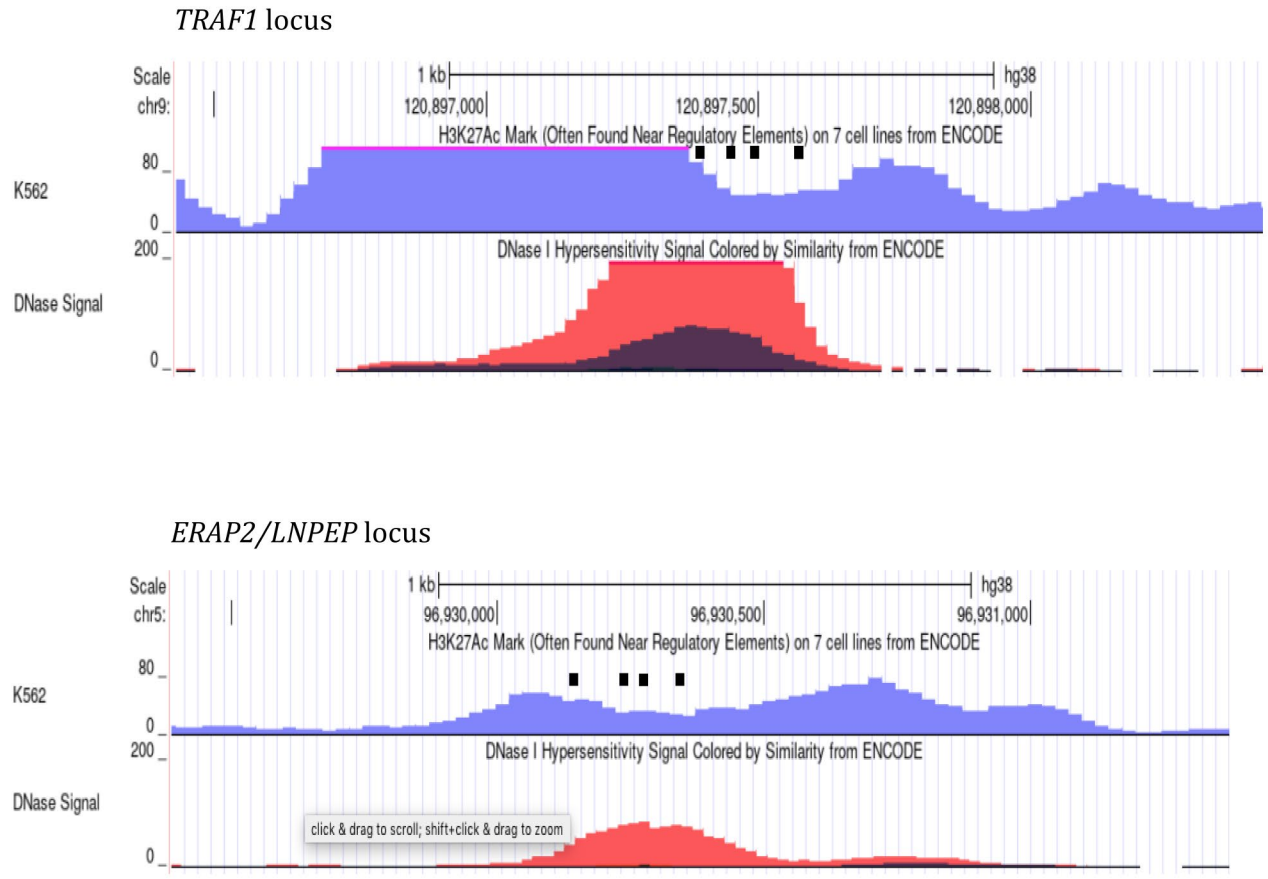

Fig. 2 Screen shots from the UCSC Genome Browser indicating the positions of gRNAs (black rectangles) used for CRISPR1 attenuation of enhancers in the *TRAF1* (top panel) and *ERAP2/LNPEP* loci (bottom panel) in K562 cells. H3K27ac ChIPseq peaks (purple) and DNase1 hypersensitive site (red) are also shown

MPRA screening identifies multiple variants with transcriptional effects

We transfected K562 cells with oligonucleotide probes representing 7,312 SNPs on the JIA risk haplotypes identified from GWAS and candidate gene studies [5, 27] as well as the Immunochip [6]. After final quality control measures, there were 5,226 sequences with sufficient representation in the MPRA library to undertake downstream analyses. Of these, 1,482 (28%) showed regulatory activity. Of these, 530 (19.8%) showed a significant difference from the common allele in unstimulated K562 cells, and 490 (18%) showed differential expression in IFNg-stimulated cells, using FC>2 and FDR=0.01 as a cut-off. After excluding SNPs within the HLA class I and class II loci (n=406), where coding functions are believed to be the most important disease drivers, we further filtered SNPs to identify those that were in open chromatin and within H3K27ac ChIPseq peaks (unstimulated cells) or open chromatin and H3me1, but not H3K27ac-marked regions for SNPs identified in exclusively in stimulated cells. Using these procedures, we identified n=42 SNPs in unstimulated K562 cells (Table 1). After stimulation with IFNγ (250 ng/ml), we identified an additional 42 SNPs that showed significant effects on gene expression that were not identified in unstimulated cells (Table 2). These findings are consistent with previously published studies that suggest that many disease-relevant alleles may exert their effects on immune cells only after those cells are activated [28].

In many cases, we identified multiple alleles on the same haplotype, although these alleles were not necessarily within the same functional element. For example, in the *TRAF1*, *LNPEP/LNPEP/ERAP2*, and *IL6R/ATP8B2* loci, we identified variants within both intergenic and intronic enhancers in unstimulated K562 cells. In IFNγ-stimulated cells, we identified variants in both intergenic and intronic enhancers in the *IL6R/ATP8B2, TYK2/ICAM3*, and *LNPEP/ERAP2* loci. In stimulated K562 cells, we also identified expression-enhancing variants that are situated within the promoter and within an exon within the *ATXN2* locus. This finding may reflect that these regions have enhancer as well as promoter/coding functions, as has been described for other genomic regions [29, 30].

Table 1 SNPs screening positive in unstimulated K562 cells ($n=42$)

SNPs within H3K27ac-marked intronic region in TRAF 1 (chr9:120,921,603 – 120,923,567)
rs1014529
rs10818485
rs7021206
SNPs within an H3K27ac-marked intergenic region in TRAF 1 (chr9:120,929,875 – 120,936,755)
rs10739578
rs2109896
rs7859805
rs10985080
SNP within H3K27ac-marked intergenic region in TNFSF1 (chr13:42,476,952 – 42,480,776)
rs2062305
SNPs within H3K27ac-marked intergenic region in IRF1 chr5:132,491,627 – 132,499,135)
rs2549004
rs2549007
rs2549009
rs2706385
rs2706386
rs41525648
SNPs within H3K27ac-marked intronic region in IL2RA (chr10:6,047,090 – 6,055,660)
rs1924138
rs791587
rs10795763
SNP within H3K27ac marked intronic region in LTBR chr12:6,384,706-6,385,089
rs10849448
SNP within H3K27ac-marked intergenic region in IL6R-ATP8B2 (chr1:154,312,218 – *154,401,421*)
rs11581043
rs4845614
rs1194591
SNP within H3K27ac marked intronic region in STAT4. chr2:191,037,761 – 191,039,047
rs1400653
SNPs within H3K27ac-marked intronic region in RMI2 (chr16:11,348,423 – 11,351,691)
rs11643024
rs2032929
rs2032931
rs2032933
rs8050084
rs9302459
rs9922058
rs9922935
SNP within H3K27ac marked intergenic region in ERAP2/LNPEP (chr5:96,929,602 – 96,933,14)
rs1216565
SNPs within H3K27ac-marked intronic region in ERAP2/LNPEP (chr5:96,967,226 – 96,969,507)
rs1559267
rs1820149
SNP within H3K27ac marked intronic region in IL10. (chr1:206,769,578 – 206,772,118)
rs1518111
SNP within H3K27ac marked intronic region in ZFPl36F1. chr14:68,793,403 – 68,794,934
rs17106304
rs2236263
SNP within H3K27ac marked intergenic region in TYK2 (chr19:10,345,527 – 10,347,710)
rs4611572
SNP within H3K27ac-marked intergenic region in CCR2 (chr3:46,321,182 – 46,323,267)
rs35675823
SNP within H3K27ac marked intronic region in TIMMDC1 (chr3:119,506,272 – 119,508,916)

Table 1 (continued)

rs4688012
SNPs within H3K27ac-marked intronic region in JAZF1 (chr7:28,124,680 – 28,152,184
rs757730
rs1635853
SNP within H3K27ac marked intronic region in IL6 (chr7:22,725,624 – 22,727,193)
rs1800797

* ENCODE ChIPseq data and human neutrophil ChIPseq data [10]

Identification of target genes

The identification of risk-enhancing alleles of JIA haplotypes may facilitate the identification of target genes, i.e., the genes whose expression levels are influenced by the risk-driving SNPs. To test this concept, we relied on the fact that most enhancers regulate genes within the same TAD [22] and that the 3D chromatin structures are identifiable from publicly available chromatin conformation data.

We tested the intergenic enhancer harboring rs10985080, located at chr9:120,896,106–120,897,428 (GRCh38/hg38) on the *TRAF1* haplotype, as well as the intergenic enhancer on the *LNPEP/ERAP2* harboring rs1216565 as described in the *Methods* section. Using publicly available HiC data from the *3D Genome Browser* (http://3dgenome.fsm.northwestern.edu) [31], we visualized heat maps defining TAD structures in K562 cells and reported by Rao et al. [32]. We identified genes within the same chromatin loops that were likely targets of these enhancers. We show HiC data derived from K562 cells in Fig. 3. For *TRAF1*, we identified *TRAF1, PHF19* and *C5* as the most likely candidate targets for this enhancer.

Within the TAD encompassing the intergenic enhancer in the *TRAF1* locus we identified 2 expressed genes, *PHF1* and *C5*. Note that *TRAF1* itself was not expressed. We therefore used CRISPRi to attenuate the function of this enhancer; sequences of the gRNAs used for this purpose are provided in Fig. 4A. Attenuation of the *TRAF1* intergenic enhancer by CRISPRi resulted in a significant ($p<0.01$) reduction in expression of the *PHF19* gene, but not *C5.* These results are shown in Fig. 4B. We next examined the intergenic enhancer harboring rs1216565, located at chr5:96,929,854–96,931,091 (GRCh38/hg38) in the *LNPEP/ERAP2* locus.

Once again, we used publicly available 3D chromatin data to identify likely targets, which included *LNPEP, ERAP2*, and *CAST*. Sequences for the gRNAs for this experiment are shown in **4 A**. As shown in Fig. 4C, attenuation of this enhancer resulted in significant reductions in the expression of both *LNPEP* and *ERAP2*, but not the adjacent gene, *CAST*. This finding is consistent with the known capacity of enhancers to regulate multiple genes within the same topologically-associated domain [22].

Corroboration of MPRA data with human genotyping-expression data

Finally, we sought to gain additional information that MPRA-identified SNPs located in the enhancers in the *TRAF1* and *LNPEP/ERAP2* locus influence gene expression of the candidate target genes in humans. We used the Genotype-Tissue Expression (GTEx) project's [33] expression quantitative trait locus (eQTL) calculator and querying whole blood expression data for this purpose (https://www.gtexportal.org/home/testyourown) to determine whether individuals harboring these SNPs do, in fact, show differential expression of the CRISPRi-identified target genes. Results of these analysis are shown in Table 3, below. SNPs highlighted in bold indicate those with greater-than-by chance likelihood that the SNP has an influence on the expression of the listed gene.

In each case, then, GTEx whole blood expression data further strengthens the predictions from the MPRA + CRISPRi analysis, as individuals carrying these alleles show the predicted alterations in the CRISPRi-identified targets when compared to individuals carrying the common allele. At the same time, these analyses demonstrate the utility of our approach in clarifying GTEx data where multiple SNPs in strong LD appear to influence the expression of a gene. Furthermore, when multiple expression-altering alleles are adjacent (and therefore in strong LD), the GTEx analysis may clarify the specific SNP(s) that exert the strongest influence on the candidate target genes. Note that for rs10985080 and rs1216565, despite significant differences from the common allele, the effect sizes are relatively small. This finding is consistent with the observations of Gasperini et al. [20], who have shown that the effect sizes for most enhancers on the genes they regulate is in the range of 15–30%.

Discussion

The field of genetics as applied to complex traits has started to move beyond the identification of genetic associations and toward the elucidation of the mechanisms through which genetic variants confer risk [34, 35]. However, a significant impediment to accomplishing this task is the fact that a strength GWAS studies, which leverage LD to identify regions conferring genetic risk, is also a weakness, in that the SNPs that tag genetic risk loci are in LD with dozens, sometimes hundreds, of other SNPs,

Table 2 SNPs screening positive in K562 stimulated by interferon gamma ($n=42$)

SNP in H3K4me1 intronic region in PTPN22 (chr1:113,800,033–113,807,167)
rs1217378
SNPs in H3K4me1 intronic region in IL6R/ATP8B2 (chr1:154,312,194 – 154,313,462)
rs11581043
rs1194608
SNP in H3K4me1 intergenic region in IL6R/ATP8B2 (chr1:154,351,989 – 154,356,495)
rs4845369
SNPs in H3K4me1 intronic region in IL10 (chr1:206,768,609 – 206,772,399)
rs1518111
rs2945417
SNP within H3K4me1 intronic region in STAT4 (chr2:191,041,359 – 191,043,004)
rs1996400
SNPs in H3K4me1 intergenic region in CCR2 (chr3:46,297,890 – 46,306,532)
rs7374671
rs4683215
rs35053103
rs2888523
rs2888524
rs34997146
rs62242985
rs6441972
SNPs in H3K4me1 intronic region in TIMMDC1 (chr3:119,394,863 – 119,412,040)
rs4687853
rs2177812
rs7610049
SNP in H3K4me1 intergenic region in LNPEP/ERAP2 (chr5:96,920,167 – 96,924,958)
rs193994
rs1216565
SNP in H3K4me1 intronic region in LNPEP/ERAP2 (chr5:96,995,366 – 96,995,917)
rs430827
SNPs in H3K4me1 intergenic region in IRF1 (chr5:132,490,541 – 132,492,289
rs41525648
rs2706385
rs2706386
SNPs in H3K4me1 intergenic region in TRAF1 (chr9:120,941,159 – 120,947,188)
rs10818481
rs2900180
SNP in H3K4me1 intronic region in IL2RA (chr10:6,040,089 – 6,040,684)
rs791592
SNPs in ATXN2
rs10849962 (Appears to be in the promoter; H3K27ac-marked region spans chr12:111,598,660 – 111,600,586)
rs6416335 (Appears to be an exonic enhancer in BRAP (chr12:111,643,075–111,643,762)
SNP in ZFP36L1 intronic region (chr14:68,794,471 – 68,796,262)
rs3742887
SNPs in H3K4me1 intergenic region in ZFP36L1 (chr14:68,800,391 – 68,801,105)
rs1595260
rs4899258
SNP in H3K4me1 intronic region in RMI2 (chr16:11,283,902 – 11,284,672)
rs34764020
rs7205578
rs6498184
SNPs in H3K4me1 intergenic region in PTPN2 (chr18:12,775,801 – 12,776,691)
rs4327116

Table 2 (continued)

rs2847279
rs9952753
SNP in H3K4me1 intronic region in PTPN2 (chr18:12,860,578 – 12,861,552)
rs515151
SNPs in H3K4me1 intronic region in TYK2 / ICAM3 (chr19:10,333,403 – 10,334,831)
rs2278442
rs2304240
SNP in H3K4me1 intronic region (chr19:10,347,076 – 10,348,009)
rs510506

** ENCODE ChIPseq data and human neutrophil ChIPseq data [10]

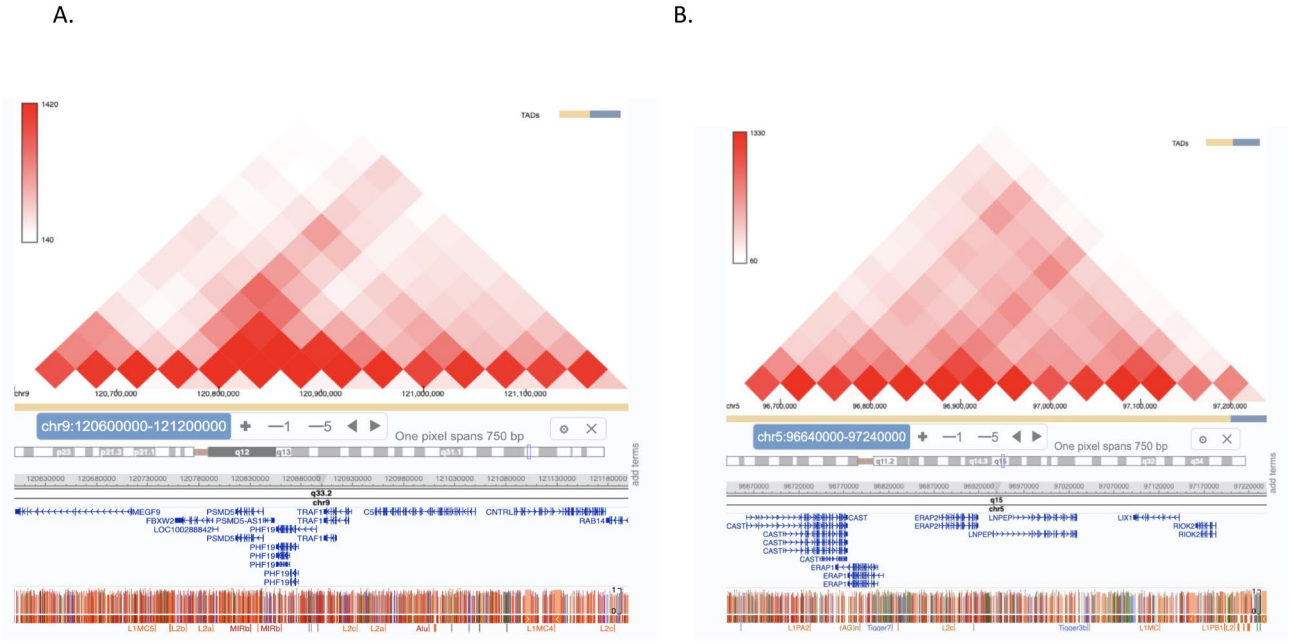

Fig. 3 Screen shots from the 3D Genome Browser and the WashU Genome browser visualizing HiC data in K562 cells. Shown are interacting regions within the *TRAF1* (A, left panel) and *LNPEP-ERAP2* (B, right panel) JIA risk loci

most of which have no influence at all on disease risk. Thus, distinguishing the true causal variants (i.e., those that exert the relevant biological effects) from the innocuous ones in which they are in LD, has been a challenge. At the same time, the discovery that, for most complex traits [8], including autoimmune diseases [9], genetic risk is likely to impinge on regulatory functions rather than the protein-coding sequences of pathology-driving genes, has complicated the search for target genes (i.e., the genes influenced by the causal variants).

In this paper, we demonstrate a systematic strategy for identifying both causal variants and their target genes on JIA risk haplotypes. We find that, by using existing chromatin data in combination with MPRA to screen for expression-altering variants, we can identify a finite number of variants that, based on their functional properties, are strong candidates as actual causal variants, as others have recently shown [36]. Subsequent identification of target genes can then be accomplished using CRISPRi approaches, especially for those variants that lie within enhancers, which are likely to be fundamental to autoimmune disease pathogenesis [37]. The CRISPRi experiments are simplified by the fact that most enhancers regulate genes within the same TAD [22], and, thus experiments can be performed in a targeted fashion rather than genome-wide. Finally, one can make a causal link between expression levels of genes identified by CRISPRi and the variants that screen positive on MPRA using GTEx whole blood expression data.

The MPRA screening yielded some surprising results. We note, for example, that there were many loci where we identified multiple expression-altering alleles within the same functional element. For example, rs2549004, rs2549007, rs2549009, rs2706385, rs2706386 and rs41525648 within a single intergenic enhancer on the *IRF1* haplotype and rs1559267, rs1820149, and rs1559267 within a single intronic enhancer on the *TRAF1* haplotype. This finding suggests that the disease-associated

A

	Name	Sequence	Position (hg38)
Negative control	Scrambled gRNA	GCACTACCAGAGCTAACTCA	Not Applicable
TRAF1 loci gRNAs	gRNA1	TATAGTTAGGGGCCTTGCAC	chr9:120897422-120897441
	gRNA2	AGTCCTCTAGTGCTATCCAG	chr9:120897508-120897527
	gRNA3	AGTGGTGGGGATGCACCACC	chr9:120897340-120897359
	gRNA4	ACCAAGCTCCTTTATAGTTA	chr9:120897434-120897453
ERAP2/LNPEP loci gRNAs	gRNA1	AAGGCTAATCCTGAATATTG	chr5: 96930107-96930126
	gRNA2	CAGGTGGTTGGTTGATATTA	chr5:96930177-96930196
	gRNA3	TGGGCCAGAACCACTCCATG	chr5:96930235-96930254
	gRNA4	GCAGTGTTAACATGATCCGA	chr5:96930332-96930351

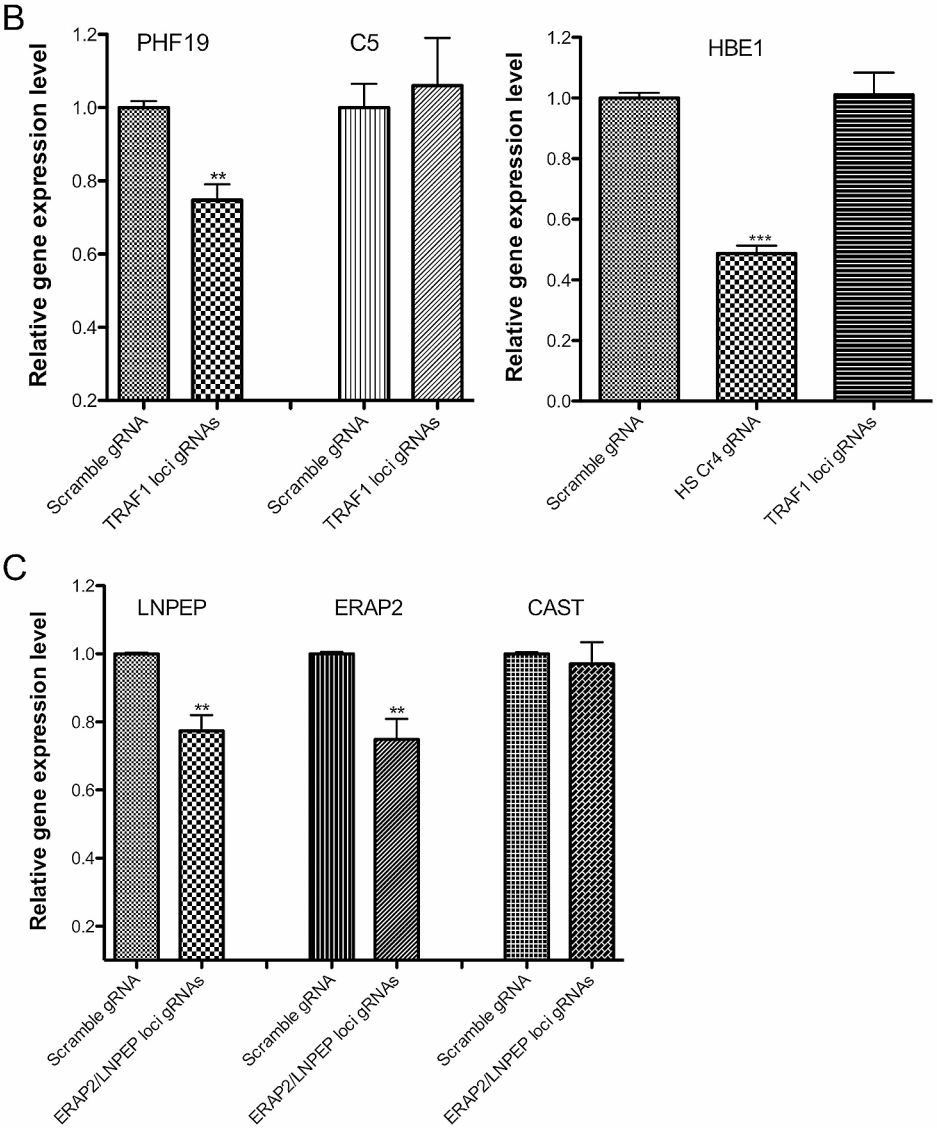

Fig. 4 Results from CRISPRi interrogation of JIA risk loci to identify target genes of intergenic enhancers harboring MPRA-screened SNPs in the *TRAF1* (rs10985080) and *LNPEP/ERAP2* (rs1216565) loci. Panel A provides the sequences and positions of the gRNAs used to interrogate these regions. Panel B shows results from the interrogation of the *TRAF1* locus. Ablation of this intergenic enhancer significantly reduces the expression of *PHF1*, which encodes a polycomb family transcriptional repressor, but not *C5*, another expressed gene within the same topologically-associated domain as this enhancer. Off-target effects were assessed by examining the effects of these same gRNAs on the expression of *HBE1*. The gRNAs showed no effects on *HBE1* expression, although specifically targeted gRNAs to an *HBE1* enhancer reduced expression by 50%. Panel C shows results from interrogation of the intergenic enhancer within the *LNPEP/ERAP2* locus. Attenuating this enhancer, which is situated between the *LNPEP* and *ERAP2* genes, significantly reduced expression of both genes. ** = $p < 0.01$. *** = $p < 0.001$

Table 3 MPRA-identified SNPs located in the enhancers in the *TRAF1* and *LNPEP/ERAP2* loci associated with gene expression levels of the candidate target genes in humans

Locus	Gene	SNP	*P*-value	*P*-value threshold	NES*
TRAF1	*PHF19*	rs10739578	0.00077	0.00021	0.065
		rs2109896	0.084	0.00021	0.036
		rs7859805	0.035	0.00021	0.044
		rs10985080	**0.000068**	0.00021	0.078
LNPEP/ERAP2	*ERAP2*	**rs1216565**	**2.4e-152**	0.00020	0.87
	LNPEP	**rs1216565**	**0.0000064**	0.00020	0.072

NES = Normalized effect size

haplotypes exert risk because they contain multiple alleles in strong LD that, together, alter immune regulatory functions. Furthermore, many of the risk haplotypes contain more than one functional element that is affected, and these different functional elements may exert their effects on different genes. There is no reason to assume, for example, that the intergenic and intronic enhancers on the *TRAF1* haplotype regulate the same genes.

Another useful fact that emerges from our data is that physical interactions between enhancers and promoters do not necessarily indicate a regulatory relationship. For example, HiC and promoter capture HiC data demonstrate physical interactions between the intergenic enhancer in *LNPEP/ERAP2* and the promoter *CAST* gene. However, attenuating this enhancer had no effect on *CAST* expression. While this finding doesn't exclude the possibility that this enhancer might work in concert with others to regulate *CAST*, it serves as a precautionary message in how we use and interpret 3D chromatin data and gene expression data from patient cells.

Our work also highlights the utility of using the *Sleeping Beauty* transposase system and hypBase vectors for gRNAs in functional genomics experiments. Genome-wide CRISPRi screens for enhancer activity have typically use lentivirus and/or plasmid vectors to attenuate enhancer function [22]. However, these assays can be vexing to perform and replicate because of the low multiplicity of infection (MOI) rates seen with such vectors. This makes it difficult to interpret different experiments or even compare replicates within a single experiment. Our approach allows stable and high levels of expression of both the epigenome editing enzyme and the gRNAs, facilitating both replication and inter-experimental comparisons (e.g., between the efficiency of intronic vs. intergenic enhancers in regulating a specific gene).

This work has several limitations, the most important of which is the use of cell lines rather than primary human cells for these assays. There is accumulating data that, based on the similarity of their chromatin to the cognate primary human cells such as Jurkat and THP-1 are suitable as models in genetic studies of human autoimmune diseases [38]. However K562, which were derived from a human myelogenous leukemia, are less like their primary human counterparts, although the TAD structures strongly resemble primary myeloid cells [12]. Recent developments, which include using a Rous sarcoma virus promoter instead of the minimal promoter we used here, suggest that MPRA assays can be performed in primary human cells, provided that a sufficient number of replicates are performed to attenuate the differences between individual donors [39]. We now have such experiments under way in our laboratory.

Another limitation of the findings here was the agnostic nature of the variant selection process. We chose variants with MAF > 1% on the JIA haplotypes identified in several different studies [6], following the approach of Tewhey et al [13]. Variants were not selected based on their presence within plausible functional chromatin or their frequency in genotyped patients with JIA. As Lu and colleagues have shown [36] using such selection methods can increase the efficiency of the MPRA.

Another important limitation to this approach is the fact that MPRA may not detect effects as they would occur in native chromatin [35, 40]. This limitation may lead to both false positives and false negatives. False positives may be reduced by using an additional criterion (or criteria) to filter MPRA-identified variants. For example, Ainsworth et al. [41] have shown that applying analyses of DNA topology, an important determinant of DNA non-coding functions, can improve the predictive value of variants screened on MPRA, since these analyses detect a feature intrinsic to native chromatin. Thus, the addition of a filtering feature that detects effects in native chromatin is likely to significantly reduce the number of false positives that emerge from MPRA alone. It should be noted that false negatives, which DNA topology analyses won't solve, is much less of a problem for the field. The JIA risk haplotypes contain > 13,000 SNPs in LD with the SNPs used to identify/tag the risk loci. The urgent need is to reduce this number so that further functional characterization can proceed in an efficient way.

Even after this process is completed, there will be further work to be done. The agnostic nature of GWAS and genetic fine mapping studies makes it impossible to identify the cells whose functions are affected by

disease-driving variants. In JIA, there are likely to be multiple cells types, in addition to CD4+T cells and neutrophils [10, 11] that are influenced by genetic variants. These include B cells [42] and monocytes [43], and possibly other cells that regulate the innate immune response, such as hepatocytes [44]. All of these cell types will need to be tested separately, and it is likely that different SNPs will be found to alter different regulatory regions in each cell type. Furthermore, we will need to study specific cell subsets (e.g., for CD4+T cells and B cells) and different activation states.

For many of these cell types, there may be serious limitations in using GTEx whole blood expression data to make the causal connection between SNPs identified on MPRA and expression levels of genes identified on CRISPRi. We note that our MPRA and CRISPRi studies were performed in myeloid K562 cells, and that neutrophils (derived from myeloid precursors) are the most abundant leukocyte in adult peripheral blood. Gene expression profiles from whole blood are strongly influenced by neutrophil expression signatures [45, 46]. It seems plausible that genetic influences that specifically affect gene expression in lymphocytes and lymphocyte subsets would be difficult to corroborate using existing GTEx data, especially given the relatively small number of genotyped subjects (fewer than 700) currently in the GTEx whole blood data set. In these cases, establishing a causal link between a given SNP and the expression of candidate target gene might require direct interrogation of the cells of interest using CRISPR/homology-directed repair strategies.

Conclusion

We describe a systematic approach for identifying both causal variants and their target genes on JIA risk haplotypes. This approach relies on knowledge of the chromatin structures that encompass the risk haplotypes as well as a massively parallel genomic assay to identify functional features that make them strong candidates for being disease-driving variants. The use of screening assays in primary human cells and the adaptation of informative cell lines can be expected to rapidly advance our understanding of genetic mechanisms that drive JIA risk.

Abbreviations

ATTC	American Tissue Type Collection
ChIPseq	Chromatin immunoprecipitation-sequencing
GFP	Green fluorescence protein
JIA	Juvenile idiopathic arthritis
LD	Linkage disequilibrium
MPRA	Massively parallel reporter assay
RA	Rheumatoid arthritis
SNP	Single nucleotide polymorphism
STR	Short tandem repeats

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12920-024-01954-z.

Supplementary Material 1

Acknowledgements

The authors wish to thank Joshua J. Breunig for providing us the hypBase vectors used in the gRNA transfections. We also thank Hannah Ainsworth, PhD, for producing *Fig. 1*.

Author contributions

KJ: Performed the experimental work and assisted in data analysis and interpretation and manuscript presentation. TL: Assisted in the design of the MPRA library and analyzed the MPRA data. Assisted in data analysis and interpretation and manuscript presentation. SK: Assisted in the preparation and quality control measures on the MPRA library. RT: Assisted in the design and quality control of the MPRA library, assisted with data analysis and interpretation and manuscript preparation. DK and YP: Provided technical advice and guidance for the CRISPRi studies. YP: Assisted in manuscript preparation. JNJ – Designed the study and assisted in the analysis and interpretation of the data. Assisted in manuscript preparation. All authors reviewed and approved the manuscript.

Funding

This work was supported by NIH grants R21-AR071878 and R21 AR076948 (JNJ), as well as R01-NS094181, R21-NS102558, and R21-NS112608 (YP). This work was also supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1TR001412 to the University at Buffalo. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

Data availability

Raw data from the MPRA sequencing has been deposited in the National Library of Medicine's Sequence Read Archive (SRA), BioProject accession number PRJNA818294.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Pediatrics, Clinical and Translational Research Center, University at Buffalo Jacobs School of Medicine School Medicine & Biomedical Sciences, 701 Ellicott St, Buffalo, NY 14203, USA
[2]Roswell Park Cancer Institute, 665 Elm St, Buffalo, NY 14203, USA
[3]Jackson Laboratories, 600 Main St, Bar Harbor, ME 04609, USA
[4]Department of Biochemistry, University at Buffalo Jacobs School of Medicine School Medicine & Biomedical Sciences, 955 Main St, Buffalo, NY 14203, USA
[5]Genetics, Genomics, & Bioinformatics Program, University at Buffalo Jacobs School of Medicine School Medicine & Biomedical Sciences, 955 Main St, Buffalo, NY 14203, USA
[6]University of Washington Rheumatology Research, 750 Republican St., E520, Seattle, WA 98109, USA

Received: 15 April 2024 / Accepted: 27 June 2024
Published online: 12 July 2024

References

1. Gortmaker SL, Sappenfield W. Chronic childhood disorders: prevalence and impact. Pediatr Clin North Am. 1984;31(1):3–18.
2. Singsen BH. Rheumatic diseases of childhood. Rheum Dis Clin North Am. 1990;16(3):581–99.
3. Glass DN, Giannini EH. Juvenile rheumatoid arthritis as a complex genetic trait. Arthritis Rheum. 1999;42(11):2261–8.
4. Prahalad S, Zeft AS, Pimentel R, Clifford B, McNally B, Mineau GP, et al. Quantification of the familial contribution to juvenile idiopathic arthritis. Arthritis Rheum. 2010;62(8):2525–9.
5. Hersh AO, Prahalad S. Immunogenetics of juvenile idiopathic arthritis: a comprehensive review. J Autoimmun. 2015;64:113–24.
6. Hinks A, Cobb J, Marion MC, Prahalad S, Sudman M, Bowes J, et al. Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. Nat Genet. 2013;45(6):664–9.
7. McIntosh LA, Marion MC, Sudman M, Comeau ME, Becker ML, Bohnsack JF, et al. Genome-Wide Association Meta-Analysis reveals Novel Juvenile Idiopathic Arthritis susceptibility loci. Arthritis Rheumatol. 2017;69(11):2222–32.
8. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190–5.
9. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015;518(7539):337–43.
10. Jiang K, Zhu L, Buck MJ, Chen Y, Carrier B, Liu T, et al. Disease-Associated single-nucleotide polymorphisms from noncoding regions in Juvenile Idiopathic Arthritis are located within or adjacent to functional genomic elements of human neutrophils and CD4 + T cells. Arthritis Rheumatol. 2015;67(7):1966–77.
11. Zhu L, Jiang K, Webber K, Wong L, Liu T, Chen Y, et al. Chromatin landscapes and genetic risk for juvenile idiopathic arthritis. Arthritis Res Ther. 2017;19(1):57.
12. Jiang K, Kessler H, Park Y, Sudman M, Thompson SD, Jarvis JN. Broadening our understanding of the genetics of Juvenile Idiopathic Arthritis (JIA): interrogation of three dimensional chromatin structures and genetic regulatory elements within JIA-associated risk loci. PLoS ONE. 2020;15(7):e0235857.
13. Tewhey R, Kotliar D, Park DS, Liu B, Winnicki S, Reilly SK, et al. Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. Cell. 2016;165(6):1519–29.
14. Wong L, Jiang K, Chen Y, Jarvis JN. Genetic insights into juvenile idiopathic arthritis derived from deep whole genome sequencing. Sci Rep. 2017;7(1):2657.
15. Throm AA, Moncrieffe H, Orandi AB, Pingel JT, Geurs TL, Miller HL et al. Identification of enhanced IFN-gamma signaling in polyarticular juvenile idiopathic arthritis with mass cytometry. JCI Insight. 2018;3(15).
16. Jarvis JN, Dozmorov I, Jiang K, Frank MB, Szodoray P, Alex P, et al. Novel approaches to gene expression analysis of active polyarticular juvenile rheumatoid arthritis. Arthritis Res Ther. 2004;6(1):R15–32.
17. Love MI, Huber W, Anders S. Moderated estimation of Fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
18. Kim D, Park Y. Molecular mechanism for the multiple sclerosis risk variant rs17594362. Hum Mol Genet. 2019;28(21):3600–9.
19. Kim D, An H, Shearer RS, Sharif M, Fan C, Choi JO, et al. A principled strategy for mapping enhancers to genes. Sci Rep. 2019;9(1):11043.
20. Kim D, An H, Fan C, Park Y. Identifying oligodendrocyte enhancers governing Plp1 expression. Hum Mol Genet. 2021;30(23):2225–39.
21. Huang X, Wilber AC, Bao L, Tuong D, Tolar J, Orchard PJ, et al. Stable gene transfer and expression in human primary T cells by the sleeping Beauty transposon system. Blood. 2006;107(2):483–91.
22. Gasperini M, Hill AJ, McFaline-Figueroa JL, Martin B, Kim S, Zhang MD et al. A genome-wide Framework for Mapping Gene Regulation via Cellular Genetic screens. Cell. 2019;176(1–2):377 – 90 e19.
23. Albers HM, Kurreeman FA, Houwing-Duistermaat JJ, Brinkman DM, Kamphuis SS, Girschick HJ, et al. The TRAF1/C5 region is a risk factor for polyarthritis in juvenile idiopathic arthritis. Ann Rheum Dis. 2008;67(11):1578–80.
24. Paladini F, Fiorillo MT, Tedeschi V, Mattorre B, Sorrentino R. The multifaceted nature of Aminopeptidases ERAP1, ERAP2, and LNPEP: from evolution to Disease. Front Immunol. 2020;11:1576.
25. Thakore PI, D'Ippolito AM, Song L, Safi A, Shivakumar NK, Kabadi AM, et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat Methods. 2015;12(12):1143–9.
26. Breunig JJ, Levy R, Antonuk CD, Molina J, Dutra-Clarke M, Park H, et al. Ets factors regulate neural stem cell depletion and gliogenesis in Ras Pathway Glioma. Cell Rep. 2015;12(2):258–71.
27. Herlin MKPM, Herlin T. Update on genetic susceptibility and pathogenesis in juvenile idiopathic arthritis. EMJ Rheumatol. 2014;1:73–83.
28. Gate RE, Cheng CS, Aiden AP, Siba A, Tabaka M, Lituiev D, et al. Genetic determinants of co-accessible chromatin regions in activated T cells across humans. Nat Genet. 2018;50(8):1140–50.
29. Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, et al. Coding exons function as tissue-specific enhancers of nearby genes. Genome Res. 2012;22(6):1059–68.
30. Ahituv N. Exonic enhancers: proceed with caution in exome and genome sequencing studies. Genome Med. 2016;8(1):14.
31. Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, et al. The 3D genome browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol. 2018;19(1):151.
32. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159(7):1665–80.
33. Kim-Hellmuth S, Aguet F, Oliva M, Munoz-Aguirre M, Kasela S, Wucher V et al. Cell type-specific genetic regulation of gene expression across human tissues. Science. 2020;369(6509).
34. Gallagher MD, Chen-Plotkin AS. The Post-GWAS era: from association to function. Am J Hum Genet. 2018;102(5):717–30.
35. Rao S, Yao Y, Bauer DE. Editing GWAS: experimental approaches to dissect and exploit disease-associated genetic variation. Genome Med. 2021;13(1):41.
36. Lu X, Chen X, Forney C, Donmez O, Miller D, Parameswaran S, et al. Global discovery of lupus genetic risk variant allelic enhancer activity. Nat Commun. 2021;12(1):1611.
37. Hui-Yuen JS, Zhu L, Wong LP, Jiang K, Chen Y, Liu T, et al. Chromatin landscapes and genetic risk in systemic lupus. Arthritis Res Ther. 2016;18(1):281.
38. Ray JP, de Boer CG, Fulco CP, Lareau CA, Kanai M, Ulirsch JC, et al. Prioritizing disease and trait causal variants at the TNFAIP3 locus using functional and genomic features. Nat Commun. 2020;11(1):1237.
39. Bourges C, Groff AF, Burren OS, Gerhardinger C, Mattioli K, Hutchinson A, et al. Resolving mechanisms of immune-mediated disease in primary CD4 T cells. EMBO Mol Med. 2020;12(5):e12112.
40. Inoue F, Kircher M, Martin B, Cooper GM, Witten DM, McManus MT, et al. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. Genome Res. 2017;27(1):38–52.
41. Ainsworth HC, Howard TD, Langefeld CD. Intrinsic DNA topology as a prioritization metric in genomic fine-mapping studies. Nucleic Acids Res. 2020;48(20):11304–21.
42. Wong L, Jiang K, Chen Y, Hennon T, Holmes L, Wallace CA, et al. Limits of Peripheral Blood mononuclear cells for gene expression-based biomarkers in Juvenile Idiopathic Arthritis. Sci Rep. 2016;6:29477.
43. Wu CY, Yang HY, Huang JL, Lai JH. Signals and mechanisms regulating monocyte and macrophage activation in the pathogenesis of Juvenile Idiopathic Arthritis. Int J Mol Sci. 2021;22(15).
44. Zhou Z, Xu MJ, Gao B. Hepatocytes: a key cell type for innate immunity. Cell Mol Immunol. 2016;13(3):301–15.
45. Jiang K, Wong L, Sawle AD, Frank MB, Chen Y, Wallace CA, et al. Whole blood expression profiling from the TREAT trial: insights for the pathogenesis of polyarticular juvenile idiopathic arthritis. Arthritis Res Ther. 2016;18(1):157.
46. Jiang K, Sawle AD, Frank MB, Chen Y, Wallace CA, Jarvis JN. Whole blood gene expression profiling predicts therapeutic response at six months in patients with polyarticular juvenile idiopathic arthritis. Arthritis Rheumatol. 2014;66(5):1363–71.

Publisher's Note