Supplemental Material

Title:
Ruminant-specific retrotransposons shape regulatory evolution of bovine immunity

Authors:
Conor J. Kelly[1], Carol Chitko-McKown[2], Edward B. Chuong[1]

Corresponding Author:
Edward B. Chuong (edward.chuong@colorado.edu)

Affiliations:
1. Department of Molecular, Cellular, and Developmental Biology and BioFrontiers Institute, University of Colorado Boulder, Boulder, CO, USA
2. USDA, ARS, Roman L. Hruska US Meat Animal Research Center (MARC), Clay Center, Nebraska, USA

This PDF file includes:
Supplemental Methods
Figs. S1 to S17
Captions for Tables S1 to S17
References

Other supplemental material for this manuscript includes the following:
Tables S1 to S17 (Excel format)

SUPPLEMENTAL METHODS

Isolation of primary monocyte and leukocytes

Monocytes and leukocytes were derived from peripheral blood mononuclear cells (PBMCs) from different individuals. 15 mL of blood were obtained from two male cattle by jugular venipuncture into 60 cc heparinized syringes. Both sampled animals are referred to as “line 999” and exhibit the following breed makeup: 1/2 Angus, 1/4 Gelbvieh, 3/16 Charolais, 1/16 Hereford. Monocytes and leukocytes were extracted from an 8.5 month old cryptorchid bull and 11 month old steer, respectively. Animal handling and isolation of bovine blood cells was approved by the U.S. Meat Animal Research Center (USMARC) Institutional Animal Care and Use Committee (#2.1).

15 mL of blood was combined with 15 mL 1× PBS and 15 mL Ficoll-Paque Plus and centrifuged for 45 minutes at 1,000 × g at room temperature. The PBMC layer was retained, washed in 50 mL of cold 1× PBS, and lysed in 1 mL Red Blood Cell Lysis Buffer for 5 minutes at 37°C and 5% CO_2. The PBMCs were washed three times with 50 mL, 10 mL, and 10 mL cold 1× PBS, respectively, and resuspended in 45 mL of FBS-free RPMI 1640. Monocytes were isolated as described in Chitko-McKown et al. 2004. PBMCs were added to treated T75 flasks and incubated at 37°C and 5% CO_2 overnight, and the non-adherent, predominantly leukocyte fraction was removed by washing with PBS. Monocytes were derived from the remaining adherent population, and leukocytes were derived from the non-adherent fraction. IFNG treatments were performed prior to separating adherent and non-adherent populations.

Lentivirus production

24 h prior to transfection 4×10^6 viable HEK293T cells were seeded in a coated 10cm dish in 10 mL Dulbecco’s DMEM, 1× penicillin-streptomycin, and 10% fetal bovine serum. Prior to transfection the media was replaced with antibiotic-free DMEM supplemented with 10% fetal bovine serum. For transfection FuGENE was used as follows: 20 ug psPax2, 2 ug pCMV-VSV-G, 16 ug pRDA_174 or pRDA_052_crRNA, 114 µL FuGENE, and Opti-MEM (up to 2.5 mL) were combined and incubated at room temperature for 20 minutes. The mixture was added dropwise to the dish containing HEK293T cells, and the cells were incubated at 37°C and 5% CO_2. After 6 h the media was replaced with DMEM supplemented with 1× penicillin-streptomycin and 10% fetal bovine serum. Virus was harvested 36 h after this media change and filtered using a 0.45 um PES filter.

Generation and analysis of CRISPR-enAsCas12a mutants

Bovine MDBK cells (gift from Sara Sawyer) were cultured in MEM supplemented with nonessential amino acids, 1× penicillin-streptomycin, and 10% fetal bovine serum. To generate an MDBK line that stably expresses *enAsCas12a*, the following procedure was used: 24 h prior to transduction, 1×10^5 viable MDBK cells were seeded into coated 6-well plates. Before transduction the media in each well was swapped with MEM supplemented with 10% fetal bovine serum and 8 ug/mL polybrene for a final volume of 2 mL after the addition of lentivirus. Each well was transduced with different volumes of pRDA_174 lentivirus (50, 200, 600, 1000 µL). After 24 h the media was replaced with fresh complete MEM. 24 h after this media change 4 ug/mL blasticidin was added to each well to select for cells stably expressing *enAsCas12a*. This selection continued for a total of 14 days, after which one well (200 µL) was chosen for further use. The presence of integrated *enAsCas12a* was validated by gDNA PCR (Supplemental Table S16). Clonal lines were isolated using array dilution method with 25% conditioned MEM, and 6 clones were screened for expression of integrated *enAsCas12a* by RT-qPCR using the Luna Universal One-Step RT-qPCR Kit according to the manufacturer’s instructions (Supplemental

Table S16). Two clones were selected for further experimentation and maintained under 2 ug/mL blasticidin selection.

To generate each MER41 deletion, clonal MDBK cells stably expressing *enAsCas12a* were transduced with pRDA_052_cRNA lentivirus containing target gRNAs, following the same procedure above (Supplemental Table S16). 48 h after transduction each well was treated with 2 ug/mL puromycin. Puromycin selection continued for a total of 7 days. Evidence of editing was confirmed using gDNA PCR. Cells that were transduced with 200 µL lentivirus were clonally expanded using the array dilution method with 25% condition MEM, and 80-120 clones were screened for homozygous deletions by PCR using both internal and flanking primer pairs at the expected deletion site. To determine deletion breakpoint sequences, PCR products flanking each deletion site were cloned into pJET and transformed into NEB 5-alpha Competent *E. coli*, and at least 2 individual colonies were sequenced using Sanger sequencing (Supplemental Table S16). Independently, to control for any variation introduced through a second clonal expansion, clonal MDBK lines stably expressed *enAsCas12a* without any gRNAs underwent an additional clonal isolation using the array dilution method. 5 control clones were selected for further experimentation.

Preparation of gRNA constructs

For each MER41 element (associated with *IFNAR2* and *LOC510185*), two 23bp gRNA sequences were designed to generate a single internal deletion that encompasses the putative STAT1 binding sites (Supplemental Table S16). Each pair of gRNAs was synthesized as a single crRNA oligonucleotide and included BsmBI sites with accompanying overlaps (bolded sequences) as explained in DeWeirdt et al. 2021:

5’-[handle]CGTCTCA**AGAT**[gRNA1][DR-wt][gRNA2][DR-1]TTTTTT***GAAT***CGAGACG[handle]-3’

Sequence was added to the 5’ and 3’ ends to help facilitate BsmBI recognition and to meet the 300bp requirement for synthesis through Twist Bioscience. Only gRNA sequences with a first percentile (see DeWeirdt et al. 2021) 5’ PAM were considered. All gRNA sequences were verified to uniquely target the locus of interest using BLAT (Kent 2002) against the bosTau9 assembly. crRNA oligonucleotides were synthesized by Twist Bioscience as a single, 300bp oligonucleotide or precloned into pTwist-Amp-High.

Each crRNA oligonucleotide was cloned into pRDA_052 following a modified protocol from (DeWeirdt et al. 2021). Briefly, crRNA oligonucleotides were cloned into pRDA_052 using 0.10 fmol crRNA oligonucleotide, 0.05 fmol pRDA_052, 2.00 µL T4 DNA Ligase Buffer, 1.00 µL T4 DNA Ligase, 1.00 µL BsmBI-v2, and nuclease-free water up to 20 µL. crRNA olignonucleotides that were precloned into pTwist-Amp-High were subcloned into pRDA_052 using 75 ng pTwist-Amp-High-crRNA, 75 ng pRDA_052, 2.00 µL T4 DNA Ligase Buffer, 1.00 µL T4 DNA Ligase, 1.00 µL BsmBI-v2, and nuclease-free water up to 20 µL. PCR cycling conditions: 60 minutes at 42°C, 5 minutes at 60°C, and 10 minutes at 65°C. 2 µL of each ligation reaction was used to transform NEB Stable Competent *E. coli*. Plasmid DNA was harvested using Zymo ZymoPURE II Plasmid Midiprep Kit, and the sequence of each construct was verified by Sanger sequencing (Quintara Biosciences, Fort Collins, CO; Genewiz, South Plainfield, NJ) (Supplemental Table S16).

ATAC-seq

ATAC-seq libraries were prepared following a published protocol (Corces et al. 2017) using 5×10^4 viable (>90%) MDBK cells. 30 minutes before harvesting, cells were treated with 200 U/mL DNase I for

30 minutes. Each 5×10^4 sample was transposed using 100 nM transposase as specified in Corces et al. 2017. Preamplifcation was performed by combining 20 µL transposed sample, 2.5 µL 25 µM i5 primer, 2.5 µL 25 µM i7 primer, and 25 µL 2x NEBNext Ultra II Q5 Master Mix. PCR cycling conditions: 30 s at 98°C, 5x (10 s at 98°C, 30 s at 63°C, 45 s at 65°C). qPCR was performed as follows: 5 µL of preamplified sample was combined with 0.5 µL 25 µM i5 primer, 0.5 µL 25 µM i7 primer, 3.75 µL nuclease-free water, 0.25 µL 25× SYBR Gold, and 5 µL 2x NEBNext Ultra II Q5 Master Mix. qPCR cycling conditions: 30 s at 98°C, 20x (10 s at 98°C, 30 s at 63°C, 45 s at 65°C). Preamplified samples were amplified for another 7 cycles. Final libraries were cleaned using a 0.5×/1.3× double-sided bead cleanup using KAPA Pure Beads and quantified using Qubit dsDNA High Sensitivity and TapeStation 4200 HSD5000. Libraries were pooled and sequenced on an Illumina NovaSeq 6000 (University of Colorado Genomics Core) as 150bp paired-end reads.

CUT&RUN

CUT&RUN pulldowns were generated using a published protocol (Janssens and Henikoff 2019; Meers et al. 2019). All buffers were prepared according to the "High Ca^{2+}/Low Salt" section of the protocol using 0.04% digitonin. 5×10^5 viable cells were used for each pulldown. The following antibodies were used: rabbit anti-mouse IgG (1:100), rabbit anti-H3K27ac (1:100), rabbit anti-pRPB1-Ser5 (1:50; POLR2A), rabbit anti-STAT1 (1:100), rabbit anti-pSTAT1-Ser727 (1:100) (Supplemental Table S16). pA-MNase (gift from Steve Henikoff) was added to each sample following primary antibody incubation at a final concentration of 700 ng/mL. Chromatin digestion, release, and extraction was carried out according to the published protocol (Janssens and Henikoff 2019; Meers et al. 2019). Pulldown success was determined by Qubit dsDNA High Sensitivity and TapeStation 4200 HSD5000 before proceeding with library preparation.

Libraries were generated using a modified protocol for use with the KAPA HyperPrep Kit. Briefly, the full volume of each pulldown (30 µL) was diluted to 50 µL in nuclease-free water, and libraries were generated following the manufacturer's protocol with the following modifications. Freshly diluted 0.200 µM single-index adapters were added to each library at a low concentration (9 nM) to minimize adapter dimer formation. Adapter-ligated libraries were treated with 0.2% SDS and 0.4 mg/mL Proteinase K for 1 h at 37°C and underwent a double-sided 0.8×/1.0× cleanup with KAPA Pure Beads. Purified, adapter-ligated libraries were amplified using the following PCR cycling conditions: 45 s at 98°C, 15x (15 s at 98°C, 10 s at 60°C), 60 s at 72°C. Amplified libraries underwent a double-sided 0.8×/1.0× cleanup. The final libraries were quantified using Qubit dsDNA High Sensitivity and TapeStation 4200 HSD5000. Libraries were pooled and sequenced on an Illumina NovaSeq 6000 (University of Colorado Genomics Core) as 150bp paired-end reads.

RT-qPCR

RNA was extracted from MDBK cells using the Zymo Quick RNA Miniprep Plus Kit, following the manufacturer's instructions. Prepared lysates were stored at -80°C until processing, and single-use aliquots of extracted RNA were stored -80°C until library preparation. Total RNA samples for each treatment condition and clonal replicate were prepared in three biological replicates. A spectrophotometer was used to determine RNA concentration and purity; all samples passed quality assessment. RNA expression levels for *CTCF*, *IL2RB*, *LOC510185*, and *IFNAR2* were quantified using the Luna Universal One-Step RT-qPCR Kit according to the manufacturer's instructions (Supplemental Table S16). In brief, for each reaction 25 ng of RNA was combined with 5 µL 2× Luna Universal One-Step Reaction Mix, 0.5 µL 20× Luna WarmStart RT Enzyme Mix, 0.4 µL 10 µM forward primer, and 0.4 µL 10 µM reverse primer. Reactions were amplified using a CFX384 Touch Real-Time PCR Detection

System (Bio-Rad) with the following PCR conditions: 10 min at 55°C, 1 min at 95°C, 40x (10 s at 95°C, 30 s at 60°C). On-target amplification was assessed by melt curve analysis. Statistical significance was assessed using a two-tailed unpaired Student's *t*-test with a threshold of $p < 0.05$.

RNA-seq

RNA was extracted from MDBK and BL3.1 cells using the Zymo Quick RNA Miniprep Plus Kit. Monocyte and leukocyte lysates were prepared using Trizol, and RNA was extracted using the Zymo Direct-zol RNA Miniprep Kit. All RNA lysates were stored at -80°C until processing, and single-use aliquots of extracted RNA were stored -80°C until library preparation. RNA integrity was quantified using High Sensitivity RNA TapeStation 4200. PolyA enrichment and library preparation was performed using the KAPA mRNA HyperPrep Kit according to the manufacturer's protocols. Briefly, either 200 or 500 ng of RNA was used as input. 1.5 µM single-index or unique dual-index adapters were added at a final concentration of 7 nM. Purified, adapter-ligated library was amplified for a total of 10 (500 ng) or 12 (200 ng) cycles following the manufacturer's protocol. The final libraries were pooled and sequenced on a NovaSeq 6000 (University of Colorado Genomics Core) as 150bp paired-end reads.

RNA-seq analysis

Adapters and low-quality reads were trimmed using BBDuk v38.05 (https://sourceforge.net/projects/bbmap/) using options 'ktrim=r k=23 mink=11 hdist=1 maq=10 qtrim=r trimq=10 tpe tbo literal=AAAAAAAAAAAAAAAAAAAAAA'. Library quality was assessed using FastQC v0.11.8 and MultiQC v1.7 (Ewels et al. 2016), and trimmed reads were aligned to the bosTau9 (ARS-UCD1.2) assembly using HISAT2 v2.1.0 (Kim et al. 2019) with options '--rna-strandness RF --no-softclip' (Supplemental Table S1). Only uniquely aligned fragments (MAPQ >= 10) were retained using samtools v1.10 (Li et al. 2009). Alignments from technical replicates were merged after alignment. For visualization aligned fragments were converted to stranded, CPM normalized bigWigs using deepTools bamCoverage v3.0.1 (Ramírez et al. 2014) with options '--normalizeUsing CPM --ignoreForNormalization ChrX ChrM'. Aligned fragments were assigned to the complete bosTau9 RefSeq gene annotation in a stranded manner using featureCounts v1.6.2 (Liao et al. 2014) with options '-s 2 -t exon -g gene_id', and differentially expressed genes between stimulated and unstimulated cells were called using DESeq2 v1.26.0 (Supplemental Table S1) (Love et al. 2014). We found that *LOC510185* is incorrectly annotated in Ensembl. By manually inspecting RNA-seq reads and running genome-guided transcriptome assembly, we confirmed that the complete RefSeq annotation for *LOC510185* (XM_010805353.2; Chr5:75584534-75591455) is correct. For visualization, $\log_2$FC values were shrunken using the apeglm function v1.8.0 (Zhu et al. 2019). Gene ontology analysis was performed using gProfiler with an FDR cutoff of 0.05 (Supplemental Table S2) (Raudvere et al. 2019).

ISGs were defined as genes with a false discovery rate of at least 0.05 and $\log_2$FC greater than 0. Nonresponsive genes were defined using the following cutoffs: 1) baseMean >= 100, FDR >= 0.90, $\log_2$FC >= -0.05, $\log_2$FC <= 0.05. Loci were collapsed to their transcriptional start site to determine relative distances from IFNG-inducible H3K27ac peaks and TEs. The Venn diagram was prepared using Intervene v0.6.4 (Khan and Mathelier 2017).

Mutant RNA-seq analysis

Mutant RNA-seq data were analyzed in a manner similar to the wildtype RNA-seq (see above) with adjustment to the differential expression analysis. Differentially expressed genes were called using DESeq2 v1.26.0 (Love et al. 2014) using a two-factor comparison incorporating both genotype (wildtype/mutant) and treatment (untreated/IFNG) information (Supplemental Table S11). All three

MER41_BT.IL2RB clones were included in the analysis. Out of the five clones harboring the MER41_BT.IFNAR2 deletion, and we selected two clones (KOs #2, #3) that showed the desired deletion to define differentially expressed genes by RNA-seq (Supplemental Fig. 13). Independent mutant clones were treated as biological replicates and reported expression differences were called comparing mutant IFNG against wildtype IFNG. As controls, we used: 1) the parent MDBK clonal line stably expressing Cas12 and 2) a MDBK/Cas12 clonal line recovered from the MER41.IFNAR2 gRNA array dilution screen. Sanger sequencing of the deletion locus in this clone revealed evidence of minor edits at the gRNA target sites but not the intended deletion of MER41_BT.IFNAR2 (Supplemental Figure S13). This clone was used as an effective control replicate. Two replicates were included for both controls and MER41_BT.IFNAR2 deletions. One replicate was included for all MER41_BT.LOC510185 deletions. For visualization, $\log_2$FC values were shrunken using the ashr function v2.2-47 (Stephens 2017). Gene loci were collapsed to their transcriptional start site to determine relative distances to each deletion site using BEDTools v2.28.0 (Quinlan and Hall 2010).

ATAC-seq analysis

Adapters and low-quality reads were trimmed using BBDuk v38.05 (https://sourceforge.net/projects/bbmap/) using options 'ktrim=r k=34 mink=11 hdist=1 tpe tbo qtrim=r trimq=10'. Trimmed reads were aligned to the bosTau9 assembly (chromosomal and mitochondrial scaffolds only) using Bowtie 2 v2.2.9 (Langmead and Salzberg 2012) with options '--end-to-end --very-sensitive -X 1000 --fr', and only uniquely mapping reads with a minimum MAPQ of 10 were retained (Supplemental Table S3). Fragments aligning to the mitochondrial genome were removed, and duplicates were removed using sambamba v0.6.9. Remaining fragments were used to call ATAC-seq peaks with an FDR < 0.05 using MACS2 v2.1.1 (Liu 2014) with options '--keep-dup all --format BAMPE'. Scores corresponding to the fraction of reads in called peak regions (FRIP) and transcriptional start site enrichment using the complete RefSeq bosTau9 annotation were calculated to assess library quality (Supplemental Table S3). One untreated replicate was removed from the analysis as it yielded significantly fewer peaks and lower FRIP and transcriptional start site enrichment scores than other samples. For intersecting with CUT&RUN data or TEs, replicate peak files (where available) were concatenated, and peaks that fall within 100 bp of another were merged using BEDTools v2.28.0 (Quinlan and Hall 2010). Normalized profiles corresponding to read coverage per 1 million reads were used for heatmap and metaprofile visualization, which were generated using deepTools v3.0.1 (Ramírez et al. 2014).

CUT&RUN analysis

Alignment and peak calling. Adapters and low-quality reads were trimmed using BBDuk v38.05 (https://sourceforge.net/projects/bbmap/) using options 'ktrim=r k=34 mink=11 hdist=1 tpe tbo qtrim=r trimq=10'. Trimmed reads were aligned to the bosTau9 assembly using BWA-MEM v0.7.15 (Li 2013), and only uniquely mapping reads with a minimum MAPQ of 10 were retained (Supplemental Table S4). Fragments aligning to the mitochondrial genome were removed, and alignments corresponding to replicate samples were merged. Aligned fragments from transcription factor pulldowns were subset using a 150 bp cutoff using deepTools v3.0.1 (Ramírez et al. 2014). Peak calling was performed using complete and size subsetted alignment files with MACS2 v2.1.1 (Liu 2014) in a two-step process where separate sets of peaks were called with 1) single-end options '--format BAM --shift=-75 --extsize=150' and 2) paired-end option '--format BAMPE'. For both modes only peaks with a *p*-value < 0.01 were retained, and all libraries were normalized against respective IgG control libraries. Independently, peaks were called from IgG control libraries using MACS2 v2.1.1 (Liu 2014) with paired-end option '--format BAMPE' for visualizing background signal enrichment. Normalized bigWigs corresponding to

read coverage per 1 million reads were used for heatmap and metaprofile visualization, which were generated using deepTools v3.0.1 (Ramírez et al. 2014). Only peaks from complete (not size-subset) transcription factor libraries were used for further analyses, although size-subset bigWigs were used for genome browser screenshots.

Identifying IFNG-inducible CUT&RUN peaks. All merged peaks for a particular pulldown (across all replicates, untreated and IFNG-stimulated) were merged, and aligned fragments were counted for each peak using BEDTools v2.28.0 (Quinlan and Hall 2010). IFNG-inducible and IFNG-downregulated peaks were called using DESeq2 v1.26.0 with an FDR < 0.05 and a $\log_2$ fold change ($\log_2$FC) > 0 and < 0, respectively (Supplemental Table S4). Nonresponsive H3K27ac peaks were defined using the following cutoffs: mean normalized count >= 100 or 500, FDR >= 0.90, $\log_2$FC >= -0.075, $\log_2$FC <= 0.075. For signal heatmaps, signal metaplots, and genome browser screenshots, differentially enriched H3K27ac peaks were centered on concatenated and merged (within 100 bp) ATAC peaks to improve resolution. Motif analysis was performed using XSTREME v5.4.1 (Grant and Bailey 2021) with options '--minw 6 --maxw 20 --streme-motifs 20 --align center' querying against the JASPAR CORE 2018 vertebrates database (Supplemental Table S5) (Fornes et al. 2020).

Enrichment near IFNG-stimulated genes. The top 750 ISGs were extracted from the larger list of 1496 ISGs sorted by descending $\log_2$FC. The absolute distance to the nearest ISG was determined for all 4525 IFNG-inducible H3K27ac peaks using BEDTools v2.28.0 (Quinlan and Hall 2010). The expected background was determined by random shuffling using BEDTools v2.28.0 (Quinlan and Hall 2010). Additionally, 4525 nonresponsive H3K27ac peaks were extracted from a larger list of 5011 peaks by random sampling. Statistical significance was determined for the first 20kb bin by Fisher's exact test using BEDTools v2.28.0 (Quinlan and Hall 2010). The analysis was repeated using two additional gene sets: 1) top 750 significantly downregulated genes, sorting by ascending $\log_2$FC and 2) randomly sampled 750 nonresponsive genes.

Activity-by-Contact (ABC) analysis

Publicly available Hi-C data from a Brangus individual was downloaded from the SRA using accession SRR6691720. Reads were aligned to the bosTau9 assembly using BWA-MEM v0.7.17 (Li 2013) with arguments '-SP5M', and the resulting bam file was converted to pairsam format using pairtools parse v0.2.2 (https://github.com/mirnylab/pairtools). Duplicate reads were marked using pairtools dedup v0.2.2 (https://github.com/mirnylab/pairtools), and only aligned fragments with pairtools classification "UU" or "UC" were retained using pairtools filter v0.2.2 (https://github.com/mirnylab/pairtools) resulting in approximately 80M pairs. A Knight Ruiz (KR)-normalized, Sau3AI restriction site-aware Hi-C matrix was prepared using juicer pre v1.22.01 (Durand et al. 2016) at 15kb resolution and fitted to a powerlaw distribution in preparation for running the ABC model (Fulco et al. 2019). The resolution (15kb) and depth (80M uniquely mapping alignment pairs) were insufficient to identify contacts at high-throughput using existing tools. However, we were able to confirm that each putative enhancer falls within a 1 Mb domain by visually inspecting the Hi-C contact matrix.

Despite low resolution and depth, we were able to apply the Hi-C matrix in conjunction with ATAC-seq and H3K27ac data from IFNG-stimulated MDBK cells to predict enhancer activity using the Activity-by-Contact (ABC) model (Fulco et al. 2019). The ABC model predicts enhancer-gene contacts by leveraging epigenomic and chromatin capture data. Each potential enhancer-gene connection is assigned an ABC interaction score that depends on the activity of the enhancer by ATAC-seq and CUT&RUN in addition to the likelihood of contact by Hi-C. Importantly, the ABC model can make

predictions for a given cell type without cell-type-specific Hi-C data, however it requires that all epigenomic data be sourced from the same cell type (Fulco et al. 2019). We ran the ABC model as previously described in (Fulco et al. 2019). In brief, we first identified candidate enhancer elements using makeCandidateRegions.py with options '--peakExtendFromSummit 250 –nStrongestPeaks 150000' and quantified activity using run.neighborhoods.py. Predicted enhancer-gene pairs were attributed an ABC interaction score using predict.py with options '--hic_type juicebox –hic_resolution 15000 –scale_hic_using_powerlaw –threshold 0.02 –make_all_putative'. Predicted enhancer regions with an ABC interaction score over 0.001 were intersected with IFNG-inducible ($log_2FC > 0$, FDR < 0.05) H3K27ac regions or randomly shuffled IFNG-inducible H3K27ac regions. Putative enhancers were further selected if they were predicted to interact with an ISG ($log_2FC > 0$, FDR < 0.05) based on the MDBK RNA-seq (Supplemental Table S6). We used relaxed thresholds to fully capture all possible enhancer-gene interactions, however we further filtered for enhancer-gene interactions with an ABC interaction score greater than 0.01. Although the MER41_BT.IFNAR2 element was predicted to interact with *IFNAR2*, there is minimal evidence for MER41_BT.IL2RB interacting with *IL2RB*, presumably due to the absence of epigenomic signal from MDBK that overlaps *IL2RB* (Supplemental Table S6).

Transposable element analysis

All TE analysis in this section was performed using the RepeatMasker annotations for bosTau9 available on UCSC (updated 2018-11-07). For motif analyses, binding motif position-weight matrices for STAT1 (MA0137.3 for Gamma Activated Site/GAS motif, MA0517.1 for the Interferon Stimulated Response Element/ISRE motif) were obtained from the JASPAR CORE 2018 vertebrate database (Fornes et al. 2020). Motif occurrences within TE consensus sequences or the bosTau9 cattle assembly were identified using FIMO v5.0.3 (Grant et al. 2011) using a *p*-value cutoff of 1×10^{-4} (TE heatmaps) or 1×10^{-3} (TE motif analysis).

Identifying TE-derived inducible enhancers. IFNG-inducible peaks were defined using H3K27ac, but H3K27ac-enriched regions are too broad (>1 kb) and often encompass multiple TEs. We refined peak locations by centering them on ATAC-seq summits (1 bp) contained within the H3K27ac-defined region using BEDTools v2.28.0 (Quinlan and Hall 2010), with the rationale that these summits represent key transcription factor binding sites within the enhancer. We then asked whether these summits overlapped an annotated TE using BEDTools v2.28.0 (Quinlan and Hall 2010) (Supplemental Table S7).

Identifying overrepresented TE families. To assess family-level enrichment, GIGGLE v0.6.3 (Layer et al. 2018) was used to create a database of all TEs in the bosTau9 genome. Merged +IFNG ATAC-seq and CUT&RUN peaks were then queried against the TE database. Results were ranked by descending GIGGLE enrichment score, and enriched TE families were identified according to the odds ratio, Fisher's two-tail *p*-value, and number of overlaps (Supplemental Table S8).

TE heatmaps. To assess ATAC-seq and CUT&RUN normalized signal pileup, a list of 101 and 11928 TEs were subset from all 5491 MER41_BT and 362131 Bov-A2 elements, respectively. TEs were retained if they: 1) overlapped merged +IFNG ATAC-seq peaks, 2) do not fall within 5kb of another ATAC-overlapping TE, 3) do not fall within 5kb of a gene TSS (RefSeq annotation downloaded from the UCSC bosTau9 assembly), and 4) overlap a putative GAS or ISRE motif as defined by FIMO (Grant et al. 2011) with a *p*-value cutoff of 1×10^{-4}. Signal from CPM normalized bigWigs was plotted over subset TEs as heatmaps using deepTools v3.0.1 (Ramírez et al. 2014).

TE motif analysis. For the heatmap visualization of motif presence, only MER41_BT that were 50% full-length relative to consensus were retained. Bov-A2 that were within 5% of full-length were retained. Repeat 5' start coordinates were recalculated based on their alignment to the consensus using the RepeatMasker annotations. For all TE motif heatmaps, motif presence was plotted using putative GAS or ISRE motifs as defined by FIMO v5.0.3 (Grant et al. 2011) with a *p*-value cutoff of 1×10^{-3}.

Enrichment near IFNG-stimulated genes. The top 750 ISGs were extracted from the larger list of 1496 ISGs sorted by descending $\log_2$FC. The absolute distance to the nearest ISG was determined for 4198 TE-derived IFNG-inducible H3K27ac peaks (centered using ATAC-seq summits as described above). The expected background was determined by randomly shuffling ATAC-centered IFNG-inducible H3K27ac peaks using BEDTools v2.28.0 (Quinlan and Hall 2010). 3996 ATAC-centered nonresponsive H3K27ac peaks were used as an additional comparison. Statistical significance was determined for the first 20kb bin by Fisher's exact test using BEDTools v2.28.0 (Quinlan and Hall 2010). The analysis was repeated using two additional gene sets: 1) top 750 significantly downregulated genes, sorting by ascending $\log_2$FC and 2) randomly sampled 750 nonresponsive genes.

Re-annotation of MER41 repeats: High quality assemblies from 30 mammalian species representing 5 distinct lineages (primate, cetartiodactyla, perissodactyla, pholidota, carnivora) were annotated using RepeatMasker v4.1.0 (Smith A, Hubley R. RepeatModeler Open-1.0. In: RepeatMasker Open-4.0) with options '-e rmblast -s -gccalc' using the RepeatMasker-Repbase (RMRB, release 20181026) library (Supplemental Table S12). We included all MER41-like families annotated in cetartiodactyla (MER41_Vpa, MER41_SS, MER41_BT, and MER41_Ttr) and included primate MER41 (MER41_Hs) as a control. For each assembly, the numbers of annotated MER41, Bov-A2, and BovB elements from 24 assemblies were visualized in a bubble plot. For MER41 and Bov-A2, only elements at least >80% (MER41) and >50% (Bov-A2) of the consensus length were considered to avoid overcounting and false positive matches. BovB elements were not filtered by length to retain copies that may have undergone 5' truncation as a consequence of the copy-and-paste mechanism (Pasquesi et al. 2018). Species divergence times and phylogeny were obtained from TimeTree (Kumar et al. 2017). To determine MER41 family relatedness, annotated MER41 elements across all cetartiodactyl families were collapsed into a list of 7556 elements, and only non-overlapping elements that were between 400 and 800 nt were extracted (n=990). Consensus sequences for a representative set of families including all cetartiodactyl MER41, carnivore MER41_Cf, and primate MER41B were also included. All elements were used to define a set of unique 6-mers, and the number of occurrences of each unique 6-mer in each sequence was counted using kmer-counter (https://github.com/alexpreynolds/kmer-counter). The resulting counts matrix was used for PCA analysis using PCAtools v2.4.0 (https://github.com/kevinblighe/PCAtools).

Analysis of TE syntenic conservation: Evolutionary lineage placements based on copy number were independently verified by analysis of presence or absence of syntenic MER41 families using the TE_Orthology script (https://github.com/4ureliek/TEorthology) (Kapusta et al. 2017). For each annotated cetartiodactyla assembly, all MER41 insertions that are greater than 200 bp and not nested in another annotated repeat were extracted including 100 nt 5' and 3' flanking sequences. This allows for greater confidence that the queried TEs belong to the family as annotated and makes for a meaningful inter-species comparison. Extracted MER41 elements were queried against the other 18 cetartiodactyla assemblies using BLASTN v2.7.1 (Camacho et al. 2009) with an *E*-value cutoff of 1×10^{-50}. Hits are filtered based on the presence of at least 50% 5' and 3' flanking sequence, and the total number and number of potentially orthologous matches are recorded (Supplemental Table S13). In

summary, this approach captures 1656 out of 1658 unique cattle MER41_BT elements (with flanking sequence) as perfect self-matches.

Detection of TE insertion polymorphisms: Publicly available whole genome, paired-end sequencing data (mean read length: 148bp) were downloaded using fasterq-dump v2.10.5. Adapters and low quality reads were trimmed using BBDuk v38.05 (https://sourceforge.net/projects/bbmap/) using options 'ktrim=r k=34 mink=11 hdist=1 tpe tbo qtrim=r trimq=10'. Trimmed reads were aligned to the bosTau9 reference assembly with the bosTau5 Y Chromosome assembly using BWA-MEM v0.7.15 (Bellott et al. 2014; Li 2013). Technical replicates were merged using Picard v2.6.0 (http://broadinstitute.github.io/picard/), and coverage was estimated using QualiMap v2.2.1 (Okonechnikov et al. 2016).

MELT v2.1.5 (Gardner et al. 2017) was used to call TE variants. Briefly, MELT extracts discordant and split reads from alignment files, queries them against a TE consensus and fixed coordinates using a prepared TE .zip file, and outputs a VCF file that provides the genotype for each sample and variant call. TE .zip files were prepared for Bov-A2, BovB, and MER41_BT using the Repbase v24.02 consensus sequences allowing up to 10 mismatches per 100 bases. TE insertions were called following the MELT-Split pipeline using option '-r 150' and supplying the estimated average coverage for each sample. TE deletions were called following the MELT-Deletion pipeline. Insertion calls were retained if they met the following criteria: within 10% of full length (Bov-A2 only), ASSESS > 3, SR > 3, no null genotypes in any sample. Deletion calls were retained as long as they were within 10% of full length (Bov-A2 only) and exhibited no null genotypes in any sample. Filtered and insertion variant calls were aggregated (Supplemental Table S14). For both the PCA and histogram, aggregated variant call genotypes were collapsed by whether or not they had a single allele supporting the variant. The PCA of the aggregated TE variant genotype calls was performed using PCAtools v2.4.0 (https://github.com/kevinblighe/PCAtools). Filtered deletion calls were called as being epigenetically marked if they overlapped any raw ATAC-seq or H3K27ac, POLR2A, STAT1, or pSTAT1 peak, and the absolute distance from each to all gene transcriptional start sites within 250kb were determined using BEDTools v2.28.0 (Supplemental Table S15) (Quinlan and Hall 2010). Alignments over filtered deletions were visualized using Samplot v1.1.6 (Belyeu et al. 2021) with option '-d 100'.

LOC510185 sequence analysis

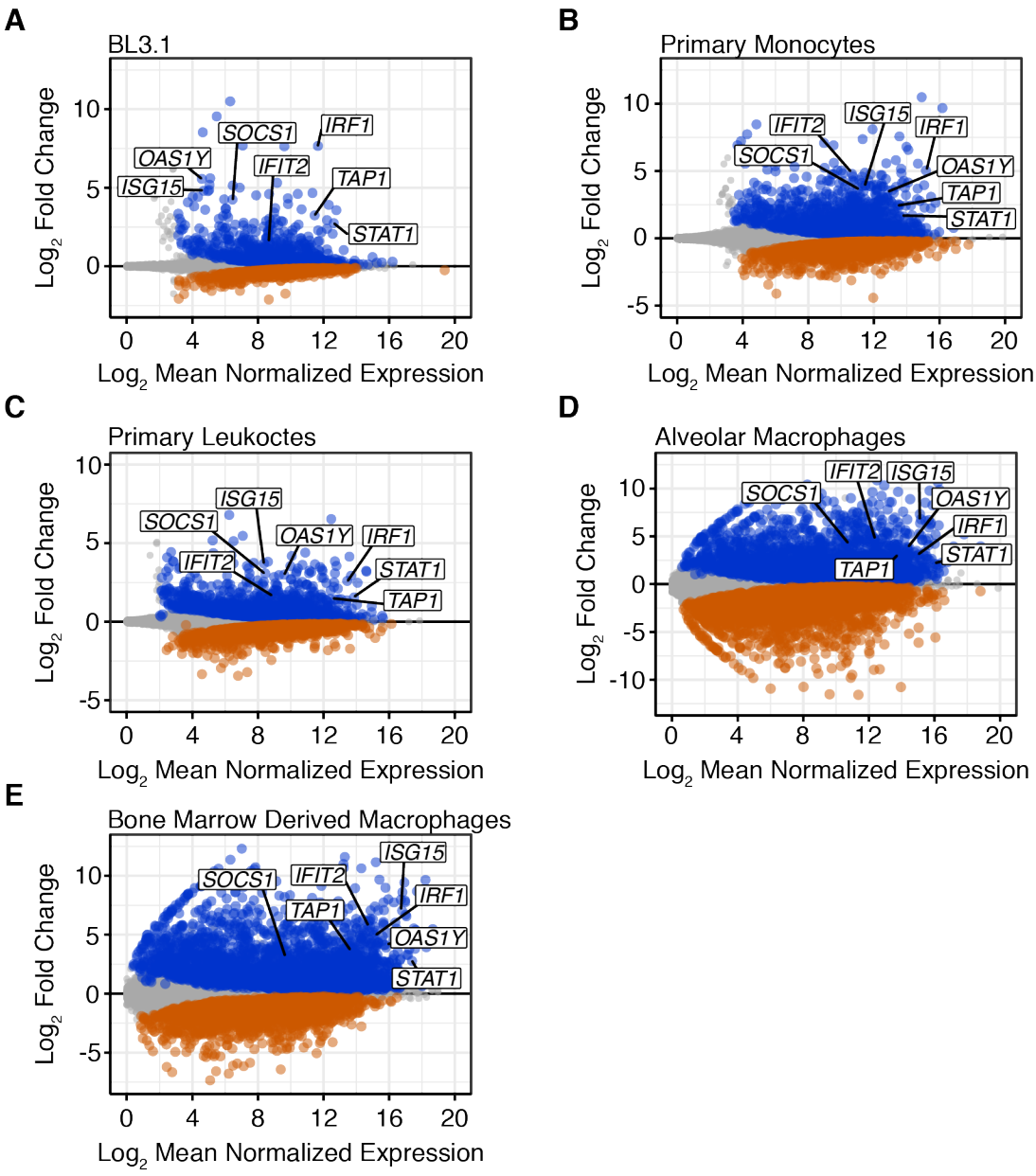
The inferred protein sequence for LOC510185 (RefSeq accession XP_010803655.1) and protein sequences for bovine (UniProt accession F1N409) and human (Uniprot P14784) IL2RB were aligned using MUSCLE v3.8.1551 (Edgar 2004). The resulting alignment was annotated according to homology to human IL2RB according to UniProt. The time of tandem duplication was estimated as described in (Hu et al. 2012). In brief, the bovine LOC510185 and IL2RB sequences were aligned using CLUSTAL O v1.2.4 (McWilliam et al. 2013). The aligned sequences were converted into their original cDNA sequences using PAL2NAL v14 (Suyama et al. 2006), which estimated synonymous (dS) and nonsynonymous (dN) substitution rates using CODEML v4.9j (Yang 2007). The time of duplication was estimated using the following equation: $T = dS/2\lambda$ ($\lambda = 7.76 \times 10^{-10}$) (Liu et al. 2006).

To identify candidate LOC510185 orthologs in cetartiodactyl species outside of cattle, we queried the bovine IL2RB amino acid sequence (UniProt accession F1N409) against a panel of 26 cetartiodactyl species using TBLASTN v2.13.0 (Gertz et al. 2006) (Supplemental Table S9). Among the species included were camelids (n=4), pig (n=1), whippomorphs (n=11), and pecorans (n=10). Mouse deer was excluded due to poor gene annotation quality. *IL2RB* orthologs were identified according to the

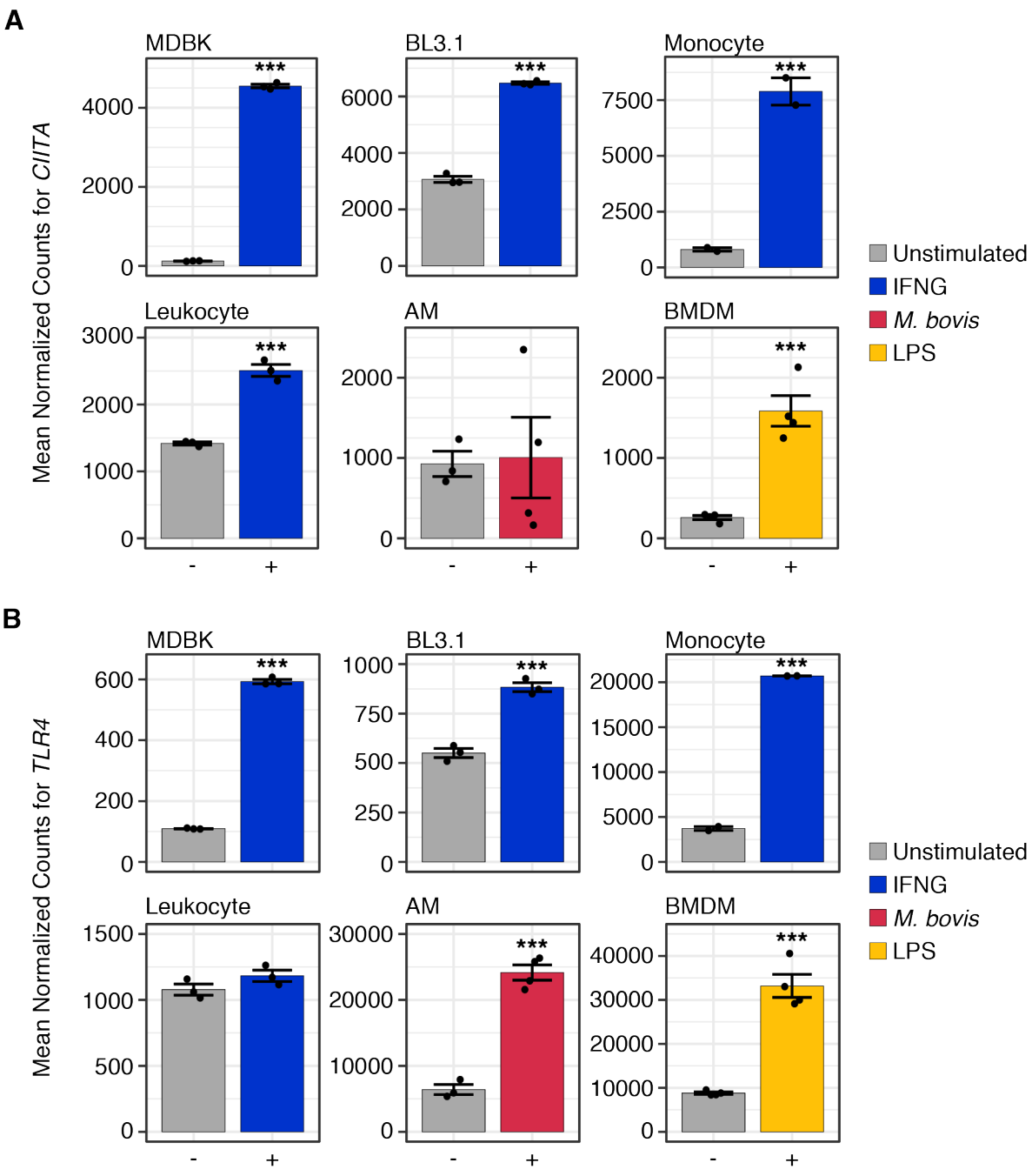
following thresholds: *E*-value < 1×10^{-100}, query coverage > 70%. Other matches were considered as *LOC510185* orthologs according to the following thresholds: *E*-value < 1×10^{-50}, query coverage < 50%, percent identity > 70%. The tandem gene duplication was present only in 9 of the 10 pecoran annotations, 8 of which fell within 35kb upstream of the *IL2RB* ortholog and were in the same orientation. The tandem gene duplication was absent in the scimitar oryx annotation, however querying the IL2RB cDNA sequence against the full assembly using BLASTN v2.7.1 (Camacho et al. 2009) revealed a high-confidence match (83% identity, 25% query, 6×10^{-161}). Repeating this analysis with the java mouse deer (*T. javanicus*) assembly revealed an *IL2RB*, but not *LOC510185*, ortholog.

Reanalysis of publicly available ChIP-seq

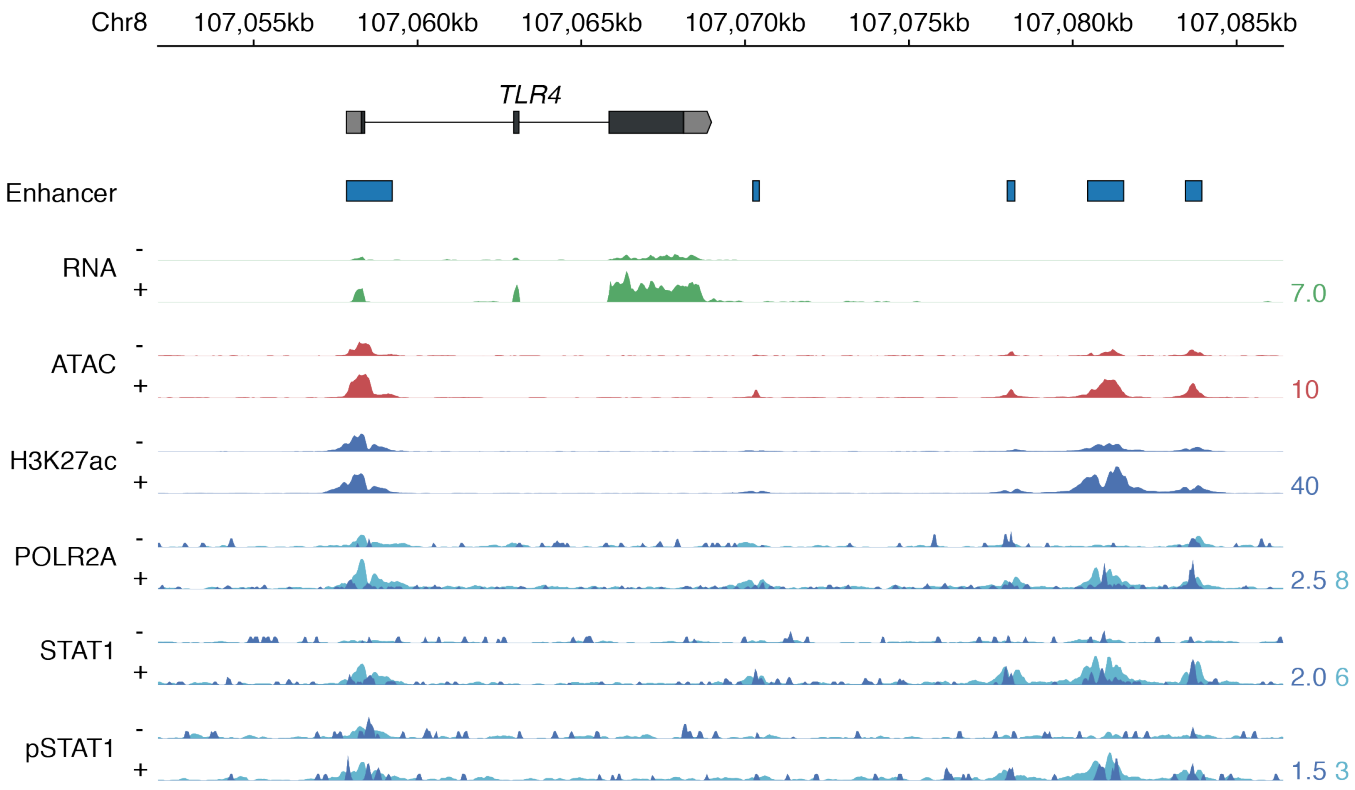
For the bigWig tracks visualized in Supplemental Fig. S17, publicly available ChIP-seq data from bovine alveolar macrophages challenged with *M. bovis* were downloaded using accession GSE116734 (Hall et al. 2019). Adapters and low-quality reads were trimmed using BBDuk v38.05 (https://sourceforge.net/projects/bbmap/) using options 'ktrim=r k=34 mink=11 hdist=1 tpe tbo qtrim=r trimq=10'. Trimmed reads were aligned to the bosTau9 assembly using BWA-MEM v0.7.15 (Li 2013), and only uniquely mapping reads with a minimum MAPQ of 10 were retained. Technical replicates were merged, and peak calling was performed using with MACS2 v2.1.1 (Liu 2014) using the default FDR cutoff of 0.05 and arguments '-B --SPMR -f BAMPE'. Normalized bigWigs corresponding to read coverage per 1 million reads were used for visualization, which were generated using deepTools v3.0.1 (Ramírez et al. 2014).



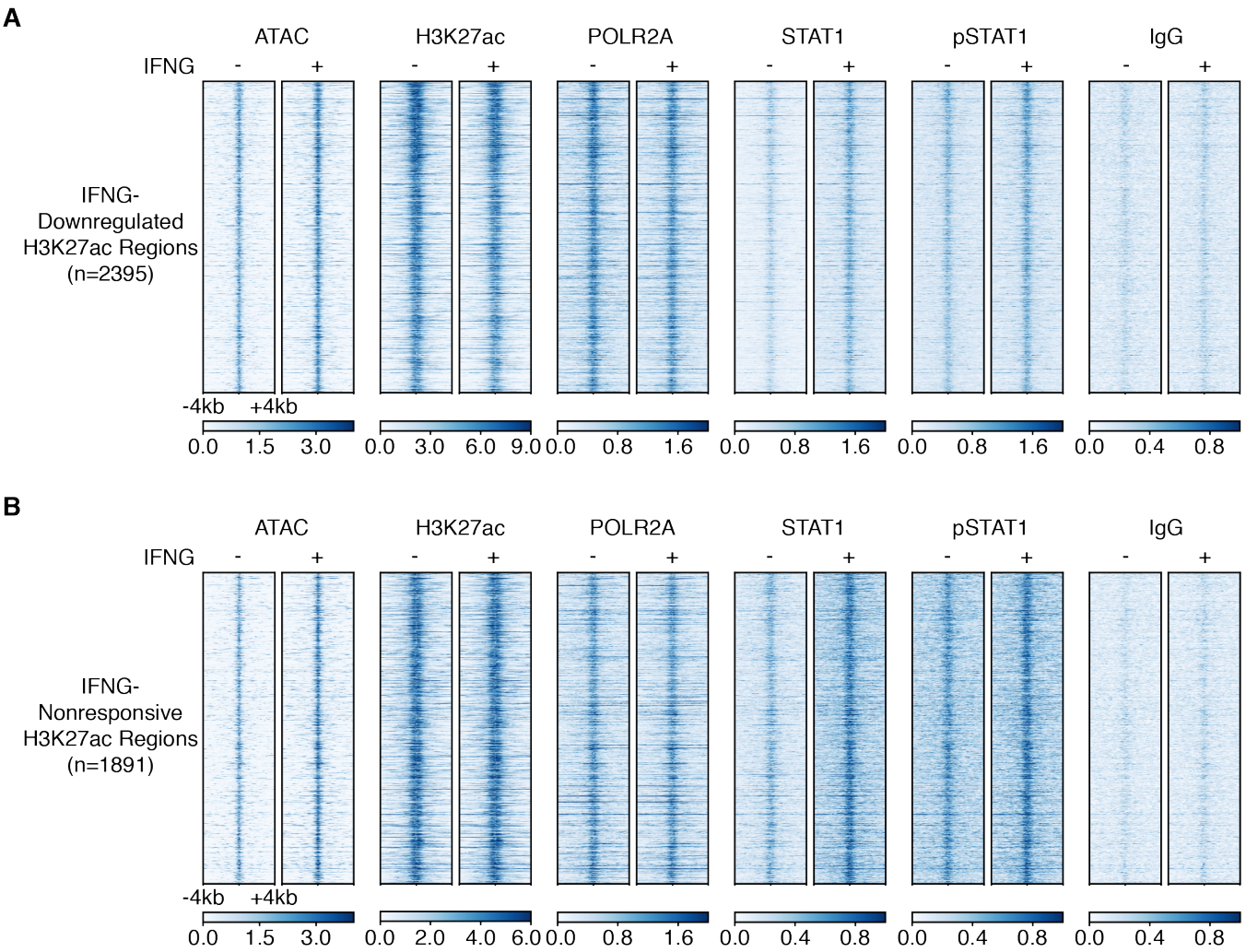
Supplemental Fig. S1. MA plot for differentially upregulated (blue) and downregulated (orange) genes expressed in (A) BL3.1 cells, (B) monocytes, (C) leukocytes, (D) alveolar macrophages, and (E) bone marrow derived macrophages. Genes with an FDR < 0.05 are shown in grey.



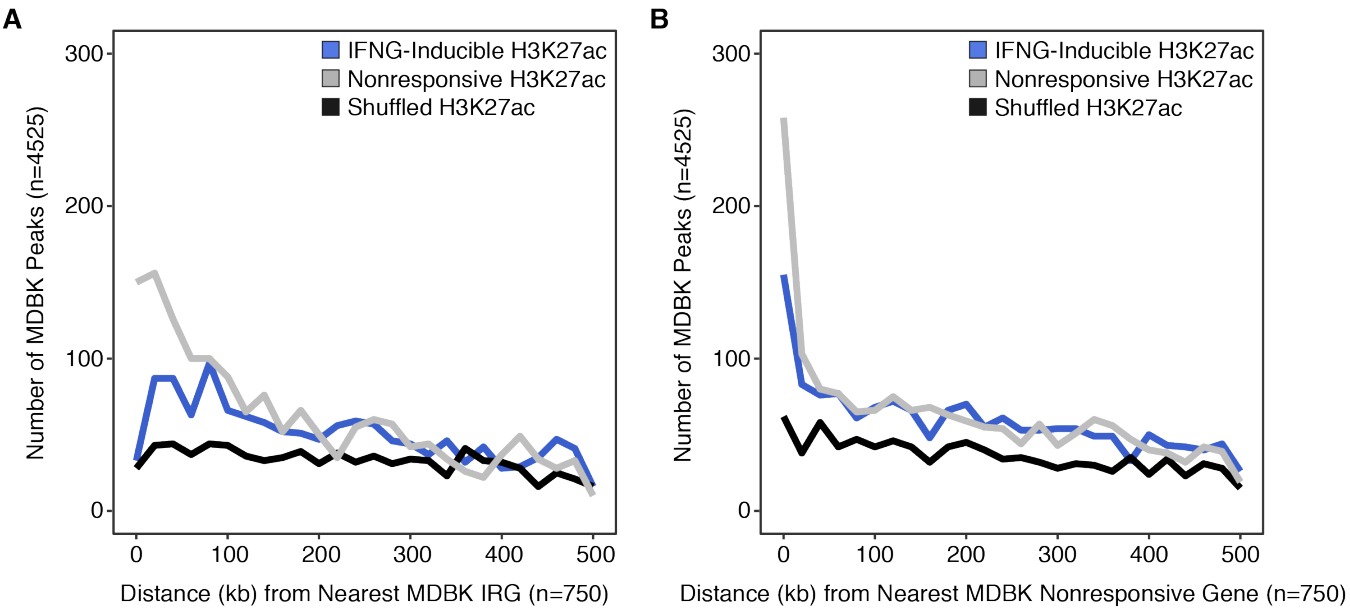
Supplemental Fig. S2. Mean DESeq2 normalized counts showing expression for (A) *CIITA* and (B) *TLR4* from wild type bovine cells. Count values are shown for MDBK (n=3), BL3.1 (n=3), monocytes (n=2), leukocytes (n=3), alveolar macrophages (n=3 untreated, n=4 stimulated), and bone marrow derived macrophages (n=4). ***FDR < 0.001. Error bars denote SEM. AM: alveolar macrophage; BMDM: bone marrow derived macrophage; FDR: false discovery rate.



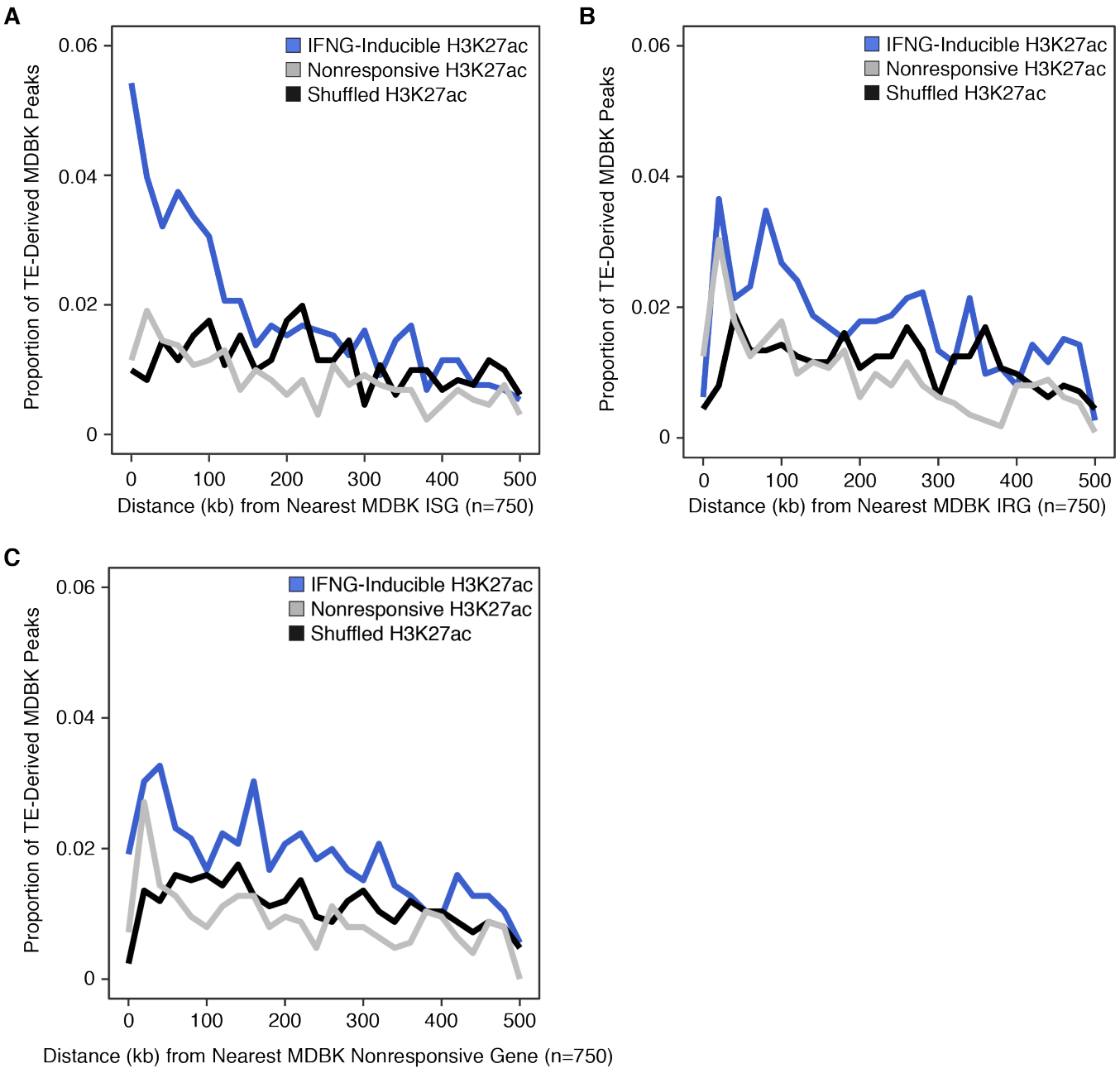
Supplemental Fig. S3. Genome browser view of *TLR4*. RNA-seq (green), ATAC-seq (red), and CUT&RUN (blue) tracks are normalized per million reads. CUT&RUN tracks for POLR2A, STAT1, and pSTAT1 pulldowns are divided by aligned fragments <= 150bp (dark blue) and > 150bp (light blue). The H3K27ac tracks correspond to all aligned fragments (dark blue). Enhancers shown represent IFNG-inducible H3K27ac peaks that have been centered by ATAC-seq summits. Values on the right of each track correspond to signal maxima.



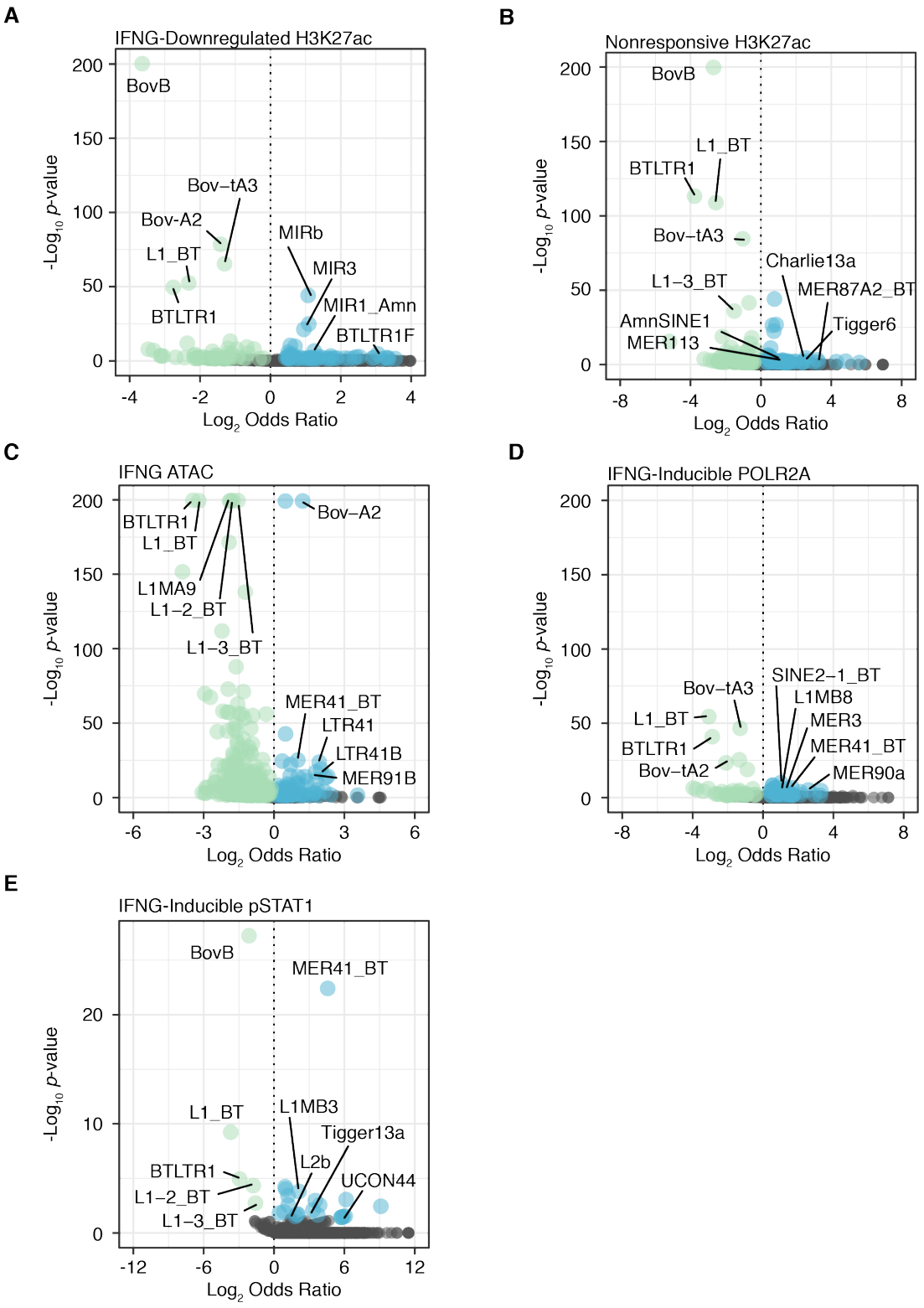
Supplemental Fig. S4. Heatmaps showing normalized ATAC-seq and CUT&RUN signal over (A) IFNG-downregulated H3K27ac peaks and (B) IFNG-nonresponsive H3K27ac peaks. Heatmap intensity values represent CUT&RUN signal per million reads. Peaks are sorted by ascending FDR.



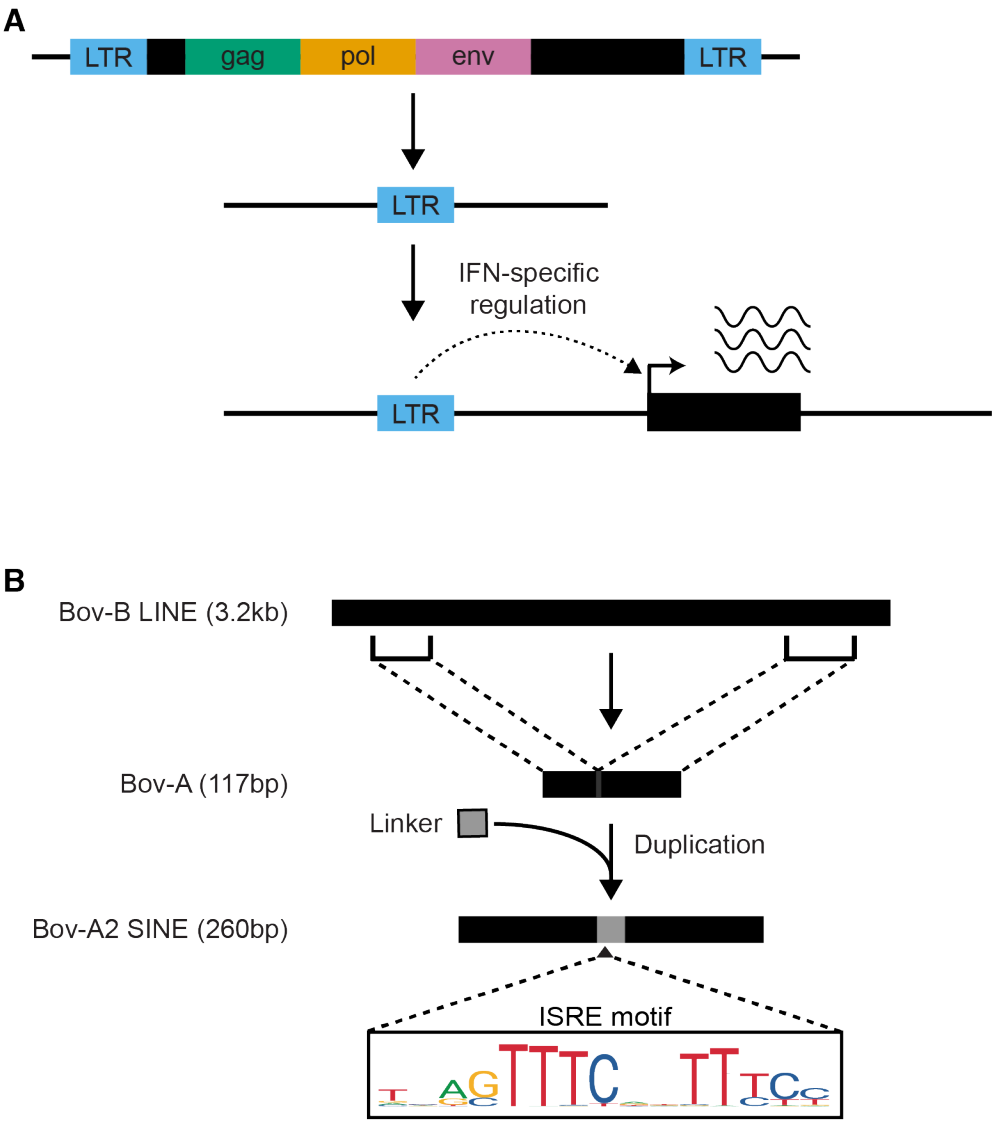
Supplemental Fig. S5. Frequency histogram of absolute distances from each MDBK IFNG-inducible (n=4525), nonresponsive (n=4525), and shuffled IFNG-inducible H3K27ac peak (n=4525) to the nearest MDBK (A) IRG (n=750) and (B) nonresponsive gene (n=750). Statistical significance of the observed enrichment within the first 20 kb of the nearest gene was assessed by Fisher's exact test. Fisher's two-tail *p*-value for IFNG-inducible, nonresponsive, and shuffled IFNG-inducible H3K27ac peaks were (A) 8.6×10^{-6}, 4.2×10^{-87}, and 5.7×10^{-1} for IRGs and (B) 3.2×10^{-12}, 8.2×10^{-58}, and 5.4×10^{-2} for nonresponsive genes. IRG: Interferon-repressed gene.



Supplemental Fig. S6. Frequency histogram of absolute distance from each TE-derived MDBK H3K27ac peak to the nearest MDBK (A) ISG (n=750), (B) IRG (n=750) and (C) nonresponsive gene (n=750). TEs marked by IFNG-inducible (n=4198), nonresponsive (n=3996), and shuffled IFNG-inducible H3K27ac (n=1903) peaks were defined by overlapping summits from merged IFNG ATAC peaks. Statistical significance of the observed enrichment within the first 20 kb of the nearest gene was assessed by Fisher's exact test. Fisher's two-tail *p*-value for IFNG-inducible, nonresponsive, and shuffled IFNG-inducible H3K27ac peaks were (A) 3.8×10^{-47}, 7.9×10^{-8}, and 1.0×10^{1} for ISGs (B) 7.8×10^{-3}, 5.5×10^{-10}, and 2.5×10^{-2} for IRGs, and (C) 9.8×10^{-9}, 2.8×10^{-8}, 5.0×10^{-3} for nonresponsive genes. IRG: Interferon-repressed gene.

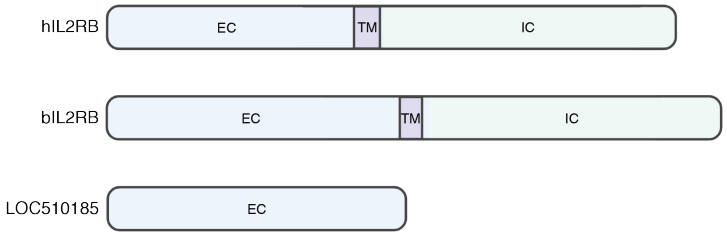


Supplemental Fig. S7. Volcano plot visualizing family-level TE enrichment for MDBK (A) IFNG-downregulated H3K27ac (n=1616), (B) nonresponsive H3K27ac (n=3891), (C) merged IFNG ATAC (n=106241), (D) IFNG-inducible POLR2A (n=1062), and (E) IFNG-inducible pSTAT1 (n=351) peaks. Statistical significance was determined using Fisher's exact test. TE families with a Fisher's two-tail *p*-value < 0.05 were defined as enriched (blue) or depleted (green) based on the reported odds ratio. TE families with a *p*-value > 0.05 are shown in grey.



Supplemental Fig. S8. (A) Each original exogenous retroviral insertion introduces two copies of the long terminal repeat (LTR) in the host genome. Over evolutionary time, ectopic recombination results in the excision of the internal ERV regions that encode for proteins necessary for retrotransposition, leaving a solitary LTR that may be co-opted as *cis*-regulatory elements through harboring binding sites for host transcription factors (Chuong et al. 2017). We propose a model whereby MER41_BT LTR elements have been co-opted as IFNG-inducible enhancers that drive cattle ISG expression through providing STAT1 binding sites. Adapted from Chuong et al. 2017. (B) Bov-A2 elements are thought to have arisen through the formation and duplication of monomeric Bov-A units derived from the 5' and 3' ends of BovB and the subsequent acquisition of a short linker, which is strongly enriched for ISRE motifs. Adapted from Onami et al. 2007.

A



B

```
hIL2RB      MAAPALSWRLPLLILLLPLATSWASAAVNG------------------------------
bIL2RB      MAVPALSCCLSLLVLLLLLAIPQTSTAVNGLSISSPGRNRSGGGTLEGHGCPTASPTRHI  60
LOC510185   MAAPALSCCLSLLILLLLLLAIPQTSTAVNGLSISSPGRNRSGGGTLEGHGCPTASPTRRI

hIL2RB      ----TSQFTCFYNSRANISCVWSQDGALQDTSCQVHAWPDRR-------RWNQTCELLPV
bIL2RB      FTAYASNITCFYNSRANVSCIWSHDEGLRATTCCLHSRQPQREVLFPTKQWNHTCELQPV  120
LOC510185   FTAYASNITCFYNSRANVSCIWSHDEGLRATTCCLHSRQPQREVLFPTKQWNHTCELQPV

hIL2RB      SQASWACNLILG-APDSQKLTTVDIVTLRVLCREGVRWRVMAIQDFKPFENLRLMAPISL
bIL2RB      RPGSWACNLVLGHSPEAQKLTVVDVMKLTVMCSEGGKWRRMMTQDIKPFDYIRLVPPHSL  160
LOC510185   RPGSWACNLVLGHSPEAQKLTIVDIMKLTVMCSEGGKWRRMMTQDIKPFDYIRLVPPHSL

hIL2RB      QVVHVETHRCNISWEISQASHYFERHLEFEARTLSPGHTWEEAPLLTLKQKQEWICLETL
bIL2RB      QVVRIETHRCNITWTVSQVSHYIQNDVEFEARLRYADHSWEDARLLTLRQNQQWISLENL  240
LOC510185   QVVHIETHRCNITWTVSQVSHYIQNDVEFEARLRYADHSWEDARLLTLRQNQQWISLENL

hIL2RB      TPDTQYEFQVRVKPLQGEFTTWSPWSQPLAFRTKPAALGKDT--IPWLGHLLVGLSGAFG
bIL2RB      APGMEYELQVRAKPRLGSHEVWSHWSQPLAFRTVPAETKKKIPPLPWLNHIFLGVGSFFG  300
LOC510185   APSMEYELQVRAKPRLGSHEVWSHWSQPLAFRTVPAVA---------LRH----------

hIL2RB      FIILVYLLINCRNTGPWLKKVLKCNTPDPSKFFSQLSSEHGGDVQKWLSSPFPSSSFSPG
bIL2RB      LVLLVYFLGSFRCIRLWLKNVLKCHIPDPSEFFSQLSSEHGGDFQKWLSSPFPSSSFSRS  360
LOC510185   -------MGSYRT----------------------------------------------

hIL2RB      GLAPEISPLEVLERD-KVTQ-LLLQQDKVPEPASLSSNHSLTSCFTNQGYFFFHLPDALE
bIL2RB      GLDPEISPLEVLDRDAKATQLLLLQQDKGSSSSAETSGHSVTSCFTNQGYFFFHLPDALE  420
LOC510185   -------------RD----------------------------------------------

hIL2RB      IEACQVYFTYDPYSEEDPDEGVAGAPTGSSPQPLQPLSGEDDAYCTFPSRDDLLLFSPSL
bIL2RB      IEACQVYFAYDPCTEE-LDEGGPRAPEGALLPPLPTPPGDDDAYCTFPPGEDLLLFSPSL  480
LOC510185   ------------------------------------------------------------

hIL2RB      LGGPSPPSTAPGGSGAGEERMPPSLQERVPRDWDPQPLGPPTPGVPDLVDFQPPPELVLR
bIL2RB      LSGPGPPNTAQWGTGAGEERLASSPQEGVPGDWTPQPLRPPALEAPDLVDLQSSPEHELG  540
LOC510185   ------------------------------------------------------------

hIL2RB      EAGEEVPDAGPREGVSFPWSRPPGQGEFRALNARLPLNTDAYLSLQELQGQDPTHLV
bIL2RB      EARERVPGPSPREGTSFPWASPPGQGQVRAPTVCLTLNTDAYLSLQELQDQDPAYLV     597
LOC510185   ------------------------------------------------------------
```

Supplemental Fig. S9. (A) Schematic showing extracellular (EC, blue), transmembrane (TM, purple), and intracellular (IC, green) domains for human IL2RB (UniProt P14784), bovine IL2RB (UniProt F1N409), and LOC510185 (RefSeq XP_010803655.1). LOC510185 domain annotation was inferred by homology. (B) Multiple sequence alignment of human IL2RB, bovine IL2RB, and LOC510185 amino acid sequences using MUSCLE (Edgar 2004). Domain boundaries are colored with respect to bovine IL2RB.

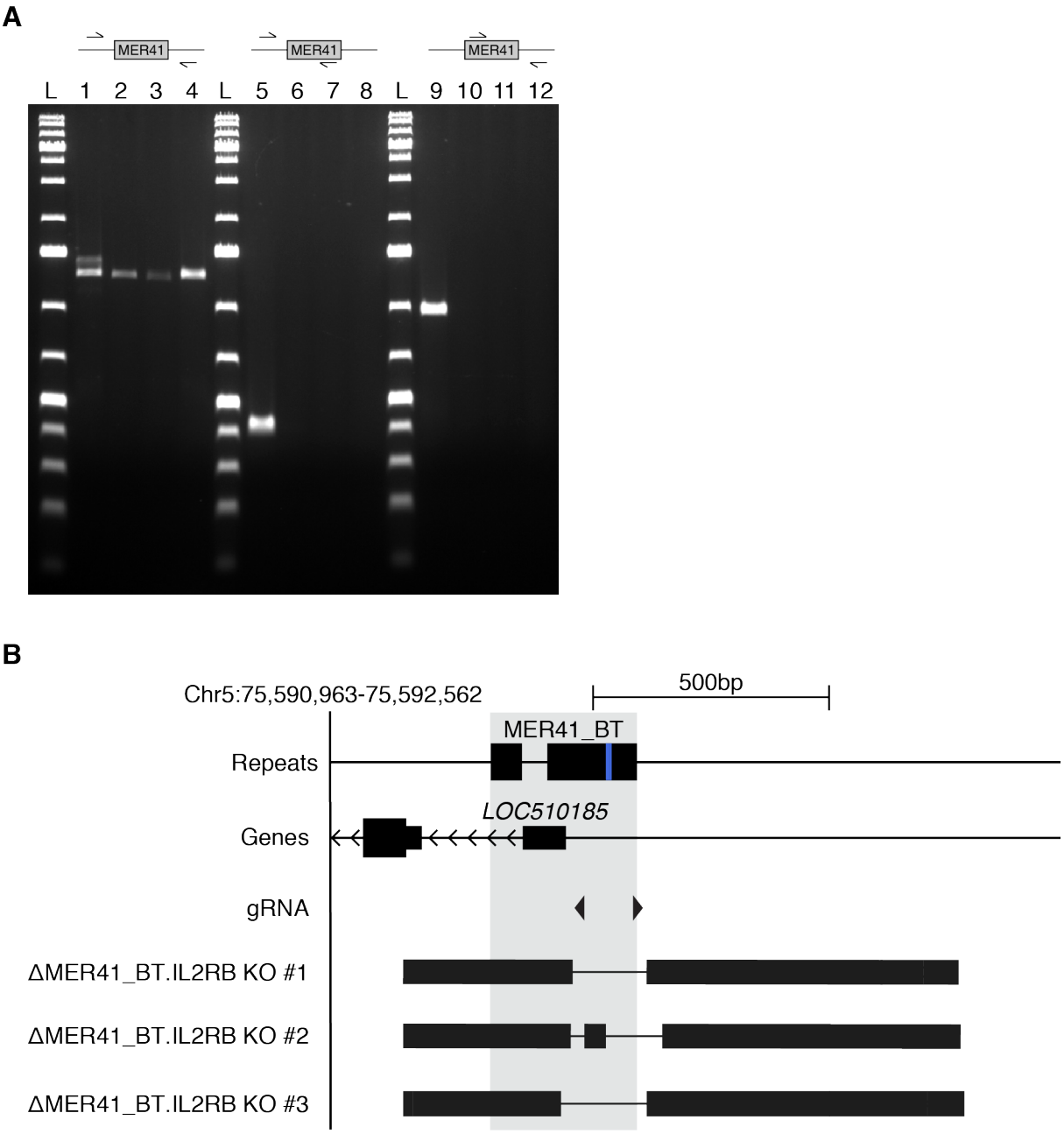
A

```
LOC510185  MAAPALSCCLSLLILLLLLAIPQTSTAVNGLSISSPGRNRSGGGTLEGHGCPTASPTRRIFTAYASNITCFYNSRANVSCIWSHDEGLRA
    IL2RB  MAVPALSCCLSLLVLLLLLAIPQTSTAVNGLSISSPGRNRSGGGTLEGHGCPTASPTRHIFTAYASNITCFYNSRANVSCIWSHDEGLRA
           **.**********:**********************************************:******************************

LOC510185  TTCCLHSRQPQREVLFPTKQWNHTCELQPVRPGSWACNLVLGHSPEAQKLTIVDIMKLTVMCSEGGKWRRMMTQDIKPFDYIRLVPPHSL
    IL2RB  TTCCLHSRQPQREVLFPTKQWNHTCELQPVRPGSWACNLVLGHSPEAQKLTVVDVMKLTVMCSEGGKWRRMMTQDIKPFDYIRLVPPHSL
           **************************************************:**:**********************************

LOC510185  QVVHIETHRCNITWTVSQVSHYIQNDVEFEARLRYADHSWEDARLLTLRQNQQWISLENLAPSMEYELQVRAKPRLGSHEVWSHWSQPLA
    IL2RB  QVVRIETHRCNITWTVSQVSHYIQNDVEFEARLRYADHSWEDARLLTLRQNQQWISLENLAPGMEYELQVRAKPRLGSHEVWSHWSQPLA
           ***:********************************************************.**************************

LOC510185  FRTVPAVALRHMGSYRTRD
    IL2RB  FRTVPAETKKKIPPLPWLN
           ****** : :::      :
```

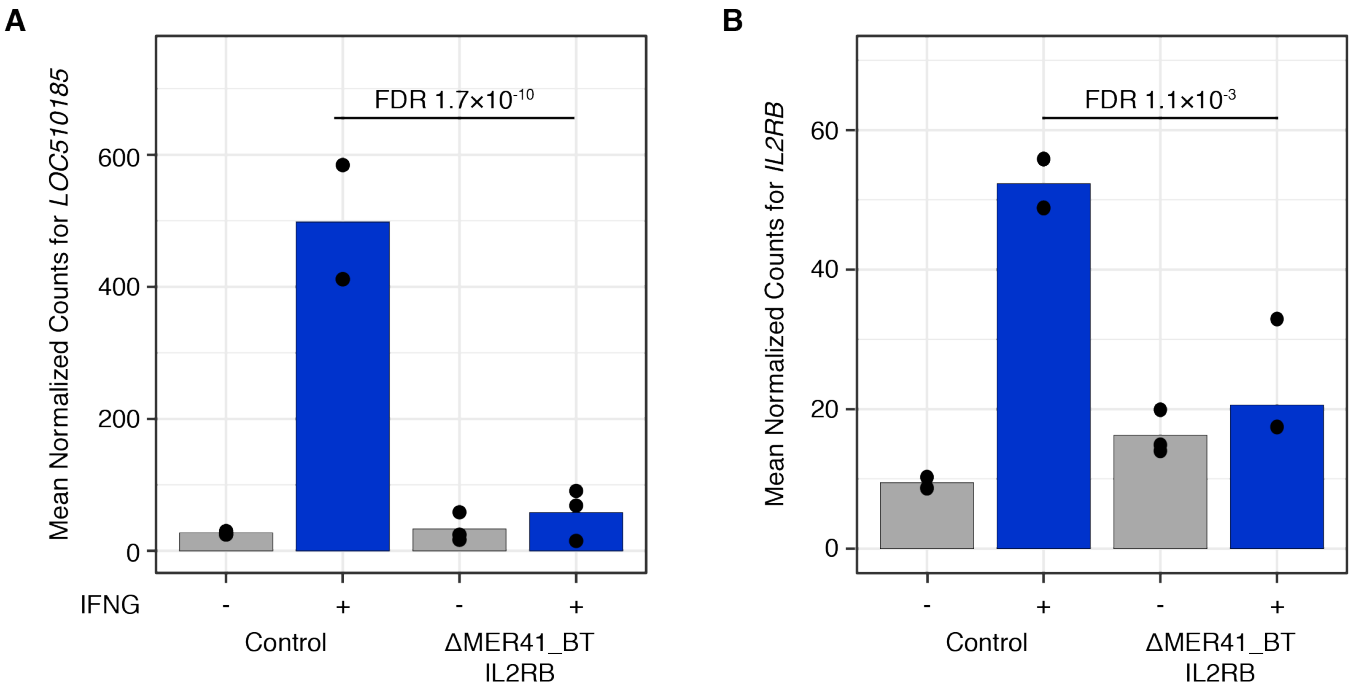
B

```
LOC510185  atggcggctccagctctgtcctgctgtctgtccctcctcatcctcctcttgctgctggccatccctcagacatctacagcagtgaatgga
    IL2RB  atggcggttccggctctgtcctgctgtctgtccctcctcgtcctcctcttgctgttggccatccctcagacatctacagcagtgaatgga
           ******* *** ************************* ************** ***********************************

LOC510185  ctgagcatctcctcacctggtaggaacaggagtggcgggggaacccttgaaggccatgggtgccctacagccagccccacacgccgcatc
    IL2RB  ctgagcatctcctcacctggtaggaacaggagtggcgggggaacccttgaaggccatgggtgccctacagccagccccacacgccacatc
           ********************************************************************************** ****

LOC510185  ttcacggcatatgcttccaacatcacatgcttctataactcgagagccaacgtctcctgcatctggagccatgatgagggcctgcgggcc
    IL2RB  ttcacggcatatgcttccaacatcacatgcttctataactcgagagccaacgtctcctgcatctggagccatgatgagggcctgcgggcc
           ********************************************************************************

LOC510185  accacctgctgcttgcacagccgtcagcctcagagagaggtgctttttcctacaaaacagtggaaccacacctgtgagctgcagccagtg
    IL2RB  accacctgctgcttgcacagccgtcagcctcagagagaggtgctttttcctacaaaacagtggaaccacacctgtgagctgcagccagtg
           ********************************************************************************

LOC510185  aggccaggatcctgggcatgcaacctggtcctgggacattctccagaagctcagaaactgaccatagtggacatcatgaaattgacggtg
    IL2RB  aggccaggatcctgggcatgcaacctggtcctgggacattctccagaagctcagaaactgaccgtagtggatgtcatgaaattgacggtg
           ************************************************************ *******  *****************

LOC510185  atgtgctctgaaggcgggaagtggaggaggatgatgacccaggatatcaagccctttgattacattcgtctggttccaccccactcactc
    IL2RB  atgtgctctgaaggcgggaagtggaggaggatgatgacccaggatatcaagccctttgattacattcgtctggttccaccccactcactc
           ********************************************************************************

LOC510185  caggtcgtccacatagagacccataggtgcaacataacctggactgtctcccaggtatcccactacatccaaaacgacgtggagtttgag
    IL2RB  caggtcgtccgcatagagacccataggtgcaacataacctggactgtctcccaggtatcccactacatccaaaacgacgtggagtttgag
           ********** ********************************************************************************

LOC510185  gcccggttaaggtacgccgaccacagctgggaggacgcccggctactgaccctcaggcagaaccagcaatggatctccctggagaatctc
    IL2RB  gcccggttaaggtacgccgaccacagctgggaggacgcccggctactgaccctcaggcagaaccagcaatggatctccctggagaatctc
           ********************************************************************************

LOC510185  gctccaagtatggaatatgagcttcaggtgcgggccaagccccggctaggcagccacgaggtttggagccactggagccagcccctggcc
    IL2RB  gctccaggcatggaatatgagcttcaggtgcgggccaagccccggctaggcagccacgaggtttggagccactggagccagcccctggcc
           ****** * ********************************************************************************

LOC510185  ttcaggacggtccctgcagttgctctgagacatatgggatcttataggaccagagac
    IL2RB  ttcaggacggtccctgcagaaaccaagaagaagatcccacctctcccttggttgaac
           ******************   *   **   * **   * **            **
```

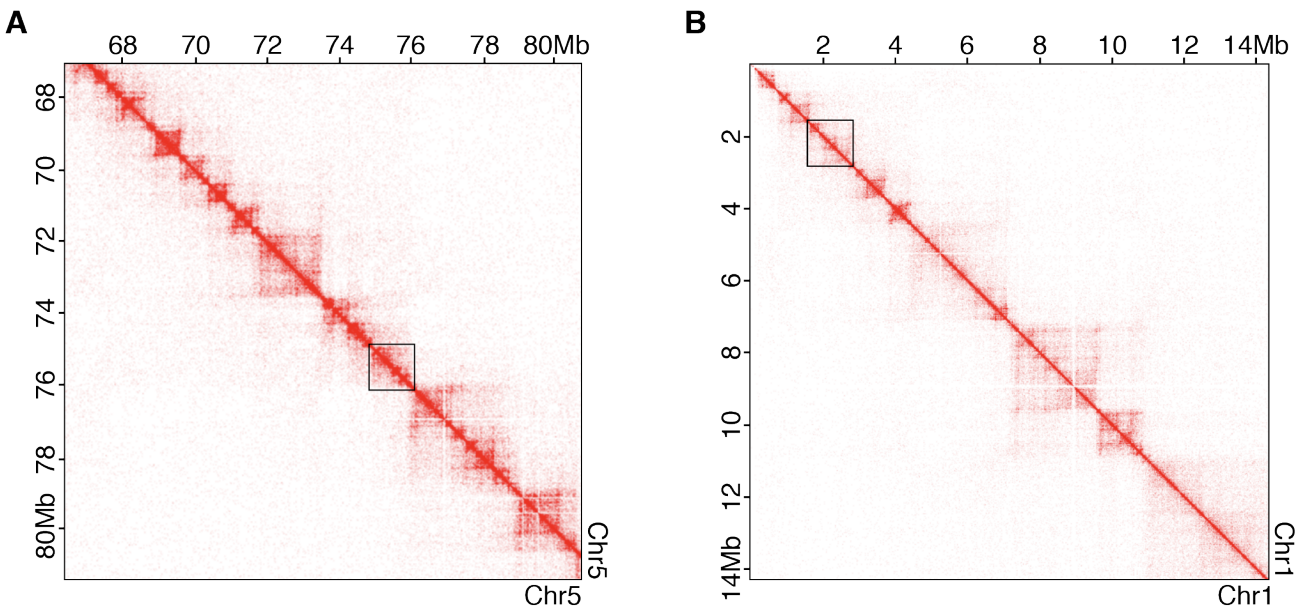
Supplemental Fig. S10. (A) Pairwise sequence alignment of bovine IL2RB (UniProt F1N409) and LOC510185 (RefSeq XP_010803655.1) amino acid sequences using Clustal Omega (McWilliam et al. 2013). The alignment was subsequently converted to the underlying (B) cDNA sequence using PAL2NAL (Suyama et al. 2006), aligned with Clustal Omega (McWilliam et al. 2013), and used to estimate the time of duplication (Supplemental Methods).



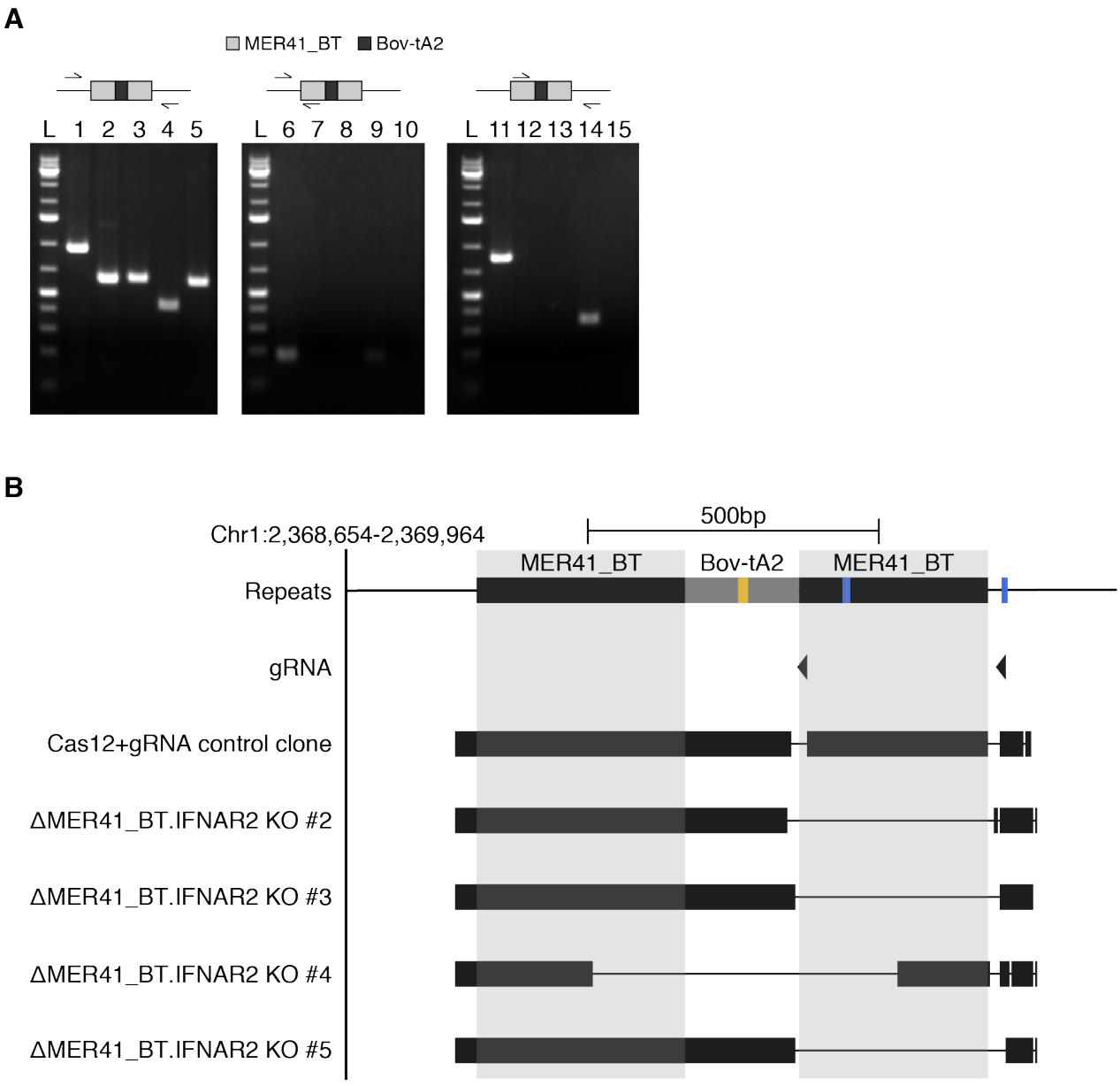
Supplemental Fig. S11. (A) Schematic (above) and agarose gel (below) showing PCR products for validating MDBK MER41_BT.IL2RB KO clones. Presence of the wild type and KO bands was validated using both flanking and internal primers as indicated with the arrows. Left: flanking only. Middle and right: flanking/internal. L: ladder. Lanes 1, 5, 9: MER41_BT.IL2RB heterozygote MDBK. Lanes 2, 6, 10: MER41_BT.IL2RB KO #1. Lanes 3, 7, 11: MER41_BT.IL2RB KO #2. Lanes 4, 8, 12: MER41_BT.IL2RB KO #3. (B) Genome browser track showing Sanger sequencing results that have been aligned to the reference assembly using BLAT (Kent 2002). The GAS motif internal to MER41_BT is shown in blue.



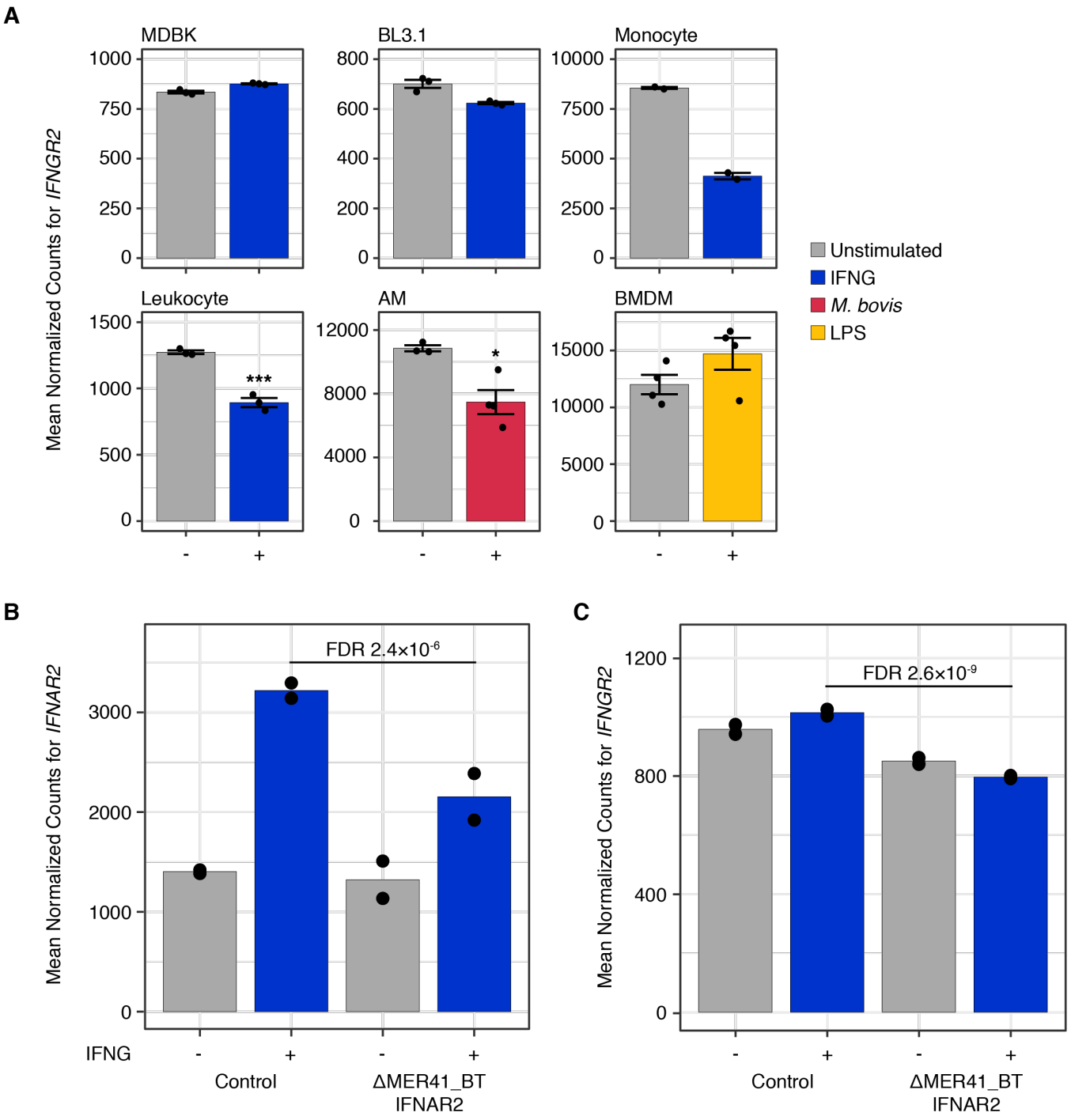
Supplemental Fig. S12. Normalized DESeq2 counts for (A) *LOC510185* and (B) *IL2RB* from MDBK Cas12 controls (n=2) and MER41_BT.IL2RB mutant clones (n=3). Data points denote clonal replicates.



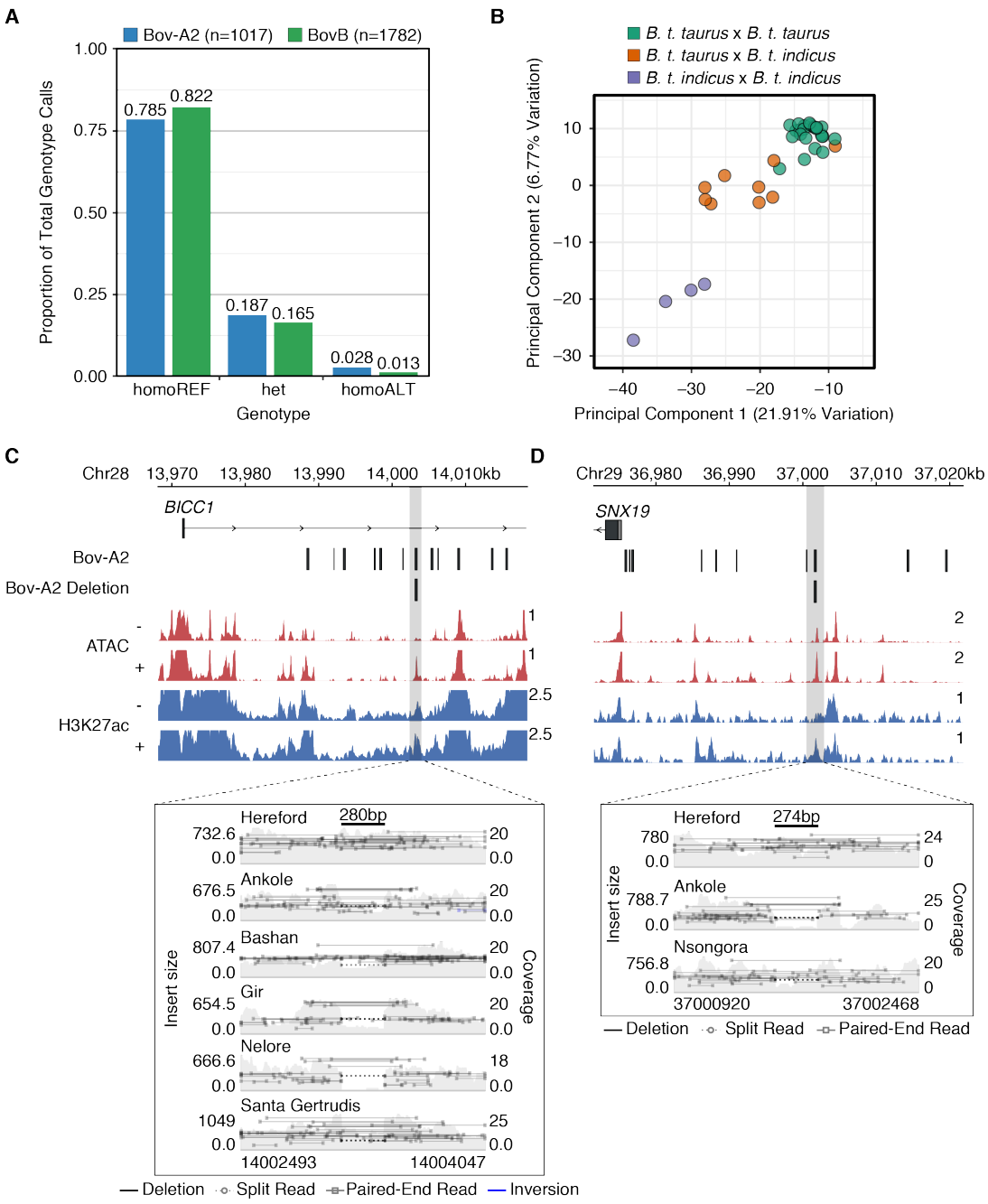
Supplemental Fig. S13. Knight-Ruiz (KR) Hi-C contact maps (25kb resolution) highlighting 1 Mb regions containing (A) MER41_BT.IL2RB (Chr5:75591298-75591607) and (B) MER41_BT.IFNAR2 (Chr1:2368871-269745). The contact matrix was visualized using Juicebox v2.1.10 (Robinson et al. 2018).



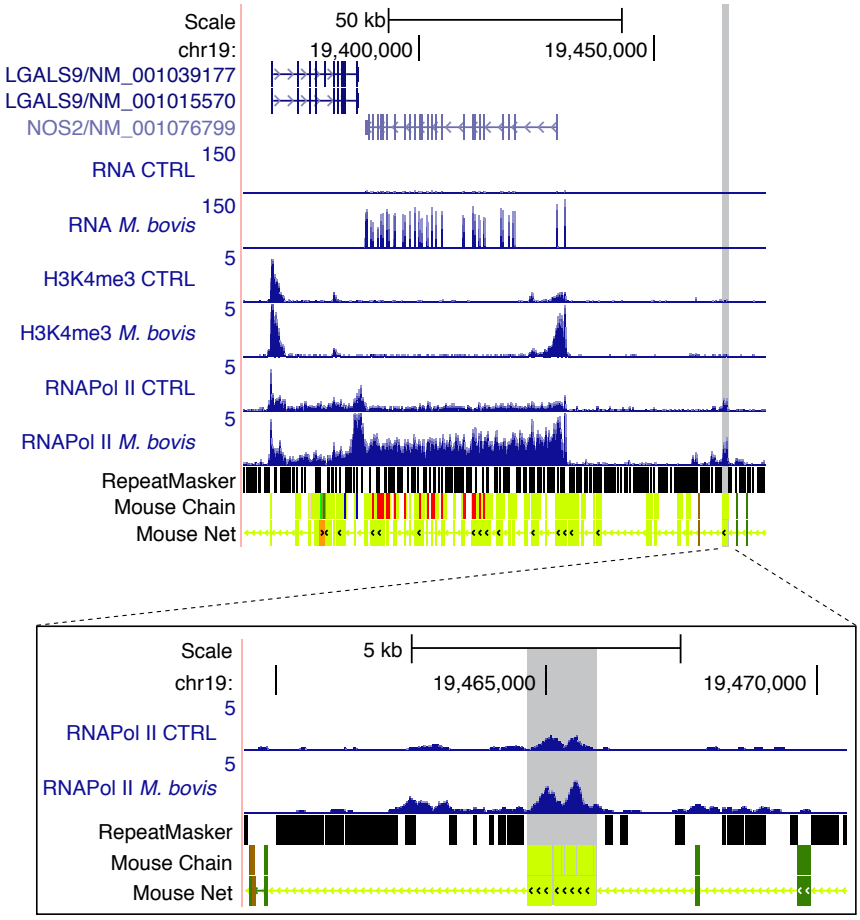
Supplemental Fig. S14. (A) Schematic (above) and agarose gel (below) showing PCR products for validating MDBK MER41_BT.IFNAR2 KO clones. Presence of the wild type and KO bands was validated using both flanking and internal primers as indicated with the arrows. Left: flanking only. Middle and right: flanking/internal. L: ladder. Lanes 1, 6, 11: Cas12+gRNA control clone. Lanes 2, 7, 12: MER41_BT.IFNAR2 KO #2. Lanes 3, 8, 13: MER41_BT.IFNAR2 KO #3. Lanes 4, 9, 14: MER41_BT.IFNAR2 #4. Lanes 5, 10, 15: MER41_BT.IFNAR2 KO #5. (B) Genome browser track showing Sanger sequencing results that have been aligned to the reference assembly using BLAT (Kent 2002). GAS motifs are shown in blue, and the ISRE motif internal to the interrupting Bov-tA2 SINE element is shown in yellow.



Supplemental Fig. S15. (A) Mean DESeq2 normalized counts showing expression for *IFNGR2* from wild type bovine cells. Count values are shown for MDBK (n=3), BL3.1 (n=3), monocytes (n=2), leukocytes (n=3), alveolar macrophages (n=3 untreated, n=4 stimulated), and bone marrow derived macrophages (n=4). ***FDR < 0.0001, *FDR < 0.01. Error bars denote SEM. (B) Normalized DESeq2 counts for *IFNAR2* from MDBK Cas12 controls (n=2) and MER41_BT.IFNAR2 mutant clones (n=2). Data points denote clonal replicates. (C) Bar plot as in (B) but for *IFNGR2* expression. AM: alveolar macrophage; BMDM: bone marrow derived macrophage; FDR: false discovery rate.



Supplemental Fig. S16. (A) Proportion of polymorphic Bov-A2 (n=1017) and BovB (n=1782) variants as a function of genotype. Variants were called using MELT (Gardner et al. 2017), and filtered insertions and deletions were aggregated into a single list by family. Polymorphic TE calls were grouped as homozygous for the reference allele (homoREF), heterozygous (het), or homozygous for the variant allele (homoALT). (B) PCA plot of polymorphic Bov-A2 and BovB elements by genotype. Individual samples (n=38) are colored by subspecies. (C) Genome browser view (top) of a putative Bov-A2 deletion (Chr28:14003130-14003410) intronic to *BICC1*. ATAC-seq and CUT&RUN tracks from MDBK are normalized per million reads. Values on the right of each track correspond to signal maxima. Variant visualization plot (bottom) was produced using Samplot (Belyeu et al. 2021) and depicts aligned fragments from six individuals over the predicted Bov-A2 deletion. (D) Genome browser view and variant visualization plot as in (C) but for a Bov-A2 element (Chr29:37001557-37001831) upstream of *SNX19* including variant plots for three individuals.



Supplemental Fig. S17. Genome browser view of the bovine *NOS2* locus. Bovine alveolar macrophage ChIP-seq data were reanalyzed using accession GSE116734 (Hall et al. 2019). ChIP-seq tracks are CPM normalized. We observed a putative, non-TE derived enhancer (grey highlight) 30kb upstream of the NOS2 transcriptional start site that is partially conserved through mouse (*Mus musculus*).

SUPPLEMENTAL TABLES

Supplemental Table S1 (separate file)
Wild type RNA-seq alignment statistics and DESeq2 results for MDBK cells, BL3.1 cells, monocytes, leukocytes, alveolar macrophages, and bone marrow derived macrophages.

Supplemental Table S2 (separate file)
gProfiler gene ontology results for wildtype RNA-seq upregulated genes.

Supplemental Table S3 (separate file)
MDBK ATAC-seq alignment statistics.

Supplemental Table S4 (separate file)
MDBK CUT&RUN alignment statistics and DESeq2 results.

Supplemental Table S5 (separate file)
CUT&RUN motif enrichment using XSTREME.

Supplemental Table S6 (separate file)
Predicted enhancer-gene contacts using the ABC model.

Supplemental Table S7 (separate file)
TE-derived IFNG-inducible, IFNG-downregulated, and nonresponsive H3K27ac peaks.

Supplemental Table S8 (separate file)
Family-level TE enrichment using GIGGLE.

Supplemental Table S9 (separate file)
Absolute distances for IFNG-inducible MER41_BT to the nearest MDBK ISG.

Supplemental Table S10 (separate file)
Summary of TBLASTN results querying cattle IL2RB against cetartiodactyls

Supplemental Table S11 (separate file)
Alignment statistics and DESeq2 results from MER41_BT.IL2RB and MER41_BT.IFNAR2 deletions in MDBK cells.

Supplemental Table S12 (separate file)
Summary of assemblies used in RepeatMasker analysis.

Supplemental Table S13 (separate file)
MER41 synteny analysis using TEOrthology.

Supplemental Table S14 (separate file)
Filtered Bov-A2, BovB, and MER41_BT insertions and deletions as called using MELT.

Supplemental Table S15 (separate file)

Absolute distances for epigenetic-marked Bov-A2 to nearest gene transcriptional start site (TSS)

Supplemental Table S16 (separate file)
Reagents and sequences used in this study.

Supplemental Table S17 (separate file)
Identifying information for all publicly available whole genome sequencing data analyzed in this study.

REFERENCES

Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho T-J, Koutseva N, Zaghlul S, Graves T, Rock S, et al. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature 508: 494–499.

Belyeu JR, Chowdhury M, Brown J, Pedersen BS, Cormier MJ, Quinlan AR, Layer RM. 2021. Samplot: a platform for structural variant visual validation and automated filtering. Genome Biol 22: 161.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10: 421.

Chitko-McKown CG, Fox JM, Miller LC, Heaton MP, Bono JL, Keen JE, Grosse WM, Laegreid WW. 2004. Gene expression profiling of bovine macrophages in response to Escherichia coli O157:H7 lipopolysaccharide. Dev Comp Immunol 28: 635–645.

Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet 18: 71–86.

Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, et al. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 14: 959–962.

DeWeirdt PC, Sanson KR, Sangree AK, Hegde M, Hanna RE, Feeley MN, Griffith AL, Teng T, Borys SM, Strand C, et al. 2021. Optimization of AsCas12a for combinatorial genetic screens in human cells. Nat Biotechnol 39: 94–104.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst 3: 95–98.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.

Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32: 3047–3048.

Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D, et al. 2020. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 48: D87–D92.

Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, Grossman SR, Anyoha R, Doughty BR, Patwardhan TA, et al. 2019. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. Nat Genet 51: 1664–1669.

Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Pittard WS, Mills RE, 1000 Genomes Project Consortium, Devine SE. 2017. The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. Genome Res 27: 1916–1929.

Gertz EM, Yu Y-K, Agarwala R, Schäffer AA, Altschul SF. 2006. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol 4: 41.

Grant CE, Bailey TL. 2021. XSTREME: Comprehensive motif analysis of biological sequence datasets. bioRxiv 2021.09.02.458722. https://www.biorxiv.org/content/10.1101/2021.09.02.458722v1 (Accessed September 24, 2021).

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. Bioinformatics 27: 1017–1018.

Hall TJ, Vernimmen D, Browne JA, Mullen MP, Gordon SV, MacHugh DE, O'Doherty AM. 2019. Alveolar Macrophage Chromatin Is Modified to Orchestrate Host Response to Mycobacterium bovis Infection. *Front Genet* **10**: 1386.

Hu R, Chi X, Chai G, Kong Y, He G, Wang X, Shi D, Zhang D, Zhou G. 2012. Genome-wide identification, evolutionary expansion, and expression profile of homeodomain-leucine zipper gene family in poplar (Populus trichocarpa). PLoS One 7: e31149.

Janssens D, Henikoff S. 2019. CUT&RUN: Targeted in situ genome-wide profiling with high efficiency for low cell numbers v3. protocols.io. https://www.protocols.io/view/cut-amp-run-targeted-in-situ-genome-wide-profiling-zcpf2vn.

Kapusta A, Suh A, Feschotte C. 2017. Dynamics of genome size evolution in birds and mammals. Proc Natl Acad Sci U S A 114: E1460–E1469.

Kent WJ. 2002. BLAT--the BLAST-like alignment tool. Genome Res 12: 656–664.

Khan A, Mathelier A. 2017. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. BMC Bioinformatics 18: 287.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37: 907–915.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol Biol Evol 34: 1812–1819.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357–359.

Layer RM, Pedersen BS, DiSera T, Marth GT, Gertz J, Quinlan AR. 2018. GIGGLE: a search engine for large-scale integrated genome analysis. Nat Methods 15: 123–126.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30: 923–930.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [q-bioGN]. http://arxiv.org/abs/1303.3997.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Liu GE, Matukumalli LK, Sonstegard TS, Shade LL, Van Tassell CP. 2006. Genomic divergences among cattle, dog and human estimated from large-scale alignments of genomic sequences. BMC Genomics 7: 140.

Liu T. 2014. Use Model-Based Analysis of ChIP-Seq (MACS) to Analyze Short Reads Generated by Sequencing Protein–DNA Interactions in Embryonic Stem Cells. In Stem Cell Transcriptional Networks: Methods and Protocols (ed. B.L. Kidder), pp. 81–95, Springer New York, New York, NY.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15: 550.

McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. 2013. Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res 41: W597–600.

Meers MP, Bryson TD, Henikoff JG, Henikoff S. 2019. Improved CUT&RUN chromatin profiling tools. Elife 8. http://dx.doi.org/10.7554/eLife.46314.

Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32: 292–294.

Onami J-I, Nikaido M, Mannen H, Okada N. 2007. Genomic expansion of the Bov-A2 retroposon relating to phylogeny and breed management. Mamm Genome 18: 187–196.

Pasquesi GIM, Adams RH, Card DC, Schield DR, Corbin AB, Perry BW, Reyes-Velasco J, Ruggiero RP, Vandewege MW, Shortt JA, et al. 2018. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. Nat Commun 9: 2774.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842.

Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 42: W187–91.

Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J. 2019. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res 47: W191–W198.

Robinson JT, Turner D, Durand NC, Thorvaldsdóttir H, Mesirov JP, Aiden EL. Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. Cell Syst 6: 256-258.

Stephens M. 2017. False discovery rates: a new deal. Biostatistics 18: 275–294.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34: W609–12.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586–1591.

Zhu A, Ibrahim JG, Love MI. 2019. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics 35: 2084–2092.