Ruminant-specific retrotransposons shape regulatory evolution of bovine immunity

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

[1]Department of Molecular, Cellular, and Developmental Biology and BioFrontiers Institute, University of Colorado Boulder, Boulder, Colorado 80309, USA; [2]USDA, ARS, Roman L. Hruska US Meat Animal Research Center (MARC), Clay Center, Nebraska 68933, USA

Cattle are an important livestock species, and mapping the genomic architecture of agriculturally relevant traits such as disease susceptibility is a major challenge in the bovine research community. Lineage-specific transposable elements (TEs) are increasingly recognized to contribute to gene regulatory evolution and variation, but this possibility has been largely unexplored in ruminant genomes. We conducted epigenomic profiling of the type II interferon (IFN) response in bovine cells and found thousands of ruminant-specific TEs including MER41_BT and Bov-A2 elements predicted to act as IFN-inducible enhancer elements. CRISPR knockout experiments in bovine cells established that critical immune factors including *IFNAR2* and *IL2RB* are transcriptionally regulated by TE-derived enhancers. Finally, population genomic analysis of 38 individuals revealed that a subset of polymorphic TE insertions may function as enhancers in modern cattle. Our study reveals that lineage-specific TEs have shaped the evolution of ruminant IFN responses and potentially continue to contribute to immune gene regulatory differences across modern breeds and individuals. Together with previous work in human cells, our findings demonstrate that lineage-specific TEs have been independently co-opted to regulate IFN-inducible gene expression in multiple species, supporting TE co-option as a recurrent mechanism driving the evolution of IFN-inducible transcriptional networks.

[Supplemental material is available for this article.]

Modern cattle were domesticated 10,000 years ago and are an important livestock species (Troy et al. 2001; Orozco-terWengel et al. 2015). Cattle exhibit extensive phenotypic diversity, with over 1000 recognized breeds and closely related subspecies. Understanding the evolution and genetic basis of bovine biology, including traits involved in production and disease susceptibility, is a major challenge in the agricultural research community. However, most bovine genomic studies have focused on the effects of single nucleotide variants or small insertions and deletions (Littlejohn et al. 2016; Bouwman et al. 2018; Xiang et al. 2021). Transposable elements (TEs) could potentially have an important role shaping bovine variation and evolution, but with a few exceptions identified by genetic trait mapping (Girardot et al. 2006; Albrecht et al. 2012; Menzi et al. 2016; Schütz et al. 2016; Liang et al. 2021; Trigo et al. 2021), the impact of TE-derived structural variation on bovine biology remains largely unexplored.

Here, we investigated the significance of TEs on bovine biology, focusing on their capacity to act as gene regulatory elements. Epigenomic studies have revealed that TEs are a major source of *cis*-regulatory elements and have been repeatedly co-opted to regulate the expression of host genes (Chuong et al. 2016; Fuentes et al. 2018; Ye et al. 2020). The cattle genome harbors numerous ruminant-specific retrotransposons including bovine endogenous retroviruses (ERVs) and the BovB retrotransposon (Garcia-Etxebarria and Jugo 2013; Ivancevic et al. 2018), which makes up over 25% of the cattle genome (Adelson et al. 2009). Ruminant-specific retrotransposons have recently been documented to show evidence of regulatory activity (Young et al. 2018; Halstead et al. 2020), but their consequences on bovine gene regulation are unknown.

To study the contribution of TEs to bovine gene regulation, we focused on the transcriptional network underlying the type II interferon (IFN) response. We previously discovered that primate-specific TEs have been co-opted as enhancers that regulate antiviral IFN-stimulated genes (ISGs) (Chuong et al. 2016). Given that innate immune responses show high interspecies variation (Shaw et al. 2017; Bush et al. 2020), we hypothesized that ruminant-specific TEs may have been independently co-opted to facilitate immune regulatory evolution in cattle. We therefore conducted epigenomic profiling of the bovine response to interferon gamma (IFNG) using MDBK cells, a model bovine cell line used to study innate immunity (Barreca and O'Hare 2004; Fay et al. 2020).

Results

Ruminant-specific retrotransposons contribute IFNG-inducible regulatory elements

To explore whether TEs contribute to bovine immune gene regulation, we conducted epigenomic profiling of the bovine type II IFN response (Fig. 1A). We used RNA-seq to profile the transcriptional response to recombinant bovine IFNG in multiple cell types, including monocytes, leukocytes, B lymphocyte line BL3.1 cells, and kidney epithelial MDBK cells. We identified a total of 4192 IFNG-stimulated genes showing inducible expression in at least one profiled cell type, with 298 shared across all cell types including canonical ISGs (Fig. 1B,C; Supplemental Fig. S1; Supplemental Tables S1, S2). In parallel, we conducted chromatin profiling of IFNG-stimulated and untreated MDBK cells using ATAC-seq and

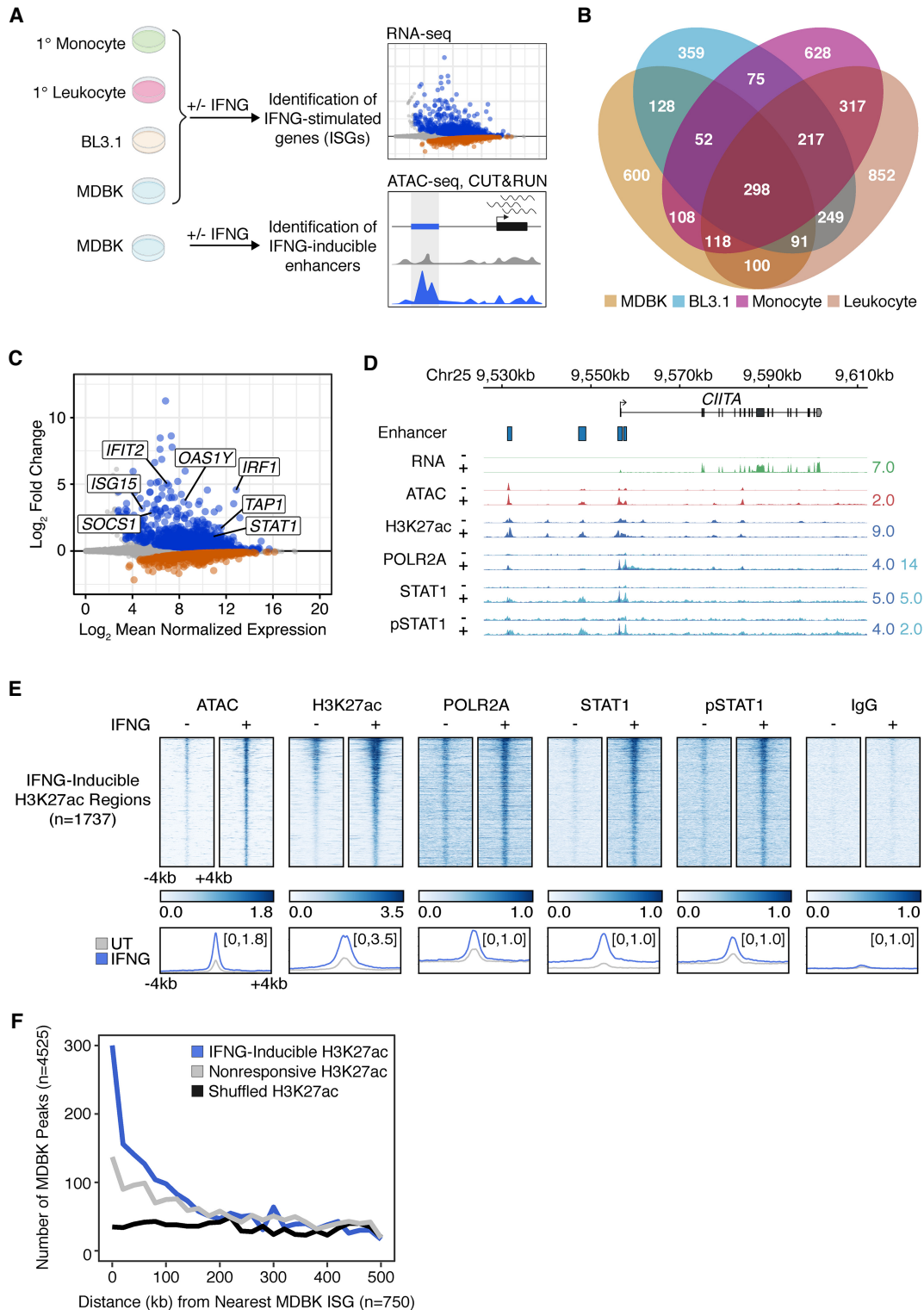**Corresponding author: edward.chuong@colorado.edu**

Figure 1. Epigenomic profiling of the bovine type II IFN response. (*A*) Schematic of experimental design. Created with BioRender.com. (*B*) Venn diagram of ISGs as defined by RNA-seq from MDBK, BL3.1, monocytes, and leukocytes. (*C*) MA plot of ISGs (blue) and IRGs (orange) from MDBK RNA-seq. Genes with an FDR < 0.05 are shown in gray, and canonical ISGs are labeled. (*D*) Genome Browser view of the *CIITA* locus. MDBK RNA-seq, ATAC-seq, and CUT&RUN tracks are CPM-normalized. CUT&RUN tracks for POLR2A, STAT1, and phosphorylated STAT1 pulldowns are colored by aligned fragments ≤150 bp (dark blue) and >150 bp (light blue). The H3K27ac tracks correspond to all aligned fragments. Predicted IFNG-inducible enhancers are shown. Signal track maxima are indicated to the *right* of each track. (*E*) Heat maps showing CPM-normalized MDBK CUT&RUN signal over IFNG-inducible H3K27ac (n = 1737) peaks sorted by ascending FDR. *Bottom* metaprofiles depict average normalized CUT&RUN signal across loci. (*F*) Frequency histogram of absolute distances from each MDBK peak to the nearest MDBK ISG. (ISG) Interferon-stimulated gene.

CUT&RUN with antibodies against H3K27ac, phosphorylated RNA polymerase II subunit A (POLR2A; Entrez gene ID 282312), total STAT1, and phosphorylated STAT1 (Supplemental Tables S3, S4). Successful antibody targeting of bovine STAT1 was validated by recovery of Gamma-interferon Activated Site (GAS) and IFN-Stimulated Response Element (ISRE) motifs in peaks called from IFNG-stimulated pulldowns (Supplemental Table S5).

We next used our MDBK CUT&RUN chromatin profiling data to predict the locations of IFNG-inducible enhancers, based on a significant increase in H3K27ac signal coverage. Examination of chromatin profiles at known ISGs such as *CIITA* confirmed IFNG-inducible signal at predicted enhancers (Fig. 1D; Supplemental Fig. S2A). We also observed robust transcriptomic and epigenomic induction of *TLR4* (Entrez gene ID 281536), which has not been widely identified as a canonical ISG in mouse and primate lineages and may represent an example of a lineage-specific ISG (Supplemental Figs. S2B, S3). Using DESeq2, we identified 1737 elements with significantly increased H3K27ac read coverage across replicates (Supplemental Table S4). Consistent with IFNG-inducible enhancer activity, these elements displayed increased levels of chromatin accessibility, binding by total and phosphorylated STAT1, and POLR2A (Fig. 1E). These trends were not observed for elements with constitutive or decreased H3K27ac signal (Supplemental Fig. S4). Our set of predicted IFNG-inducible elements were also enriched for GAS and ISRE motifs predicted to bind STAT and IRF transcription factors (Supplemental Table S5), consistent with their activation by IFNG stimulation (Ivashkiv 2018). To ask whether these IFNG-inducible elements are enriched near ISGs, we focused on matched MDBK RNA-seq data, as previous work by others in mouse and human has shown that the cellular response to IFNG is cell type–specific (van Boxel-Dezaire and Stark 2007). Inducible elements are localized near MDBK ISGs (Fig. 1F) but not interferon-repressed genes (IRGs) or nonresponsive genes (Supplemental Fig. S5), compared to noninducible enhancers or random genomic regions, showing strong enrichment within 20 kb of ISGs ($P = 3.8 \times 10^{-300}$) compared to nonresponsive elements ($P = 2.0 \times 10^{-58}$) and randomly shuffled genomic regions ($P = 0.72$). Additionally, we leveraged publicly available Hi-C data from a Brangus cattle individual (Low et al. 2020) along with ATAC-seq and H3K27ac CUT&RUN data from MDBK cells to predict enhancer–gene interactions using the Activity-by-Contact (ABC) model (Supplemental Methods; Fulco et al. 2019). Using a more relaxed set of 5933 IFNG-inducible H3K27ac regions, 2415 (40.7%) were identified as potential enhancers for 1392 ISGs (Supplemental Table S6). These analyses indicate that our epigenomic data set successfully captured inducible genes and regulatory elements underlying the type II IFN response in MDBK cells.

We next asked what fraction of predicted IFNG-inducible regulatory elements were derived from TEs. We defined IFNG-inducible regulatory elements based on H3K27ac signal and used open chromatin peak summits to refine the location of regulatory elements due to the broad nature of H3K27ac enrichment (Supplemental Methods). Out of 4198 IFNG-inducible regulatory elements, we found that 1658 (39.5%) have peak summits residing within ruminant-specific TEs (Fig. 2A; Supplemental Table S7), which is in line with previous epigenomic analyses in human and mouse cells (Sundaram et al. 2014; Chuong et al. 2016). In contrast, fewer (1065/5042, or 21.1%) of noninducible or down-regulated elements were derived from TEs, suggesting that TEs play a pronounced role in shaping the epigenome during immune stimulation. TE-derived inducible enhancers also showed strong colocalization with ISGs (Supplemental Fig. S6), consistent with a role in inducible gene regulation.

We next used GIGGLE (Layer et al. 2018) to determine whether any specific TE families were significantly overrepresented within the set of predicted IFNG-inducible enhancers (Methods). We found that the MER41_BT and Bov-A2 TE families were strongly overrepresented within regions showing IFNG-inducible levels of H3K27ac and STAT1 (Fig. 2B,C; Supplemental Fig. S7; Supplemental Table S8). At a genome-wide level, MER41_BT and Bov-A2 elements show inducible levels of H3K27ac, as well as increased binding of POLR2A and STAT1 (Fig. 2D). These families contain perfect and partial matches to GAS or ISRE binding motifs within their consensus sequences and in most intact extant copies (Fig. 2E,F), indicating that MER41_BT and Bov-A2 elements are an abundant source of STAT1 binding sites in the bovine genome.

Co-option of the MER41_BT elements to regulate IFNG-inducible immune gene expression

We next focused on investigating the regulatory impact of MER41_BT, a ruminant-specific ERV represented by 5491 LTR elements in the cattle genome (Supplemental Fig. S8A). A previous study has shown that some MER41_BT elements harbor binding sites for several pluripotency factors and are enriched for open chromatin in bovine eight-cell embryos (Halstead et al. 2020). However, the activity of endogenous MER41_BT elements has not previously been implicated in bovine immunity. We previously demonstrated that the MER41_BT consensus could drive IFNG-inducible reporter expression in HeLa cells; however, we now sought to determine whether endogenous MER41_BT elements exhibit robust regulatory potential in the context of the bovine genome (Chuong et al. 2016). Of the 5491 MER41_BT elements annotated in the reference genome, 1071 contain at least one significant match to the GAS motif ($P < 1 \times 10^{-4}$). Our epigenomic analysis of MDBK cells revealed that many MER41_BT elements showed canonical signs of IFNG-inducible enhancer activity, including inducible levels of H3K27ac, POLR2A, STAT1, and phosphorylated STAT1 (Fig. 2D). A total of 107 MER41_BT elements show some evidence of inducible regulatory activity in MDBK cells, 45 of which are located within 250 kb of an ISG (Supplemental Table S9). These observations suggest that a subset of MER41_BT elements may functionally contribute to the regulation of bovine ISGs.

We next asked whether individual MER41_BT elements function to regulate important immune genes in bovine cells. We first investigated a MER41_BT element located 41 kb upstream of *IL2RB* (Entrez gene ID 104968644), designated MER41_BT.IL2RB, which encodes a membrane-bound receptor for the interleukin 2 (IL2) and interleukin 15 (IL15) signaling pathways. *IL2RB* exhibits inducible expression in IFNG-stimulated MDBK cells and monocytes in addition to LPS-stimulated bone marrow–derived macrophages (BMDMs), suggesting that this element may regulate *IL2RB* (Fig. 3A). In addition to potentially acting as an enhancer for *IL2RB*, the MER41_BT element appears to act as a promoter for a bovine-specific transcript annotated in RefSeq as *LOC510185* (Entrez gene ID 510185). Consistent with the predicted inducible activity of the MER41 element, the *LOC510185* transcript exhibits IFNG-inducible expression in multiple bovine cell lines and primary cells, including monocytes, leukocytes, the BL3.1 B lymphocyte cell line, and MDBK cells (Fig. 3B,C). Additionally, *LOC510185* is robustly expressed in LPS-stimulated BMDMs and alveolar macrophages infected with *Mycobacterium bovis* (Fig. 3B; Young et al.
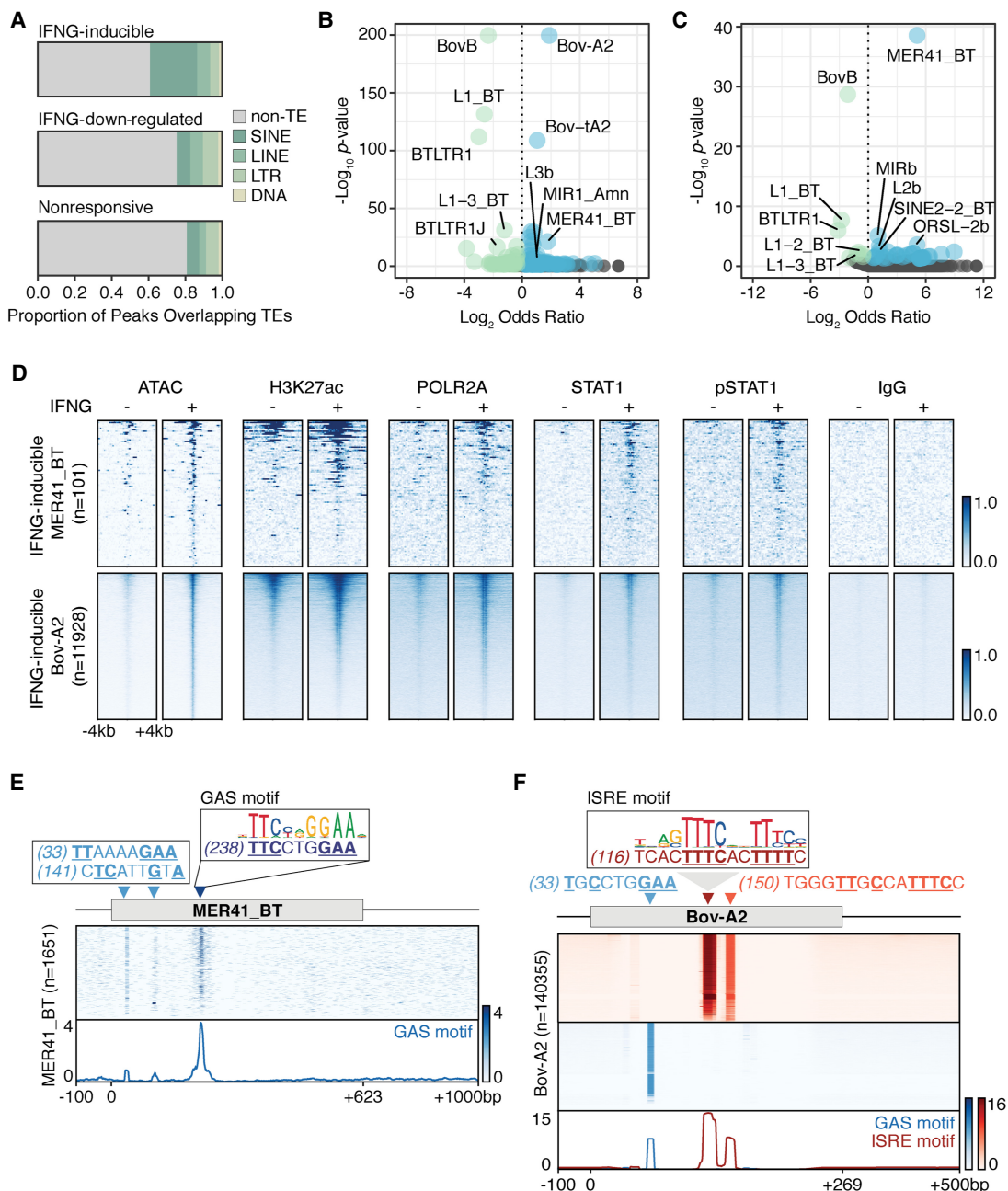
Figure 2. Contribution of ruminant-specific TEs to the IFNG-inducible regulatory landscape. (*A*) Proportion of IFNG-inducible (n = 4198), IFNG-down-regulated (n = 1548), and nonresponsive H3K27ac peaks (n = 3494) where the internal ATAC-seq summit overlaps a TE. (*B*) Volcano plot visualizing family-level TE enrichment for IFNG-inducible H3K27ac. TE families with a Fisher's two-tailed *P*-value < 0.05 were defined as enriched (blue) or depleted (green) based on the reported odds ratio. Nonsignificant families are shown in gray. (*C*) Family-level TE enrichment for IFNG-inducible STAT1 peaks, as in *B*. (*D*) Heat maps showing normalized MDBK ATAC-seq and CUT&RUN signal over IFNG-inducible MER41_BT (n = 101) and Bov-A2 (n = 11928) elements sorted by descending mean CPM-normalized signal. (*E*) Schematic of IFNG-associated motifs within the MER41_BT consensus and extant sequences (n = 1651) filtered for those that are at least 50% of full length relative to the consensus and sorted by descending CPM-mean normalized signal. *Bottom* metaprofile represents average signal across all elements. (*F*) Schematic of motifs as in *E* but across 140355 Bov-A2 elements, filtered for those within 5% of full length relative to the consensus. Heat map intensity corresponds to motif matches based on the log likelihood ratio. Partial matches are shown in light blue or light red, and the numbers in parentheses represent the position of each predicted motif in the consensus sequence.

2018; Hall et al. 2020). Whereas the function of *LOC510185* is unknown, it arose from a recent tandem segmental duplication of the *IL2RB* locus, and its predicted open reading frame retains only the extracellular ligand binding domain of IL2RB (Supplemental Fig. S9). We aligned the IL2RB and LOC510185 protein sequences

and calculated the estimated time of duplication using the rate of nonsynonymous mutation (Supplemental Fig. S10). This revealed that the duplication occurred ~34.1 mya, suggesting that the duplication occurred after the divergence of Tragulidae and Pecora (CI 40–48 mya) (Supplemental Methods; Kumar et al.
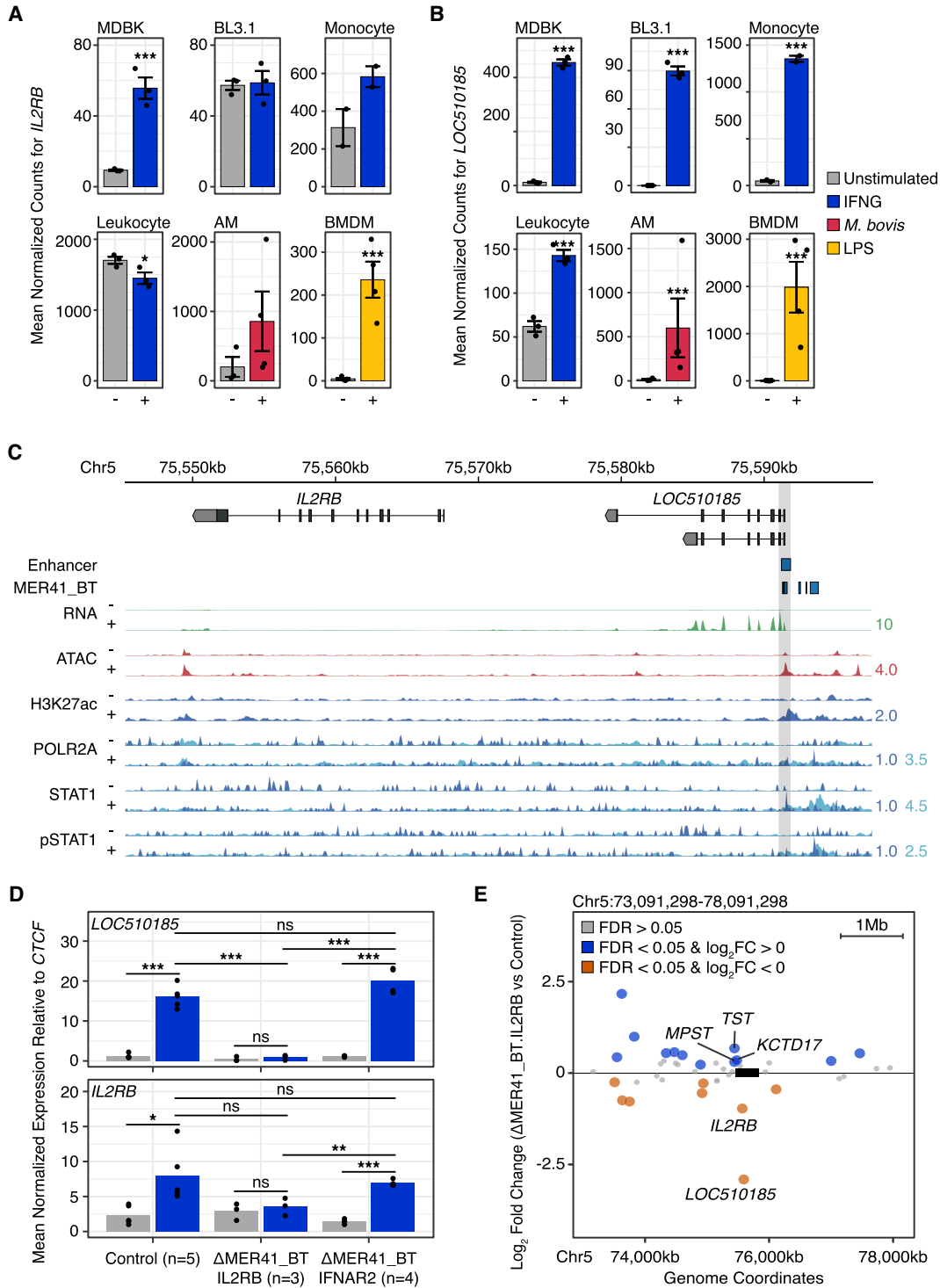
Kelly et al.

Figure 3. Co-option of a MER41_BT element for *IL2RB* and *LOC510185* regulation. (*A*) DESeq2-normalized counts showing immune-stimulated expression for *IL2RB* from MDBK (n = 3), BL3.1 (n = 3), monocytes (n = 2), leukocytes (n = 3), alveolar macrophages (AM, n = 3 untreated, n = 4 stimulated), and bone marrow–derived macrophages (BMDM, n = 4). Treatments are indicated by color. (***) FDR < 0.0001, (*) FDR < 0.01. Error bars denote SEM. (*B*) Normalized counts as in *A* but for *LOC510185*. (*C*) Genome Browser view of *IL2RB* and *LOC510185*. MDBK RNA-seq, ATAC-seq, and CUT&RUN tracks are CPM-normalized. CUT&RUN signal profile tracks for POLR2A, STAT1, and pSTAT1 pulldowns were generated using subnucleosomal fragments ≤150 bp (dark blue) as well as larger fragments >150 bp (light blue). MER41_BT.IL2RB (Chr 5: 75,591,298–75,591,607) is highlighted in gray. Predicted IFNG-inducible enhancers are shown. Values on the *right* of each track correspond to signal maxima. (*D*) RT-qPCR of *LOC510185* and *IL2RB* expression levels in control (n = 5), MER41_BT.IL2RB (n = 3), and MER41_BT.IFNAR2 (n = 4) MDBK clonal cells after IFNG treatment. Data points denote clonal replicates. (*) P < 0.05, (**) P < 0.001, (***) P < 0.0001, Student's t-test. (*E*) RNA-seq gene expression differences caused by the MER41_BT.IL2RB deletion within a 5-Mb window centered on the deletion site (box not to scale), after IFNG-stimulation. Significantly up-regulated (blue) and down-regulated (orange) genes within 500 kb of the element are labeled. (AM) Alveolar macrophage, (BMDM) bone marrow–derived macrophage.

1478 Genome Research
www.genome.org

2017). To this end, we confirmed the presence of a *LOC510185* ortholog in pecorans but not camelids, pig, or whippomorphs (Supplemental Table S10). The widespread inducible expression of *LOC510185* in bovine cells suggests it may function as a bovine-specific factor involved in IL2 or IL15 signaling.

To experimentally characterize the potential promoter and enhancer activity of MER41_BT.IL2RB, we used CRISPR-Cas12a to delete the element in MDBK cells and determine its impact on gene expression. Although MDBK cells are an immortalized kidney cell line where *IL2RB* is only weakly expressed, transcriptomic evidence from primary immune cells suggests that inducible *IL2RB* and *LOC510185* expression is physiologically relevant. To specifically test regulatory activity, we deleted a region immediately upstream of the predicted *LOC510185* transcriptional start site that harbors a STAT1 binding site (Methods). We recovered three sequence-validated independent clones carrying homozygous deletions of MER41_BT.IL2RB (Supplemental Fig. S11) and profiled the IFNG responses for each clone using RT-qPCR and RNA-seq (Fig. 3D; Supplemental Fig. S12; Supplemental Table S11). We observed a reduction in IFNG-stimulated expression levels of both *IL2RB* (qPCR $P = 0.11$; RNA-seq FDR = 0.0011) and *LOC510185* ($P = 9.8 \times 10^{-5}$; FDR = 1.7×10^{-10}) upon deletion of MER41_BT.IL2RB (Fig. 3D,E; Supplemental Fig. S12). Deletion of another MER41 element upstream of the *IFNAR2* locus had no impact on *IL2RB* or *LOC510185* expression (Fig. 3D). Clonal cells harboring the MER41_BT.IL2RB deletion still exhibit increased *IFNAR2* expression in response to IFNG by RT-qPCR, indicating that the global response to IFNG remained intact (Fig. 4C). Despite seeing a significant reduction in *IL2RB* expression in IFNG-stimulated cells harboring the MER41_BT.IL2RB deletion, the ABC model did not identify the MER41_BT.IL2RB element as a potential regulator for *IL2RB* (Supplemental Table S6). However, we visually confirmed that MER41_BT.IL2RB and IL2RB fall within the same 1-Mb domain by Hi-C (Supplemental Fig. S13A). Together, these results show that a specific MER41_BT element functions as an IFNG-inducible promoter of *LOC510185* and may function as an IFNG-inducible enhancer of *IL2RB* in MDBK cells. However, inducible *IL2RB* expression in primary bone marrow–derived macrophages and MDBK cells (Fig. 3A), albeit statistically significant, is relatively low and may not indicate functional protein expression. Although we wished to confirm the function of MER41_BT.IL2RB in cells expressing functional IL2RB, we were unable to identify a suitable cell line. Further investigation is required to determine whether MER41_BT.IL2RB functionally regulates IL2RB and LOC510185 protein expression.

We next characterized a separate MER41_BT element located 19 kb upstream of the gene *IFNAR2* (MER41_BT.IFNAR2), which encodes a membrane-bound receptor mediating type I interferon signaling. Bovine *IFNAR2* shows IFNG-inducible expression in leukocytes, MDBK cells, and BL3.1 cells (Fig. 4A). *IFNAR2* expression is also induced in alveolar macrophages challenged with *M. bovis* and LPS-stimulated BMDMs (Fig. 4A). The predicted upstream MER41_BT.IFNAR2 element shows inducible enhancer activity in MDBK cells and is the top predicted enhancer for *IFNAR2* according to the ABC model (Fig. 4B; Supplemental Table S6). Additionally, MER41_BT.IFNAR2 and *IFNAR2* appear to fall in the same 1-Mb chromatin domain by visual inspection of the Hi-C contact map (Supplemental Fig. S13B). To test whether this element regulates bovine *IFNAR2*, we used CRISPR-Cas12a to delete a region containing a STAT1 binding motif within the MER41_BT.IFNAR2 element (Supplemental Fig. S14). We isolated four independent clones carrying the GAS deletion and profiled the IFNG response

in each clone using RT-qPCR (Fig. 2C). We also performed RNA-seq on two of these four clones (Supplemental Fig. S15B; Supplemental Table S11). We found that *IFNAR2* was still inducible by IFNG stimulation but showed significantly reduced expression compared to control cells ($P = 0.0044$; FDR = 2.4×10^{-6}) and cells independently harboring the MER41_BT.IL2RB deletion (Fig. 4C,D; Supplemental Fig. S15B). We also found that the nearby interferon gamma receptor gene *IFNGR2* showed significantly reduced expression under IFNG stimulation (FDR = 2.6×10^{-8}) (Fig. 4D; Supplemental Fig. S15A,C), indicating that deletion of the MER41_BT element affected the expression of multiple genes. We did not observe complete ablation of inducible *IFNAR2* or *IFNGR2* expression, potentially due to the presence of nearby IFNG-inducible enhancers that could compensate for the deletion (Fig. 4B). These results demonstrate that MER41_BT.IFNAR2 functions as an enhancer that contributes to IFNG-inducible regulation of *IFNAR2* and *IFNGR2* in MDBK cells.

TEs contribute to evolutionary diversification of ruminant immune responses

Having established that a subset of ruminant-specific TEs function to regulate ISGs in bovine cells, we next examined their impact on ruminant genome evolution. The MER41_BT family was originally annotated as a bovine-specific ERV belonging to the MER41-like family of ERVs, which are present in several other lineages, including primates and carnivores (Bao et al. 2015). MER41-like ERVs from different lineages exhibit high sequence similarity, consistent with independent germline integrations by a retroviral lineage that underwent multiple cross-species transmissions (Zhuo and Feschotte 2015). To better resolve the timing of MER41_BT emergence, we re-annotated repeats in the assembled reference genomes of 30 mammals including 19 cetartiodactyl species (Supplemental Table S12), using a uniform set of MER41-like consensus sequences (Methods). Based on the presence and abundance in each species clade, we estimated when each MER41 family originated (Fig. 5A). This confirmed that primate MER41 and bovine MER41_BT arose from separate germline integration events and revealed that MER41_BT elements are present in all Cetruminantia species. MER41_BT elements derive from an ancestral MER41-like retrovirus that integrated into an ancestral cetartiodactyl genome roughly 70 mya, prior to the divergence of Camelidae and Artiofabula (Fig. 5B). This estimate coincides with the emergence of primate MER41 elements after the divergence of anthropoid primates from prosimian lineage roughly 70 mya, consistent with a scenario where primate and ruminant MER41 derive from a common ancestral retroviral lineage.

We next used synteny analysis to determine whether individual MER41_BT insertions were conserved across Cetruminantia species. We asked whether each annotated element in the cattle genome, along with flanking sequence, was present in the genomes of select cetartiodactyl species (Supplemental Methods). This revealed that the majority of insertions were poorly conserved across cetruminant species (Supplemental Table S13), suggesting that MER41_BT may have continued to propagate in host genomes as cetruminant lineages diverged. However, further investigation will be required to more directly identify clade-specific insertion sites to assess the extent of MER41 diversification. Given our evidence that MER41_BT influences gene regulation in cattle, it is possible that lineage-specific insertions of MER41_BT elements have contributed to gene regulatory diversification across cetruminant lineages.
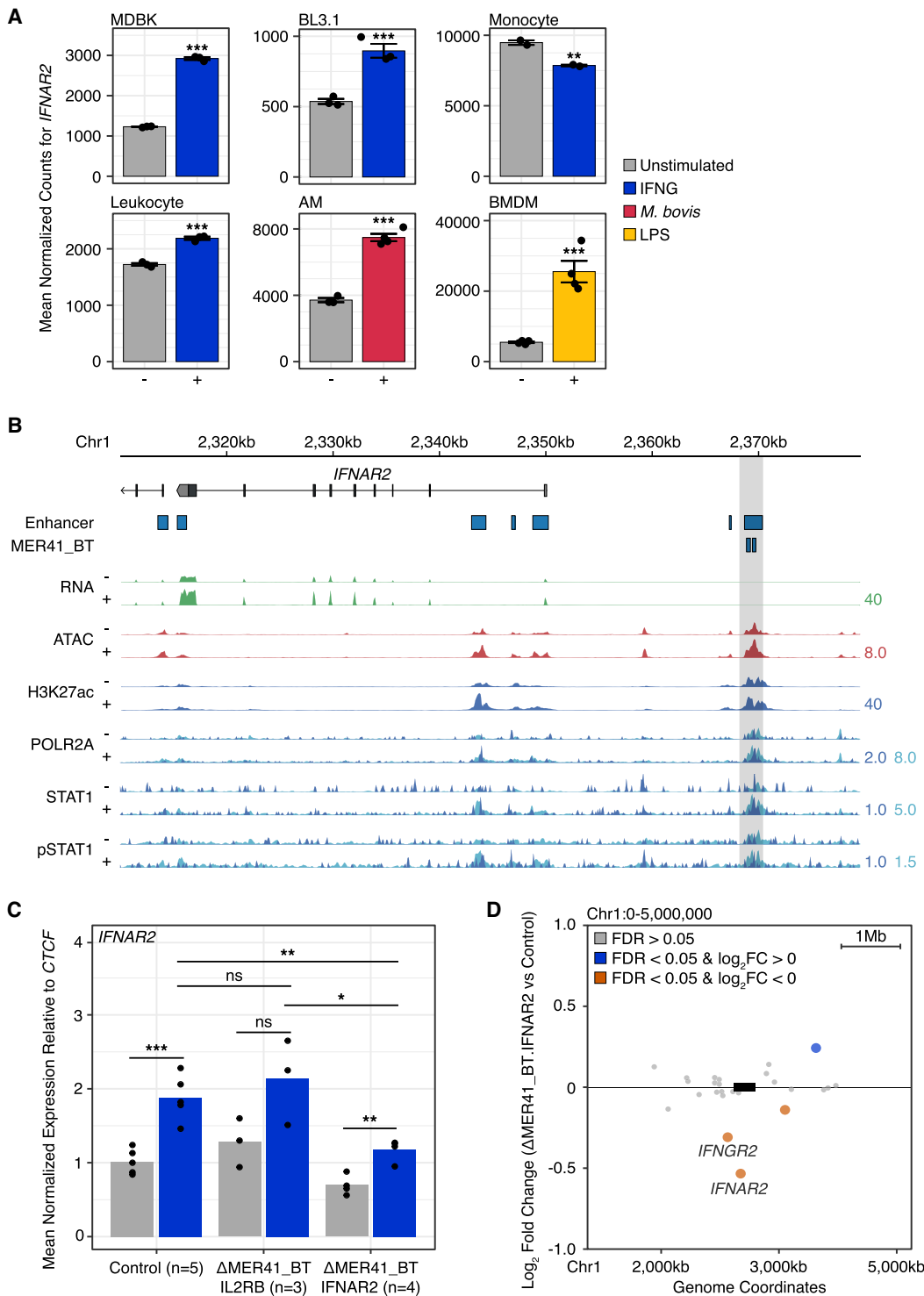
Figure 4. Co-option of a MER41_BT element for *IFNAR2* regulation. (*A*) Mean DESeq2-normalized counts showing expression for *IFNAR2* from wild-type bovine cells. (***) FDR < 0.0001, (**) FDR < 0.001. Error bars denote SEM. (*B*) Genome Browser view of *IFNAR2*. MDBK RNA-seq, ATAC-seq, and CUT&RUN tracks are normalized per million reads. CUT&RUN tracks for POLR2A, STAT1, and pSTAT1 pulldowns are divided by aligned fragments ≤150 bp (dark blue) and >150 bp (light blue). IFNG enhancers shown indicate predicted IFNG-inducible H3K27ac peaks. MER41_BT.IFNAR2 (Chr 1: 2,368,871–2,369,745) is highlighted in gray. Predicted IFNG-inducible enhancers are shown. Values on the *right* of each track correspond to signal maxima. (*C*) qPCR of *IFNAR2* expression levels in control (n = 5), MER41_BT.IL2RB (n = 3), and MER41_BT.IFNAR2 (n = 4) MDBK clonal cells after IFNG treatment. Data points denote clonal replicates. (*) $P < 0.05$, (**) $P < 0.001$, (***) $P < 0.0001$, Student's t-test. (*D*) RNA-seq gene expression differences caused by the MER41_BT.IFNAR2 deletion within a 5-Mb window centered on the deletion site (box not to scale), after IFNG stimulation. Significantly up-regulated (blue) and down-regulated (orange) genes within 500 kb of the element are labeled.
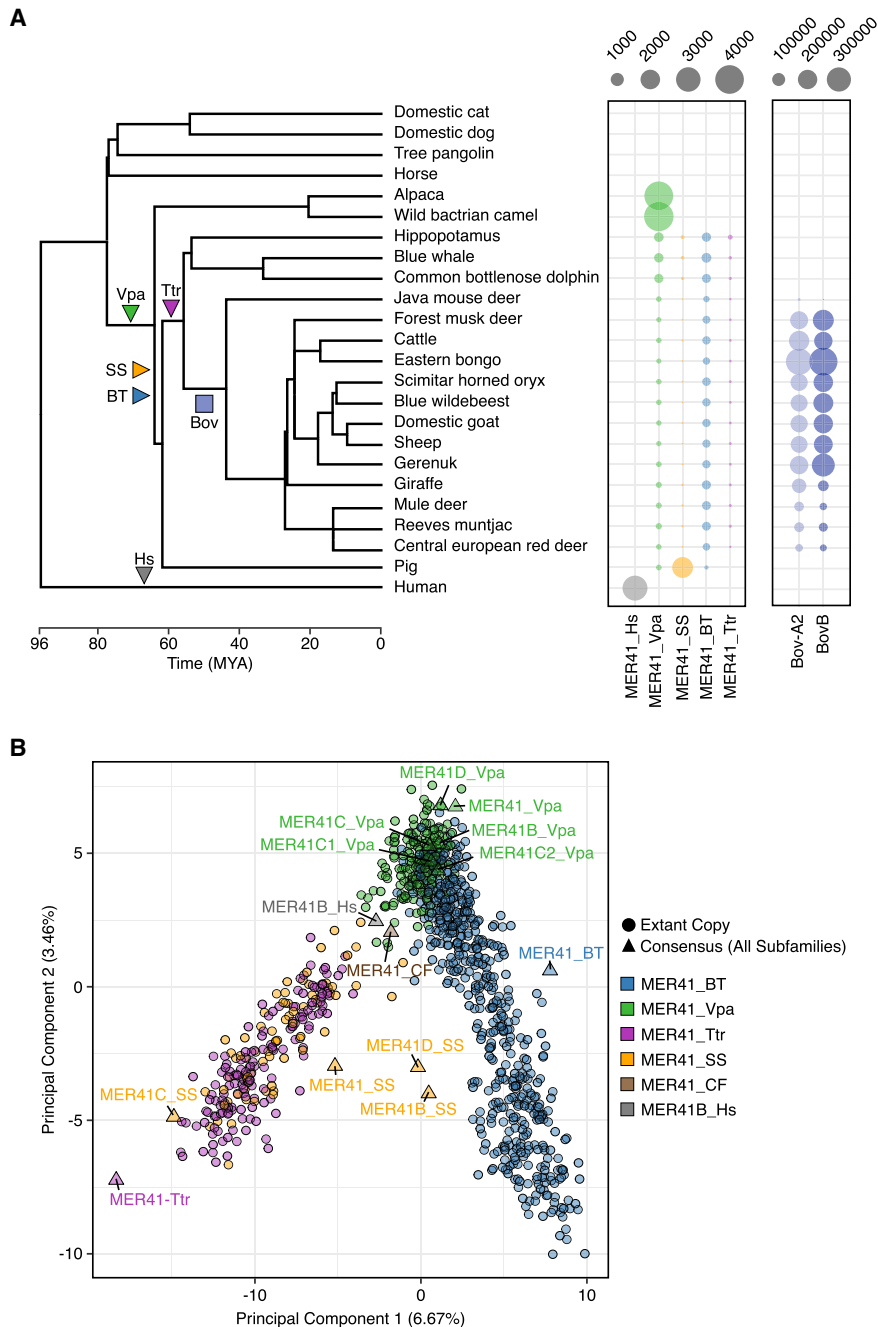
Figure 5. Mapping TE family evolution in Cetartiodactyla. (*A*) Number of MER41, Bov-A2, and BovB elements identified across 24 mammalian assemblies using RepeatMasker. Bubble size corresponds to the number of TEs annotated in each assembly. Phylogeny was obtained from TimeTree (Kumar et al. 2017) and depicts species divergence. Estimated insertion times are indicated on the phylogeny as triangles (MER41) and a square (Bov-A2 and BovB). (*B*) Principal component analysis (PCA) plot of extant copies of bovine MER41 elements (circles) and cetartiodactyl, carnivore (MER41_CF), and human (MER41B_Hs) MER41 consensus sequences (triangles) colored by superfamily. Consensus sequences are named based on the species in which they were discovered. (Hs) *Homo sapiens*, (Vpa) *Vicugna pacos*, (Ttr) *Tursiops truncatas*, (SS) *Sus scrofa*, (CF) *Canis familiaris*, (BT) *Bos taurus*.

Next, we examined the evolutionary history of Bov-A2 elements, which our analysis identified as another major source of predicted IFNG-inducible enhancers. Bov-A2 elements are SINE-like elements derived from the BovB retrotransposon roughly 20–40 mya and constitute one of the most abundant sequences in bovine

genomes (Nijman et al. 2002; Onami et al. 2007; Nilsson et al. 2012). Bov-A2 elements have been predicted to harbor binding sites for immune-related transcription factors STAT1, STAT3, NF-kB, and IRF1 (Dekel et al. 2015; Young et al. 2018). However, analysis in previous studies was limited to few Bov-A2 elements, and it remains unclear whether Bov-A2 may act globally as functional regulators of the innate immune response. There are over 360,000 Bov-A2 sequences in the cattle genome, 280,000 of which contain at least one GAS or ISRE motif. In MDBK cells, Bov-A2 and other BovB derivatives represent 32.5% of all predicted IFNG-inducible regulatory elements, or 82.3% of all TE-derived inducible elements (Supplemental Table S7). Compared to MER41_BT, Bov-A2 elements were active more recently, as evident through previous studies that identified cross-species differences in Bov-A2 insertions (Nijman et al. 2002; Nilsson et al. 2012; Dekel et al. 2015; Young et al. 2018). Additionally, Bov-A2 elements are hypomethylated in cattle sperm and harbor binding sites for host zinc finger proteins, suggesting that they may be recently or still active (Zhou et al. 2020). Taken together, this indicates that Bov-A2 has the potential to drive functional regulatory variation in immunity; however, no study to date has directly examined whether Bov-A2 insertions are polymorphic within modern bovid species.

We investigated whether Bov-A2 elements are a source of modern variation by using MELT (Gardner et al. 2017) to identify polymorphic mobile element insertions in a diverse panel of 38 cattle whole-genome sequences (Tsuda et al. 2013; Shin et al. 2014; Stothard et al. 2015; Bickhart et al. 2016; Heaton et al. 2016; Chen et al. 2018; Workman et al. 2018). Across all individuals, our analysis revealed 1017 Bov-A2 elements with read-supported evidence for insertional polymorphism, including 740 nonreference insertions and 277 deletions relative to the reference Hereford *bosTau9/ARS-UCD1.2* assembly (Fig. 6A; Supplemental Fig. S16A; Supplemental Table S14). We also observed 1782 polymorphic BovB elements, which are the proposed originator of Bov-A2 elements (Fig. 6A; Supplemental Table S14; Nijman et al. 2002; Onami et al. 2007; Nilsson et al. 2012). As expected, we did not identify any insertional polymorphisms of the more ancient MER41_BT family (Fig. 6A; Supplemental Table S14). Many Bov-A2 and BovB polymorphisms were present in more than one individual (Fig. 6B) and principal components analysis
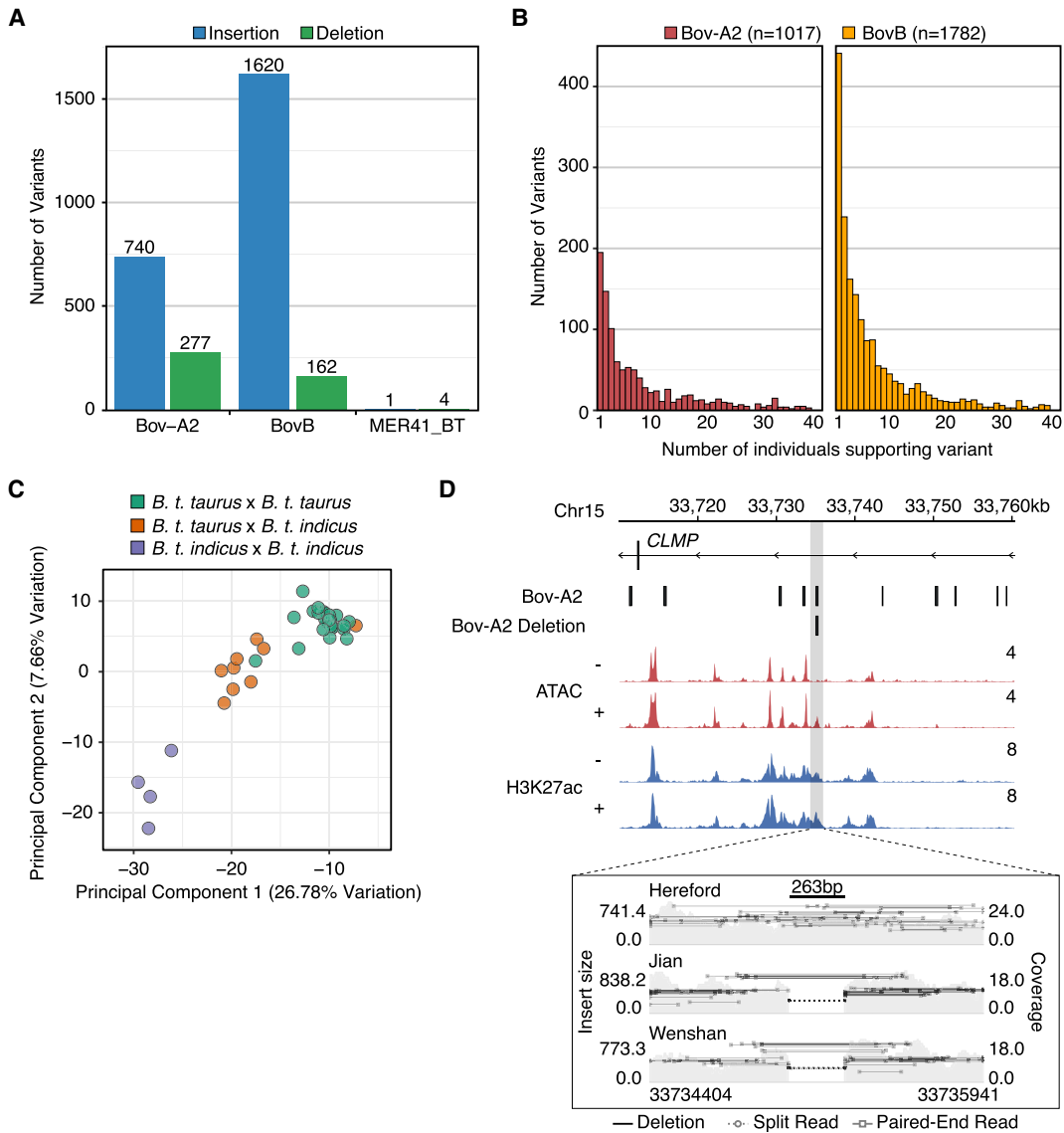
Figure 6. Detection of Bov-A2 insertion polymorphisms. (*A*) Number of polymorphic Bov-A2, BovB, and MER41_BT insertions (blue) and deletions (green) for each TE family. (*B*) Histogram of polymorphic Bov-A2 variants as a function of the number of individuals supporting each variant. Individuals were defined as supporting a variant if they showed evidence of at least one variant allele. (*C*) Histogram of polymorphic BovB variants as a function of number of individuals supporting each variant. (*D*) Genome Browser view of a putative Bov-A2 deletion (Chr 15: 33,735,041–33,735,304) intronic to *CLMP*. 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 was produced using Samplot (Belyeu et al. 2021) and depicts aligned fragments from three individuals over the predicted Bov-A2 deletion.

of cattle genomes based on Bov-A2 and BovB insertion status corresponded with subspecies (Fig. 6C; Supplemental Fig. S16B). Collectively, our analysis indicates that Bov-A2 insertions are a current source of structural genetic variants in modern cattle individuals.

Finally, we asked whether any polymorphic Bov-A2 insertions show evidence of regulatory activity. Based on our epigenomic data from MDBK cells, we identified 17 Bov-A2 elements with regulatory activity that were present in the Hereford reference genome but missing in at least one allele in our analysis (Supplemental Table S15). This included a Bov-A2-derived element located near the bovine gene *CLMP*, which was present in the Hereford reference genome and MDBK cells but missing in Angus individuals (Fig. 6D). Additionally, we identified Bov-A2-de-

rived deletions in the first intron of *BICC1* and 26 kb upstream of *SNX19* that showed evidence of IFNG-inducible regulatory activity (Supplemental Fig. S16C,D). Although more work is necessary to causally link Bov-A2 polymorphic loci to functional variation, our observations suggest that TE insertional polymorphisms may be an underappreciated class of regulatory variants that contribute to differences in immune gene expression.

Discussion

Infectious diseases have a large economic impact on the cattle industry, and understanding the genetic architecture of bovine immune responses is key for improving treatments and breeding by

genomic selection (Keirn et al. 2001; Stear et al. 2001; Thompson-Crispi et al. 2014; Mallard et al. 2015). To this end, thousands of individual genomes from diverse cattle breeds and other ruminant species have been sequenced to date (Hayes and Daetwyler 2019). However, the majority of association studies have focused on SNPs, with structural variants only recently gaining attention (Liu et al. 2010; Chen et al. 2017; Hu et al. 2020). Lineage-specific TEs are a prominent source of genomic variation that has been implicated in trait diversity in many domesticated species (Lisch 2013; Wragg et al. 2013; Langevin et al. 2018; Bannasch et al. 2021), but there are currently only a handful of examples of TE insertions affecting ruminant variation, including cholesterol deficiency and coat color (Menzi et al. 2016; Liang et al. 2021). Our epigenomic approach uncovered TEs as a source of IFNG-inducible enhancers that regulate bovine ISGs, with some that are still polymorphic in modern cattle. Our findings warrant further examination of TE insertions as a class of variants to be considered in bovine genome-wide association studies, potentially by adopting new graph-based reference genomes that encompass bovine structural variants (Crysnanto et al. 2021).

By investigating TE-mediated regulation in a less well-studied organism, our study also uncovered a striking example of parallel evolution, involving the independent co-option of the MER41 family of retroviruses in both primates and ruminants. Primate MER41 and ruminant MER41_BT share high sequence identity but originate from independent germline integration events, consistent with separate germline integrations by a retroviral lineage that underwent multiple cross-species transmissions (Zhuo and Feschotte 2015). Therefore, we speculate that descendants of an ancestral MER41 retrovirus lineage integrated into the genomes of multiple mammalian lineages in a time period of 60–70 mya, including primates and Cetartiodactyla. Our study found that MER41 elements integrated into separate lineages have been convergently co-opted for similar regulatory functions regulating host IFN responses. It is possible that the sequences of ancient MER41 retroviruses may have had regulatory properties, such as inducible activation, that predisposed them to successful endogenization and eventual domestication for host function.

Our study also revealed pervasive enhancer activity at Bov-A2 elements, which are among the most abundant sequences in cattle genomes. Bov-A2 elements are SINEs derived from the BovB retrotransposon, which integrated into ruminant genomes via horizontal transfer from a reptile (Ivancevic et al. 2018). Individual Bov-A2 elements have previously been characterized as a source of functional variation across pecoran species, functioning as a source of LPS-inducible promoter sequences. One such element has been shown to function as the promoter for the cattle *NOS2* locus, conferring LPS-inducible expression in cattle that we found is unexplainable by the presence of a cattle-specific enhancer (Supplemental Fig. S17; Young et al. 2018). Our study reveals that thousands of Bov-A2 elements may possess IFNG-inducible regulatory activity. Together with our analysis confirming that Bov-A2 insertions are still polymorphic in modern cattle, these findings implicate an important role for TEs in contributing to both ruminant evolution and modern genetic variation. Although further experimental investigation will be necessary to link TE-derived regulatory variants to phenotypic differences, our findings suggest that TEs may underlie differences in immune gene regulation across bovine individuals and populations.

Together with our previous work in human cells (Chuong et al. 2016), our discovery that lineage-specific TEs have been independently co-opted to regulate ISGs in multiple species draws par-

allels to the finding that ERVs have been repeatedly co-opted to mediate placentation in mammals (Dupressoir et al. 2012; Chuong et al. 2013; Dunn-Fletcher et al. 2018; Sun et al. 2021). A growing number of case examples have been discovered where unrelated TEs have been co-opted for convergent functions in multiple species, suggesting that TEs may have a propensity to facilitate the evolution of certain biological processes, particularly those involving genetic conflicts. Studies investigating the epigenetic regulation of IFN responses in diverse vertebrate species may therefore uncover a widespread role for TEs in mediating the evolution of innate immune regulatory networks.

Methods

For a detailed list of all reagents, primer sequences, and gRNA sequences used in this study refer to Supplemental Table S16. A more detailed summary of the approaches used in this study can be found in Supplemental Materials.

Profiling cell line immune responses

Bovine MDBK cells (gift from Sara Sawyer) were cultured in MEM supplemented with nonessential amino acids, 1× penicillin-streptomycin (P/S), and 10% fetal bovine serum (FBS) and routinely passaged using accutase. BL3.1 cells were cultured in RPMI 1640 medium supplemented with 1× P/S and 10% FBS. All IFNG treatments were performed using 100 ng/mL recombinant bovine IFNG. Cell lines were cultured at 37°C and 5% CO_2. Stranded, poly(A)-enriched RNA-seq libraries were performed on untreated and IFNG-stimulated MDBK, BL3.1, monocyte, and leukocyte cells using the KAPA mRNA HyperPrep kit according to the manufacturer's instructions. ATAC-seq and CUT&RUN was performed on untreated and IFNG-stimulated MDBK cells as previously described (Corces et al. 2017; Meers et al. 2019). Paired-end 150-bp sequences were generated on an Illumina NovaSeq 6000 (University of Colorado Genomics Core). Sequences were aligned to the bovine genome (bosTau9) with HISAT2 v2.1.0 (Kim et al. 2019), BWA-MEM v0.7.15 (Li 2013), or Bowtie 2 v2.2.9 (Langmead and Salzberg 2012) and filtered based on mapping score (MAPQ > 10). ISGs were identified using DESeq2 v1.26.0 (Love et al. 2014) with fragments assigned to the complete bosTau9 RefSeq annotation. ATAC-seq and CUT&RUN peak calling were performed using MACS2 v2.1.1 (Liu 2014). IFNG-stimulated enhancers were defined as regions marked by IFNG-inducible H3K27ac that overlap ATAC-seq signal from IFNG-stimulated MDBK cells. Normalized bigWig files corresponding to read coverage per 1 million reads were used for heat map and metaprofile visualization using deepTools v3.0.1 (Ramírez et al. 2014). IFNG induction was confirmed by enrichment of GAS and ISRE motifs using XSTREME v5.4.1 (Grant and Bailey 2021). To determine whether IFNG-inducible enhancers are enriched around ISGs, IFNG-inducible enhancers were defined by the absolute distance to the nearest MDBK ISG, and statistical significance was determined for the first 20-kb bin by Fisher's exact test using BEDTools v2.28.0 (Quinlan and Hall 2010).

Transposable element analysis

To identify TE families that are globally enriched for predicted IFNG-inducible enhancers, we used the existing RepeatMasker annotations for bosTau9, available on the UCSC Genome Browser (updated 2018-11-07). TE-derived, IFNG-inducible enhancers were defined using H3K27ac peaks centered on overlapping ATAC-seq summits (1 bp). To assess enrichment at the family-level, GIGGLE v0.6.3 (Layer et al. 2018) was used to query ATAC-seq and CUT&RUN peaks against a database of all annotated TEs. TE

heat maps were generated using a list of MER41_BT and Bov-A2 elements overlapping IFNG-inducible enhancers that harbor 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 TEs as heat maps using deepTools v3.0.1 (Ramírez et al. 2014). To assess enrichment of IFNG-inducible TEs around ISGs, IFNG-inducible TEs were defined by the absolute distance to the nearest MDBK ISG, and statistical significance was determined for the first 20-kb bin by Fisher's exact test using BEDTools v2.28.0 (Quinlan and Hall 2010).

Generation and analysis of CRISPR-enAsCas12a mutants

MDBK cells stably expressing *enAsCas12a* were prepared according to a previously published protocol (Supplemental Methods; DeWeirdt et al. 2021). *enAsCas12a* expression in clonal MDBK cells was validated by RT-qPCR using the Luna Universal One-Step RT-qPCR kit following the manufacturer's instructions. For each MER41 element (associated with *IFNAR2* and *LOC510185*), two 23-bp gRNA sequences were designed to generate a single internal deletion that encompasses the putative STAT1 binding sites. Each crRNA oligonucleotide was cloned into pRDA_052 (Addgene) following a modified protocol from DeWeirdt et al. (2021) (Supplemental Methods). To generate each MER41 deletion, MDBK cells stably expressing *enAsCas12a* were transduced with lentivirus containing the gRNAs and underwent puromycin selection (2 ug/mL); 80–120 clones were screened for homozygous deletions by PCR using internal and flanking primer pairs at the expected deletion site. Breakpoint sequences were confirmed by Sanger sequencing (Quintara Biosciences; Genewiz). Independently, clonal MDBK lines stably expressing *enAsCas12a* without gRNAs underwent an additional clonal isolation process. Five control clones were selected for further experimentation.

Re-annotation of MER41 repeats

High-quality assemblies from 24 mammalian species representing five distinct lineages (primate, Cetartiodactyla, Perissodactyla, Pholidota, Carnivora) were annotated using RepeatMasker v4.1.0 (Smit et al. 2019) using the RepeatMasker-Repbase (RMRB, release 20181026) library (Supplemental Table S12). For each assembly, the numbers of annotated, length-filtered MER41, Bov-A2, and BovB elements were visualized in a bubble plot (Supplemental Methods). 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 filtered by length and collapsed into a list of 990 elements. Consensus sequences for a representative set of families were also included. The number of occurrences of each unique 6-mer in each sequence was counted using kmer-counter (https://github.com/alexpreynolds/kmer-counter) and used for principal component analysis (PCA).

Detection of TE insertion polymorphisms

Publicly available whole-genome, paired-end sequencing data were aligned to the bosTau9 reference assembly with the bosTau5 Y Chromosome assembly using BWA-MEM v0.7.15 (Supplemental Table S17; Li 2013; Bellott et al. 2014). TE variants were called using MELT v2.1.5 (Gardner et al. 2017) and filtered according to length, genotype, confidence score, and number of supporting split reads (Supplemental Materials). Filtered deletion and insertion variant calls were aggregated and summarized based on whether or not they had at least one supporting allele prior to PCA. Candidate Bov-A2 deletion variants were classified as putative regulatory elements if they overlapped any ATAC-seq or CUT&RUN peaks and fell within 250 kb of a gene TSS (Supplemental Table S15). Alignments over filtered deletions were visualized using Samplot v1.1.6 (Belyeu et al. 2021).

External data sets

Publicly available data were downloaded from public repositories using fasterq-dump from the NCBI SRA Toolkit. RNA-seq data sets were obtained from the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJEB22535 and from the NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE116734. Whole-genome sequencing data sets were obtained from the NCBI BioProject database under acccesion numbers PRJDB2660, PRJEB1829, PRJNA176557, PRJNA210519, PRJNA277147, PRJNA324822, PRJNA379859, and PRJNA325058.

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE185082. All code is available as Supplemental Code and on GitHub (https://github.com/coke6162/bovine_TE_evolution).

Competing interest statement

The authors declare no competing interests.

References

Adelson DL, Raison JM, Edgar RC. 2009. Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome. *Proc Natl Acad Sci* **106:** 12855–12860. doi:10.1073/pnas.0901282106

Albrecht E, Komolka K, Kuzinski J, Maak S. 2012. Agouti revisited: transcript quantification of the ASIP gene in bovine tissues related to protein expression and localization. *PLoS One* **7:** e35282. doi:10.1371/journal.pone.0035282

Bannasch DL, Kaelin CB, Letko A, Loechel R, Hug P, Jagannathan V, Henkel J, Roosje P, Hytönen MK, Lohi H, et al. 2021. Dog colour patterns explained by modular promoters of ancient canid origin. *Nat Ecol Evol* **5:** 1415–1423. doi:10.1038/s41559-021-01524-x

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* **6:** 11. doi:10.1186/s13100-015-0041-9

Barreca C, O'Hare P. 2004. Suppression of herpes simplex virus 1 in MDBK cells via the interferon pathway. *J Virol* **78:** 8641–8653. doi:10.1128/JVI.78.16.8641-8653.2004

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. doi:10.1038/nature13206

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. doi:10.1186/s13059-021-02380-5

Bickhart DM, Xu L, Hutchison JL, Cole JB, Null DJ, Schroeder SG, Song J, Garcia JF, Sonstegard TS, Van Tassell CP, et al. 2016. Diversity and population-genetic properties of copy number variations and multicopy genes in cattle. *DNA Res* **23**: 253–262. doi:10.1093/dnares/dsw013

Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, Sahana G, Govignon-Gion A, Boitard S, Dolezal M, et al. 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat Genet* **50**: 362–367. doi:10.1038/s41588-018-0056-5

Bush SJ, McCulloch MEB, Lisowski ZM, Muriuki C, Clark EL, Young R, Pridans C, Prendergast JGD, Summers KM, Hume DA. 2020. Species-specificity of transcriptional regulation and the response to lipopolysaccharide in mammalian macrophages. *Front Cell Dev Biol* **8**: 661. doi:10.3389/fcell.2020.00661

Chen L, Chamberlain AJ, Reich CM, Daetwyler HD, Hayes BJ. 2017. Detection and validation of structural variations in bovine whole-genome sequence data. *Genet Sel Evol* **49**: 13. doi:10.1186/s12711-017-0286-5

Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y, Hu S, Huang S, Zhang H, Zheng Z, et al. 2018. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat Commun* **9**: 2337. doi:10.1038/s41467-018-04737-0

Chuong EB, Rumi MAK, Soares MJ, Baker JC. 2013. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet* **45**: 325–329. doi:10.1038/ng.2553

Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**: 1083–1087. doi:10.1126/science.aad5497

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. doi:10.1038/nmeth.4396

Crysnanto D, Leonard AS, Fang Z-H, Pausch H. 2021. Novel functional sequences uncovered through a bovine multiassembly graph. *Proc Natl Acad Sci* **118**: e2101056118. doi:10.1073/pnas.2101056118

Dekel Y, Machluf Y, Ben-Dor S, Yifa O, Stoler A, Ben-Shlomo I, Bercovich D. 2015. Dispersal of an ancient retroposon in the TP53 promoter of Bovidae: phylogeny, novel mechanisms, and potential implications for cow milk persistency. *BMC Genomics* **16**: 53. doi:10.1186/s12864-015-1235-8

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. doi:10.1038/s41587-020-0600-6

Dunn-Fletcher CE, Muglia LM, Pavlicev M, Wolf G, Sun M-A, Hu Y-C, Huffman E, Tumukuntala S, Thiele K, Mukherjee A, et al. 2018. Anthropoid primate–specific retroviral element THE1B controls expression of *CRH* in placenta and alters gestation length. *PLoS Biol* **16**: e2006337. doi:10.1371/journal.pbio.2006337

Dupressoir A, Lavialle C, Heidmann T. 2012. From ancestral infectious retroviruses to bona fide cellular genes: role of the captured *syncytins* in placentation. *Placenta* **33**: 663–671. doi:10.1016/j.placenta.2012.05.005

Fay PC, Cook CG, Wijesiriwardana N, Tore G, Comtet L, Carpentier A, Shih B, Freimanis G, Haga IR, Beard PM. 2020. Madin-Darby bovine kidney (MDBK) cells are a suitable cell line for the propagation and study of the bovine poxvirus lumpy skin disease virus. *J Virol Methods* **285**: 113943. doi:10.1016/j.jviromet.2020.113943

Fuentes DR, Swigut T, Wysocka J. 2018. Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. *eLife* **7**: e35989. doi:10.7554/eLife.35989

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. doi:10.1038/s41588-019-0538-0

Garcia-Etxebarria K, Jugo BM. 2013. Evolutionary history of bovine endogenous retroviruses in the Bovidae family. *BMC Evol Biol* **13**: 256. doi:10.1186/1471-2148-13-256

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. doi:10.1101/gr.218032.116

Girardot M, Guibert S, Laforet M-P, Gallard Y, Larroque H, Oulmouden A. 2006. The insertion of a full-length *Bos taurus* LINE element is responsible for a transcriptional deregulation of the Normande *Agouti* gene. *Pigment Cell Res* **19**: 346–355. doi:10.1111/j.1600-0749.2006.00312.x

Grant CE, Bailey TL. 2021. XSTREME: comprehensive motif analysis of biological sequence datasets. bioRxiv doi:10.1101/2021.09.02.458722v1

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017–1018. doi:10.1093/bioinformatics/btr064

Hall TJ, Vernimmen D, Browne JA, Mullen MP, Gordon SV, MacHugh DE, O'Doherty AM. 2020. Alveolar macrophage chromatin is modified to orchestrate host response to *Mycobacterium bovis* infection. *Front Genet* **10**: 1386. doi:10.3389/fgene.2019.01386

Halstead MM, Ma X, Zhou C, Schultz RM, Ross PJ. 2020. Chromatin remodeling in bovine embryos indicates species-specific regulation of genome activation. *Nat Commun* **11**: 4654. doi:10.1038/s41467-020-18508-3

Hayes BJ, Daetwyler HD. 2019. 1000 Bull Genomes Project to map simple and complex genetic traits in cattle: applications and outcomes. *Annu Rev Anim Biosci* **7**: 89–102. doi:10.1146/annurev-animal-020518-115024

Heaton MP, Smith TPL, Carnahan JK, Basnayake V, Qiu J, Simpson B, Kalbfleisch TS. 2016. Using diverse U.S. beef cattle genomes to identify missense mutations in *EPAS1*, a gene associated with pulmonary hypertension. *F1000Res* **5**: 2003. doi:10.12688/f1000research.9254.2

Hu Y, Xia H, Li M, Xu C, Ye X, Su R, Zhang M, Nash O, Sonstegard TS, Yang L, et al. 2020. Comparative analyses of copy number variations between *Bos taurus* and *Bos indicus*. *BMC Genomics* **21**: 682. doi:10.1186/s12864-020-07097-6

Ivancevic AM, Kortschak RD, Bertozzi T, Adelson DL. 2018. Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. *Genome Biol* **19**: 85. doi:10.1186/s13059-018-1456-7

Ivashkiv LB. 2018. IFNγ: signalling, epigenetics and roles in immunity, metabolism, disease and cancer immunotherapy. *Nat Rev Immunol* **18**: 545–558. doi:10.1038/s41577-018-0029-z

Keirn SC, Freeman AE, Kehrli ME. 2001. Genetic control of disease resistance and immunoresponsiveness. *Vet Clin North Am Food Anim Pract* **17**: 477–493. doi:10.1016/S0749-0720(15)30002-5

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. doi:10.1038/s41587-019-0201-4

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol* **34**: 1812–1819. doi:10.1093/molbev/msx116

Langevin M, Synkova H, Jancuskova T, Pekova S. 2018. Merle phenotypes in dogs - *SILV* SINE insertions from Mc to Mh. *PLoS One* **13**: e0198536. doi:10.1371/journal.pone.0198536

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359. doi:10.1038/nmeth.1923

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. doi:10.1038/nmeth.4556

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN].

Liang D, Zhao P, Si J, Fang L, Pairo-Castineira E, Hu X, Xu Q, Hou Y, Gong Y, Liang Z, et al. 2021. Genomic analysis revealed a convergent evolution of LINE-1 in coat color: a case study in water buffaloes (*Bubalus bubalis*). *Mol Biol Evol* **38**: 1122–1136. doi:10.1093/molbev/msaa279

Lisch D. 2013. How important are transposons for plant evolution? *Nat Rev Genet* **14**: 49–61. doi:10.1038/nrg3374

Littlejohn MD, Tiplady K, Fink TA, Lehnert K, Lopdell T, Johnson T, Couldrey C, Keehan M, Sherlock RG, Harland C, et al. 2016. Sequence-based association analysis reveals an *MGST1* eQTL with pleiotropic effects on bovine milk composition. *Sci Rep* **6**: 25376. doi:10.1038/srep25376

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. Kidder BL), pp. 81–95. Springer, New York.

Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, Mitra A, Alexander LJ, Coutinho LL, Dell'Aquila ME, et al. 2010. Analysis of copy number variations among diverse cattle breeds. *Genome Res* **20**: 693–703. doi:10.1101/gr.105403.110

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550. doi:10.1186/s13059-014-0550-8

Low WY, Tearle R, Liu R, Koren S, Rhie A, Bickhart DM, Rosen BD, Kronenberg ZN, Kingan SB, Tseng E, et al. 2020. Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle. *Nat Commun* **11**: 2071. doi:10.1038/s41467-020-15848-y

Mallard BA, Emam M, Paibomesai M, Thompson-Crispi K, Wagter-Lesperance L. 2015. Genetic selection of cattle for improved immunity and health. *Jpn J Vet Res* **63**: S37–S44.

Meers MP, Bryson TD, Henikoff JG, Henikoff S. 2019. Improved CUT&RUN chromatin profiling tools. *eLife* **8:** e46314. doi:10.7554/eLife.46314

Menzi F, Besuchet-Schmutz N, Fragnière M, Hofstetter S, Jagannathan V, Mock T, Raemy A, Studer E, Mehinagic K, Regenscheit N, et al. 2016. A transposable element insertion in *APOB* causes cholesterol deficiency in Holstein cattle. *Anim Genet* **47:** 253–257. doi:10.1111/age.12410

Nijman IJ, van Tessel P, Lenstra JA. 2002. SINE retrotransposition during the evolution of the Pecoran ruminants. *J Mol Evol* **54:** 9–16. doi:10.1007/s00239-001-0012-2

Nilsson MA, Klassert D, Bertelsen MF, Hallström BM, Janke A. 2012. Activity of ancient RTE retroposons during the evolution of cows, spiral-horned antelopes, and Nilgais (Bovinae). *Mol Biol Evol* **29:** 2885–2888. doi:10.1093/molbev/mss158

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. doi:10.1007/s00335-007-9000-1

Orozco-terWengel P, Barbato M, Nicolazzi E, Biscarini F, Milanesi M, Davies W, Williams D, Stella A, Ajmone-Marsan P, Bruford MW. 2015. Revisiting demographic processes in cattle with genome-wide population genetic analysis. *Front Genet* **6:** 191. doi:10.3389/fgene.2015.00191

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841–842. doi:10.1093/bioinformatics/btq033

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–W191. doi:10.1093/nar/gku365

Schütz E, Wehrhahn C, Wanjek M, Bortfeld R, Wemheuer WE, Beck J, Brenig B. 2016. The Holstein Friesian lethal haplotype 5 (HH5) results from a complete deletion of *TBF1M* and cholesterol deficiency (CDH) from an ERV-(LTR) insertion into the coding region of *APOB*. *PLoS One* **11:** e0154602. doi:10.1371/journal.pone.0154602

Shaw AE, Hughes J, Gu Q, Behdenna A, Singer JB, Dennis T, Orton RJ, Varela M, Gifford RJ, Wilson SJ, et al. 2017. Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. *PLoS Biol* **15:** e2004086. doi:10.1371/journal.pbio.2004086

Shin D-H, Lee H-J, Cho S, Kim HJ, Hwang JY, Lee C-K, Jeong J, Yoon D, Kim H. 2014. Deleted copy number variation of Hanwoo and Holstein using next generation sequencing at the population level. *BMC Genomics* **15:** 240. doi:10.1186/1471-2164-15-240

Smit A, Hubley R, Green P. 2019. RepeatMasker Open-4.0. http://repeatmasker.org.

Stear MJ, Bishop SC, Mallard BA, Raadsma H. 2001. The sustainability, feasibility and desirability of breeding livestock for disease resistance. *Res Vet Sci* **71:** 1–7. doi:10.1053/rvsc.2001.0496

Stothard P, Liao X, Arantes AS, De Pauw M, Coros C, Plastow GS, Sargolzaei M, Crowley JJ, Basarab JA, Schenkel F, et al. 2015. A large and diverse collection of bovine genome sequences from the Canadian Cattle Genome Project. *Gigascience* **4:** 49. doi:10.1186/s13742-015-0090-5

Sun M-A, Wolf G, Wang Y, Senft AD, Ralls S, Jin D, Dunn-Fletcher CE, Muglia LJ, Macfarlan TS. 2021. Endogenous retroviruses drive lineage-specific regulatory evolution across primate and rodent placentas. *Mol Biol Evol* **38:** 4992–5004. doi:10.1093/molbev/msab223

Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T. 2014. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res* **24:** 1963–1976. doi:10.1101/gr.168872.113

Thompson-Crispi KA, Sargolzaei M, Ventura R, Abo-Ismail M, Miglior F, Schenkel F, Mallard BA. 2014. A genome-wide association study of immune response traits in Canadian Holstein cattle. *BMC Genomics* **15:** 559. doi:10.1186/1471-2164-15-559

Trigo BB, Utsunomiya ATH, Fortunato AAAD, Milanesi M, Torrecilha RBP, Lamb H, Nguyen L, Ross EM, Hayes B, Padula RCM, et al. 2021. Variants at the *ASIP* locus contribute to coat color darkening in Nellore cattle. *Genet Sel Evol* **53:** 40. doi:10.1186/s12711-021-00633-2

Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P, Chamberlain AT, Sykes BC, Bradley DG. 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410:** 1088–1091. doi:10.1038/35074088

Tsuda K, Kawahara-Miki R, Sano S, Imai M, Noguchi T, Inayoshi Y, Kono T. 2013. Abundant sequence divergence in the native Japanese cattle Mishima-Ushi (*Bos taurus*) detected using whole-genome sequencing. *Genomics* **102:** 372–378. doi:10.1016/j.ygeno.2013.08.002

van Boxel-Dezaire AHH, Stark GR. 2007. Cell type-specific signaling in response to interferon-γ. In *Interferon: the 50th anniversary* (ed. Pitha PM), pp. 119–154. Springer Berlin Heidelberg, Berlin, Heidelberg.

Workman AM, Chitko-McKown CG, Smith TPL, Bennett GL, Kalbfleisch TS, Basnayake V, Heaton MP. 2018. A bovine CD18 signal peptide variant with increased binding activity to *Mannheimia hemolytica* leukotoxin. *F1000Res* **7:** 1985. doi:10.12688/f1000research.17187.1

Wragg D, Mwacharo JM, Alcalde JA, Wang C, Han J-L, Gongora J, Gourichon D, Tixier-Boichard M, Hanotte O. 2013. Endogenous retrovirus EAV-HP linked to blue egg phenotype in Mapuche fowl. *PLoS One* **8:** e71393. doi:10.1371/journal.pone.0071393

Xiang R, MacLeod IM, Daetwyler HD, de Jong G, O'Connor E, Schrooten C, Chamberlain AJ, Goddard ME. 2021. Genome-wide fine-mapping identifies pleiotropic and functional variants that predict many traits across global cattle populations. *Nat Commun* **12:** 860. doi:10.1038/s41467-021-21001-0

Ye M, Goudot C, Hoyler T, Lemoine B, Amigorena S, Zueva E. 2020. Specific subfamilies of transposable elements contribute to different domains of T lymphocyte enhancers. *Proc Natl Acad Sci* **117:** 7905–7916. doi:10.1073/pnas.1912000117

Young R, Bush SJ, Lefevre L, McCulloch MEB, Lisowski ZM, Muriuki C, Waddell LA, Sauter KA, Pridans C, Clark EL, et al. 2018. Species-specific transcriptional regulation of genes involved in nitric oxide production and arginine metabolism in macrophages. *Immunohorizons* **2:** 27–37. doi:10.4049/immunohorizons.1700073

Zhou Y, Liu S, Hu Y, Fang L, Gao Y, Xia H, Schroeder SG, Rosen BD, Connor EE, Li C-J, et al. 2020. Comparative whole genome DNA methylation profiling across cattle tissues reveals global and tissue-specific methylation patterns. *BMC Biol* **18:** 85. doi:10.1186/s12915-020-00793-5

Zhuo X, Feschotte C. 2015. Cross-species transmission and differential fate of an endogenous retrovirus in three mammal lineages. *PLoS Pathog* **11:** e1005279. doi:10.1371/journal.ppat.1005279