Supplemental Material

Table of Contents

Sequencing, assembly and annotation of the genomes of three wild ruminants.....1
Resolution of the ruminant phylogeny.....1
Gene family expansions and contractions in the cetartiodactyl lineage.....2
Selective sequence constraint in cetartiodactyl species.....3
Functional sequence constraint in cetartiodactyl species.....4
References.....5
Supplemental Tables.....6
Supplemental Figures.....10

Sequencing, assembly and annotation of the genomes of three wild ruminants

We obtained high-molecular weight DNA from cell cultures of giraffe, gemsbok, and Indian muntjac from the Frozen Zoo (San Diego, USA). For giraffe, various insert size libraries were used to generate 296.23 Gbp of raw data, of which 171.09 Gbp were of high-quality, representing 68.4× genome coverage. SOAPdenovo produced a final assembly of 2.55 Gbp, consistent with the k-mer estimate of 2.52 Gbp; contig and scaffold N50 were 21.78 kbp and 2.99 Mbp, respectively. Similar results were obtained for gemsbok with 179.65 Gbp high-quality data, representing 61.95× genome coverage; a final assembly size of 2.90 Gbp, similar to the estimated size of 3.21 Gbp, and with a contig and scaffold N50 of 17.22 kbp and 1.48 Mbp, respectively. For Indian muntjac, 177.45 Gbp high-quality data were generated, representing 61.19× genome coverage. The final assembly size of the Indian muntjac genome was 2.88 Gbp, similar to the k-mer estimate of 2.76 Gbp; with a contig and scaffold N50 of 30.51 kbp and 1.07 Mbp, respectively (Suppl. Table 1). To assess the assembly qualities, we aligned ~35× randomly selected high-quality short insert size reads to their respective assemblies. Between 95.3% and 98.95% reads were successfully mapped, covering 97.52-98.88% of the assemblies (Suppl. Table 1).

Resolution of the ruminant phylogeny

A gene family is a group of gene paralogs and orthologs descended from a single gene in the last common ancestor of the targeted species. In this study, we used the TreeFam methodology (Li et al. 2006) to define gene families in 16 mammalian genomes using newly defined or already

available gene annotations (cattle, sheep, goat, Père David's deer, giraffe, gemsbok, Indian muntjac, yak, Tibetan antelope, Minke whale, pig, camel, horse, mouse, rat, and human). We applied the same pipeline and parameters that were used previously (Kim et al. 2011). A total of 16,148 gene families of which 1,327 are single-copy orthologous families were obtained.

We used the single-copy (orthologous) gene families to reconstruct the phylogenetic tree of these 16 mammals. Codon 1, 2, 3 and 1+2 sequences were extracted from coding sequences (CDS) alignments and used as input for building trees, along with protein and CDS sequences. Then, we used RAxML (Stamatakis et al. 2005) to build phylogenetic trees under GTR+gamma for nucleotide sequences and JTT+gamma model for protein sequences. We assessed the branch reliability by using 1,000 bootstrap replicates. Mouse and rat genomes were used to root the trees.

We concatenated the single-copy gene families to estimate the divergence times based on the topology obtained in the phylogenetic analysis. PAML mcmctree (Yang and Rannala 2006) was used to determine split times with the approximate likelihood calculation method. PAML baseml (Yang 2007) was initially applied to compute the alpha parameter under REV substitution model and substitution rate per time unit. Then the gradient (g) vector and Hessian (H) matrix were estimated, which described the shape of the log-likelihood surface around MLE of branch lengths. Tracer (http://beast.bio.ed.ac.uk/) was applied to check convergence.

Using the TreeFam methodology, we identified 16,148 gene families across 16 mammalian genomes, including 9 ruminant species. From these gene families, 1,327 were single-copy orthologous families. We used these single-copy families and RAxML under GTR/JTT+gamma models based on CDS, peptide, codon1, 2, 3 and 1+2 sequences to estimate the tree topology of the studied species. All models resulted in the same topology with a bootstrap value of 100 in all nodes after 1,000 replicates (Suppl. Fig. 1). By concatenating the single-copy families we estimated the divergence times based on the previously obtained topology (Fig. 1). Our results indicate that cetartiodactyls diverged from their most recent ancestor with equids ~77 million years ago (Mya). Within Cetartiodactyla, tylopods (camels) split from the rest ~60 Mya, Suina (pigs) diverged from whales and ruminants ~54 Mya, while ruminants separated from cetaceans ~47 Mya, coinciding with the middle-Eocene period (38-48 Mya). Within ruminants, our analysis suggests that giraffes and cervids shared a common ancestor ~21 Mya, while bovids form a monophyletic clade, that diverged from the rest ~22 Mya.

Gene family expansions and contractions in the cetartiodactyl lineage

Gene family expansion analysis was performed using CAFE (Hahn et al. 2005). In CAFE, a random birth and death model was proposed to study gene gain and loss in gene families across the

previously defined phylogenetic tree. The global parameter λ, that described both gene birth (λ) and death (μ= -λ) rate across all branches on the tree for all gene families, was estimated using maximum likelihood. A conditional p-value was calculated for each gene family, and families with conditional p-values < 0.05 were considered to have a significantly accelerated rate of expansion or contraction.

A total of 43 gene families were expanded in the ruminant ancestor. From these, 37 continued to further expand in the ruminant clade, including PAGs, MOGATs, and CATHLs (Fig. 1). These 37 families were enriched for the GO functional categories of *defense response, antigen processing and presentation*, and *aspartic-type endopeptidase activity*, among others (GO enrichment test, FDR < 0.05; Suppl. Table 6). In the bovid ancestor six gene families were contracted, while 40 gene families were further expanded, of which 22 continued to expand in the descendant species, including CD1s and SPLUNC2, related to *immune response, antigen processing and presentation via MHC class Ib,* and *lipid binding.* Sixty gene families were expanded in the cattle lineage after the split of bovids, including genes related to *lipid metabolic process* and *digestion* (GO enrichment test, FDR < 0.05; Suppl. Table 6).

Although EBRs were found in gene-dense regions (1.2× fold enrichment, FDR = 0.02), gene family member expansions and contractions were not significantly enriched in any types of EBRs. Four (out of 43) ruminant expansions were located +/- 50 kbp of ruminant-specific EBRs (Suppl. Fig. 6). These families included the pregnancy-associated glycoproteins (PAGs) in BTA29, olfactory receptors (OR2M@ in BTA7 and OR4C@ in BTA15 both near pecoran-specific EBRs), and scavenger receptors (such as SSC5D) close to a ruminant-specific EBR in BTA18. Only five of the bovidae-specific expansions were found close to bovidae EBRs, including Kelch-like family (KLHL@), keratin associated proteins (KRTAP@), zinc fingers, SET domain without mariner transposase function family (SETMAR), and olfactory receptors. In the cattle lineage 6/60 expanded gene families were located near cattle EBRs, including UBE2@, MAGE@, TSPYL@, CT47, TRGC, and RPL9. Interestingly, SETMAR family was expanded both in the bovidae ancestor and in the cattle lineage and in both lineages members of this gene family were located near EBRs. We did not observe any association between gene family contractions and EBRs in the ancestral genomes with only 2/235 contracted families in the cattle lineage being found close to cattle-specific EBRs (OR4@ and RPLP@).

Selective sequence constraint in cetartiodactyl species

Conserved non-coding elements (CNEs) were defined in three lineages: i) mammalian CNEs, ii) cetartiodactyl-specific CNEs, and iii) ruminant-specific CNEs (Suppl. Table 7). We associated the CNEs to their neighboring genes following the proximal distance rule implemented in GREAT (McLean et al. 2010) software stating that 'gene regulatory domains extend two directions from the proximal

promoter of the nearest gene (-5 kbp/ +1 kbp from the transcription starting site), but no more than 1Mbp'. Then, we performed functional enrichment analysis using as a background list those genes with at least one CNE in their regulatory domain. Earlier studies on CNE gain in the primate and rodent lineage found that CNEs that were recruited near genes, which were not previously associated with CNEs, were enriched in nervous system development; while CNEs that were added near genes that were already flanked by CNEs, were related to transcriptional regulation (Takahashi and Saitou 2012).The GO analysis indicated that ruminant-specific CNEs formed close to genes that already had older CNEs (mammalian and/or cetartiodactyl) in their close proximity were associated to the functional terms *metabolic process* and *antigen processing and presentation*; while ruminant-specific CNEs recruited near genes without older CNE were related to *sensory perception* (FDR < 0.05). Whereas CNEs overlapping ruminant TEs were found in close proximity of genes related to *animal organ development* and *cell differentiation* (GO hypergeometric test, FDR < 0.05).

Previous studies have shown that some CNEs originated from retrotransposons that have been exapted and are under selective constraint. We found that Eulor, MERs, and UCONs TE families have the highest levels of sequence constraint of all the TEs families in all species analyzed (18.97%–56.41% of these TEs are under evolutionary constraint as defined by overlap with mammalian CNEs); while ruminant-specific CNEs were found enriched in retrotransposons (ERVs, LINEs, and SINEs), particularly in ruminant-specific TEs (LTR31B_BT, SINE2-1_BT, and L1-2_BT, fold enrichment of 3.94, 1.6 and 1.2, respectively, FDR < 0.05, Suppl. Table 9).

Functional sequence constraint in cetartiodactyl species

After defining ancestral mammalian, cetartiodactyl, and ruminant putative enhancers, we analyzed their TE content. We found that ancestral mammalian enhancers were enriched in ancient MIR elements, with ~50% of them containing MIRb, MIR3, and MIRc (permutation test, FDR = 0.0085). In contrast, only ~20% of all cattle enhancers contained MIRs, but >40% contained ruminant-specific TEs (Bov-tA2 and SINE2-2_BT, FDR = 0.0013) and other non-LTRs (L2s, FDR = 0.0014). When we compared the TE content in enhancers near ruminant-lineage EBRs to enhancers far from these EBRs, we found that L1-2_BT was significantly enriched in enhancers near these EBRs (2.39× fold-enrichment, FDR = 0.03).

Association of genomic and epigenomic features with EBRs

To determine the cutoff distance between a given genomic or epigenomic feature from EBRs, we performed a permutation test using the Genomic Association Tester (GAT) (Heger et al. 2013). Four distances were analyzed, including 50, 100, and 200 kbp up to 1 Mbp from the EBR boundaries,

as well features within the EBRs (Suppl. Fig. 3). The 50 kbp extension was chosen because it showed a shift in the enrichment for most of the features.

Distribution of gene expression values associated to EBRs

Taking into account that the number of EBRs is relatively low, and to assess the robustness of our conclusions, we plotted the distribution of expression for genes near EBRs and for a randomly selected set of 120 genes after re-sampling 2,000 times in other parts of the genome. We then looked specifically at the expression levels of these genes in human, pig, and cattle and compared them to the average gene expression of genes near EBRs (Suppl. Fig. 8).

References

Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N. 2005. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res* **15**: 1153-1160.

Heger A, Webber C, Goodson M, Ponting CP, Lunter G. 2013. GAT: a simulation framework for testing the association of genomic intervals. *Bioinformatics* **29**: 2046-2048.

Kim EB, Fang X, Fushan AA, Huang Z, Lobanov AV, Han L, Marino SM, Sun X, Turanov AA, Yang P et al. 2011. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* **479**: 223-227.

Li H, Coghlan A, Ruan J, Coin LJ, Hériché J-K, Osmotherly L, Li R, Liu T, Zhang Z, Bolund L et al. 2006. TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res* **34**: D572-580.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**: 495-501.

Stamatakis A, Ludwig T, Meier H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**: 456-463.

Takahashi M, Saitou N. 2012. Identification and characterization of lineage-specific highly conserved noncoding sequences in Mammalian genomes. *Genome Biol Evol* **4**: 641-657.

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

Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol* **23**: 212-226.

Supplemental Tables

Supplemental Table 1. Summary statistics of the newly sequenced ruminant genomes.

Statistic	Giraffe	Gemsbok	Indian muntjac
Genome size (Gbp)	2.55	2.90	2.88
Contig N50 (kbp)	21.78	17.22	30.51
Scaffold N50 (Mbp)	2.99	1.48	1.07
No. annotated genes	21,621	23,125	23,643
Genes with functional annotation (%)	87.77	86.52	84.96
Genome coverage of TEs (%)	39.80	42.57	41.71
BUSCO complete mammalian genes*	3,895 (94.9%)	3,807 (92.8%)	3,821 (93.1%)
BUSCO partial mammalian genes*	114 (2.8%)	129 (3.1%)	153 (3.7%)

* Total number of mammalian BUSCOs is 4,104 genes.

Supplemental Table 2. Summary statistics of the length of pair-wise homologous synteny blocks (HSBs) and mammalian msHSBs.

Species	Type	Mean (bp)	Max. (bp)	Coverage cattle genome (%)
Yak	chromosomes	22,358,796	124,328,324	98.80
Goat	chromosomes	16,776,129	121,234,144	98.00
Giraffe	scaffolds	2,123,567	20,556,850	95.43
Indian muntjac	scaffolds	1,341,571	33,839,161	80.33
Sheep	chromosomes	17,445,406	121,212,901	98.65
Tibetan antelope	chromosomes	21,422,960	134,206,916	98.67
Minke whale	scaffolds	3,873,017	35,655,134	96.45
Camel	scaffolds	1,708,782	13,668,333	88.05
Pig	chromosomes	6,165,742	81,203,868	95.82
Alpaca	scaffolds	3,766,751	41,450,703	96.76
Dog	chromosomes	7,234,192	53,969,409	97.52
Horse	chromosomes	7,712,024	77,637,799	97.61
Human	chromosomes	6,414,013	90,014,617	97.28
Rhesus macaque	chromosomes	6,757,564	90,096,702	97.68
Mouse	chromosomes	5,764,780	65,767,983	94.55
Chimp	chromosomes	6,619,217	78,765,857	96.67
Rat	chromosomes	5,758,665	65,742,198	94.24
Mammalian msHSBs	NA	1,484,900	15,926,968	76.28

Supplemental Table 3. Reorganized RACFs of ancestors and their mappings against the cattle genome (Separate excel file).

Supplemental Table 4. Classification of ruminant lineage EBRs using the placement of cattle BACs in chevrotain and giraffe metaphase spreads (Separate excel file).

Supplemental Table 5. Rearrangement rates of the nodes leading to extant ruminant species.

Node	No. EBRs	Rate (EBRs/My)
Cattle	33	11.07*
Bovidae	3	1.16
Bovinae	31	2.09
Caprinae	7	3.75*
Caprinae + Antilopidae	2	0.87*
Cervidae	3	4.09*
Cetacea + Ruminantia	11	2.28
Cetartiodactyla	13	1.27
Giraffidae + Cervidae	4	0.79*
Pecora	25	1.19
Ruminantia	33	6.60*
Ruminantia or Pecora	20	NA
Suina + Cetacea + Ruminantia	10	2.14
Tylopoda	17	0.59*
Mean	NA	2.24

*Rate statistically different from mean, p-value < 0.05 after FDR correction.

Supplemental Table 6. Gene family expansions and contractions in the lineages from the cetartiodactyl ancestor. The results for each branch are presented in separate excel sheets, including a representative gene symbol, the gene ontology (GO) annotation, and the number of gene members in the 16 mammalian species analysed (Separate excel file).

Supplemental Table 7. Summary statistics of the conserved elements longer than 50 bp detected in the cattle genome.

Classification	No. elements	Percentage of defined CEs	Max length (bp)	Coverage of cattle genome (%)
CEs defined using a multiple alignment of 15 mammalian species	1,050,968	--	6,173	6.30
CEs defined using a multiple alignment of 9 ruminant species	1,590,503	100.00	8,909	11.16
CNEs defined using a multiple alignment of 9 ruminant species	850,432	53.47	3,770	5.70
CNEs overlapping TEs	137,284	8.63	1,257	0.61
Mammalian CNEs	545,561	34.30	2,469	2.34
Cetartiodactyl CNEs	179,415	11.28	1,849	2.79
Ruminant-specific CNEs	122,619	7.71	713	0.57
Ruminant-specific CNEs overlapping TEs	25,926	1.63	682	0.10

CEs: conserved elements; CNEs: conserved non-coding elements; TEs: transposable elements.

Supplemental Table 8. Enrichments results of the association of transposable elements with three types of evolutionary breakpoint regions (EBRs) (Separate excel file).*Origin of TE obtained from RepBase and Adelson et al 2009.

Supplemental Table 9. Association of conserved non-coding elements (CNEs) with transposable elements (TEs) (Separate excel file).

Supplemental Table 10. Summary statistics of the enhancers defined as peaks of H3K27Ac in the cattle genome (raw data obtained from Villar et al. 2015)**.**

Classification	No.	Percentage of all enhancers
Detected in all mammals	232	0.74
Detected in pig, whale, and cattle	481	1.53
Detected only in cattle	15,387	49.04
Not assigned	15,272	48.69

Supplemental Table 11. Association of types of conserved non-coding elements (CNEs) and functional enhancers with mammalian msHSBs.

Feature	Fold enrichment	Log2 Fold	P-value	FDR
Bovid CNEs	1.0578	0.0811	0.0001	0.0001
Ruminant CNEs	1.0748	0.1041	0.0001	0.0001
Cetartiodactyl CNEs	1.1182	0.1612	0.0001	0.0001
Mammalian CNEs	1.1312	0.1779	0.0001	0.0001
Cattle enhancers	1.0031	0.0045	0.3000	0.3000
Cetartiodactyl enhancers	1.0027	0.0039	0.4612	0.4612
Mammalian enhancers	1.1619	0.2165	0.0001	0.0001

Supplemental Table 12. Branch of origin of the transcription factor binding sites (TFBSs) in the three types of enhancers.

Branch of origin	Cattle		Cetartiodactyl		Mammalian		Total	
	No.	Freq.	No.	Freq.	No.	Freq.	No.	Freq.
Cattle	284,175	0.080	9,096	0.060	5,804	0.052	299,075	0.078
Bovinae	708,782	0.199	28,701	0.188	18,135	0.163	755,618	0.197
Bovidae	294,634	0.083	11,831	0.078	7,304	0.066	313,769	0.082
Ruminantia	1,191,962	0.334	49,575	0.325	31,673	0.285	1,273,211	0.332
Cetace + Ruminantia	348,050	0.098	14,165	0.093	9,120	0.082	371,335	0.097
Suina + Cetacea + Ruminantia	92,608	0.026	5,236	0.034	3,427	0.031	101,271	0.026
Cetartiodactyla	63,346	0.018	3,136	0.021	2,250	0.020	68,732	0.018
Mammalia	585,243	0.164	30,657	0.201	33,475	0.301	649,375	0.169
Total	3,568,800	1.000	152,397	1.000	111,188	1.000	3,832,387	1.000

Supplemental Table 13. Motifs of the transcription factor binding sites (TFBSs) found in enhancers near three different types of EBRs (Separate excel file).

Supplemental Table 14. Branch of origin of the transcription factor binding sites (TFBSs) found in enhancers near three different types of EBRs.

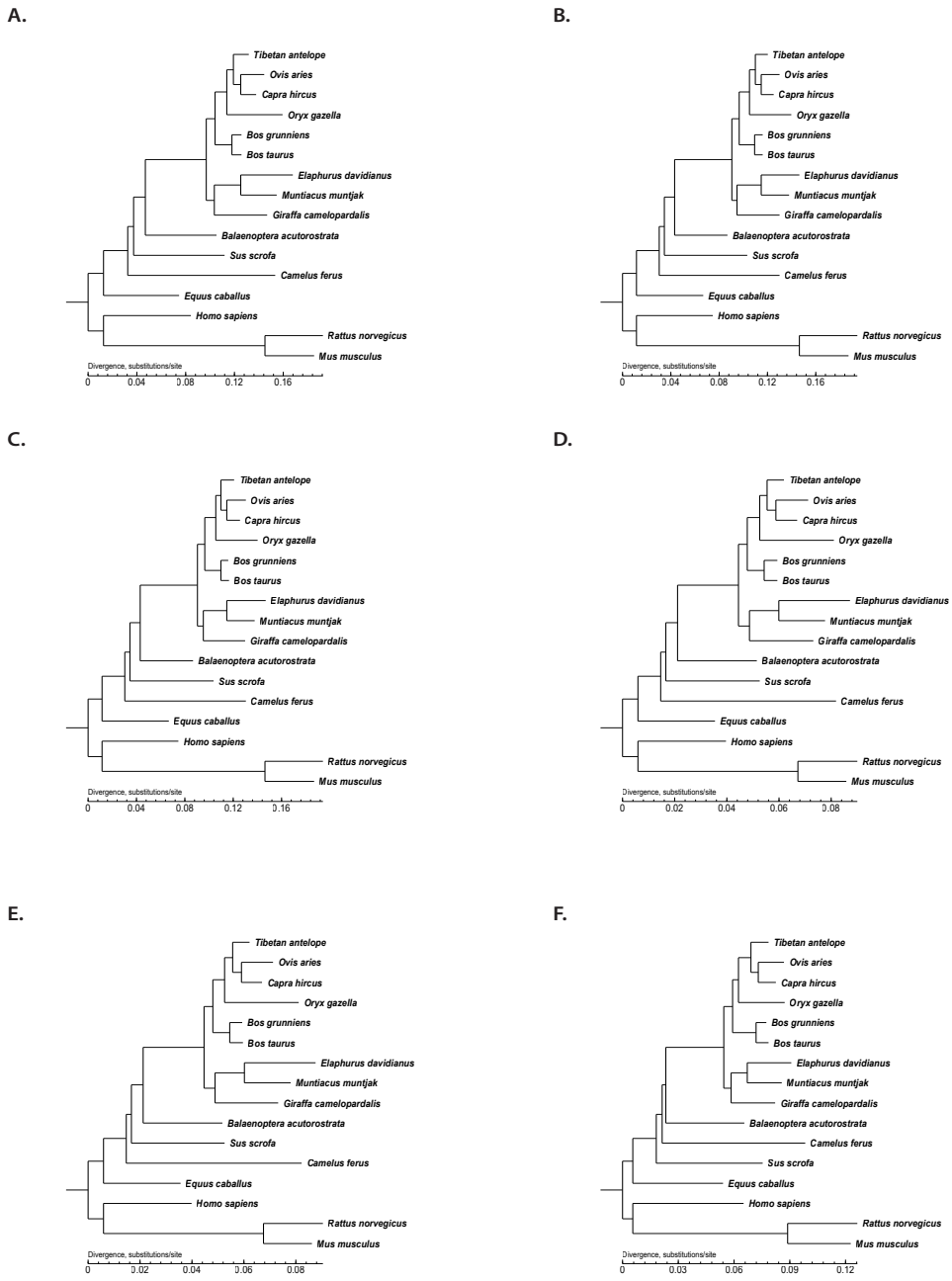
Branch of origin	Bovid-cattle EBRs		Ruminant EBRs		Cetartiodactyl EBR		Not close to EBRs		Total	
	No.	Freq.	No.	Freq.	No.	Freq.	No.	Freq.	No.	Freq.
Cattle	1,345	0.128	1,545	0.092	471	0.071	295,714	0.078	299,075	0.078
Bovinae	2,206	0.210	3,465	0.206	1,257	0.191	748,690	0.197	755,618	0.197
Bovidae	943	0.090	1,587	0.094	611	0.093	310,628	0.082	313,769	0.082
Ruminantia	3,184	0.303	6,017	0.358	2,611	0.396	1,261,398	0.332	1,273,210	0.332
Cetacea + Ruminantia	966	0.092	1,405	0.084	674	0.102	368,290	0.097	371,335	0.097
Suina + Cetacea + Ruminantia	283	0.027	347	0.021	118	0.018	100,523	0.026	101,271	0.026
Cetartiodactyla	175	0.017	268	0.016	74	0.011	68,215	0.018	68,732	0.018
Mammalia	1,408	0.134	2,178	0.130	782	0.119	645,007	0.170	649,375	0.169
Total	10,510	1.000	16,812	1.000	6,598	1.000	3,798,465	1.000	3,832,385	1.000

Supplemental Table 15. Association of transcription factor binding sites (TFBSs) present in enhancers near EBRs with transposable elements (TEs).

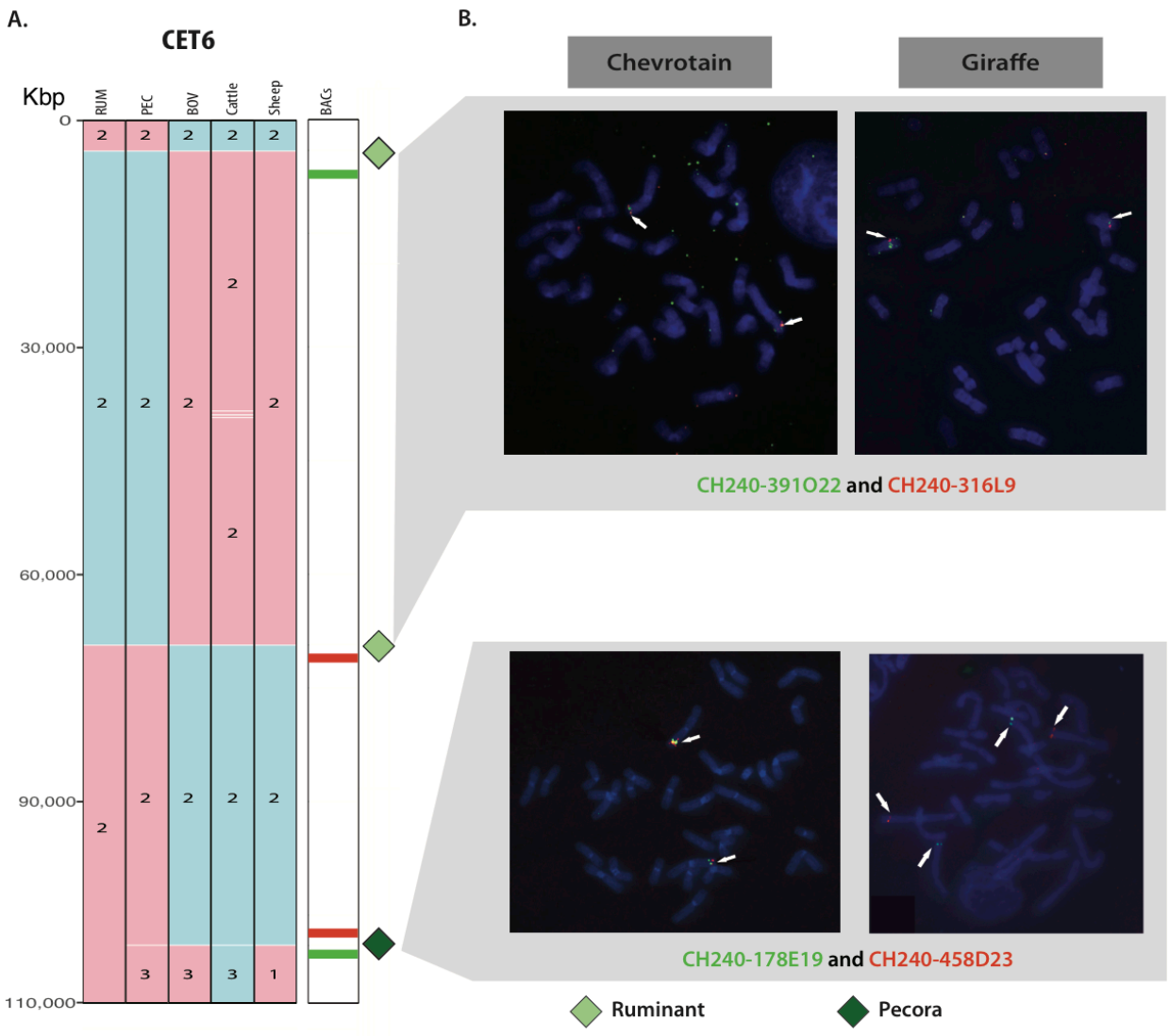
TE	Fold	Log2Fold	P-value	FDR	Percentage of TEs with TFBSs	Origin of TE
HAL1-3A	3.66	1.87	0.001	0.0027	43.99	Mammal
MER54B	2.41	1.27	0.001	0.0027	26.36	Mammal
MER5C1	2.01	1.01	0.001	0.0027	17.54	Unknown
MamGypLTR1c	1.89	0.92	0.001	0.0027	20.42	Mammal
LTR16B	1.84	0.88	0.001	0.0027	19.22	Unknown
LTR16A2	1.83	0.87	0.001	0.0027	20.18	Unknown
L1MA5	1.78	0.83	0.001	0.0027	18.02	Unknown
LTR11B_BT	1.77	0.82	0.001	0.0027	17.99	Ruminant
LTR39C2_BT	1.73	0.79	0.001	0.0027	17.93	Ruminant
MamRep1879	1.66	0.73	0.001	0.0027	15.13	Mammal
MER20	1.62	0.70	0.001	0.0027	16.48	Unknown
SINE2-2_BT	1.62	0.70	0.001	0.0027	17.33	Ruminant
BOV-A2	1.51	0.59	0.001	0.0027	15.40	Ruminant
LTR75_BT	1.50	0.59	0.001	0.0027	15.64	Ruminant
LTR41	1.44	0.53	0.001	0.0027	13.85	Unknown
ERV1-1-LTR_BT	1.38	0.47	0.001	0.0027	13.70	Ruminant
LTR33	1.35	0.43	0.001	0.0027	13.75	Unknown
LTR50	1.34	0.43	0.001	0.0027	12.64	Unknown
SINE2-3_BT	1.32	0.40	0.001	0.0027	13.79	Ruminant
LTR33A	1.30	0.38	0.001	0.0027	13.20	Unknown
L1_BT	1.30	0.38	0.001	0.0027	12.45	Ruminant
L1_Art	1.28	0.36	0.001	0.0027	13.25	Ruminant
LTR78B	1.27	0.35	0.001	0.0027	13.61	Unknown
CHRL1_BT	1.27	0.35	0.001	0.0027	13.45	Ruminant
L1M2	1.20	0.27	0.001	0.0027	11.55	Unknown
CHRL	1.12	0.16	0.001	0.0027	11.60	Ruminant

Supplemental Figures

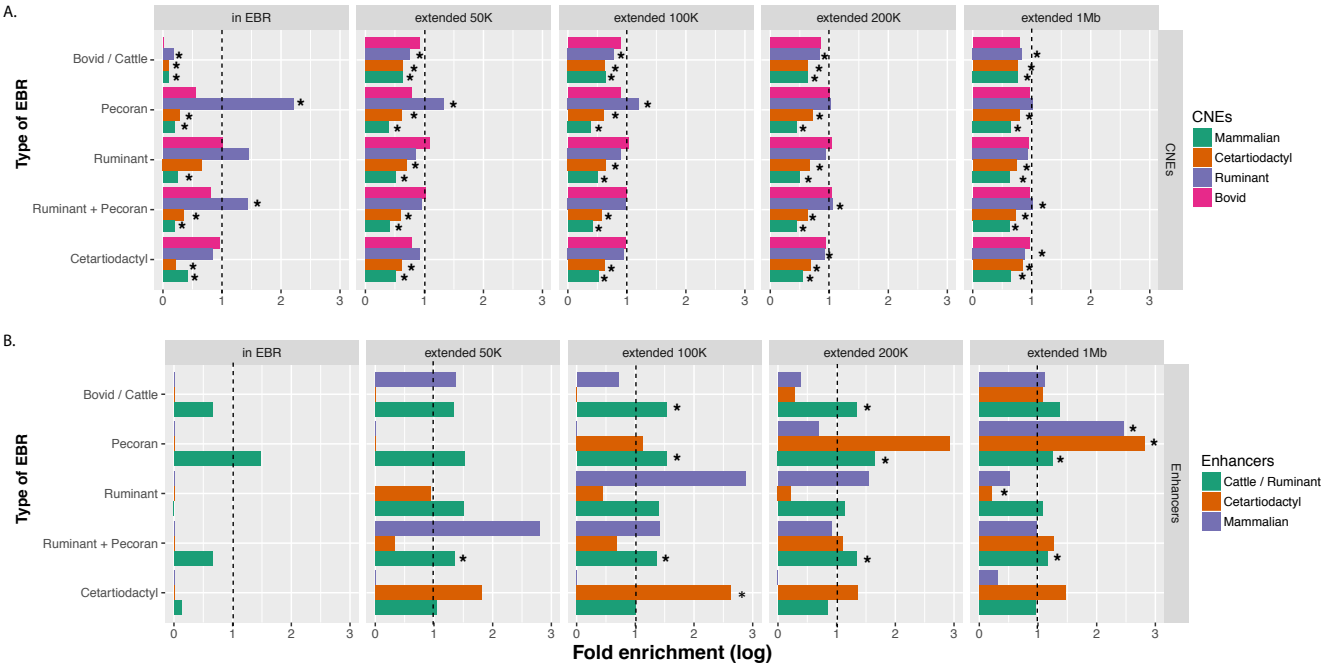
Supplemental Figure 1. Reconstructed phylogenetic trees. A) Maximum likelihood (ML) under the model JTT+gamma using protein sequences, B) ML under GTR+gamma using codon sequences, C) ML under the HKY85+gamma using codon sequences, D) ML under HKY85+gamma using phase 1 codon sequences, E) ML under GTR+gamma using phase 1 codon sequences, F) Neighbour joining under JTT+gamma using protein sequences.



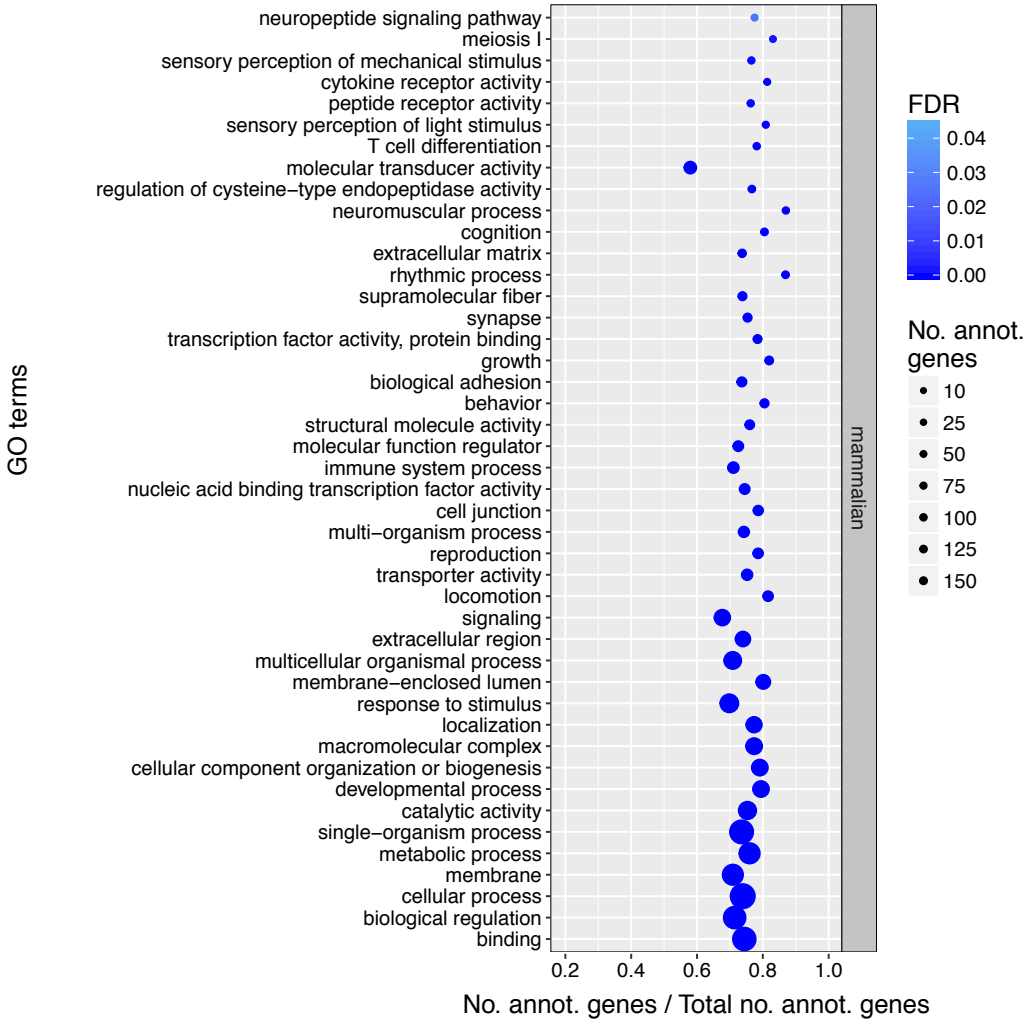
Supplemental Figure 2. Cetartiodactyl Ancestor chromosome 6 (CET6). A. The ancestral ruminant (RUM), pecoran (PEC), and bovid (BOV) chromosomes together with cattle and sheep chromosomes are shown as tracks in the CET6. Blue and red blocks define syntenic fragments in "+" or "-" orientation, respectively compared to the CET6, with chromosome numbers inside the blocks. BAC track shows the position of four BACs in CET6. The colors correspond to the same labelling in panel B. Light green diamonds demarcate ruminant-specific EBR, while dark green diamond shows a pecoran-specific EBR. **B.** Dual color FISH results using two cattle BACs flanking each EBR on chevrotain and giraffe metaphase spreads, with white arrows pointing to the signal of hybridization. The top panel shows mapping of a ruminant-specific EBR, since the hybridization pattern is the same in chevrotain and giraffe metaphases; while the bottom panel shows a pecoran-specific EBR, because the hybridization pattern is different in giraffe compared to chevrotain and ruminant ancestor.



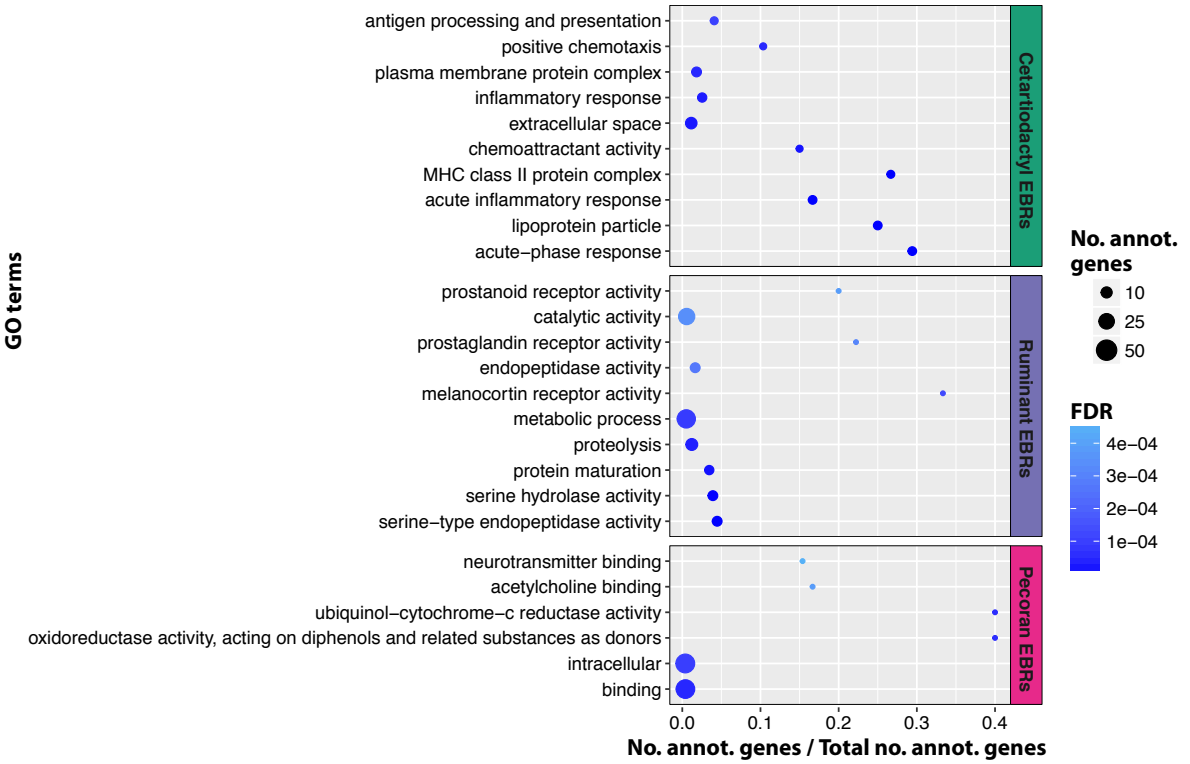
Supplemental Figure 3. Association of types of EBRs with conserved non-coding elements (CNEs) and functional enhancers. **A.** Fold enrichment of the CNEs inside EBRs, 50 kbp, 100 kbp, 200 kbp, and 1 Mbp surrounding the different types of EBRs. **B.** Fold enrichment of the functional enhancers. Asterisks show the statistically significant enrichments (FDR < 0.05). Dotted lines demarcate a fold enrichment of 1.



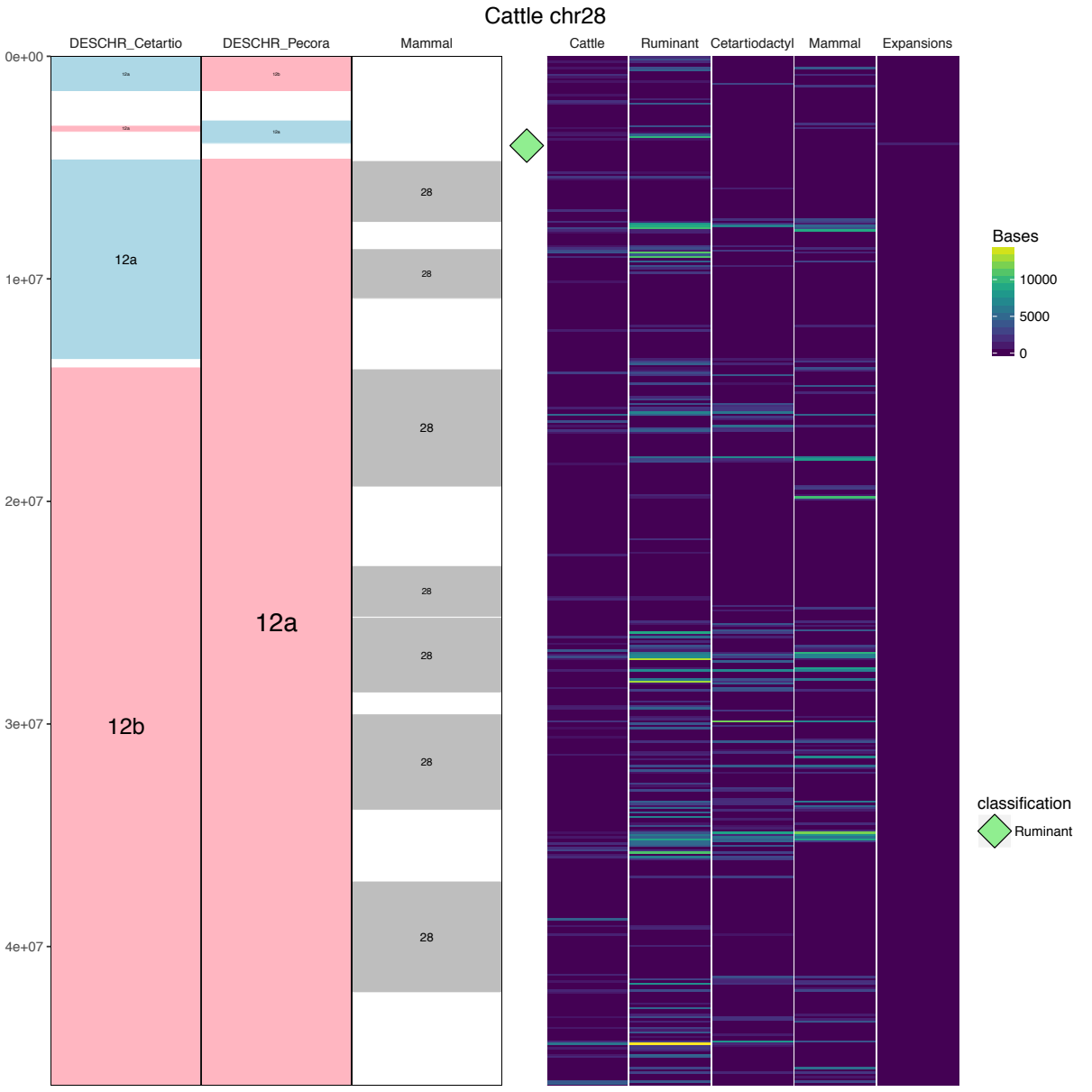
Supplemental Figure 4. Gene Ontology enrichment analysis of genes in msHSBs. Bubble size depicts the number of genes annotated in each GO term. Bubble shade represents the *p*-value with darker shades for lower *p* values. The *x*-axis shows the ratio of genes annotated for each GO term in the analysed list versus the background list (*p* value < 0.05; FDR < 5%).



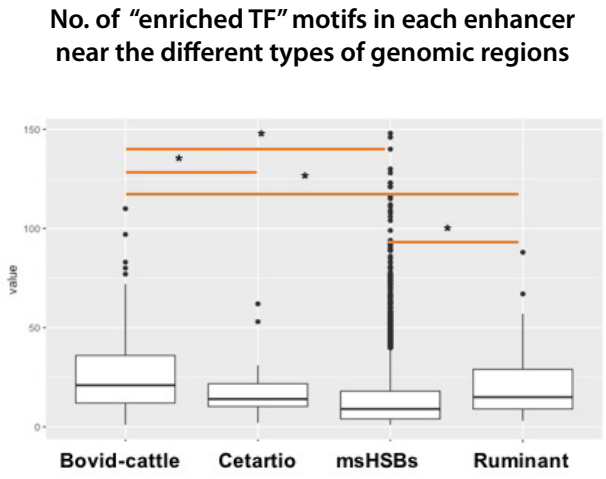
Supplemental Figure 5. Gene Ontology enrichment analysis of genes near EBRs. Bubble size depicts the number of genes annotated in each GO term. Bubble shade represents the *p*-value with darker shades for lower *p* values. The *x*-axis shows the ratio of genes annotated for each GO term in the analysed list versus the background list (*p* value < 0.05; FDR < 5%).



Supplemental Figure 6. Comparative organization of the reconstructed cetartiodactyl and pecoran ancestral chromosomes with the cattle genome as a reference. Order and orientation of syntenic fragments are visualized using the Evolution Highway comparative chromosome browser. Blue and pink colors represent orientation of blocks relative to the reference, with blue indicating the same orientation, and pink indicating the opposite orientation. Pink does not always indicate an inversion because the orientation of RACFs is randomly chosen during the reconstruction. The number within each block represents a chromosome number for a reconstructed ancestor, and a lower-case letter indicates a fragment of the chromosome. Light grey blocks indicate mammalian multispecies homologous synteny blocks (msHSBs). Diamonds indicate the position of EBRs, with light and dark green showing EBRs classified as ruminant or pecoran, respectively, using FISH data. Heatmaps show the density of ruminant, cetartiodactyl, and mammalian enhancers, and ruminant gene family expansions in windows of 100 kbp along cattle chromosomes (example of chromosome 28 in cattle, and the rest in a separate file).



Supplemental Figure 7. Number of 25 TF motifs in enhancers near each type of EBRs or msHSBs.
* p-value < 0.05.



Supplemental Figure 8. Gene expression levels in human, cattle, and pig liver for genes near EBRs.
Gene expression levels for genes near EBRs and randomly selected genes after re-sampling 120 genes 2,000 times are plotted for human, pig, cattle and the overall expression of all species included in the analysis. **Wilcoxon test p-value 0.04.