

RESEARCH ARTICLE

Open Access



# Comparative epigenomics reveals the impact of ruminant-specific regulatory elements on complex traits

Siqian Chen<sup>1</sup>, Shuli Liu<sup>1,2</sup>, Shaolei Shi<sup>1</sup>, Yifan Jiang<sup>1</sup>, Mingyue Cao<sup>1</sup>, Yongjie Tang<sup>1</sup>, Wenlong Li<sup>1</sup>, Jianfeng Liu<sup>1</sup>, Lingzhao Fang<sup>3,4\*</sup>, Ying Yu<sup>1\*</sup>  and Shengli Zhang<sup>1^</sup>

## Abstract

**Background:** Insights into the genetic basis of complex traits and disease in both human and livestock species have been achieved over the past decade through detection of genetic variants in genome-wide association studies (GWAS). A majority of such variants were found located in noncoding genomic regions, and though the involvement of numerous regulatory elements (REs) has been predicted across multiple tissues in domesticated animals, their evolutionary conservation and effects on complex traits have not been fully elucidated, particularly in ruminants. Here, we systematically analyzed 137 epigenomic and transcriptomic datasets of six mammals, including cattle, sheep, goats, pigs, mice, and humans, and then integrated them with large-scale GWAS of complex traits.

**Results:** Using 40 ChIP-seq datasets of H3K4me3 and H3K27ac, we detected 68,479, 58,562, 63,273, 97,244, 111,881, and 87,049 REs in the liver of cattle, sheep, goats, pigs, humans and mice, respectively. We then systematically characterized the dynamic functional landscapes of these REs by integrating multi-omics datasets, including gene expression, chromatin accessibility, and DNA methylation. We identified a core set ( $n = 6359$ ) of ruminant-specific REs that are involved in liver development, metabolism, and immune processes. Genes with more complex cis-REs exhibited higher gene expression levels and stronger conservation across species. Furthermore, we integrated expression quantitative trait loci (eQTLs) and GWAS from 44 and 52 complex traits/diseases in cattle and humans, respectively. These results demonstrated that REs with different degrees of evolutionary conservation across species exhibited distinct enrichments for GWAS signals of complex traits.

**Conclusions:** We systematically annotated genome-wide functional REs in liver across six mammals and demonstrated the evolution of REs and their associations with transcriptional output and conservation. Detecting lineage-specific REs allows us to decipher the evolutionary and genetic basis of complex phenotypes in livestock and humans, which may benefit the discovery of potential biomedical models for functional variants and genes of specific human diseases.

<sup>^</sup>Shengli Zhang is deceased.

\*Correspondence: lingzhao.fang@qgg.au.dk; yuying@cau.edu.cn

<sup>1</sup> Key Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture and Rural Affairs & National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China

<sup>4</sup> Center for Quantitative Genetics and Genomics (QGG), Aarhus University, Aarhus, Denmark

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



**Keywords:** Regulatory elements, Ruminant evolution, Liver, Epigenetic regulation, GWAS enrichment

## Background

Over the past decade, genome-wide association studies (GWAS) have successfully discovered hundreds of thousands of genetic variants associated with complex traits and diseases in both human and livestock species [1–3]. As the majority of these variants are located in noncoding regions [4], it is challenging to understand how they impact complex phenotypes. Previous studies have illustrated that trait-associated variants are significantly enriched in regulatory regions, such as promoters and enhancers, in well-studied species (e.g., humans and mice) [5, 6]. Currently, global efforts such as the Functional Annotation of Animal Genomes (FAANG) initiative and the Farm animal Genotype-Tissue Expression (FarmGTEx) project are working to uncover basic knowledge of genomic function and regulation in livestock species [7–9]. However, a comprehensive atlas of regulatory elements (REs) is still lacking for most livestock species, which limits our understanding of the functional biology of species evolution and restricts the genetic improvement of complex traits in livestock. As abundant terrestrial herbivores [10], ruminants, such as cattle, sheep, and goats, have a unique history of species differentiation and play an important economic role in modern animal husbandry. Therefore, a comprehensive comparison of REs between major ruminants and other species will provide novel insights into functional genome evolution specific to ruminants. Moreover, it will allow us to explore the genetic basis underlying complex traits of economic value in these farm animal species.

The emergence of cross-species comparative epigenomics has provided a new method for both elucidating genomic evolution and identifying potential functional noncoding variants associated with complex traits and diseases [11]. By comparing the chromatin landscape of primary aortic endothelial cells isolated during the acute NF- $\kappa$ B response among humans, mice, and cattle, Alizada et al. found that inflammatory- and cardiovascular-associated genetic variants discovered by GWAS were significantly enriched in the species-conserved regulatory regions nearby NF- $\kappa$ B target genes [12]. In addition, by cross-species mapping of epigenomic marks, Liu et al. found that the genetic control of immune and reproductive traits is conserved to a certain degree between humans and cattle [13]. These findings indicate that evolutionarily conserved REs play key roles in shaping complex phenotypes across species [14, 15]. Although previous studies have investigated the evolution of the transcriptome (e.g., long noncoding RNAs) in ruminants

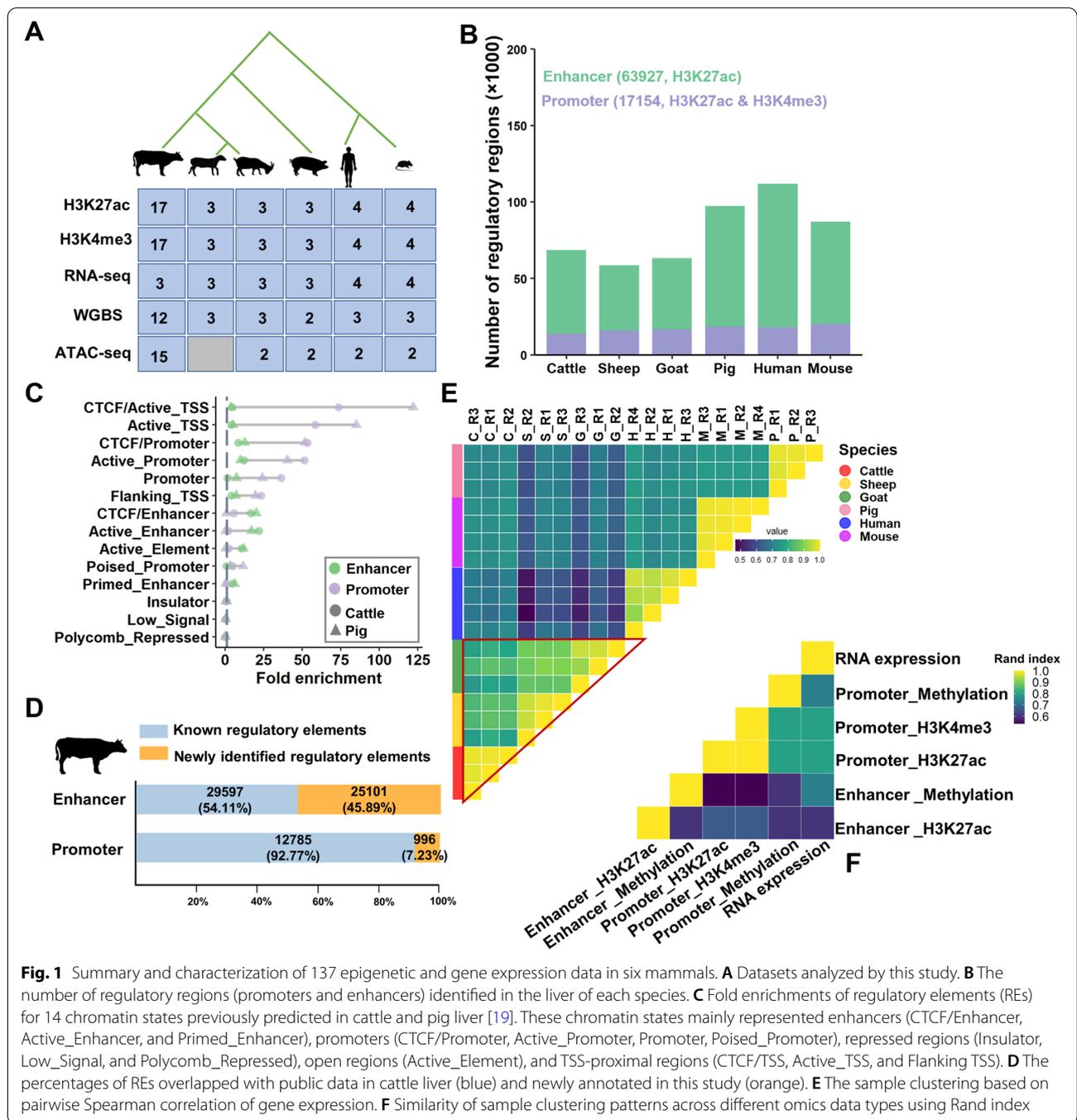
[8, 16], a comprehensive comparison of epigenetic regulation and its potential impacts on other molecular phenotypes and complex traits is still lacking.

Here, by using the liver as a representative tissue, we systematically detected and functionally characterized the epigenomic landscapes and explored the dynamics of REs across three ruminant species (i.e., cattle, sheep, and goat) and three non-ruminant species (i.e., pig, mouse, and human). We annotated an average of 81,081 REs (17,154 and 63,927 promoters and enhancers, respectively) across six species by integrating 137 multi-omics datasets, including epigenetic data such as histone modifications, gene expression, chromatin accessibility, and DNA methylation (Additional file 1: Fig. S1). By detecting lineage-specific REs and associating them with expression quantitative traits loci (eQTLs) and large-scale GWAS datasets from 44 and 52 complex traits in cattle and humans, respectively, we further explored how comparative epigenomics across species could help us understand the evolutionary and genetic mechanism of complex phenotypes. Overall, our study provides a valuable resource for REs in ruminants and highlights the key roles of conserved functional elements in complex traits in both human and livestock species.

## Results

### Overview of multi-omics datasets

To study the epigenomic changes during ruminant evolution, we performed chromatin immunoprecipitation sequencing (ChIP-seq) for H3K27ac and H3K4me3 in the liver of cattle, sheep, and goats (Fig. 1A). In total, we generated 17 ChIP-seq datasets, and each species had three biological replicates. We also generated nine RNA sequencing (RNA-seq) datasets and nine whole genome bisulfite sequencing (WGBS) datasets in the same liver samples. We further retrieved a total of 41 public datasets including ChIP-seq datasets for H3K27ac and H3K4me3, WGBS datasets, and RNA-seq datasets, of three non-ruminant (i.e., pigs, humans, and mice) livers. Each species had at least two biological replicates (Additional file 2: Table S1A-C) [17, 18]. We have processed all the data using the same data analysis pipeline to make human, mouse, and pig datasets (previously generated) comparable to those (newly generated) of three ruminants. Furthermore, we also collected datasets from seven other cattle tissues to investigate the dynamic epigenetic landscape across tissues [19]. Overall, we uniformly analyzed 35 new genome-wide omics datasets from three ruminant livers and integrated them



with 102 previously published datasets. We obtained over 25 billion mapped reads with an average mapping rate of 91.24% after filtering low-quality reads (Fig. 1A and Additional file 2: Table S1A-D).

Through signal saturation analysis, we found that 20 million reads were required to reach the saturation of consistent peak detection for H3K27ac and H3K4me3 in single-end ChIP-seq samples, while 37.5 million reads

were required for paired-end ChIP-seq samples (Additional file 1: Fig. S2A and S2B). We detected 66,000–108,000 (mean = 83,338) H3K27ac-enriched regions and 18,000–29,000 (mean = 23,576) H3K4me3-enriched regions in liver ( $q < 0.01$ , Additional file 1: Fig. S2C and S2D). Furthermore, we defined two categories of REs, (1) promoters, which were simultaneously marked by H3K4me3 and H3K27ac, and (2) enhancers, which were

only enriched for H3K27ac. We identified an average of 81,081 REs per species, including 63,927 enhancers and 17,154 promoters (Fig. 1B and Additional file 3: Table S2). Moreover, we found that enhancers exhibited higher tissue specificity compared to promoters. For example, we found that a majority of (85.94%) enhancers in the liver exhibited tissue specificity, while only 22.92% of promoters did (Additional file 1: Fig. S2E). We observed that 78.09% of all promoters were located around (distance  $\leq$  5kb) transcriptional start sites (TSSs), whereas the majority (77.88%) of enhancers were distal to TSSs (distance  $>$  5kb) (Additional file 1: Fig. S3A), which was consistent with previous findings [17]. We calculated the enrichment fold of REs for 14 chromatin states previously predicted by ChromHMM [19] and observed that these REs were significantly enriched for the corresponding chromatin states (Fig. 1C and Additional file 1: Fig. S3B and S3C). For instance, enhancers were significantly enriched for “Active\_Enhancer” and “CTCF/Enhancer” (enrichment fold = 21.78 and 16.61, respectively). We further validated that over 70% and 45% of our newly detected promoters and enhancers overlapped with REs identified using publicly available datasets in the liver (Fig. 1D and Additional file 1: Fig. S3D and S3E) of cattle, pigs, and mice [17, 19, 20]. Overall, these results indicate the high reliability of the REs identified in this study. Notably, we also newly identified REs in the liver that had not been annotated in previous studies. For example, 7.23% and 45.89% of all cattle promoters and enhancers were newly identified in this study, respectively (Fig. 1D).

To evaluate the evolution of epigenomic marks and gene expression across these six species, we performed hierarchical clustering based on epigenomic mark signal intensities and expression levels of 9796 one-to-one orthologous genes. As expected, we observed that the three ruminants were clustered together, consistent across all six omics data types (Fig. 1E and Additional file 1: Fig. S4). This relationship pattern across species was consistent in terms of gene expression and epigenomic marks, which were measured by the pairwise Rand index (Fig. 1F). These observations reflected the consistent effects of epigenomic marks, gene expression, and genome during species divergence.

#### Co-evolution of epigenomic regulatory and gene expression

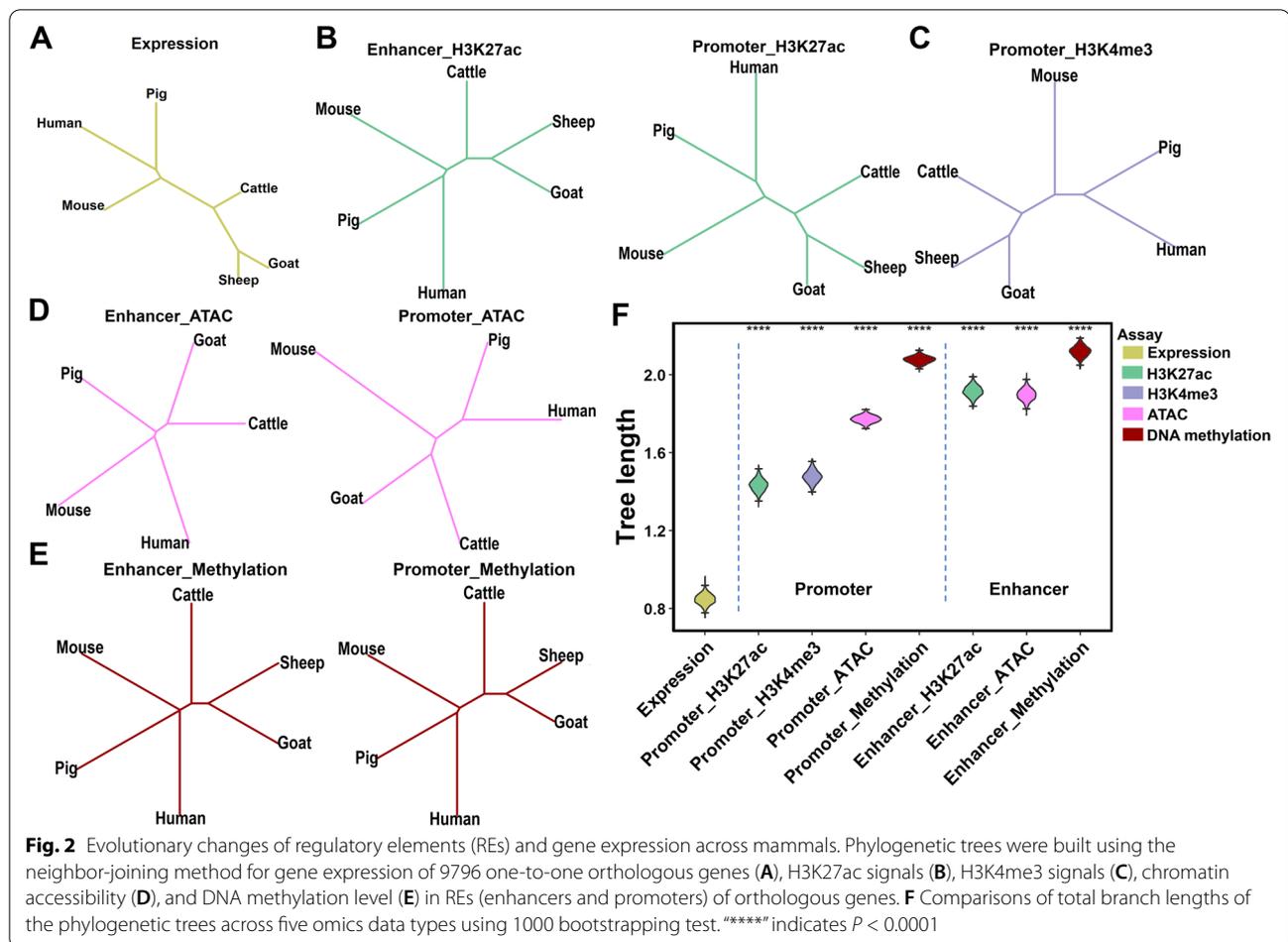
To obtain a global view of the evolution of REs and gene expression, we constructed phylogenetic trees for five distinct omics data types, including ChIP-seq for H3K27ac and H3K4me3, RNA-seq (gene expression), WGBS (DNA methylation), and ATAC (chromatin accessibility) (Fig. 2A–E). All eight phylogenetic trees were in agreement with the known genome phylogeny across

six species [21]. We observed two major mammalian lineages (ruminants and non-ruminants), followed by the separation among the ruminant lineages (Bovidae and Caprinae). This implies co-evolution and interplay among different functional genome elements and DNA sequences during mammalian evolution.

According to bootstrapping analysis, the total length of the phylogenetic trees varied widely across these omics data types (Fig. 2F), reflecting their differences in evolution rates. Notably, the branch of the gene expression tree was significantly ( $P < 0.0001$ ) shorter than those of another seven epigenomic phylogenetic trees, indicating that gene expression levels were more conserved than their regulatory elements during mammalian evolution. Furthermore, promoters were highly conserved compared to distal enhancers [17, 22]. Compared to histone modifications (H3K27ac and H3K4me3) and chromatin accessibility, DNA methylation in REs evolved faster (Fig. 2F) [23].

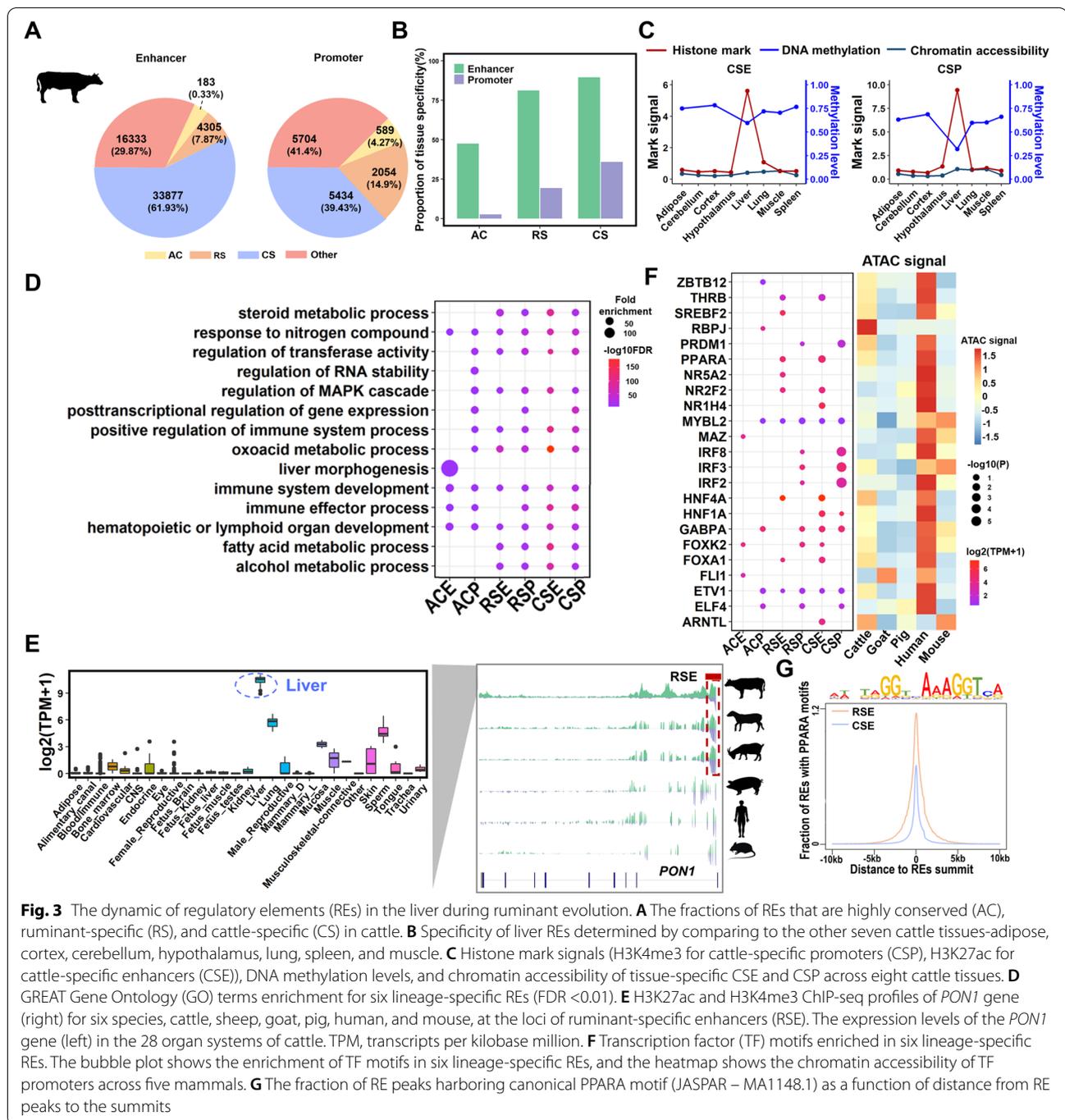
#### Dynamic landscape of hepatic REs during ruminant evolution

To further investigate the epigenomic molecular mechanisms underlying ruminant evolution, particularly in cattle, we divided cattle REs into three main categories using cattle as the “anchor” species based on their absence/presence among six species, i.e., highly conserved (AC), ruminant-specific (RS), and cattle-specific (CS) REs (Additional file 1: Fig. S5A). We identified 772 AC-REs across six mammals, including 183 enhancers and 589 promoters (Fig. 3A and Additional file 1: Fig. S5B–F). Consistent with the previous observation that promoters had a high conservation level than enhancers in the liver across primates [17], we observed that  $\sim$ 4.27% of promoters were AC-Promoters, whereas only  $\sim$ 0.33% of enhancers were AC-Enhancers. By comparing with the REs detected in another seven bovine tissues [19], we observed that REs in the liver with higher lineage-specificity exhibited higher tissue specificity, which was consistent for both enhancers and promoters. For instance, by comparing with the other seven tissues, we found that 47.54% and 89.61% of highly conserved enhancers (AC-Enhancers) and cattle-specific enhancers (CS-Enhancers) were specific in the liver, respectively (Fig. 3B). In general, we found that enhancers are more tissue-specific than promoters, consistent with the hypothesis that enhancers are important regulators of tissue-specific gene expression and are highly related to the function of the respective tissues [15, 24]. In addition, liver-specific REs exhibited hypomethylation and high expression of their target genes specifically in the liver (Fig. 3C and Additional file 1: Fig. S6 and S7) [25, 26].



Based on gene ontology (GO) enrichment analysis, we found that the putative target genes of different lineage-specific REs were significantly ( $FDR < 0.01$ ) enriched for distinct biological processes (Fig. 3D and Additional file 4: Table S3). Genes linked to AC-REs were significantly involved in fundamental and developmental biological functions (e.g., regulation of RNA stability and liver morphogenesis), whereas those linked to RS- and CS-REs were enriched in metabolic and immune processes (e.g., steroid metabolic processes and positive regulation of immune system processes). For instance, a liver-specific ruminant-specific enhancer (RS-Enhancer) located within the *PONI* gene was shared across three ruminants (Fig. 3E), which plays an important role in inhibiting low-density lipoprotein oxidation and has rapidly evolved in the ancestor of ruminants [27, 28]. Moreover, *PONI* had high expression levels specifically in adult livers of cattle (Fig. 3E) [25]. To investigate potential transcription factors (TFs) change during ruminant evolution, we performed TF

motif enrichment analysis for lineage-specific REs. The enriched TFs had similar biological functions as genes linked to the corresponding lineage-specific REs and exhibited matched expression levels and chromatin accessibility across species (Fig. 3F and Additional file 5: Table S4). For example, *RBPJ* and *MAZ*, which are involved in intrahepatic bile duct development and transcription initiation [29–32], were significantly enriched in AC-REs, whereas *HNF4A*, *PPARA*, and *FOXA1* which are associated with hepatocyte proliferation, glucagon biosynthesis, and lipid metabolism [33–35] were enriched in RS-Enhancers and CS-Enhancers. We further observed that RS-Enhancers contained more liver-related motifs than CS-Enhancers, as exemplified by the fraction of motif for the *PPARA* transcription factor (Fig. 3G). Moreover, we found specific enrichment of *IRF* motif families in ruminant-specific promoters (RS-Promoters) and cattle-specific promoters (CS-Promoters). Overall, we provide a valuable resource for lineage-specific REs, which play an



**Fig. 3** The dynamic of regulatory elements (REs) in the liver during ruminant evolution. **A** The fractions of REs that are highly conserved (AC), ruminant-specific (RS), and cattle-specific (CS) in cattle. **B** Specificity of liver REs determined by comparing to the other seven cattle tissues—adipose, cortex, cerebellum, hypothalamus, lung, spleen, and muscle. **C** Histone mark signals (H3K4me3 for cattle-specific promoters (CSP), H3K27ac for cattle-specific enhancers (CSE)), DNA methylation levels, and chromatin accessibility of tissue-specific CSE and CSP across eight cattle tissues. **D** GREAT Gene Ontology (GO) terms enrichment for six lineage-specific REs (FDR <0.01). **E** H3K27ac and H3K4me3 ChIP-seq profiles of *PON1* gene (right) for six species, cattle, sheep, goat, pig, human, and mouse, at the loci of ruminant-specific enhancers (RSE). The expression levels of the *PON1* gene (left) in the 28 organ systems of cattle. TPM, transcripts per kilobase million. **F** Transcription factor (TF) motifs enriched in six lineage-specific REs. The bubble plot shows the enrichment of TF motifs in six lineage-specific REs, and the heatmap shows the chromatin accessibility of TF promoters across five mammals. **G** The fraction of RE peaks harboring canonical PPARA motif (JASPAR – MA1148.1) as a function of distance from RE peaks to the summits

important role in multiple fat metabolism processes, immune function, and hepatocyte development.

**The number of REs contributes to gene expression conservation across species**

To detect how the evolution of enhancers and promoters regulates interspecies changes in gene expression during ruminant evolution, we generated RNA-seq datasets to

quantify gene expression levels from a total of six species. We then grouped genes into four categories: (1) genes with both enhancers and promoters (Both); (2) genes with only promoters (Promoter); (3) genes with only enhancers (Enhancer), and (4) genes with none of the defined REs (None). Firstly, we found that “Both” and “Promoter” genes had the highest conservation, followed by “Enhancer” genes, while “None” genes showed the

lowest conservation (Fig. 4A and Additional file 1: Fig. S8). The phyloP sequence conservation score (Fig. 4B) and probability of loss-of-function intolerance (pLI) score (Fig. 4C) of “Both” and “Promoter” genes were also higher than that of the other two gene sets. Next, we tested the relationship between the number of REs and gene expression levels. Consistent with previous studies [36, 37], genes with more REs tended to have higher gene expression levels (Additional file 1: Fig. S9 and S10), illustrating that the majority of the REs identified in this study showed an additive effect on gene expression levels. Moreover, we observed that the number of REs of a gene could also influence interspecies conservation of gene expression. Genes associated with multiple REs (i.e., enhancers  $\geq 3$ ; promoters  $\geq 2$ ) exhibited significant transcriptional conservation, compared to those with no or fewer REs (i.e., enhancers  $\leq 2$ ; promoters  $\leq 1$ ) (Wilcoxon signed-rank test: enhancers  $P = 0.0354$  and promoters  $P = 6.104e-05$ ; Fig. 4D). Overall, these results further support that the redundancy of REs may contribute to buffering in transcriptional evolution and regulatory innovation [36–38].

To determine whether the conserved REs contribute to the evolutionary stability of gene expression across species, we compared the evolutionary stability of genes with similar expression levels, but differing in the absence or presence of conserved REs. Genes with AC-REs exhibited higher conservation than those without (Wilcoxon signed-rank test: AC-Enhancers  $P = 0.04126$  and AC-Promoters  $P = 1.221e-04$ , respectively; Fig. 4E). These results indicate that conserved REs contribute to the evolutionary conservation of gene expression.

#### REs are enriched for genomic variants and QTLs of complex traits

To further investigate the relationship between characterized REs and genomic mutations, we calculated the enrichment of REs for single-nucleotide polymorphisms (SNPs) and copy number variable regions (CNVRs) in cattle obtained from the NCBI database of SNP (dbSNP) and Animal Omics Database (AOD) [39]. We found that promoters had higher enrichment for CNVRs and SNPs than enhancers, while lineage-specific REs had a lower enrichment for CNVRs than conserved ones (Fig. 5A). These observations were consistent with previous CNV annotation in humans that CNVRs prefer to be near

promoters and away from enhancers [40, 41]. Moreover, the density of SNPs increased in the proximity of REs (Additional file 1: Fig. S11A), particularly for CS-Promoters. We performed enrichment analyses using eQTLs in cattle livers with lineage-specific REs [9]. All six types of REs were significantly enriched for eQTLs, with the enrichment fold ranging from 1.41 to 2.03 ( $P < 0.05$ , 1000 bootstrapping test). These eQTLs exhibited higher enrichment in promoters of matching lineage-specific types than in enhancers, as previously reported in humans (Fig. 5B) [42]. This result indicates that lineage-specific REs may have played important roles in regulating gene expression during mammalian evolution.

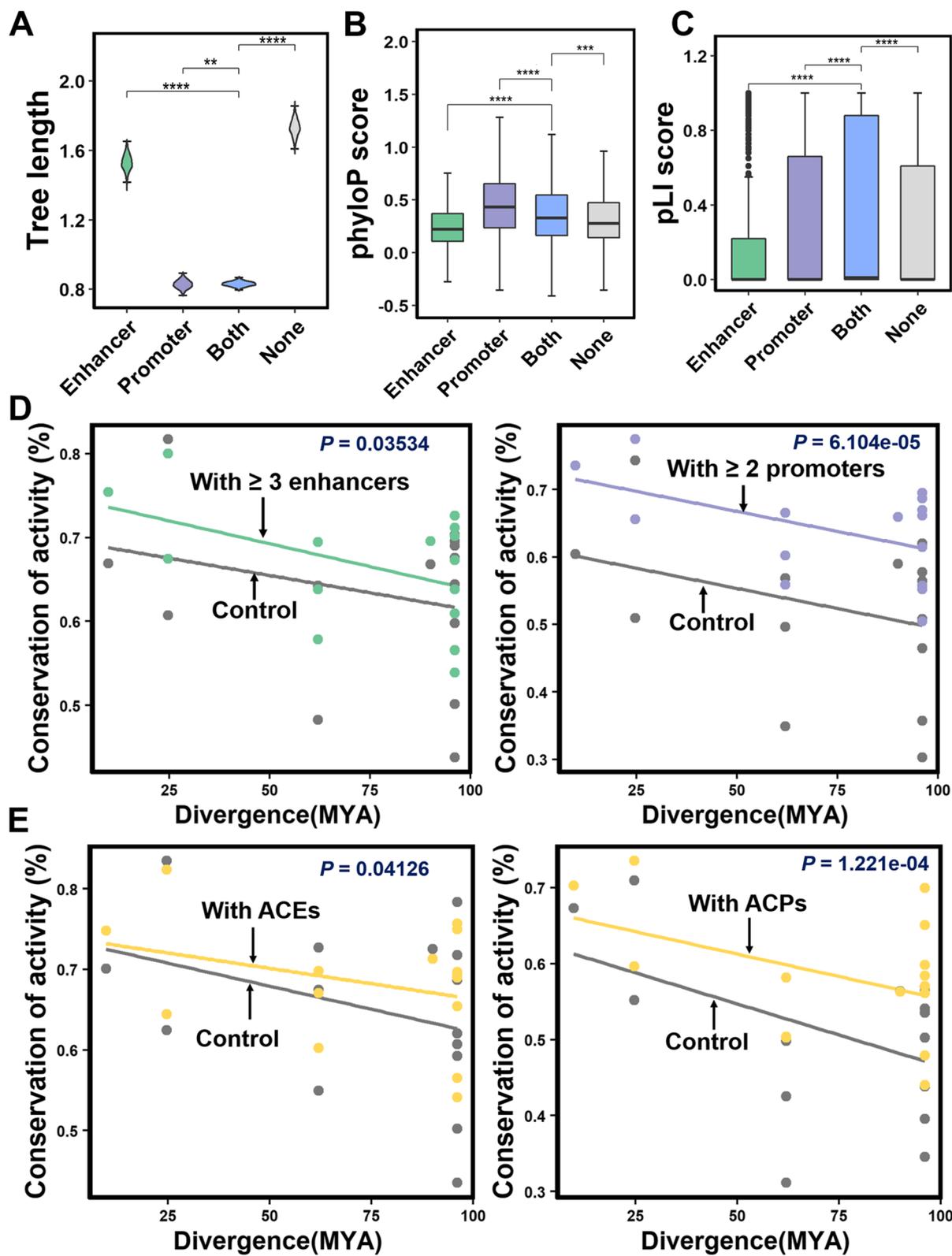
By overlapping REs with QTLs of 489 different cattle traits from the cattle QTL database (QTLdb), we observed that all six types of REs showed enrichments for QTLs of six trait categories (Fig. 5C and Additional file 6: Table S5A-F). The top enriched complex traits were associated with liver function, such as milk alpha-S1-casein content and milk cis-fatty acid content. Body condition score and body weight (24 months) were enriched for AC-Promoters, which was consistent with a previous study on beef cattle that suggested the liver is the relevant tissue of feed efficiency and is associated with daily weight gain through regulating metabolism processes [44]. Moreover, SNPs inside REs showed significantly stronger associations (as represented by the  $P$ -value) than those outside REs for both cattle and human traits, such as protein percentage in cattle and alkaline phosphatase in humans (Fig. 5D and Additional file 1: Fig. S11B). In summary, these observations further illustrate that lineage-specific REs are hotspots of causative mutations for complex traits.

#### Enrichment analysis of REs for GWAS signals of complex traits in cattle and humans

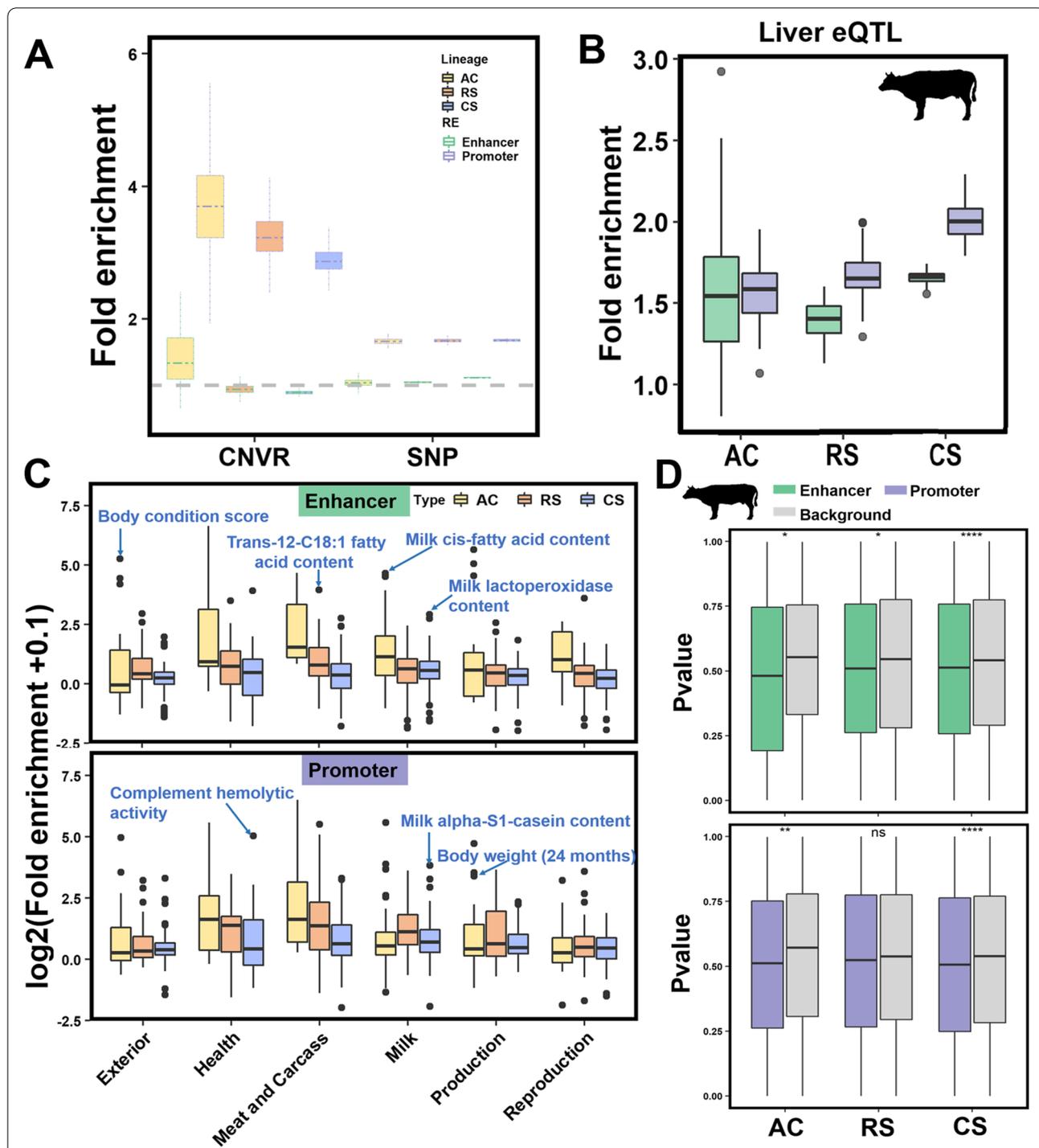
To further investigate how the conservation of REs shapes the genetics of complex traits in cattle and humans, we conducted GWAS signal enrichment analyses of the characterized REs using GWAS summary statistics from 44 and 52 complex traits in cattle and humans, respectively (Fig. 6A, B, Additional file 7: Table S6, and Additional file 8: Table S7). For human traits, we also conducted the heritability enrichment analysis using the stratified linkage disequilibrium score regression (LDSC) [45], and found that the results from

(See figure on next page.)

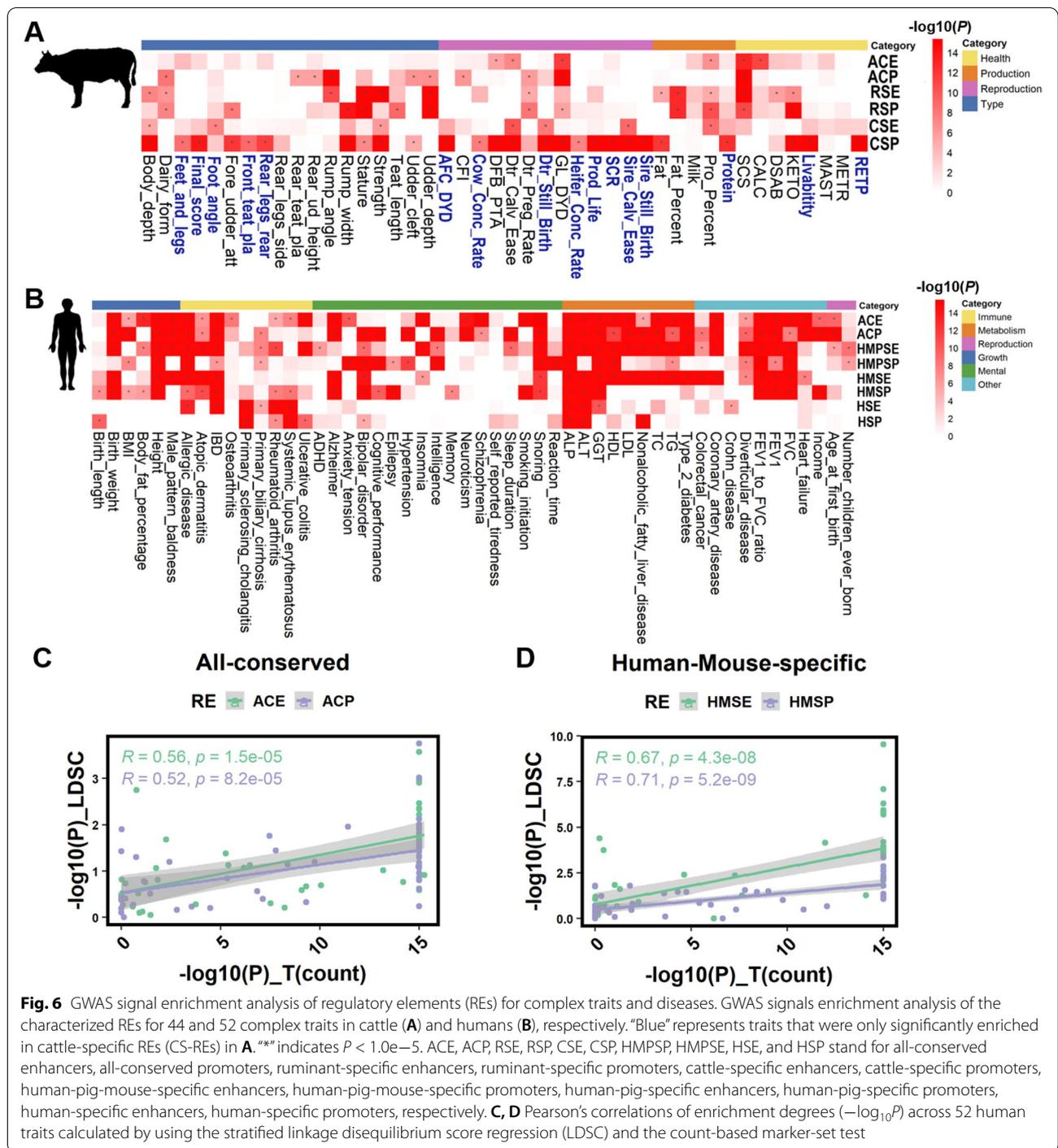
**Fig. 4** Regulatory elements (REs) drive interspecies transcriptional conservation during evolution. **A** Comparison of total branch lengths for the phylogenetic trees across four categories of one-to-one orthologues genes based on the 1000 bootstrapping test. Enhancer, Promoter, Both, and None indicate genes with only enhancers, only promoters, both enhancers, and promoters, without REs, respectively. Box plots showing the median phastCons scores (**B**) and pLI score (**C**) for four categories of one-to-one orthologues genes. **D** The number of associated enhancers and promoters contributes to interspecies transcriptional stability. **E** The association of gene stability and conserved REs, including highly conserved enhancers (ACE) and highly conserved promoters (ACP). “\*\*\*” indicates  $P < 0.01$ ; “\*\*\*\*\*” indicates  $P < 0.001$ ; “\*\*\*\*\*” indicates  $P < 0.0001$



**Fig. 4** (See legend on previous page.)



**Fig. 5** Regulatory elements (REs) are enriched for genomic variants of gene expression and complex traits. The fold enrichment of REs for genomic variants [39] (A), expression QTLs (eQTLs) detected from cattle liver [9] (B), and six categories of QTLs (including 489 complex traits) downloaded from the cattle QTL database [43] (C), based on 1000 bootstrapping test. D The P values of SNPs inside and outside of REs from GWAS summary datasets for protein percentage in cattle [3]. “\*” indicates  $P < 0.05$ ; “\*\*” indicates  $P < 0.01$ ; “\*\*\*\*” indicates  $P < 0.0001$ ; “ns” indicates  $P > 0.05$



**Fig. 6** GWAS signal enrichment analysis of regulatory elements (REs) for complex traits and diseases. GWAS signals enrichment analysis of the characterized REs for 44 and 52 complex traits in cattle (A) and humans (B), respectively. “Blue” represents traits that were only significantly enriched in cattle-specific REs (CS-REs) in A. \*\*\* indicates  $P < 1.0e-5$ . ACE, ACP, RSE, RSP, CSE, CSP, HMPSP, HMPSE, HSE, and HSP stand for all-conserved enhancers, all-conserved promoters, ruminant-specific enhancers, ruminant-specific promoters, cattle-specific enhancers, cattle-specific promoters, human-pig-mouse-specific enhancers, human-pig-mouse-specific promoters, human-pig-specific enhancers, human-pig-specific promoters, human-specific enhancers, human-specific promoters, respectively. C, D Pearson’s correlations of enrichment degrees ( $-\log_{10}P$ ) across 52 human traits calculated by using the stratified linkage disequilibrium score regression (LDSC) and the count-based marker-set test

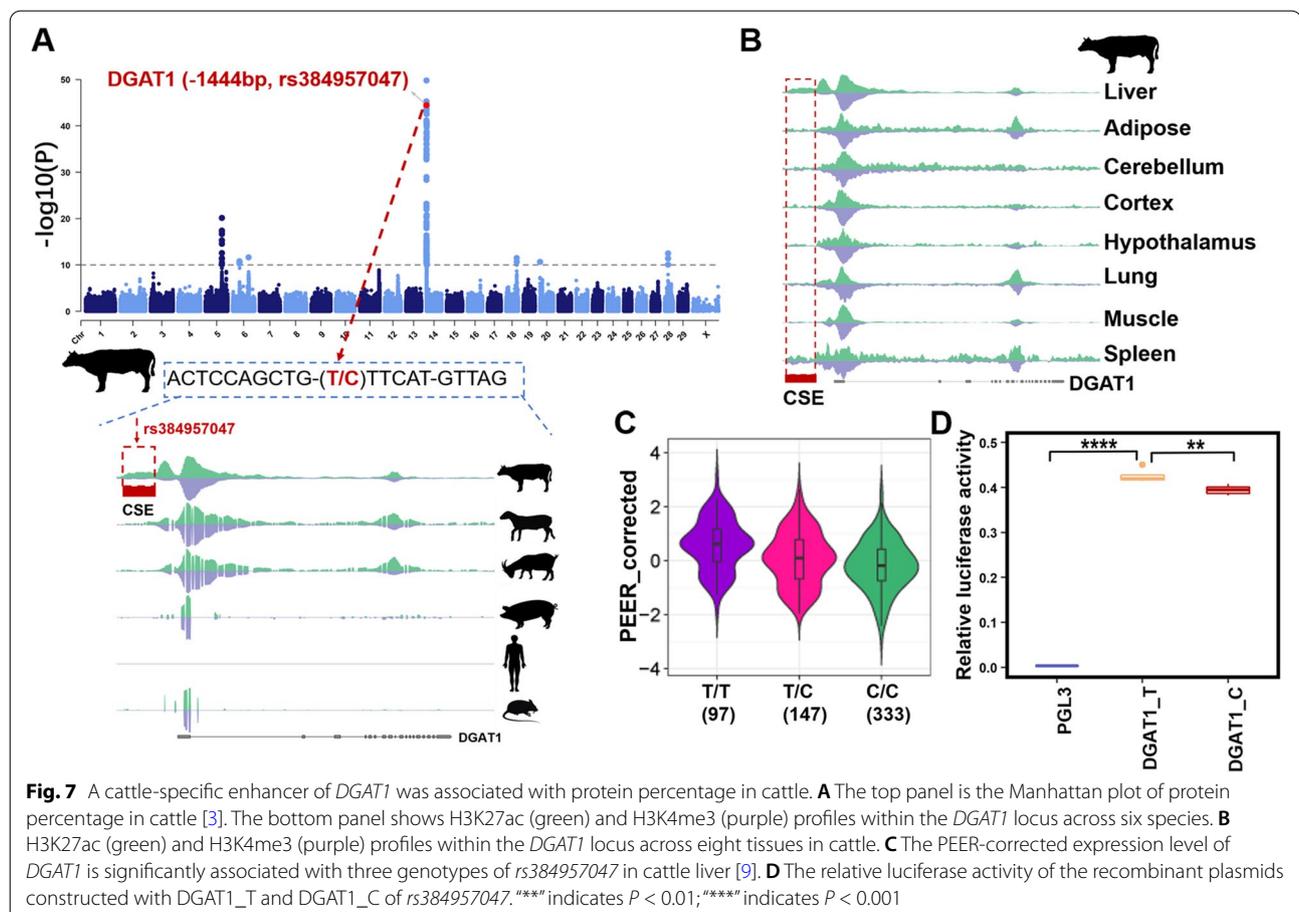
LDSC were significantly and positively correlated with those from the count-based test (Fig. 6C, D). We did not conduct the similar analysis as the matched linkage disequilibrium (LD) and minor allele frequency (MAF) were not available for the Holstein population being analyzed. It would be of interest to consider the LD and genetic

relatedness of individuals when conducting the heritability enrichment analysis in livestock, once the individual genotypes and phenotypes are available.

In general, CS-/RS-REs showed a higher enrichment for many cattle complex traits than conserved ones, whereas the opposite trends were observed for human

complex traits, where human-specific REs showed a lower enrichment for human complex traits than conserved ones. Notably, we found that highly conserved REs were enriched for similar complex traits in humans and cattle. For instance, AC-Promoters were significantly enriched for somatic cell score (SCS) ( $P < 1.00e-15$ ) in cattle and primary biliary cirrhosis ( $P < 3.51e-08$ ) in humans (Fig. 6A, B). This was consistent with a previous study, which reported that conserved regulatory regions enriched for SNPs associated with inflammatory traits in humans [12]. Furthermore, we noted that GWAS signals for some complex traits were regulated by both AC- and RS-/CS- specific REs (Fig. 6A), such as protein percentage in cattle. These results suggest that for some complex traits, cross-species comparison of the functional genome may help us interpret the biological mechanisms underlying complex traits, and then enhance genetic improvement in livestock species. We also found that GWAS signals for 16 out of 44 complex traits were only significantly enriched in CS-REs, including five body type traits, eight reproductive traits, one production trait, and two healthy traits (Fig. 6A). This may be due to the

strong artificial selection of complex traits of economic value in cattle. Interestingly, we noted that lineage-specific enhancers were enriched for GWAS signals of both protein percentage and ketosis in cattle (Fig. 6A). For instance, a liver-specific CS-Enhancer within *DGAT1* harbored risk SNPs associated with protein percentage (*rs384957047*,  $P < 1.80e-18$ ), which is located in the upstream of *DGAT1* [3, 46] (Fig. 7A, B). *DGAT1* is necessary for triacylglycerol synthesis and has been reported to be associated with milk production, milk fat composition, metabolizability, and N efficiency in cattle [47–49]. The expression level of *DGAT1* was significantly regulated by *rs384957047* in cattle liver, which is a fine-mapped *cis*-eQTL ( $P < 1.14e-11$ , Fig. 7C). Moreover, transcription factor binding prediction analysis showed that TFs were changed before and after the SNP *rs384957047* mutation in the upstream of *DGAT1* (Additional file 9: Table. S8). We further constructed a luciferase reporter experiment and found that luciferase activity was significantly changed with different alleles of *rs384957047* (T > C) in *DGAT1* (*t*-test:  $P < 0.01$ , Fig. 7D and Additional file 10: Table. S9). Overall, these observations suggest that



classifying REs by evolutionary conservation can improve our understanding of the genetic mechanisms underlying complex traits and target causal variants in both livestock and humans.

## Discussion

Gene regulatory divergence and its associations with complex traits have been extensively studied in humans [50]. However, little is known about the mechanisms underlying the regulatory roles of epigenetic variation in gene expression patterns and phenotypic divergence during ruminant evolution. Here, we provide a comprehensive insight into the comparative epigenomics of the liver across three ruminants (cattle, sheep, and goats) and three non-ruminants (pigs, humans, and mice). By integrating histone signals (HK4me3 and H3K27ac), transcriptomes, chromatin accessibility, and DNA methylation, we found that genes linked to RS-REs were significantly enriched in immune processes, including immune effector processes and positive regulation of the immune system processes. Moreover, we found specific enrichment of IRF motif families in RS-REs. The liver plays an essential role in many immune processes. For example, hepatocytes are responsible for the production of 80%–90% of the circulating innate immunity proteins in the body, and the liver contains a large number of resident immune cells. Therefore, we inferred that RS-REs may be important for the ruminant inflammatory process. However, further functional experiments are required to validate the immune functions of these RS-REs [51, 52]. By combining with large-scale GWAS summary datasets for humans and cattle, we further demonstrated the relationships between complex traits and conserved/lineage-specific REs, and provided an important resource for narrowing down the range of causative mutations. Collectively, our results provide insight into the ruminant regulatory landscape and evolution. Our study also demonstrates that the datasets used for this study from public databases provide valuable functional information for the genetic breeding of livestock. Moreover, we found that 70%–90% and 9%–45% of the currently detected promoters and enhancers in humans and mice overlapped with those previously reported (Additional file 1: Fig. S12) [17, 53, 54], implying a high concordance between the current study and previous reports, especially for promoters. However, further functional validation of those enhancers required experimental validations (e.g., enhancer reporter vectors [55]) either *in vivo* or *in vitro*.

Cross-species epigenomic comparison has been widely used to investigate the evolutionary basis of REs and their impacts on species-specific complex traits (e.g., cognitive traits in humans) [11, 56]. Fang et al. observed that conserved hypomethylated regions between humans and

cattle play an important role in immune response and are significantly enriched for GWAS signals of immune-related traits in both species [56], indicating that cross-species comparison of epigenome could contribute to narrowing down genetic variants of complex traits and diseases. Castelijns et al. demonstrated that human-specific REs were involved in oligodendrocyte function postnatally and were changed in the brain of autism patients, by comparing the H3K27ac mark in the brain across humans, chimpanzee, marmoset, macaque, and mice [11]. Here, through integrating with large-scale eQTLs and GWAS in both humans and cattle, we demonstrated how cross-species comparison of functional genomes can help us understand the evolutionary and genetic basis of complex traits/diseases. Overall, these observations suggest the importance of using cross-species datasets to decipher the mechanisms of GWAS loci.

Our integrative analysis of histone modification marks, DNA methylation, gene expression, and chromatin accessibility shed light on the regulatory patterns of co-evolution during ruminant evolution. We observed that the phylogenetic relationships of species based on epigenomic signal intensity and gene expression levels were consistent with that inferred from the DNA sequence, reflecting a co-evolution pattern of the functional genome and DNA sequence across species. Previous studies have proposed the notion that changes in regulatory elements affect gene expression levels and then lead to phenotypic diversity across species [36]. Zhou et al. reported that 40% of the variance in gene expression of lymphoblastoid cell lines could be explained jointly by evolutionary changes of five marks (i.e., H3K4me1, H3K4me3, H3K27ac, and H3K27me3, and RNA polymerase II) across primates [57]. These observations imply a synergistic effect of epigenomic marks on gene expression during mammalian evolution. However, such interaction effects among different types of epigenomic marks need to be further explored. We observed higher conservation of gene expression than regulatory marks. This may due to (1) REs found in this study could still contain false positives and there are still missing REs, (2) RNA-seq only measures gene expression changes at the gene level but not transcript level, as REs changes might be contributing to more varied transcript-level expression changes. Overall, a better understanding of buffering mechanisms underlying gene expression might help us interpret how evolutionary changes in epigenetic regulation contribute to gene expression variation across species.

In this study, we observed significant enrichment of GWAS signals in AC-REs across six mammals, which is consistent with the findings of Liu et al. [13]. For instance, AC-Enhancers were significantly enriched

for immune-related traits in both cattle (i.e., SCS) and humans (i.e., primary biliary cirrhosis). Notably, the majority of complex traits were enriched in AC-REs in humans, especially for metabolic diseases, such as non-alcoholic fatty liver disease and alanine aminotransferase. The strong association between metabolic diseases and AC-REs suggests that important metabolism-related mutations may exist in ancient REs and also imply that cattle may serve as a biomedical model for studying human metabolic diseases.

A previous study on humans also found that human-specific regulatory gains were significantly enriched for REs depleted in autism spectrum disorder and were associated with susceptibility to neural diseases [11]. We also found that CS-REs were significantly enriched for GWAS signals of many complex traits in cattle. This may be because that most of these analyzed phenotypes (e.g., milk, protein, and fat yield) in cattle are economic traits that are under intensive artificial selection [58], while most analyzed phenotypes in humans are complex diseases that are more likely to be under natural selection [59]. Moreover, all the GWAS traits being analyzed here are somehow related to milk production in Holsten dairy cattle. The enrichment of GWAS signals of cattle traits in CS-REs may also imply the importance of studying cattle functional genomes to enhance their genetic improvement. In particular, we identified a potential causative variant for the protein percentage of milk, which is located in CS-Enhancer. It might influence the protein percentage of milk might by regulating the expression of *DGATI* (the upstream of 1444bp away). The liver plays an important role in the synthesis and secretion of lipoproteins to provide the mammary gland with cholesterol and triglycerides for milk production during lactation in cows. Previous studies also observed that inhibition of *DGATI* expression in bovine hepatocytes reduced triacylglycerol accumulation and significant effects of *DGATI* on milk production traits [60, 61]. Moreover, the expression of *DGATI* in the liver was significantly associated with milk production according to transcriptome-wide association studies in cattle [9, 62]. We found that the allele T of rs384957047 increased the expression level and transcriptional activity of *DGATI* compared with allele C, which implied that the allele T of rs384957047 was associated with increased milk production. However, further analysis is needed to determine the functional impact of rs384957047 on milk production traits. Of note, the limitation of the current study is that only three ruminants were examined. In the future, other Bovidae species (e.g., yak and water buffalo) will be required to further explore whether this enhancer of *DGATI* only exists in cattle. Overall, the link between cattle functional genome evolution and complex traits

may reflect cattle-specific transcriptional programs and provide priority candidate regions for identifying causative variants of complex traits.

## Conclusions

Collectively, our study systematically annotated and compared regulatory elements in the liver across three economically important ruminants (i.e., cattle, sheep, and goats) and three non-ruminant mammals (i.e., pigs, mice, and humans). By integrating large-scale GWAS summary datasets in cattle and humans, we demonstrated the importance of cross-species comparison of REs in understanding the evolution of the functional genome, and genetic mechanisms underlying complex traits/diseases in livestock and humans.

## Methods

### Animal and tissue collection

In this study, a total of nine liver samples were collected from adult healthy individuals, including three cattle (3–4 years old; Holstein, healthy), three sheep (2–3 years old; Texel, healthy), and three goats (2–3 years old; Yunnan black goat, healthy). All samples were collected immediately postmortem and stored in liquid nitrogen. All animal procedures were performed according to protocols of the Institutional Animal Care and Use Committee (IACUC) at China Agricultural University. The ChIP-seq, WGBS, and RNA-seq liver assays were constructed from the same samples. For pig, mouse, and human data, only samples from adult and healthy individuals were considered.

### ChIP-seq datasets

We conducted ChIP according to published protocols [63]. The frozen liver samples were first powdered and washed with phosphate-buffered saline (PBS) buffer solution, and it was then vigorously shaken with 1% formaldehyde solution for 20 min at room temperature to cross-linking. Next, the samples were incubated at room temperature for 10 min by adding 250 mM glycine. The cross-linked liver samples were homogenized in a Dounce homogenizer and rinsed twice with PBS to solubilize DNA-protein complexes. The chromatin was sheared to an average size of 300 bp with sonication and centrifuged at 4 °C for 10 min to remove the pellet. Chromatin extracts were used to perform ChIP experiments using antibodies against H3K27ac (Abcam ab4729) and H3K4me3 (Abcam ab85850). We also constructed control experiments (input libraries) to assess potential artifacts related to the shearing of DNA and amplification, where DNA has been cross-linked and sonicated chromatin without immunoprecipitation. ChIP sequencing libraries were prepared from input or ChIP-input DNA

in 96-well plates and sequenced on an Illumina HiSeq 2500 with paired-end 150 bp reads by Novogene (Beijing, China).

The public liver ChIP-seq (H3K27ac and H3K4me3) datasets were obtained from the National Center for Biotechnology Sequence Read Archive (NCBI SRA) database, including human (PRJEB6906) [17], mouse (PRJEB28147), and pig (PRJEB28147) datasets [18].

#### ChIP-seq data mapping and quality evaluation

We used Trim Galore v0.4.0 and FastQC v0.11.2 software to trim the raw reads and obtain sequence quality reports. The clean reads were mapped to the corresponding reference genome of each species using BWA-MEM software v0.7.17 [64] with the default parameters: ARS-UCD1.2, Oar\_rambouillet\_v1.0, ARS1, Sscrofa11.1, GRCh38, GRCm39. Mapped reads were further filtered using SAMtools view utilities (1.10) [65] with the parameters “-q 1.” Duplicated alignment reads were removed using the Picard Tools v2.25.0 (<https://github.com/broadinstitute/picard>) with the parameter “REMOVE\_DUPLICATES = true.” To evaluate the enrichment efficiency of ChIP-seq, we examined all libraries of the relative strand cross-correlation coefficient (RSC) and normalized strand cross-correlation coefficient (NSC) with phantompeakqualtools v1.2.2 [66], where an RSC >1 and NSC >1.1 indicate that the effect of antibody enrichment was acceptable in ChIP-seq. Detailed ChIP-seq quality statistics are shown in Additional file 2: Table S1A.

#### ChIP-seq signal saturation and peak calling

To exclude the influence of sequencing depth and ensure that the ChIP-seq signal was saturated in all libraries, we subsampled mapped the reads from each library (H3K27ac, H3K4me3, and input libraries), starting with 5 million reads and a step of 2.5million. For the single-end sequencing data of public ChIP-seq data, we have a maximum subsample of 30 million reads. We subsampled a maximum of 40 million reads from our paired-end ChIP-seq datasets. For each subsample, we identified narrow peaks using MACS2 v2.1.1 [67] with options “-q 0.01.” Reproducible peaks were detected in at least two biological replicates and at least 50% overlap of their length using intersectBed from bedtools v2.30.0 [68]. Consensus peaks were obtained by merging all reproducible peaks of all biological replicates for further analysis. The sequencing results revealed that H3K27ac and H3K4me3 reached saturation of consistent peak detection at 20 million reads in the single-end ChIP-seq, while H3K27ac and H3K4me3 reached saturation at 37.5 million reads in the paired-end ChIP-seq (Additional file 1: Fig. S2). Subsequently, we used the subsample of H3K27ac and H3K4me3 with 20 million reads in the single-end

ChIP-seq and with 37.5 million reads in the paired-end ChIP-seq for all further analyses. For one H3K4me3 library in pigs and one input library in humans, we used all the reads instead of subsampled, as their deduplicated read counts were less than 20 million.

#### Identification of regulatory regions

Within each species, the consensus peaks of H3K27ac and H3K4me3 overlapped to define regulatory regions enriched for H3K27ac only or both using intersectBed from bedtools v2.30.0. We defined two categories of regulatory regions: (1) promoter, which was simultaneously marked by H3K4me3 and H3K27ac with at least 50% overlap of their length, and (2) enhancer, which contained H3K27ac peaks and did not overlap with H3K4me3 enriched regions. Histone modifications could be used to identify putative REs (i.e., enhancers and promoters) at a genome scale. However, these REs are required for further functional validation, such as using enhancer reporter vectors either in transfected cells or transgenic animals (e.g., mice) [69, 70]. The statistics of regulatory regions are shown in Additional file 3: Table S2.

#### Comparison with those previously detected

To validate the accuracy of REs mapping, we downloaded H3K27ac/H3K4me3 ChIP-seq data for the livers of cattle (PRJEB6906, PRJNA665199, PRJNA665216 and PRJEB41939) [17, 19, 20], pigs (PRJNA597497) [15], and mice (PRJEB6906) [17] from the SRA database (Additional file 11: Table S10). The REs were identified using the aforementioned methods. Mouse and human REs identified in the current study were compared to REs annotated in Ensembl version 103 [53] and VISTA [54] (downloaded July 28, 2022). The overlap of REs identified in the current study and those confirmed in the previous study was calculated using the intersect command in BEDTools v2.30.0 with parameters “-f 0.5 -F 0.5 -e.” Moreover, we obtained the chromatin states of cattle [19], pigs [19], humans [5], and mice [71] in the liver and performed enrichment analysis with the 1000 permutation test using the R package regioneR [72].

#### Whole genome alignment and cross-species comparisons

Cross-species comparisons were performed using 46 eutherian mammals Enredo-Pecan-Ortheus (EPO) multiple genome alignments available from ENSEMBL Compara API (v103) [73, 74]. To further identify the highly conserved REs across the six species, we used cattle as the central species and then aligned other species to cattle in a pairwise manner. REs were considered highly conserved if they overlapped with another regulatory region in all species with a minimum length of 50%. Similarly, lineage-specific REs were defined for

ruminants using the methods as those used to identify highly conserved regions. REs were compared between the reference species (cattle) and species in the other clade using EPO multiple genome alignments. Lineage-specific REs were highly conserved across species within the clade (i.e., cattle, sheep, and goats for the ruminant branch) but not in other species.

#### Identification of species-specific REs

The species-specific REs were determined as previously described [18]. The REs were defined as species-specific ( $N_{si}$ , see Eq. 2) if the DNA sequence could not be aligned to any other species, or if RE activity could not be detected in any other species with the underlying DNA sequence alignable across species.

$$N_i = N_{Ni} + N_{Li} + N_{Ci} \quad (1)$$

$$N_{si} = N_{Ni} + N_{Li} \quad (2)$$

where  $N_i$  = number of REs (promoters and enhancers) in the species  $i$  ( $i$ =cattle, sheep, goats, pigs, humans, and mice);  $N_{Ni}$  = number of REs in species  $i$  without DNA sequence alignment to any other species. REs without DNA sequence alignment were defined as REs that could not be mapped to orthologous regions using 46 eutherian mammals Enredo-Pecan-Ortheus (EPO) multiple genome alignments in ENSEMBL (v103);  $N_{Li}$  = number of REs in species  $i$  without sharing of regulatory activity in any other species but with DNA sequence alignment across many species;  $N_{Ci}$  = number of REs in species  $i$  with shared regulatory activity in at least one other species;  $N_{si}$  = number of species-specific REs in the species  $i$ .

#### RNA sequencing library construction

Total RNA was extracted using TRIzol method from frozen liver samples of three ruminant species: cattle, sheep, and goats. mRNA was enriched using magnetic beads with Oligo (dT) and sheared. The cDNA was synthesized using random hexamers, end-repaired, and ligated to A base and sequencing adapters. Then, cDNA libraries were obtained by polymerase chain reaction (PCR) amplification and sequenced using the Illumina HiSeq X Ten platform with 150bp-paired reads (Novogene Inc., Beijing, China). The liver RNA-seq datasets were obtained from the NCBI SRA database, including PRJEB13074 for humans [36], PRJEB33381 for mice, and PRJEB33381 for pigs [18].

#### RNA sequencing mapping and gene expression quantification

The adapter sequences and low-quality bases were filtered using Trimmomatic v0.39 [75]. Detailed RNA-seq quality statistics are shown in Additional file 2: Table S1B. The trimmed reads were mapped to the Ensembl 103 version of reference genomes with annotated genes using STAR aligner v 2.7.6 [76] with the following options: “--outFilterMismatchNmax 3 --outFilterType BySJout --quantMode GeneCounts --outFilterMismatchNoverLmax 0.04 --outSAMtype SAM SortedByCoordinate.” Gene expression levels were quantified as counts using featureCounts [77] and were normalized by Trimmed Mean of M value (TMM) with edgeR [78]. Genes with TMM-normalized counts per million (CPM) of more than 1 in 20% of the samples were used for further analysis.

#### WGBS library construction and analysis

The genomic DNA of liver samples from three ruminants (cattle, sheep, and goats) was extracted according to the TIANamp Genomic DNA Kit protocol (TIANGEN, Beijing, China). The qualified DNA was fragmented to 300bp and subjected to terminal repairing and an addition at the 3'-terminus. Bisulfite conversion of DNA was performed using the ZYMO EZ DNA Methylation-Gold Kit. DNA fragments were amplified using PCR and selected for library fragment size. The quantified library was sequenced on the Illumina HiSeq X Ten machine (PE-150 bp FC; Annoroad, Beijing, China). The liver WGBS datasets were obtained from the NCBI SRA database as well, including PRJNA287622 for humans [79], PRJNA416505 for mice [80], and PRJNA357500 for pigs [81]. The cattle WGBS datasets were obtained from the NCBI SRA database under PRJNA612978 and PRJNA417285 [26, 82], representing five tissues, i.e., adipose, cortex, lung, muscle, and spleen.

Raw reads were trimmed using Trim Galore v0.4.0 with default parameters. The clean reads were mapped to the reference genome using bismark v0.23.0 with default parameters and were deduplicated with the deduplicate\_bismark options [83]. CpGs with a coverage of at least five were used for further analysis. The methylation levels of CpG sites were calculated using MethPipe v5.0.0, according to the number of methylated Cs in reads at the position corresponding to the site divided by the total of methylated Cs and unmethylated Ts mapping to that position [84].

#### Assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) processing

We obtained 23 public ATAC-seq datasets from SRA databases, and the accession numbers are listed in

Additional file 2: Table S1D. ATAC-seq raw reads were filtered using Trim Galore v0.4.0 with the following parameters “-q 25 -length 25 -e 0.1 --stringency 4.” The clean reads were mapped to the reference genome, and PCR duplicates were eliminated using Bowtie v2.4.2 and Picard Tools v 2.25.0. ATAC-seq peaks were called using MACS2 v 2.1.1 with the following parameters “--nomodel --shift 37 -B --SPMR --ext 73 --pval 1e-2 --call-summits.” The replicated peaks that overlapped by at least 50% in the two samples were identified using intersectBed from bedtools v2.30.0.

#### Assignment of REs to putative target genes and transcriptome divergence

Putative target genes of the REs were identified using methods similar to the GREAT v4.0.4 [85]. The regulatory association domains of genes were defined as 5kb upstream and 1kb downstream from their TSSs. The REs were then linked to their putative target genes if they overlapped ( $\geq 1$ bp) with regulatory association domains of genes using the intersect command in BEDTools v2.30.0. We divided genes into two groups according to the number of REs, such as single (e.g., the number of promoters  $\leq 1$ ) vs. multiple (e.g., the number of promoters  $\geq 2$ ). When estimating the relative conservation of genes with different numbers of REs, we considered single and multiple genes with matched expression levels using the MatchIt [86] library in R with the option “caliper=0.5.” The transcriptome evolutionary divergence was measured by Spearman’s correlation coefficients of expression levels of orthologous genes between pairs of species. Divergence time was obtained from Timetree (<http://www.timetree.org/>).

#### Regulatory and transcriptome phylogenies

We constructed phylogenetic trees based on pairwise distance matrices using the neighbor-joining method in the ape R library [87]. We obtained 9796 1:1 orthologous genes from Ensembl release 103 ([http://asia.ensembl.org/info/genome/compara/homology\\_method.html](http://asia.ensembl.org/info/genome/compara/homology_method.html)) [88], which showed CPM > 1 in more than 20% of the samples. We then calculated gene expression levels and four marker (H3K4me3, H3K27ac, chromosomal accessibility, DNA methylation levels) signal intensities in REs (promoters and enhancers) of orthologous genes. Pairwise distance matrices were estimated as  $1 - \rho$  ( $\rho$  is Spearman’s correlation coefficient of species). The branch length of phylogenetic trees was assessed with 1000 bootstrap analyses.

#### Gene function and TF motif analyses

Gene function enrichment analyses were performed for genes associated with lineage-specific REs using GREAT

v4.0.4 for biological process ontologies. The regulatory domain was defined with the parameters: 5kb upstream and 1kb downstream from TSSs (extending up to 50kb in both directions for the regulatory domains of nearest genes). Significantly enriched terms were determined based on FDR < 0.05. We then performed motif enrichment analysis using HOMER v4.11 with the default motif database [89]. To scan lineage-specific peaks (RS-Enhancers and CS-Enhancers) for the PPARA motifs, we performed RSAT matrix-scan [90] with the matrix of PPARA (MA1148.1) from the JASPAR database (<https://jaspar.genereg.net/>) with the following parameters: Markov order:1, weight score  $\geq 1$ .

#### GWAS enrichment analysis based on REs

Most complex traits are polygenic, and recent studies have reported some differences in the effects of different functional regions on traits [13, 91]. To further understand the regulatory mechanism of economically important traits and diseases in livestock and humans, we obtained GWAS summary datasets for 44 complex traits of 27,214 Holstein bulls with 3,148,506 SNPs [3, 46]. For the human GWAS summary datasets, we collected GWAS summary datasets for 52 complex traits with an average SNPs of 10,846,391 and an average individual of 264,890. The details of the human GWAS summary datasets are summarized in Additional file 12: Table S11. We added 50-kb windows around REs to include the important cis-regulatory variants. We then employed a count-based marker-set test approach with the QGG R package to examine whether GWAS signals were enriched in candidate regulatory regions [92].

$$T_{\text{count}} = \sum_{j=1}^c I(t_i < t_o)$$

where  $T_{\text{count}}$  is the summary statistics for each candidate regulatory region,  $c$  is the number of SNPs in the given region, for instance, AC-Enhancers.  $I$  is an indicator function that takes value one when  $t_i < t_o$ , and we choose  $t_o = 0.01$  as the cut-off. The significance of enrichment was calculated using the hypergeometric test [93].  $P < 1.0e-5$  was set as the threshold of the significant enrichment. Detailed descriptions can be found in Fang et al. [25].

#### Dual-luciferase reporter assays

Here, 293T cells were seeded in 96-well plates and transfected until the cell density reaches 50%–70%. Three recombinant plasmids were constructed, namely DGAT1\_T, in which the SNP of DGAT1 was allele T; DGAT1\_C, in which the SNP of DGAT1 was allele C; and the control plasmid pGL3-basic. MluI and XhoI were

identified as the restriction sites of *DGAT1* (Additional file 1: Fig. S13). The target fragment sequence contains 30bp upstream of rs384957047 and the sequence to the transcription start site of *DGAT1*, including enhancer and promoter. Six hours after transfection of 293T cells, fresh medium was replaced, and 48 h after transfection, samples were collected to detect the luciferase activities using a microplate system. The transfection test was repeated at least three times, and each germplasm plasmid was equipped with two replicates.

### Other downstream bioinformatics analyses

The chromatin states of four mammal (cattle, pigs, humans, and mice), were obtained from previous studies [5, 19, 71]. Enrichment analysis of chromatin state was performed using the *regioner* library in R with 1000 times permutation tests. We conducted genomic variants (i.e., SNP, CNVR, QTL, and eQTL) enrichment analyses using *regioner* library in R (Permutation test: 1000). The genomic variants were downloaded from public databases, including NCBI dbSNP, cattle QTLdb (Release 45, August 23, 2021), AOD [39], and a previous study [9]. We calculated the histone and ATAC signal intensities using *deepTools* v3.1.1 with the following parameters “--normalizeUsing CPM.”

### Abbreviations

ChIP-seq: Chromatin immunoprecipitation followed by sequencing; RNA-seq: RNA sequencing; ATAC-seq: Assay for transposase-accessible chromatin with high-throughput sequencing; RE: Regulatory element; eQTL: Expression quantitative trait locus; WGBS: Whole genome bisulfite sequencing; GWAS: Genome-wide association studies; CNVR: Copy number variable region; SNP: Single-nucleotide polymorphism; FAANG: Functional Annotation of Animal Genomes consortium; FarmGTEx: Farm animal Genotype-Tissue Expression project; TSS: Transcription start sites; GO: Gene Ontology; TF: Transcriptional factor; TPM: Transcripts per kilobase million; EPO: Enredo-Pecan-Ortheus; LDSC: Stratified linkage disequilibrium score regression; LD: Linkage disequilibrium; MAF: Minor allele frequency; SCS: Somatic cell score; RSC: Relative strand cross-correlation coefficient; NSC: Normalized strand cross-correlation coefficient; CPM: Counts per million.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-022-01459-0>.

**Additional file 1: Figure S1.** The overview design of this study. **Figure S2.** General characteristics of ChIP-seq, ATAC-seq, and DNA methylation data in six mammals. **Figure S3.** The assessment of the mapping accuracy for putative regulatory elements (REs). **Figure S4.** Cluster analysis of the six mammals across multi-omics. **Figure S5.** Dynamic changes of REs across species. **Figure S6.** Distributions of averaged epigenomic mark signals of four different types of regulatory elements (REs) across eight tissues in cattle. **Figure S7.** The expression levels (transcripts per million, TPM) of putative target genes of six types of lineage-specific regulatory elements across cattle tissues. **Figure S8.** Phylogenetic trees of four categories of genes. **Figure S9.** The number of enhancers contributes to increasing gene expression levels. **Figure S10.** The number of promoters contributes to increasing gene expression levels. **Figure S11.** SNPs were enriched in regulatory elements (REs). **Figure S12.** Comparison of human and mouse

regulatory regions in the current study with those annotated in Ensembl and VISTA. **Figure S13.** The diagram of luciferase reporters used in this study. The red boxes represent the restriction sites of *DGAT1*.

**Additional file 2: Table S1A.** Sequencing and mapping statistics of ChIP-seq, **Table S1B.** Sequencing and mapping statistics of RNA-seq, **Table S1C.** Sequencing and mapping statistics of WGBS, **Table S1D.** Sequencing and mapping statistics of ATAC-seq. **Table S1E.** The summary of human GWAS datasets.

**Additional file 3: Table S2.** Numbers and definitions of regulatory regions.

**Additional file 4: Table S3.** Significant gene ontology (GO) terms (Binom-FdrQ < 1.0e-4) for six categories of lineage-specific REs.

**Additional file 5: Table S4.** Motif enrichment analysis for six categories of lineage-specific REs.

**Additional file 6: Table S5.** The summary of QTL enrichment folds in cattle.

**Additional file 7: Table S6.** The summary of significant GWAS signal enrichment in cattle.

**Additional file 8: Table S7.** The summary of significant GWAS signal enrichment in humans.

**Additional file 9: Table S8.** Transcription factor binding prediction analysis for rs384957047 mutation in the upstream region of *DGAT1*.

**Additional file 10: Table S9.** The luciferase activities with a different allele of rs384957047 (T > C) in *DGAT1*.

**Additional file 11: Table S10.** The public ChIP-seq data from the liver of cattle, pig and mouse to validate regulatory regions in the current study.

**Additional file 12: Table S11.** The summary of human GWAS datasets.

### Acknowledgements

We thank the contributions of all the co-authors and reviewers.

### Authors' contributions

S.Z., Y.Y., and L.F. designed the study; S.C., S.L., and S.S. performed in sample collection and data generation; S.S. and Y.J. contributed to the computational analysis; M.C. participated in dual-luciferase reporter assays and Y.T. conducted the transcriptome data analysis. W.L. supported the WGBS analysis. S.C. performed the data analysis and wrote the manuscript. Y.Y., L.F., J.L., and S.L. revised the manuscript. All authors read and approved the final manuscript.

### Author's information

Twitter handle: @LingzhaoF (Lingzhao Fang).

### Funding

This work was financially supported by Major special projects of Ministry of Science and Technology during the Fourteen Five-year Plan Period (2021YFD1200903, 2021YFD1200900), the National Natural Science Foundation of China (32072718), the NSFC-PSF Joint Project (31961143009), China Agriculture Research System of MOF and MARA, and the Seed Fund (CAU).

### Availability of data and materials

All data generated or analyzed during this study are included in this published article, its supplementary information files and publicly available repositories. All raw and processed sequencing data for three ruminant liver samples generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO) database under accession numbers GSE206184 [94], GSE206511 [95], and GSE206736 [96]. The public ChIP-seq, RNA-seq, WGBS, and ATAC-seq are downloaded from SRA databases, and the accession numbers can be found in Additional file 2: Table S1 and Additional file 11: Table S10. The chromatin states of cattle, pig, mouse, and human liver are publicly available at [http://farm.cse.ucdavis.edu/~ckern/Nature\\_Communications\\_2020/](http://farm.cse.ucdavis.edu/~ckern/Nature_Communications_2020/), [https://egg2.wustl.edu/roadmap/web\\_portal/](https://egg2.wustl.edu/roadmap/web_portal/) and <https://www.encodeproject.org/search/?searchTerm=ChromHMM+Zhiping+Weng>. Cattle QTLs are downloaded from cattle QTL database (<https://www.animalgenome.org/cgi-bin/QTLdb/BT/index>, Release 45, August 23, 2021). Cattle CNVRs and phyloP sequence

conservation score (phyloP100way.cattle.score) are found in <http://222.90.83.22:88/code/index.php/main>. The pLI score is publicly available at <http://hgdownload.soe.ucsc.edu/gbdb/hg19/gnomAD/pLI/>. Cattle liver eQTLs can be available at <https://cgtex.roslin.ed.ac.uk/>. The cattle gene atlas is publicly available at <http://cattlegeneatlas.roslin.ed.ac.uk>. The GWAS summary statistics for cattle complex traits are publically available through Figshare (<https://figshare.com/s/ea726fa95a5bac158ac1>; <https://figshare.com/s/94540148512dddf7ed32>). Details of human GWAS summary statistics are summarized in Additional file 12: Table S11. The scripts for enrichment analysis are available at the zenodo website (<https://doi.org/10.5281/zenodo.7275349>) [97]. REs annotated in this study have been deposited in the figshare database (<https://doi.org/10.6084/m9.figshare.20402178.v1>) [98].

## Declarations

### Ethics approval and consent to participate

The investigation was approved by Animal handling and the collection of animal samples was conducted under protocols approved by the Institutional Animal Care and Use Committee (IACUC) at China Agricultural University (AW90211202-1-1). All participants provided informed consent.

### Consent for publication

All co-authors have reviewed the manuscript and agreed with its submission.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Key Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture and Rural Affairs & National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China. <sup>2</sup>School of Life Sciences, Westlake University, Hangzhou, China. <sup>3</sup>MRC Human Genetics Unit at the Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK. <sup>4</sup>Center for Quantitative Genetics and Genomics (QGG), Aarhus University, Aarhus, Denmark.

Received: 25 April 2022 Accepted: 7 November 2022

Published online: 08 December 2022

## References

- Deelen J, Uh H-W, Monajemi R, van Heemst D, Thijssen PE, Böhringer S, et al. Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways. *AGE*. 2013;35(1):235–49.
- Al-Mamun HA, Kwan P, Clark SA, Ferdosi MH, Tellam R, Gondro C. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. *Genet Sel Evol*. 2015;47(1):66.
- Jiang J, Cole JB, Freebern E, Da Y, VanRaden PM, Ma L. Functional annotation and Bayesian fine-mapping reveals candidate genes for important agronomic traits in Holstein bulls. *Commun Biol*. 2019;2(1):212.
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science (New York, NY)*. 2012;337(6099):1190–5.
- Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317–30 <https://www.nature.com/articles/nature14248>. [https://egg2.wustl.edu/roadmap/web\\_portal](https://egg2.wustl.edu/roadmap/web_portal). Accessed 3 Jul 2021.
- Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204–13.
- Clark EL, Archibald AL, Daetwyler HD, Groenen MAM, Harrison PW, Houston RD, et al. From FAANG to fork: application of highly annotated genomes to improve farmed animal production. *Genome Biol*. 2020;21(1):285.
- Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, et al. Multi-species annotation of transcriptome and chromatin structure in domesticated animals. *BMC Biol*. 2019;17(1):108.
- Liu S, Gao Y, Canela-Xandri O, Wang S, Yu Y, Cai W, et al. A multi-tissue atlas of regulatory variants in cattle. *Nat Genet*. 2022; <https://www.nature.com/articles/s41588-022-01153-5>. <https://cgtex.roslin.ed.ac.uk/> Accessed 21 Oct 2021.
- IUCN. The IUCN Red List of Threatened Species, Version 2017-3. 2017 [www.iucn.org](http://www.iucn.org).
- Castelijns B, Baak ML, Timpanaro IS, Wiggers CRM, Vermunt MW, Shang P, et al. Hominin-specific regulatory elements selectively emerged in oligodendrocytes and are disrupted in autism patients. *Nat Commun*. 2020;11(1):301.
- Alizada A, Khyzha N, Wang L, Antounians L, Chen X, Khor M, et al. Conserved regulatory logic at accessible and inaccessible chromatin during the acute inflammatory response in mammals. *Nat Commun*. 2021;12(1):567.
- Liu S, Yu Y, Zhang S, Cole JB, Tenesa A, Wang T, et al. Epigenomics and genotype-phenotype association analyses reveal conserved genetic architecture of complex traits in cattle and human. *BMC Biol*. 2020;18(1):80.
- Gorkin DU, Barozzi I, Zhao Y, Zhang Y, Huang H, Lee AY, et al. An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature*. 2020;583(7818):744–51.
- Zhao Y, Hou Y, Xu Y, Luan Y, Zhou H, Qi X, et al. A compendium and comparative epigenomics analysis of cis-regulatory elements in the pig genome. *Nat Commun*. 2021;12(1):2217.
- Bush SJ, Muriuki C, McCulloch MEB, Farquhar IL, Clark EL, Hume DA. Cross-species inference of long non-coding RNAs greatly expands the ruminant transcriptome. *Genet Sel Evol*. 2018;50(1):20.
- Villar D, Berthelot C, Aldridge S, Rayner Tim F, Lukk M, Pignatelli M, et al. Enhancer Evolution across 20 Mammalian Species. *Cell*. 2015;160(3):554–66 SRA, <https://www.ncbi.nlm.nih.gov/bioproject/PRJEB6906>(2015).
- Roller M, Stamper E, Villar D, Izuogu O, Martin F, Redmond AM, et al. LINE retrotransposons characterize mammalian tissue-specific and evolutionarily dynamic regulatory regions. *Genome Biol*. 2021;22(1):62 SRA, <https://www.ncbi.nlm.nih.gov/bioproject/PRJEB28147>(2020). <https://www.ncbi.nlm.nih.gov/bioproject/PRJEB33381>(2019).
- Kern C, Wang Y, Xu X, Pan Z, Halstead M, Chanthavixay G, et al. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nature. Communications*. 2021;12(1):1821 SRA, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA665199>. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA665216>(2020).
- Prowse-Wilkins CP, Wang J, Xiang R, Garner JB, Goddard ME, Chamberlain AJ. Putative causal variants are enriched in annotated functional regions from six bovine tissues. *Front Genet*. 2021;12 SRA, <https://www.ncbi.nlm.nih.gov/bioproject/PRJEB41939>(2020).
- dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PCJ, Yang Z. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc R Soc B*. 2012;279:3491–500.
- Vermunt MW, Tan SC, Castelijns B, Geeven G, Reinink P, de Bruijn E, et al. Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nat Neurosci*. 2016;19(3):494–503.
- Xiao S, Xie D, Cao X, Yu P, Xing X, Chen C-C, et al. Comparative epigenomic annotation of regulatory DNA. *Cell*. 2012;149(6):1381–92.
- Ong C-T, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet*. 2011;12(4):283–93.
- Fang L, Cai W, Liu S, Canela-Xandri O, Gao Y, Jiang J, et al. Comprehensive analyses of 723 transcriptomes enhance genetic and biological interpretations for complex traits in cattle. *Genome Res*. 2020;30(5):790–801 <https://doi.org/10.1101/gr.250704.119>. <http://cattlegeneatlas.roslin.ed.ac.uk/>. Accessed 25 Jul 2021.
- Zhou Y, Liu S, Hu Y, Fang L, Gao Y, Xia H, et al. Comparative whole genome DNA methylation profiling across cattle tissues reveals global and tissue-specific methylation patterns. *BMC Biol*. 2020;18(1):85 SRA, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA612978> (2020).
- Rosenblat M, Gaidukov L, Khersonsky O, Vaya J, Oren R, Tawfik DS, et al. The catalytic histidine dyad of high density lipoprotein-associated serum paraoxonase-1 (PON1) is essential for PON1-mediated inhibition of low

- density lipoprotein oxidation and stimulation of macrophage cholesterol efflux. *J Biol Chem.* 2006;281(11):7657–65.
28. Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, et al. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science.* 2019;364(6446):eaav6202.
  29. Her S, Claycomb R, Tai TC, Wong DL. Regulation of the rat phenylethanolamine N-methyltransferase gene by transcription factors Sp1 and MAZ. *Mol Pharmacol.* 2003;64(5):1180.
  30. Robson-Dixon ND, Garcia-Blanco MA. MAZ elements alter transcription elongation and silencing of the fibroblast growth factor receptor 2 exon IIIb. *J Biol Chem.* 2004;279(28):29075–84.
  31. Zong Y, Panikkar A, Xu J, Antoniou A, Raynaud P, Lemaigre F, et al. Notch signaling controls liver development by regulating biliary differentiation. *Development.* 2009;136(10):1727–39.
  32. Walter TJ, Vanderpool C, Cast AE, Huppert SS. Intrahepatic bile duct regeneration in mice does not require Hnf6 or Notch signaling through Rbpj. *Am J Pathol.* 2014;184(5):1479–88.
  33. Bonzo JA, Ferry CH, Matsubara T, Kim J-H, Gonzalez FJ. Suppression of hepatocyte proliferation by hepatocyte nuclear factor 4a in adult mice. *J Biol Chem.* 2012;287(10):7345–56.
  34. Heddad Masson M, Poisson C, Guérandel A, Mamin A, Philippe J, Gosmain Y. Foxa1 and Foxa2 regulate  $\alpha$ -cell differentiation, glucagon biosynthesis, and secretion. *Endocrinology.* 2014;155(10):3781–92.
  35. Hong F, Pan S, Guo Y, Xu P, Zhai Y. PPARs as nuclear receptors for nutrient and energy metabolism. *Molecules.* 2019;24(14):2545.
  36. Berthelot C, Villar D, Horvath JE, Odom DT, Flicek P. Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nat Ecol Evol.* 2018;2(1):152–63 SRA, <https://www.ncbi.nlm.nih.gov/bioproject/PRJEB13074> (2017).
  37. Jin L, Tang Q, Hu S, Chen Z, Zhou X, Zeng B, et al. A pig BodyMap transcriptome reveals diverse tissue physiologies and evolutionary dynamics of transcription. *Nat Commun.* 2021;12(1):3715.
  38. Danko CG, Choate LA, Marks BA, Rice EJ, Wang Z, Chu T, et al. Dynamic evolution of regulatory element ensembles in primate CD4(+) T cells. *Nat Ecol Evol.* 2018;2(3):537–48.
  39. Huang Y, Li Y, Wang X, Yu J, Cai Y, Zheng Z, et al. An atlas of CNV maps in cattle, goat and sheep. *Sci China Life Sci.* 2021;64(10):1747–64.
  40. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature.* 2006;444(7118):444–54.
  41. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. *Nature.* 2010;464(7289):704–12.
  42. Ota M, Nagafuchi Y, Hatano H, Ishigaki K, Terao C, Takeshima Y, et al. Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell.* 2021;184(11):3006–3021.e3017.
  43. Hu Z-L, Park CA, Reecy JM. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.* 2019;47(D1):D701–10.
  44. Sun H-Z, Zhao K, Zhou M, Chen Y, Guan LL. Landscape of multi-tissue global gene expression reveals the regulatory signatures of feed efficiency in beef cattle. *Bioinformatics.* 2019;35(10):1712–9.
  45. He Y, Hariharan M, Gorkin DU, Dickel DE, Luo C, Castanon RG, et al. Spatiotemporal DNA methylome dynamics of the developing mouse fetus. *Nature.* 2020;583(7818):752–9.
  46. Freebern E, Santos DJA, Fang L, Jiang J, Parker Gaddis KL, Liu GE, et al. GWAS and fine-mapping of livability and six disease traits in Holstein cattle. *BMC Genomics.* 2020;21(1):41.
  47. Kuehn C, Edel C, Weikard R, Thaller G. Dominance and parent-of-origin effects of coding and non-coding alleles at the acylCoA-diaclylglycerol-acyltransferase (DGAT1) gene on milk production traits in German Holstein cows. *BMC Genet.* 2007;8:62.
  48. Schennink A, Stoop WM, Visker MHPW, Heck JML, Bovenhuis H, Van Der Poel JJ, et al. DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. *Anim Genet.* 2007;38(5):467–73.
  49. van Gastelen S, Visker MHPW, Edwards JE, Antunes-Fernandes EC, Hetingtinga KA, Alferink SJJ, et al. Linseed oil and DGAT1 K232A polymorphism: effects on methane emission, energy and nitrogen metabolism, lactation performance, ruminal fermentation, and rumen microbial composition of Holstein-Friesian cows. *J Dairy Sci.* 2017;100(11):8939–57.
  50. Douglas AT, Hill RD. Variation in vertebrate cis-regulatory elements in evolution and disease. *Transcription.* 2014;5(3):e28848.
  51. Gao B. Basic liver immunology. *Cell Mol Immunol.* 2016;13(3):265–6.
  52. Kubes P, Jenne C. Immune Responses in the Liver. *Annu Rev Immunol.* 2018;36(1):247–77.
  53. Zerbino DR, Johnson N, Juetteman T, Sheppard D, Wilder SP, Lavidas I, et al. Ensembl regulation resources. *Database.* 2016;2016:bav119.
  54. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* 2007;35(suppl\_1):D88–92.
  55. Inoue F, Ahituv N. Decoding enhancers using massively parallel reporter assays. *Genomics.* 2015;106(3):159–64.
  56. Fang L, Zhou Y, Liu S, Jiang J, Bickhart DM, Null DJ, et al. Comparative analyses of sperm DNA methylomes among human, mouse and cattle provide insights into epigenomic evolution and complex traits. *Epigenetics.* 2019;14(3):260–76.
  57. Zhou X, Cain CE, Myrthil M, Lewellen N, Michelini K, Davenport ER, et al. Epigenetic modifications are associated with inter-species gene expression variation in primates. *Genome Biol.* 2014;15(12):547.
  58. García-Ruiz A, Cole John B, VanRaden PM, Wiggans George R, Ruiz-López Felipe J, Van Tassell CP. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc Natl Acad Sci.* 2016;113(28):E3995–4004.
  59. Zeng J, de Vlaming R, Wu Y, Robinson MR, Lloyd-Jones LR, Yengo L, et al. Signatures of negative selection in the genetic architecture of human complex traits. *Nat Genet.* 2018;50(5):746–53.
  60. Yang W, Wang S, Looor JJ, Lopes MG, Zhao Y, Ma X, et al. Role of diacylglycerol O-acyltransferase (DGAT) isoforms in bovine hepatic fatty acid metabolism. *J Dairy Sci.* 2022;105(4):3588–600.
  61. Bovenhuis H, Visker MHPW, van Valenberg HJF, Buitenhuis AJ, van Arendonk JAM. Effects of the DGAT1 polymorphism on test-day milk production traits throughout lactation. *J Dairy Sci.* 2015;98(9):6572–82.
  62. Xiang R, Fang L, Liu S, Liu GE, Tenesa A, Gao Y, et al. Genetic score omics regression and multi-trait meta-analysis detect widespread cis-regulatory effects shaping bovine complex traits. *bioRxiv.* 2022. <https://doi.org/10.1101/2022.07.13.499886>.
  63. Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods.* 2009;48(3):240–8.
  64. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv e-prints* 2013;arXiv:1303.3997.
  65. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England).* 2009;25(16):2078–9.
  66. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 2012;22(9):1813–31.
  67. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):R137.
  68. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841–2.
  69. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature.* 2009;457(7231):854–8.
  70. Dailey L. High throughput technologies for the functional discovery of mammalian enhancers: new approaches for understanding transcriptional regulatory network dynamics. *Genomics.* 2015;106(3):151–8.
  71. van der Velde A, Fan K, Tsuji J, Moore JE, Purcaro MJ, Pratt HE, et al. Annotation of chromatin states in 66 complete mouse epigenomes during development. *Commun Biol.* 2021;4(1):239 <https://www.nature.com/articles/s42003-021-01756-4>. <https://www.encodeproject.org/search/?searchTerm=ChromHMM+Zhiping+Weng>. Accessed 27 Oct 2021.
  72. Gel B, Diez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics.* 2015;32(2):289–91.
  73. Herrero J, Muffato M, Beal K, Fitzgerald S, Gordon L, Pignatelli M, et al. Ensembl comparative genomics resources. *Database (Oxford).* 2016;2016:baw053.

74. Paten B, Herrero J, Beal K, Fitzgerald S, Birney E, Enredo and Pecan: Genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res.* 2008;18(11):1814–28.
75. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* (Oxford, England). 2014;30(15):2114–20.
76. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* (Oxford, England). 2013;29(1):15–21.
77. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2013;30(7):923–30.
78. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* (Oxford, England). 2010;26(1):139–40.
79. Li X, Liu Y, Salz T, Hansen KD, Feinberg A. Whole-genome analysis of the methylome and hydroxymethylome in normal and malignant lung and liver. *Genome Res.* 2016;26(12):1730–41 SRA, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA287622> (2015).
80. Grimm SA, Shimbo T, Takaku M, Thomas JW, Auerbach S, Bennett BD, et al. DNA methylation in mice is influenced by genetics as well as sex and life experience. *Nat Commun.* 2019;10(1):305 SRA, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA416505> (2017).
81. Li Y, Fang C, Fu Y, Hu A, Li C, Zou C, et al. A survey of transcriptome complexity in *Sus scrofa* using single-molecule long-read sequencing. *DNA Res.* 2018;25(4):421–37 SRA, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA357500> (2016).
82. Zhou Y, Connor EE, Bickhart DM, Li C, Baldwin RL, Schroeder SG, et al. Comparative whole genome DNA methylation profiling of cattle sperm and somatic tissues reveals striking hypomethylated patterns in sperm. *Gigascience.* 2018;7(5):gy039 RA, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA417285> (2017).
83. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics.* 2011;27(11):1571–2.
84. Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J, et al. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One.* 2013;8(12):e81148.
85. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* 2010;28(5):495–501.
86. Ho D, Imai K, King G, Stuart EA. MatchIt: nonparametric preprocessing for parametric causal inference. *J Stat Softw.* 2011;42(8):1–28.
87. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics.* 2004;20(2):289–90.
88. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, et al. Ensembl 2021. *Nucleic Acids Res.* 2021;49(D1):D884–91.
89. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;38(4):576–89.
90. Turatsinze J-V, Thomas-Chollier M, Defrance M, van Helden J. Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules. *Nat Protoc.* 2008;3(10):1578–88.
91. Xiang R, Ivd B, MacLeod IM, Hayes BJ, Prowse-Wilkins CP, Wang M, et al. Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proc Natl Acad Sci U S A.* 2019;116(39):19398–408.
92. Rohde PD, Demontis D, Cuyabano BCD, The Genomic Medicine for Schizophrenia G, Børglum AD, Sørensen P. Covariance association test (CVAT) identifies genetic markers associated with Schizophrenia in functionally associated biological processes. *Genetics.* 2016;203(4):1901–13.
93. Sørensen IF, Edwards SM, Rohde PD, Sørensen P. Multiple trait covariance association test identifies gene ontology categories associated with chill coma recovery time in *Drosophila melanogaster*. *Sci Rep.* 2017;7(1):2413.
94. Yu Y, Chen S. Dynamic changes of gene expression in the liver during ruminant evolution. *Gene expression omnibus (GEO).* 2022. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE206184>.
95. Yu Y, Chen S. Comparative analyses of DNA methylomes in the liver among three ruminants. *Gene expression omnibus (GEO).* 2022. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE206511>.
96. Yu Y, Chen S. Dynamics of regulatory elements in the liver during ruminant evolution. *Gene expression omnibus (GEO).* 2022. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE206736>.
97. Chen S. Enrichment analysis. Zenodo. 2022. <https://doi.org/10.5281/zenodo.7275349>.
98. Chen S. Regulatory elements of six species. figshare. 2022. <https://doi.org/10.6084/m9.figshare.20402178.v1>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

