Allele-biased expression of the bovine APOB gene associated with the cholesterol deficiency defect suggests cis-regulatory enhancer effects of the LTR retrotransposon insertion

Doreen Becker¹, Rosemarie Weikard¹, Annika Heimes¹, Frieder Hadlich¹, Harald M. Hammon², Marie M. Meyerholz²,³, Wolfram Petz³, Holm Zerbe³, Hans-Joachim Schuberth¹, Martina Hoedemaker⁵, Marion Schmicke⁶, Susanne Engelmann⁷,⁸ & Christa Kühn¹,⁹

The insertion of an endogenous retroviral long terminal repeat (LTR) sequence into the bovine apolipoprotein B (APOB) gene is causal to the inherited genetic defect cholesterol deficiency (CD) observed in neonatal and young calves. Affected calves suffer from developmental abnormalities, symptoms of incurable diarrhoea and often die within weeks to a few months after birth. Neither the detailed effects of the LTR insertion on APOB expression profile nor the specific mode of inheritance nor detailed phenotypic consequences of the mutation are undisputed. In our study, we analysed German Holstein dairy heifers at the peak of hepatic metabolic load and exposed to an additional pathogen challenge for clinical, metabolic and hepatic transcriptome differences between wild type (CDF) and heterozygote carriers of the mutation (CDC). Our data revealed that a divergent allele-biased expression pattern of the APOB gene in heterozygous CDC animals leads to a tenfold higher expression of exons upstream and a decreased expression of exons downstream of the LTR insertion compared to expression levels of CDF animals. This expression pattern could be a result of enhancer activity induced by the LTR insertion, in addition to a previously reported artificial polyadenylation signal. Thus, our data support a regulatory potential of mobile element insertions. With regard to the phenotype generated by the LTR insertion, heterozygote CDC carriers display significantly differential hepatic expression of genes involved in cholesterol biosynthesis and lipid metabolism. Phenotypically, CDC carriers show a significantly affected lipomobilization compared to wild type animals. These results reject a completely recessive mode of inheritance for the CD defect, which should be considered for selection decisions in the affected population. Exemplarily, our results illustrate the regulatory impact of mobile element insertions not only on specific host target gene expression.

¹Institute of Genome Biology, Research Institute for Farm Animal Biology (FBN), Dummerstorf, Germany. ²Institute of Nutritional Physiology, Research Institute for Farm Animal Biology (FBN), Dummerstorf, Germany. ³Clinic for Ruminants with Ambulatory and Herd Health Services, Centre for Clinical Veterinary Medicine, Ludwig-Maximilians-University Munich, Oberschleißheim, Germany. ⁴Institute for Immunology, University of Veterinary Medicine Hannover, Hannover, Germany. ⁵Clinic for Cattle, University of Veterinary Medicine Hannover, Hannover, Germany. ⁶Faculty of Natural Sciences III, Martin-Luther University Halle-Wittenberg, Halle, Germany. ⁷Institute for Microbiology, Technical University Braunschweig, Brunswick, Germany. ⁸Microbial Proteomics, Helmholtz Centre for Infection Research, Brunswick, Germany. ⁹Agricultural and Environmental Faculty, University of Rostock, Rostock, Germany. *email: kuehn@fbn-dummerstorf.de
but also on global transcriptome profiles with subsequent biological, functional and phenotypic consequences in a natural in-vivo model of a non-model mammalian organism.

In 2015, a genetic defect termed Cholesterol Deficiency (CD, https://omia.org/OMIA001965/9913/) was described in the worldwide Holstein Friesian cattle population. CD-affected neonatal calves suffered from insufficient development in combination with symptoms of incurable diarrhoea and generally died within the first 6 months after birth. Concentrations of cholesterol and triglycerides in the blood were markedly reduced in affected individuals, indicating a disorder of lipid metabolism. Pedigree analysis of affected calves initially suggested an autosomal recessive mode of inheritance, and the insertion of a transposable long terminal repeat (LTR) element in exon 5 of the apolipoprotein B (APOB) gene was found as causal mutation. The encoded APOB protein is a crucial key component for the transport of lipid molecules, including cholesterol, through the body and into cells. Initial results indicated that the insertion comprised a 1299 bp solo-long terminal repeat (LTR) insertion, which was predicted to result in a premature stop codon causing a protein that is 97% truncated compared to its wild-type sequence. Causal mutation data were later updated by a report describing that indeed a full-length bovine endogenous retroviral K element (ERV-K) had integrated into the coding part of the gene at position 11:77,891,739. The ERV-K was flanked by two identical full-length LTRs, and the open reading frames within the retroposon genes were altered (loss-of-function). Menzi et al. provided data that indicates a complete lack of transcribed sequence for the APOB gene downstream of the LTR insertion, and Harland reported a polyadenylation signal in the 5’ERV-LTR, which presumably generates a premature poly(A) tail and subsequent termination of transcription. ERV-LTRs are known to exert regulatory activities as reviewed by Chuong et al., and due to their autonomous promoters, they can generate chimeric transcripts and readthrough transcription. However, many studies on the regulatory effects of LTRs are based on bioinformatic prediction from functional genomic studies or on CRISPR-mediated modification of (cancer) cell lines. There are few studies on the effects of LTRs in their natural in-vivo context.

In addition to its molecular genetic background, also pathophysiological consequences of the CD mutation are controversially discussed. Recent data challenge the recessive inheritance mode of CD. Häfliger et al. suggested that the disorder is most likely dominantly inherited with incomplete penetrance in heterozygous CDC carriers (nomenclature defect according to the World Holstein Friesian Federation, http://www.whff.info/documentation/documents/GenericTraitsandCarrierCodes.pdf). But although Gross et al. found differences in cholesterol and lipoprotein concentrations between CDC and homozygous wild type (CDF) animals, CDC and CDF cows did not differ in other metabolic parameters, milk yield or fertility. Furthermore, the question arose, whether cholesterol synthesis and turnover are indeed impaired as a consequence of the CD mutation or whether rather the pathophysiological effect of the mutation is only induced via malabsorption and impaired transportation of lipid components due to a lack of APOB to form lipoprotein complexes.

Thus, to explore the molecular consequences of the LTR insertion on APOB expression and subsequent effects on lipid metabolism, we selected an experimental design, which presented a substantial challenge to APOB regulation and lipid metabolism. The lipoprotein metabolism of cows is particularly critical in the first weeks after parturition due to a severe negative energy balance associated with a substantial lipomobilization, and even more so in situations of pathogen-challenges. Thus, to explore the molecular regulatory consequences of the ERV-LTR insertion on APOB expression and subsequent effects on lipid metabolism, our study monitored hepatic expression pattern and biochemical phenotypes of CDC and CDF early lactation cows after an intramammary challenge with relevant mastitis pathogens (Staphylococcus (S.) aureus or Escherichia (E.) coli).

Material and methods
Selection of animals. The cows investigated in this study were part of a large network project to evaluate genetic predisposition for disease susceptibility in Holstein Friesian cows. Details on the selection and management of the animals have been previously described. The final cohort comprised 35 animals, seven heterozygous CDC carriers and 28 homozygous wild type CDF animals. Homozygous mutant CDS cows were not purchased from dairy cow farms across Germany prior to first calving and were brought to the Clinic for Cattle at the University of Veterinary Medicine Hannover (TiHo) for a highly standardized challenge experiment. Only CDC cows were recruited that did not show manifest clinical symptoms of the CD defect before first calving (e.g., poor development, diarrhoea unresponsive to treatment). For the cow cohort, the experiment was performed under the reference number 33.12-42502-04-15/2024 approved by the Lower Saxony Federal State Office for Consumer Protection and Food Safety. This study was also submitted to and approved by the ethics committees of the Leibniz Institute for Farm Animal Biology and the University of Veterinary Medicine Hanover, respectively. An independent previous calf experiment had been conducted at the Educational and Research Centre for Animal Husbandry, Hoogüt Neumuehle, Germany. It had been permitted by the local department for animal welfare affairs (Landesuntersuchungsamt, Koblenz, Germany) (23 177-07/G 13-20-069). All ethical evaluations were performed as required by the German Animal Care law and associated legislative regulations. The reporting in the manuscript follows the recommendations in the ARRIVE guidelines. This includes a comprehensive description of experimental animals and experimental procedures as provided.

Challenge experiment. The 35 animals were challenged in an intramammary infection model essentially as described in detail by Rohmeier et al. At 36±3 days after parturition, 24 healthy cows (four CDC, 20 CDF) were challenged with 10,000 CFU (colony forming units) S. aureus\textsubscript{5/27} each in both hindquarters of the mam-
mary gland. Eleven healthy cows (four CDC, eight CDF) were challenged with 500 CFU *E. coli*<sub>1003</sub> in one hind-quarter of the mammary gland<sup>22</sup>. A control udder quarter of each cow was infused with sterile sodium chloride solution.

**Clinical examination.** Before and after the experimental challenge, cows were comprehensively monitored for clinical and subclinical diseases as described previously<sup>18–20</sup>. Particular attention was paid to signs of inflammation as well as to biochemical indicators of impaired energy and fat metabolism. To this end, blood serum concentration of beta-hydroxybutyrate (BHB) (mmol/L) and nonesterified fatty acids (NEFA; μmol/L) and blood plasma concentrations of insulin and IGF-1 were determined in weekly intervals starting 3 weeks before parturition until the end of the experiment as described by Meyerholz et al.<sup>19</sup> in 3 days intervals starting 3 weeks before parturition until the end of the experiment. Measuring plasma GH concentration was performed with an enzyme-linked immunosorbent assay<sup>23</sup> with modifications and adaptations according to Kawashima et al.<sup>24</sup> and Meyerholz et al.<sup>25</sup>.

For statistical data analyses, the GLIMMIX procedure of SAS 9.4.1 (SAS Institute Inc., Cary, NC) was applied. Prior to evaluation, we log10-transformed NEFA, BHB, insulin, GH and IGF-1 data. For milk yield, we considered energy-corrected milk as described by Meyerholz et al.<sup>19</sup>. All statistical analyses were performed using SAS 9.4.1. The general linear mixed model implemented in the GLIMMIX procedure included fixed effects for day and group, the interaction of day and group and a random sire effect: $G: y = \beta_0 + \beta_1 \cdot \text{day} + \beta_2 \cdot \text{group} + \beta_3 \cdot \text{day} \times \text{group} + \epsilon$. For the effect of day, we had eight categories for days relative to parturition: $-18 \pm 3; -11 \pm 3; -4 \pm 3; 1; 4 \pm 3; 11 \pm 3; 18 \pm 3; 25 \pm 3; 32 \pm 3$, the group comprised either CDC or CDF. Furthermore, covariance structure comprised the repeated subject cow in compound symmetry. For those data measured before and after calving, separate models were calculated.

At the end of the experiment (*S. aureus*: 96 h after infection; *E. coli*: 24 h after infection), the animals were stunned with a penetrating captive bolt pistol, immediately followed by exsanguination via longitudinal section of the jugular veins and carotid arteries. Liver tissue from the *Lobus caudatus* and mammary gland tissue were immediately shock frozen in liquid nitrogen and subsequently stored at $-80^\circ$C.

**Transcriptome analysis by RNA sequencing.** RNA isolation from liver was conducted as described by Heimes et al.<sup>26</sup>. A second DNase digestion step was added if contamination with genomic DNA was detected by PCR<sup>27</sup>. Repeatedly, RNA concentration and purity were measured on a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA). RNA integrity was quantified on the Bioanalyzer 2100 (Agilent Technologies, Böblingen, Germany).

Finally, a strand-specific RNAseq library preparation protocol with poly(A) selection (TruSeq Stranded mRNA LP, Illumina, San Diego, CA) was applied to the samples. The RNAseq libraries were again monitored for quality on the Bioanalyzer 2100 and were paired-end sequenced on the Illumina HiSeq 2500 system (Illumina, San Diego, CA) for two times 90 base pairs.

For data management, Linux, and R scripts<sup>28</sup> and SAMtools<sup>29</sup> were used. Quality of the raw reads was checked with FastQC version 0.11.5<sup>30</sup> and MultiQC version 1.4<sup>31</sup>. Subsequently, adapters were clipped with Cutadapt version 1.12<sup>32</sup>, and low qualities bases were eliminated with QualityTrim<sup>33</sup>. The reads were aligned to the bovine reference genome ARS1.2 with Ensembl 95 reference annotation with Hisat2 version 2.1.0<sup>34</sup>. The featureCounts option of the subread package version 1.6.2<sup>35</sup> was used for strand-aware read counting. For differential expression analysis in liver samples, we applied DESeq2 version 1.26.0<sup>36</sup> in a model fitting the pathogen used for the PCR27. Repeatedly, RNA concentration and purity were measured on a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA). RNA integrity was quantified on the Bioanalyzer 2100 (Agilent Technologies, Böblingen, Germany).

For statistical data analyses, the GLIMMIX procedure of SAS 9.4.1 (SAS Institute Inc., Cary, NC) was applied. Prior to evaluation, we log10-transformed NEFA, BHB, insulin, GH and IGF-1 data. For milk yield, we considered energy-corrected milk as described by Meyerholz et al.<sup>19</sup>. All statistical analyses were performed using SAS 9.4.1. The general linear mixed model implemented in the GLIMMIX procedure included fixed effects for day and group, the interaction of day and group and a random sire effect: $G: y = \beta_0 + \beta_1 \cdot \text{day} + \beta_2 \cdot \text{group} + \beta_3 \cdot \text{day} \times \text{group} + \epsilon$. For the effect of day, we had eight categories for days relative to parturition: $-18 \pm 3; -11 \pm 3; -4 \pm 3; 1; 4 \pm 3; 11 \pm 3; 18 \pm 3; 25 \pm 3; 32 \pm 3$, the group comprised either CDC or CDF. Furthermore, covariance structure comprised the repeated subject cow in compound symmetry. For those data measured before and after calving, separate models were calculated.

At the end of the experiment (*S. aureus*: 96 h after infection; *E. coli*: 24 h after infection), the animals were stunned with a penetrating captive bolt pistol, immediately followed by exsanguination via longitudinal section of the jugular veins and carotid arteries. Liver tissue from the *Lobus caudatus* and mammary gland tissue were immediately shock frozen in liquid nitrogen and subsequently stored at $-80^\circ$C.

**Transcriptome analysis by RNA sequencing.** RNA isolation from liver was conducted as described by Heimes et al.<sup>26</sup>. A second DNase digestion step was added if contamination with genomic DNA was detected by PCR<sup>27</sup>. Repeatedly, RNA concentration and purity were measured on a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA). RNA integrity was quantified on the Bioanalyzer 2100 (Agilent Technologies, Böblingen, Germany).

Finally, a strand-specific RNAseq library preparation protocol with poly(A) selection (TruSeq Stranded mRNA LP, Illumina, San Diego, CA) was applied to the samples. The RNAseq libraries were again monitored for quality on the Bioanalyzer 2100 and were paired-end sequenced on the Illumina HiSeq 2500 system (Illumina, San Diego, CA) for two times 90 base pairs.

For data management, Linux, and R scripts<sup>28</sup> and SAMtools<sup>29</sup> were used. Quality of the raw reads was checked with FastQC version 0.11.5<sup>30</sup> and MultiQC version 1.4<sup>31</sup>. Subsequently, adapters were clipped with Cutadapt version 1.12<sup>32</sup>, and low qualities bases were eliminated with QualityTrim<sup>33</sup>. The reads were aligned to the bovine reference genome ARS1.2 with Ensembl 95 reference annotation with Hisat2 version 2.1.0<sup>34</sup>. The featureCounts option of the subread package version 1.6.2<sup>35</sup> was used for strand-aware read counting. For differential expression analysis in liver samples, we applied DESeq2 version 1.26.0<sup>36</sup> in a model fitting the pathogen used for the challenge and the carrier status at the CD locus as fixed effects. Finally, lists of differentially expressed genes (DEGs) were used for enrichment analyses in DAVID<sup>37</sup> and Ingenuity Pathway Analysis (IPA, Qiagen, Hilden, Germany) for the identification of significantly enriched GO terms, KEGG and IPA pathways and predicted IPA upstream regulators with a threshold for significance of $p_{adj} < 0.05$.

**APOB expression analysis.** For all individuals, read counts were determined for all APOB gene exons separately as well as combined for exons 1–4 (exons upstream of the LTR insertion site) and exons 5–29 (within and downstream of the LTR insertion site) using the featureCounts option of the subread package (version 1.6.2)<sup>35</sup>. The resulting counts were normalized for transcriptome-wide total read counts for each animal and the length of the respective exons. The resulting normalized read counts were screened for differences between CDC and CDF animals in a linear model using the lm function of the R packages lm4 (v. 1.1-25) and lsmeans (v.2.30-0). The fixed effect of pathogen type used for the challenge (*S. aureus*, *E. coli*) as well as the fixed effect of the APOB genotype (CDC, CDF) were fitted into the model.

To evaluate, whether both APOB haplotypes contributed equally to the increased APOB exon 1–4 expression in CDC animals, or whether there was an allelic imbalance, we took advantage of SNPs in exon 3 at position 77,887,611 and exon 13 at position 77,899,208 on BTA11 (NC_037338.1) identified after visual inspection of aligned reads via the Integrative Genomics Viewer (IGV, v2.7.2). To enable unbiased genotyping at the variant positions, genomic DNA was isolated from liver and mammary gland samples. From those samples, a genomic fragment including the entire exon 3 as well as a genomic fragment spanning exon 13 were amplified by PCR for genotyping the SNPs at positions 77,887,611 and 77,899,208, respectively. PCR amplification from genomic DNA was conducted with primers listed in Supplementary Table 1. The amplified PCR fragments were sequenced with the primers used for PCR amplification in forward and reverse directions. The obtained sequence profiles were aligned to the APOB mRNA (XM_024999521.1) and genomic (NC_037338.1) reference sequences and were visually inspected with Bioedit v7.2.5.0 to obtain the animals’ genotypes for the respective positions.

Furthermore, all 35 animals were genotyped with the Illumina Bovine HD chip (Illumina, San Diego/USA) comprising 777,692 SNPs to establish genotypes for the genomic target region on BTA11 upstream, downstream
and within the APOB locus. SNPs in the target region were filtered for minor allele frequency (MAF > 0.05), deviation from Hardy–Weinberg equilibrium (p > 0.05) and GenTrain Score (GT > 0.7). Due to the fact, that some CDF and CDC heifers had a common sire, which had a confirmed CDC carrier status, we were able to conclude on the haplotype phases of the sire and the inherited haplotypes of the respective offspring within the APOB gene (see Fig. 1). Due to the very small genomic distance between exon 3 and exon 13, we assumed that double recombination events would be extremely unlikely. The haplotype analysis showed that the mutated CD allele was on the same haplotype as the “C” allele at position 77,887,611 in exon 3 and the “T” allele at position 77,899,208 in exon 13.

This information was used for an allele biased differential expression analysis for the positions 11:77,887,611 and 11:77,899,208. To this end, we established the total number of reads spanning position 11:77,887,611, which carried either the non-reference allele “C” or the reference allele “T” by extracting read coverage information from samtools mpileup (version 1.8) with subsequent Linux scripts. An analogous analysis was done for position 11:77,899,208 by calculating the reads covering the non-reference allele “T” or the reference allele “C”. We calculated the ratio of reads with the two alternative alleles (C to T for 77,887,611, T to C for 77,899,208) for both positions. Across all animals with a heterozygous genotype C/T at 77,887,611 or 77,899,208, we tested, if this ratio differed between CDC and CDF animals with a simple linear model fitting the CD genotype as a fixed effect. Furthermore, a t-test conducted in R (version 3.6.2) was applied to test whether the ratio differed significantly from the expected value of 1.

Whole genome resequencing. To obtain further sequence information for the APOB region, genomic DNA from two CDC and CDF animals from the pathogen challenge experiment as well as two CDC calves from a previous independent experiment unrelated for at least three generations were used for whole genome resequencing (Novogene, Beijing, China) after PCR free library preparation on a NovaSeq flowcell with 2 × 150 bp sequencing. The reads obtained were processed according to the protocol for the 1000 bulls genome project. This comprised quality trimming with Trimmomatic (v. Trimmomatic-0.38) and alignment to the bovine genome assembly ARS1.2 with bwa and read sorting via samtools (version 1.8). PCR duplicates were marked and removed via Picard tools (Picard Toolkit. 2018. Broad Institute, GitHub Repository, http://broadinstitute.github.io/picard/; Broad Institute, version). GATK (GenomeAnalysisTK-3.8-1-0-gf15c1c3ef) was used for base recalibration (option BaseRecalibrator) and variant calling (option HaplotypeCaller). The respective pipeline code is provided as Supplementary File Text 1. The resulting genotype files were further processed with PLINK (version 1.9, bcftools (version 1.14) and in-house Linux scripts. In addition, we had available a further Holstein × Charolais F₂ individual with hepatic transcriptome data (SAMEA6031983, PRJEB34570) and whole genome sequence data (1000 bulls genome project, Run 7), which also had a CDF wild type status.

To exclude other variants on the affected haplotype that could be causative for the allele-biased expression of the APOB gene alternative to the LTR retrotransposon insertion, we screened the region 500 kb upstream of the APOB gene LTR insertion. To this end, we filtered SNPs and small indels on BTA11 from position 11:77,385,988 to 77,891,739 bp (NC_037338.1) in the six resequenced animals using bcftools version 1.9. We postulated that all CDF animals should be heterozygous for the variant causal for generating the haplotype-biased gene expression, whereas the variant should be homozygous for the CDF animals of the dataset because there was no indication of an allele-biased expression in CDF animals.
Differentially expressed genes.

A total of 13,213 annotated genes expressed with an FPKM > 1 in at least four samples, which were subjected to subsequent differential expression analysis. The expression analysis revealed a total of 13,213 annotated genes expressed with an FPKM > 1 in at least four samples, which were subjected to subsequent differential expression analysis.

**Results**

**Differential hepatic transcriptome expression in CDC carrier cows.** The hepatic transcriptome analysis by RNAseq generated 3.8 billion reads (0.109 million reads per sample). RNAseq datasets are submitted to the ENA repository (https://www.ebi.ac.uk/ena). The data analysis revealed one gene dosage missing, the gene, which was found to be significantly lower expressed in CDC cows compared to CDF cows in the global transcriptome analysis (Table 1). As expected under the hypothesis of the CD defect allele in CDC animals.

### Top 10 significantly DEGs (sorted by *p*adj, ascending) in the differential expression analysis CDC cows versus CDF cows.

| Gene symbol | Gene name                      | log2FoldChange | *p*adj  |
|-------------|--------------------------------|----------------|---------|
| APOB        | Apolipoprotein                  | −0.95          | 5.29E−09|
| GSTM1, ENSBTAG0000035763 | Glutathione S-Transferase Mu 1 | −1.83          | 5.16E−05|
| CYP3A5      | Cytochrome P450 Family 3 Subfamily A Member 5 | −1.38 | 8.18E−05|
| GSTM1, ENSBTAG0000017765 | Glutathione S-Transferase Mu 1 | −2.29 | 8.18E−05|
| GSTM3       | Glutathione S-Transferase Mu 3 | −1.37          | 9.00E−05|
| SULT2A1     | Sulfotransferase Family 2A Member 1 | −0.61 | 0.0024|
| EBPL         | Emopamil Binding Protein Like   | −1.00          | 0.0027|
| ELOVL6      | ELOVL fatty acid elongase 6    | −1.29          | 0.0041|
| CYP7A1      | Cytochrome P450 Family 7 Subfamily A Member 1 | −3.08 | 0.0047|
| DBI         | Acyl-CoA-binding protein        | −0.90          | 0.0053|

**Pathways and GO terms enriched for differentially expressed genes.** Enrichment analysis with the Database for Annotation, Visualization and Integrated Discovery (DAVID) investigating the collection of genes differentially expressed between CDC and CDF animals (Table 1). All respective genes consistently displayed a lower expression in CDC heterozygote carriers compared to CDF homozygous wild type animals (Supplementary Table 2) showing that the LTR insertion is associated with significantly altered hepatic expression.
animals, also supported evidence for impaired cholesterol synthesis, because the DEGs in this pathway are predominantly linked to cholesterol, steroid and lipid synthesis (e.g. CYP2C19).

The analysis of potential upstream regulators predicted a list of more than 400 potentially regulating molecules (Supplementary Table 5). However, analogously to the pathway enrichment analysis, the most significant potential regulators are metabolites, receptors or transcription regulators known for their role in lipid metabolism.

**APOB differential expression analyses.** To gain further insight into the consequences of structural alterations of the APOB gene in CDC carrier animals, we took a closer look at the respective chromosomal region (BTA11: 77,885,935–77,927,967 bp) with the Integrative Genomics Viewer version 2.7.2 [48]. As shown in Fig. 3, the expression level in the fifth exon of the APOB gene drops abruptly and substantially in CDC cows.

In the liver of the studied cows, the read count level differed substantially between the section of the APOB gene upstream of the LTR insertion site in exon 5 and the region downstream of this site. Across exons 1–4, the expression was more than tenfold higher (FPKM 258.7 (se 15.89)) across CDC heterozygous carriers compared

**Figure 2.** Ingenuity pathway Cholesterol Biosynthesis Superpathway significantly enriched of genes differentially expressed in the hepatic transcriptome between CDC and CDF cows.
to wild-type animals (FPKM 21.4 (se 8.28), \( p = 7.113 \times 10^{-15} \), Supplementary Figure 2). Specifically, the expression for each of the APOB exons upstream of the LTR retrotransposon insertion (exon 1–4) was significantly higher in CDC heterozygous carriers compared to wild-type animals (Fig. 4) including the alternatively spliced exon 2. This also applied to the expression of exon 5, which displayed the same statistically significantly higher expression in CDC compared to CDF animals as seen for exons 1–4 (\( p = 7.70 \times 10^{-13} \), Fig. 4).

An expression pattern of APOB exons 1–5 similar to the one found in the studied cows was analogously observed in an independent study, which monitored the jejunal mucosa transcriptome of young calves39. CDC calves showed a specific drop of read coverage at the insertion point of the LTR retrotransposon (Supplementary Figure 3), and for exons 1–5, the CDC calves had a significantly higher expression than the CDF wild type calves (Supplementary Figure 4).

Compared to the hepatic APOB expression pattern for exons 1–5, however, in the studied cows the situation was reversed for exons 6–29: the CDC animals displayed about only half the expression level of the CDF animals (Fig. 4). Assuming that the APOB gene expression was abolished downstream of the LTR insertion for the affected haplotype, the expression data for this part of the APOB gene would be consistent with the hypothesis that only one active gene copy is present in the CDC animals, or a premature polyadenylation signal was introduced by the ERV-K-LTR sequence insertion, as described by Harland8. The higher expression of the exons 1–4 could have resulted from a general activation of APOB transcription via a regulatory feedback loop in an attempt to compensate for the overall insufficient APOB expression. This would affect the transcription from both APOB gene copies. To evaluate a putative allele-biased expression and, thus, a potential direct regulatory effect of the LTR retrotransposon insertion, we monitored the number of reads covering SNPs in exon 3 (at position 11:77,887,611) upstream of the insertion and in exon 13 (at position 11:77,998,208) downstream of the

Figure 3. Integrative Genomics Viewer screenshot of APOB exon 5 liver expression profiles. At the top an unaffected CDF animal, below four CDC cows (note the different y-axis scales).

Figure 4. Exon-wise differential hepatic expression analysis of the APOB gene between CDC and CDF lactating cows. CDC denotes heterozygous carriers of the CD mutation; CDF animals are homozygous wild type. E-numbers indicate the analysed exons. Blue bars represent the expression levels of CDC animals; yellow bars represent those of CDF individuals. Indicated are lsmeans ± standard error. \( p \) values in italic represent the statistical significance of the differential expression between CDC and CDF. The red box in the gene model indicates the exon 5 affected by the LTR insertion.
insertion. From genotypes in variant positions that are located upstream, within and downstream of the APOB gene, which had been determined from Illumina HD genotype data in a half-sib offspring cohort from a CDC carrier sire, we could conclude that the C allele (alternative APOB allele to the ARS1.2 genome assembly) at position 11:77,887,611 was on the same haplotype as the LTR insertion in exon 5 of the APOB gene in CDC animals. Comparing the ratio of read counts for the C and T allele in heterozygous animals at position 11:77,887,611, we found that the ratio for the CDF animals was close to the expected ratio of 1 ranging from 0.74 to 1.73 (lsmean 0.873, se 2.60, Fig. 5A). In contrast, the heterozygous CDC animals displayed a statistically significant (p = 8.863 e-07) higher ratio ranging from 35.0 to 82.27 (lsmean 53.669, se 5.46, Fig. 5A). This demonstrates a clear allele-biased expression of the APOB exon 3, which is located upstream of the LTR insertion.

We also performed the same analysis, which was conducted for position 11:77,887,611 (SNP in exon 3), also in the genomic region downstream of the LTR insertion for position 11:77,998,208 (SNP in exon 13). It has to be noted that the haplotype carrying the LTR retrotransposon insertion has a T (reference allele in the ARS1.2 reference genome assembly) in position 11:77,998,208. Compared to position 11:77,887,611, the ratio of reads from the two alternative haplotypes was in opposite direction in the heterozygous CDC animals (Fig. 5B). In the CDC animals, for 11:77,998,208, the number of read counts from the LTR carrying haplotype was much lower compared to the read counts from the wild type haplotype. Almost no reads were observed originating from the LTR retrotransposon-carrying haplotype. In contrast, the ratio of read counts for the CDF animals was much higher and closer to the expected ratio of 1. There was a highly significant difference in the ratio of read counts in CDC animals compared to the read count ratio for the two alternative haplotypes in the CDF cows (lsmean 0.0104, se 0.134, for CDC vs. lsmean 1.5498, se 0.069, for CDF, p = 1.583 e-10, Fig. 5B).

Detecting genetic variants in the APOB genomic region. Whole genome resequencing obtained a coverage between 12.2 and 16.8 fold per sample. In the interval 11:77,385,988–77,891,739, we detected 3530 heterozygous or homozygous alternative allele positions for the six samples in our dataset. We wanted to explore if other variants than the LTR retrotransposon insertion could potentially be causal for the observed allele biased expression in exon 1–4 of the APOB gene. Any potential variant would have to be heterozygous in all CDC animals of our study, and the CDF animals, which did not show an allele biased expression, would have to be homozygous. In the genomic region upstream of the APOB transcription start site, no variant fulfilled this requirement until the boundaries of the next upstream neighbouring gene, TDRD15. The closest position to the APOB gene, which was heterozygous in all four CDC cows and homozygous in both CDF animals, was at 11:77,792,966, which is almost 100 kb distant from the APOB gene and upstream of the transcription start site of TDRD15.

Finally, we also inspected loci in the Farm Animal GTex dataset (49, https://cgtex.roslin.ed.ac.uk/downloads), which were described as cis-acting eQTL for the APOB gene. In liver tissue, no variant within 1 Mb upstream or downstream of the APOB gene showed a nominal p value < 10^{-4}, and even those three with a p-value < 10^{-3} were located more than 200 kb downstream to the 3’ end of the APOB gene.

Clinical and physiological effects of the CD carrier status. In the first 35 days of lactation, CDC animals had a significantly lower milk yield compared to CDF animals, but nevertheless showed a stronger mobilization of body mass indicated as loss of body weight after calving (Fig. 6A,B). This increased loss of body weight was not accompanied by a difference in the body condition score (Fig. 6C), which indicates an increased
intraabdominal lipomobilization in CDC cows. For the CDC cows, increased lipomobilization known to be characterized by elevated NEFA levels was confirmed by a significantly higher level of NEFA in blood serum before and after calving (Fig. 6D). But given the decrease in body mass and a lower milk yield, the animals were presumably less capable of utilizing the mobilized resources. Despite the obvious differences in metabolism, there were no significant differences in IGF1 (Fig. 6E), GH and insulin level in blood plasma (Fig. 6F) after calving suggesting that the somatotropic axis of CDC cows was not directly affected by a copy of the LTR retrotransposon insertion.

All three CDC cows challenged intramammary with E. coli developed acute clinical mastitis during the course of the animal experiment, while only four out of eight CDF cows showed a severe clinical response. Although the proportion of acute clinical mastitis cases was nominally higher in CDC compared to CDF animals, this difference was not statistically significant (p = 0.40). When challenged with S. aureus, no difference was seen in the incidence of clinical mastitis between the CDC (one out of four) and CDF group (three out of 20, p = 1).

Discussion

The phenotypic data on the heterozygous carrier CDC animals as well as the results from the transcriptome analysis indicated a phenotypic effect from a single copy of the APOB gene carrying the LTR retrotransposon insertion. Across all exons of the APOB gene, we observed about 50% decreased expression in CDC carriers compared to the CDF homozygous wild type animals. This is in line with results from an abrogated APOB expression in homozygous CDS carriers and would be expected due to the lack of one copy of a fully functional gene. However, our data highlighted a strong discrepancy between the expression level for exons 6–29 downstream of the insertion compared to exons 1–5. The lower expression level for exons 6–29 downstream of the insertion could well be explained by the termination of transcription via a polyadenylation signal in the inserted LTR retrotransposon leading to the premature attachment of a poly(A) tail as described by Harland. The first obvious explanation for the increased expression of exons 1–5 in the APOB gene would be a regulatory compensatory loop for the lack of an intact APOB protein resulting from the impaired expression of exons 6–29, which encode more than 95% of the APOB protein. However, a respective general compensatory mechanism should affect the expression regulation for both APOB copies regardless of the LTR retrotransposon insertion, e.g. via increased transcription factor activity or recruitment in the promoter region of the APOB gene. To evaluate this hypothesis, we tested the expression output from both haplotypes separately. For the allele-biased expression analysis, we took benefit from variant positions in the coding part of the APOB gene as well as SNPs upstream and downstream of the gene body. After establishing the APOB haplotype carrying the LTR retrotransposon insertion, we were subsequently able to determine the haplotype origin of reads spanning variant positions in APOB exons 3 and 13.

The respective data in our study demonstrated that the enhanced expression of APOB exons 1–5 in CDC animals seems to be attributed to the haplotype carrying the APOB LTR retrotransposon insertion, which created a very distinct allelic expression bias. Our data were obtained via preparation of a stranded RNAseq library with

Figure 6. Clinical, biochemical and endocrinological characteristics of CDC cows compared to wild type CDF cows. (A) Energy-corrected milk, (B) body weight relative to weight, (C) body condition score, (D) blood serum concentration for non-esterified fatty acids, (E) blood plasma concentrations of IGF-1, (F) blood plasma concentrations of insulin. Data points indicate mean and standard error of the mean.
poly(A) selection from RNA meticulously cleaned of any DNA contamination. This should avoid the issues with transposable element transcription detection due to pervasive transcription and provide reliable information on structure and source of transcripts, particularly with respect to transposable element molecules.

In summary, the APOB transcription data might be interpreted as a specific cis-regulatory enhancer effect exerted by the LTR retrotransposon insertion in the APOB gene (Fig. 7).

According to Harland, the LTR retrotransposon inserted in exon 5 of the APOB gene belongs to the ERV2-1-BT_LTR group of the bovine endogenous retrovirus (ERV) family K. ERVKs comprise a subset of bovine retroviruses and seem to be currently active in the bovine genome. For the human genome, HERV-K transposable elements have been found to be enriched with proximal and distal enhancer element motifs compared to the whole genome sequence. Transposable elements, and LTR retrotransposons in particular, are known for their regulatory activities as reviewed by Chuong et al. The ERV-K LTR insertion in the bovine APOB gene comprises two full-length LTR flanking regions. Within LTRs, their U3 region is known to contain enhancer elements, which activate and drive viral transcription and can also modulate expression of nearby cellular genes. LTRs are involved, e.g., in a regulatory dysfunction leading to leukaemia and are postulated for being responsible for ectopic CYP19A1 expression associated with the henny feathering phenotype in chicken. Recently, the advances in CRISPR technology-enabled progress in experimental validation of the regulatory function of LTR retrotransposons, which previously relied mainly on, e.g., selection signatures and chromatin features. Deniz et al. demonstrated via CRISPR-mediated targeted transposable element deletion in cancer cells that an LTR located proximal to the APOCI gene has an enhancer function for APOCI expression and also seems to exert a regulatory function on the downstream APOE gene. However, there are still very few studies demonstrating a regulatory function of LTR retrotransposons in vivo, particularly in natural models and non-model organisms. Our data suggest that the bovine APOB gene LTR retrotransposon insertion indeed is an in vivo model for a remarkable enhancer activity of ERV elements. Due to the widespread occurrence of LTR retrotransposon insertions, as reviewed by Thompson et al. and particularly due to the biological activity of the ERV-K type retroviruses, they might be a major source of diversity in gene regulation in the bovine genome.

Based on the gene expression data of our study alone, it could not be formally excluded that other variants on the LTR retrotransposon carrying haplotype could be responsible for the enhancer-like effects in our data set. To explore this option, we analysed transcriptomic and whole genome resequencing data from the neighbourhood of the APOB gene. No gene in the region comprising 2 Mb upstream or downstream of the APOB gene showed a significantly differential expression between CDC and CDF animals, which excludes an element with general enhancer activity in this region on the LTR insertion carrying haplotype. Since 500 kb upstream of the APOB gene, no other variant fulfilled the requirements of (i) heterozygosity in all resequenced CDC animals and (ii) homozygosity of the resequenced CDF animals, we conclude that other elements with specific cis activity for APOB expression regulation on the LTR insertion carrying haplotype are very unlikely.

The APOB gene is predominantly expressed in the liver and small intestine with the alternative predominant isoforms in the small intestine (APOB-48) and liver (APOB-100) as reviewed by Whitfield et al. This is confirmed by whole transcriptome RNAseq data analyses of blood and mammary gland samples from CDC and CDF cows in this study, which did not show indication of APOB gene expression, data not shown. The APOB gene is known for a tissue-specific posttranscriptional mRNA editing in the small intestine, which creates a premature stop codon and is responsible for producing the truncated APOB-48 protein in intestine tissue. The signature of this regulatory editing process was also observed in the jejunal mucosa samples of our study (Supplementary Figure 5) and might have contributed to the divergent APOB expression pattern in the 3’ exons.

**Figure 7.** Model for modulated APOB expression due to ERV-K LTR insertion via combined enhancer activity and premature poly (A) signal insertion. Yellow and blue boxes indicate the expression levels for APOB exons 1–6 originating from the wild type (WT) haplotype or the haplotype with LTR insertion in CDC heterozygous carrier cows and homozygous CDF wild type cows. Red line indicates the promoter region of the APOB gene.
between CDC hepatic and intestinal samples relative to CDF samples (Fig. 4, Supplementary Figure 4). In the liver, the APOB-100 protein plays an essential role in the endocytosis of LDL complexes via a ligand function for the cellular LDL receptor. The smaller APOB-48 protein is required for building chylomicrons in the intestine and consequently for lipid uptake from the diet (reviewed by Whitfield et al.65). Due to its essential role in lipid metabolism, alterations in the APOB gene expression are likely to exert function beyond direct modulation of lipid intake.

Our transcriptomic data from the differential expression analysis confirm that the ERV-K LTR insertion indirectly affects the regulation of cholesterol synthesis in addition to its direct effect on APOB transcription. Earlier studies identified that CDC animals show a decreased plasma cholesterol level11, 12. Gross et al.12 suggested that this results from a lack of APOB protein with the consequence of building less HDL, LDL and/or VLDL complexes containing cholesterol. Thus, the authors postulated that the decreased blood cholesterol levels would not reflect the true cholesterol abundance in the CDC animals. However, hepatic transcriptomes of CDC animals in our study show that important genes encoding enzymes for the final step of cholesterol synthesis are significantly lower expressed compared to CDF wild type animals. Therefore, the decreased plasma cholesterol level in CDC animals carrying the LTR insertion might also result from a decreased cholesterol synthesis in addition to a lack of cholesterol transport capacity. Reduced cholesterol synthesis, as indicated by our data, could explain the detrimental effects observed in heterozygous CDC individuals as described by Häfliger et al.12. In addition, cholesterol is the starting point of estrogen biosynthesis, and our data indicate that also this biological pathway is significantly affected in CDC carriers, which could result in impaired reproduction of CDC carriers.

The phenotypic and the transcriptomic data of our study indicated an effect of a single copy of the LTR retrotransposon carrying the APOB gene. This is in line with reports from Häfliger et al.12 and confirms that the CD defect does not follow a fully recessive mode of inheritance. As indicated by the effects on milk yield, body weight loss, body condition score, and NEFA, the CDC animals mobilize more internal abdominal fat, however, at a decreased performance level. In cattle, excess energy is primarily stored as triglycerides, which are stored in the adipose tissue and are mobilized when the energy intake does not meet the energy demand63, 64. NEFA is one of the most important blood indicators of lipomobilization in ruminants and an indicator of negative energy balance in postpartum dairy cattle15, 65. CDC animals have a significantly higher level of NEFA compared to CDF animals, but a significantly lower milk yield, whereas the BCS change is comparable between both groups. This suggests that the animals were possibly not fully capable to utilize the mobilized lipid resources.

There was no statistically significant difference in the incidence of clinical mastitis in CDC animals compared to CDF animals. However, the number of CDC animals challenged with E. coli (N = 3) or S. aureus (N = 4) was relatively small in comparison to the number of challenged CDF animals (N = 8 and N = 20, respectively), which could reduce the statistical power. To our knowledge, there is no study that compared the incidence of clinical mastitis between CDC and CDF animals, although Cole et al.14 reported a small favourable effect on genetic merit of the CD mutation carrying haplotype for somatic cell score, an indicator trait for mastitis incidence.

There is upcoming evidence for structural variants playing a substantial role in gene expression diversity66. This is in line with the in-vivo experimental data obtained in this study. The gene regulatory effects of the ERV-K LTR beyond the decreased transcript level downstream of its insertion in the APOB gene demonstrate the strong potential of mobile elements to modify gene expression and complex phenotypes. Even though Scott et al.66 found that mobile elements were only weakly enriched for eQTL signals, future investigations of functional diversity within populations should direct specific attention to this class of genetic variants and their potential regulatory effects, because mobile element catalogues and graph-based genome exploration13, 16 are only starting to be available, particularly for non-model organisms.

In conclusion, using transcriptomic and whole-genome sequencing data in combination with clinical data of CDC and CDF cows under pathogen challenge during the early first lactation, we demonstrated a divergent allele-biased expression pattern of the APOB gene in CDC animals. This expression pattern might be a consequence of enhancer activity induced by the LTR insertion in the APOB gene, in addition to an already reported artificial polyadenylation signal within the insertion. Furthermore, the LTR insertion significantly alters the hepatic expression of genes involved in lipid metabolism and cholesterol biosynthesis. CDC animals showed an indication of impaired lipid utilization compared to CDF animals. In addition, the APOB gene LTR insertion might represent an in vivo model demonstrating the regulatory function of a genomic mobile element insertion on gene expression and subsequent consequences.

Data availability
RNA-Seq datasets are publicly available from the ENA data portal (https://www.ebi.ac.uk/ena) (Project Numbers PRJEB33849, PRJEB24380). Whole genome resequencing data have been submitted to the ENA repository (https://www.ebi.ac.uk/ena) under Project Number PRJEB51478.

Received: 1 April 2022; Accepted: 31 July 2022
Published online: 05 August 2022

References
1. Kipp, S. et al. In Proceedings of the 2015 Interbull Meeting Vol. Interbull Bull 49 49–53 (Orlando, Florida, USA, 2015).
2. Saleem, S. et al. Technical note: The role of circulating low-density lipoprotein levels as a phenotypic marker for Holstein cholesterol deficiency in dairy cattle. J. Dairy Sci. 99, 5545–5550. https://doi.org/10.3168/js.2015-10805 (2016).
3. Gross, J. J. et al. Rapid communication: Cholesterol deficiency-associated APOB mutation impacts lipid metabolism in Holstein calves and breeding bulls. J Anim Sci 94, 1761–1766. https://doi.org/10.2527/jas.2016-0439 (2016).
4. Mock, T. et al. Clinicopathological phenotype of autosomal recessive cholesterol deficiency in Holstein cattle. J. Vet. Intern. Med. 30, 1369–1375. https://doi.org/10.1111/jvim.13976 (2016).
5. Kipp, S. et al. Identification of a haplotype associated with cholesterol deficiency and increased juvenile mortality in Holstein cattle. *J. Dairy Sci.* 99, 8913–8931. https://doi.org/10.3168/jds.2016-11118 (2016).

6. Schütz, E. et al. The Holstein Friesian Lethal Haploptote 5 (HH5) Results from a Complete Deletion of TBF1M and Cholesterol Deficiency (CDH) from an ERV-(LTR) Insertion into the Coding Region of APOB. *PLoS ONE* 11, e0154602. https://doi.org/10.1371/journal.pone.0154602 (2016).

7. Menzi, F. et al. A transposable element insertion in APOB causes cholesterol deficiency in Holstein cattle. *Anim. Genet.* 47, 253–257. https://doi.org/10.1111/age.12419 (2016).

8. Harland, C. *Germline Mutations in Bos taurus* PhD thesis, Université de Liège, Liège, Belgique (2018).

9. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* 18, 71–86. https://doi.org/10.1038/nr.2016.139 (2017).

10. Lanciano, S. & Cristofari, G. Measuring and interpreting transposable element expression. *Nat. Rev. Genet.* 21, 721–736. https://doi.org/10.1038/s41598-022-17798-5 (2022).

11. Basiel, B. L., Macrina, A. L. & Dechow, C. D. Cholesterol deficiency carriers have lowered serum cholesterol and perform well at an elite cattle show. *JDS Commum.* 1, 6–9 (2020).

12. Gross, J. J.

13. Kipp, S.

14. Cole, J. B., Null, D. J. & VanRaden, P. M. Phenotypic and genetic effects of recessive haplotypes on yield, longevity, and fertility. *J. Dairy Sci.* 99, 7274–7288. https://doi.org/10.3168/jds.2015-10777 (2016).

15. Contreras, G. A., O’Boyle, N. J., Herdt, T. H. & Sordillo, L. M. Lipomobilization in periparturient dairy cows influences the composition of plasma nonesterified fatty acids and leukocyte phospholipid fatty acids. *J. Dairy Sci.* 93, 2508–2516. https://doi.org/10.3168/jds.2009-2876 (2010).

16. Sordillo, L. M. Nutritional strategies to optimize dairy cattle immunity. *J. Dairy Sci.* 99, 4976–4982. https://doi.org/10.3168/jds.2015-10354 (2016).

17. Heimes, A. et al. Cows selected for divergent mastitis susceptibility display a differential liver transcriptome profile after experimental Staphylococcus aureus mammary gland inoculation. *J. Dairy Sci.* 103, 6364–6373. https://doi.org/10.3168/jds.2019-17612 (2020).

18. Heimes, A. et al. Characterization of functional traits with focus on udder health in heifers with divergent paternally inherited haplotypes on BTA18. *BMC Vet. Res.* 15, 241. https://doi.org/10.1186/s12917-019-1988-4 (2019).

19. Hafliger, I. M.

20. Meyerholz, M. M.

21. Kipp, S.

22. Rohmeier, L.

23. Meyerholz, M. M.

24. Kawashima, C.

25. Weikard, R., Goldammer, T., Eberlein, A. & Kuehn, C. Novel transcripts discovered by mining genomic DNA from defined regions of *Escherichia coli*, but not *Staphylococcus aureus* triggers an early increased expression of factors contributing to the innate immune defense in the udder of the cow. *Vet. Res.* 39, 18. https://doi.org/10.1051/vetres:2007057 (2008).

26. Roh, S. G. et al. Effects of muscarinic and adrenergic blockade on growth hormone secretion induced by growth hormone-releasing peptide-2 (GHRP-2) in ovariectomized ewes. *Endocr. J.* 44, 861–865 (1997).

27. Kawashima, C. et al. Metabolic profiles in ovulatory and anovulatory primiparous dairy cows during the first follicular wave postpartum. *J. Reprod. Dev.* 53, 113–120. https://doi.org/10.1262/jrd.18105 (2007).

28. Meyerholz, M. M. et al. Serum IGFBP4 concentration decreased in dairy heifers towards day 18 of pregnancy. *J. Vet. Sci.* 16, 413–421. https://doi.org/10.4142/jvs.2015.16.4.413 (2015).

29. Heimes, A. et al. Hepatic transcriptome analysis identifies divergent pathogen-specific targeting-strategies to modulate the innate immune system in response to intramammary infection. *Front. Immunol.* 11, 715. https://doi.org/10.3389/fimmu.2020.00715 (2020).

30. Weikard, R., Goldammer, T., Eberlein, A. & Kuehn, C. Novel transcripts discovered by mining genomic DNA from defined regions of bovine chromosome 6. *BMC Genom.* 10, 186. https://doi.org/10.1186/1471-2164-10-186 (2009).

31. Beever, J. E., Fisher, S. R., Guerin, G. & Lewin, H. A. Mapping of eight human chromosome 1 orthologs to cattle Chromosomes 3 and 16. *Mamm Genome* 8, 533–536 (1997).

32. Danecek, P. et al. Twelve years of SAMTools and BCTools. *Gigascience* 10, giab008. https://doi.org/10.1093/gigascience/giab008 (2021).

33. FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

34. MultiQC. http://multiqc.info/.

35. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10. https://doi.org/10.14866/ej.17.1.200 (2011).

36. QualiTrimm. https://bitbucket.org/arobinson/qualitriss.

37. Perta, M., Kim, D., Perta, G. M., Leck, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. https://doi.org/10.1038/nprot.2016.095 (2016).

38. Liao, Y., Smyth, G. K. & Shi, W. The R package Rsnehead is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 47, e47. https://doi.org/10.1093/nar/gkj114 (2019).

39. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 530. https://doi.org/10.1186/s13059-014-0550-8 (2014).

40. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protoc.* 4, 44–57 (2009).

41. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* 27, 2897–2903. https://doi.org/10.1093/bioinformatics/btt509 (2011).

42. Hammon, H. M. et al. Different milk bloters have substantial effects on the jejunal mucosal immune system of pre-weaning calves, as demonstrated by whole transcriptome sequencing. *Sci Rep* 8, 1693. https://doi.org/10.1038/s41598-018-19954-2 (2018).

43. Bouwman, A. C. et al. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat. Genet.* 50, 362–367. https://doi.org/10.1038/s41588-018-0056-5 (2018).

44. Rosen, B. D. et al. De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* 9, giza021. https://doi.org/10.1093/gigascience/giza021 (2020).

45. Li, H. et al. The sequence alignment/map format and SAMTools. *Bioinformatics* 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352 (2009).

46. McKenna, A. et al. The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. https://doi.org/10.1101/gr.107524.110 (2010).

47. Puczell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. https://doi.org/10.1016/S0002-9297(07)00835-8 (2007).

48. Nolte, W. et al. Biological network approach for the identification of regulatory long non-coding RNAs associated with metabolic efficiency in cattle. *Front. Genet.* 10, 1130. https://doi.org/10.3389/fgene.2019.01130 (2019).
46. Daetwyler, H. D. et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* **46**, 858–865 (2014).
47. Howe, K. L. *et al.* *Ensembl* 2021. *Nucleic Acids Res.* **49**, D884–D891. https://doi.org/10.1093/nar/gkaa942 (2021).
48. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26. https://doi.org/10.1038/nbt.1754 (2011).
49. Liu, S. *et al.* A comprehensive catalogue of regulatory variants in the cattle transcriptome. *BioRxiv*, https://doi.org/10.1101/2020.11.12.406280 (2020).
50. Grummer, R. R. Etiology of lipid-related metabolic disorders in periparturient dairy cows. *J. Dairy Sci.* **76**, 3882–3896. https://doi.org/10.3168/dsj.2002-0302(93)77729-2 (1993).
51. Singh, K. *et al.* Differential, tissue-specific transcriptional regulation of apolipoprotein B secretion by transforming growth factor beta. *J. Biol. Chem.* **277**, 39513–39524. https://doi.org/10.1074/jbc.M205513200 (2002).
52. Garcia-Etxebarria, K. & Jugo, B. M. Evolutionary history of bovine endogenous retroviruses in the Bovidae family. *BMC Evol. Biol.* **13**, 256. https://doi.org/10.1186/s12862-013-0256 (2013).
53. Jansz, N. & Faulkner, G. J. Endogenous retroviruses in the origins and treatment of cancer. *Genome Biol.* **22**, 147. https://doi.org/10.1186/s13059-021-02357-4 (2021).
54. Derse, D. & Casey, J. W. Two elements in the bovine leukemia virus long terminal repeat that regulate gene expression. *Science* **231**, 1437–1440. https://doi.org/10.1126/science.2306624 (1986).
55. Khoury, G. & Gruss, P. Enhancer elements. *Cell* **33**, 313–314. https://doi.org/10.1016/0092-8674(83)90410-5 (1983).
56. Deniz, O. *et al.* Endogenous retroviruses are a source of enhancers with oncogenic potential in acute myeloid leukaemia. *Nat. Commun.* **11**, 3506. https://doi.org/10.1038/s41467-020-17206-4 (2020).
57. Li, J. Y. *et al.* Characterization of the endogenous retrovirus insertion in CYP19A1 associated with henny feathering in chicken. *Mol. DNA** **10**, 38. https://doi.org/10.1186/s13100-019-01811-4 (2019).
58. Thompson, P. J., Macfarlan, T. S. & Lorincz, M. C. Long terminal repeats: From parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol. Cell** **62**, 766–776. https://doi.org/10.1016/j.molcel.2016.03.029 (2016).
59. Xue, B., Sechi, L. A. & Kelvin, D. J. Human endogenous retrovirus K (HML-2) in health and disease. *Front. Microbiol.* **11**, 1690. https://doi.org/10.3389/fmicb.2020.01690 (2020).
60. Whitfield, A. J., Barrett, P. H. R., Van Bockxmeer, F. M. & Burnett, J. R. Lipid disorders and mutations in the APOB gene. *Clin. Chem.* **50**, 1725–1732. https://doi.org/10.1373/clinchem.2004.038026 (2004).
61. Brodhagen, J. *et al.* Development and evaluation of a milk protein transcript depletion method for differential transcriptome analysis in mammary gland tissue. *BMC Genom.* **20**, 400. https://doi.org/10.1186/s12864-019-5781-3 (2019).
62. Chen, Z. G., Eggeman, T. L. & Patterson, A. P. ApoB mRNA editing is mediated by a coordinated modulation of multiple apoB mRNA editing enzyme components. *Am. J. Physiol. Gastrointes. Liver Physiol.* **292**, G53–G65. https://doi.org/10.1152/ajpgi.00118.2006 (2007).
63. Scott, A., Chiang, C. & Hall, I. M. Structural variants are a major source of gene expression differences in humans and often affect multiple nearby genes. *Genome Res.* **31**, 2249–2257 (2021).
64. Crysman, D., Leonard, A. S., Fang, Z. H. & Pausch, H. SINE jumping contributes to large-scale polymorphisms in the pig genomes. *Mol. DNA** **12**, 17. https://doi.org/10.1186/s13100-021-00246-y (2021).

**Acknowledgements**

The authors would like to thank the laboratory staff at the FBN Dummerstorf as well as the PhD-students of the ChronMast project at the TiHo Hanover. We also thank the Förderverein Bioökonomieforschung for continuous support.

**Author contributions**

R.W., H.M.H., W.P., H.Z., M.H., H.J.S., M.S., S.E., and C.K. designed research; A.H., M.M.M., D.B. and C.K. performed research; A.H., R.W., D.B. and C.K. analyzed data and wrote the paper. All authors read and approved the final manuscript.

**Funding**

Open Access funding enabled and organized by Projekt DEAL.

**Competing interests**

The authors declare that they do not have any competing financial and/or non-financial interests in relation to the work described. This work is supported by Funds of the German Government’s Special Purpose Fund held at Landwirtschaftliche Rentenbank. The funding body had no role in the design of the study, collection, analysis, and interpretation of data nor in writing the manuscript. Furthermore, the funders had no role in the decision to publish the manuscript.

**Additional information**

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-17798-5.

**Correspondence** and requests for materials should be addressed to C.K.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
