Received: 17 May 2022 | Accepted: 20 July 2022

DOI: 10.1111/age.13246

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A novel missense variant in *endothelin-2* (*EDN2*) causes a growth and respiratory lethal syndrome in bovine

Paulina G. Eusebi | **Óscar Cortés** | **Elisabeth Contreras** | **Javier Cañón** | **Susana Dunner** | **Natalia Sevane**

Departmento de Producción Animal, Facultad de Veterinaria, Universidad Complutense, Madrid, Spain

Correspondence
Natalia Sevane, Departmento de Producción Animal, Facultad de Veterinaria, Universidad Complutense, Madrid, Spain.
Email: nsevane@ucm.es

Abstract

The high level of fragmentation of the Spanish Lidia cattle breed, divided into lineages called 'castas' and into herds within lineages based on reproductive isolation, increases the risk of homozygosity and the outbreak of recessive genetic defects. Since 2004, an increasing number of calves have been identified in a Lidia herd with signs of severe growth retardation, respiratory alterations and juvenile lethality, which constitutes a novel inherited syndrome in cattle and was subsequently termed growth and respiratory lethal syndrome. We performed a genome-wide association study on a cohort of 13 affected calves and 24 putative non-carrier parents, mapping the disease to a wide 6 cM region on bovine chromosome 3 ($p<10^{-7}$). Whole genome re-sequencing of three affected calves and three putative non-carrier parents identified a novel missense variant (c.149G>A|p.Cys50Tyr) in exon 2 of the *endothelin 2* (*EDN2*) gene. Bioinformatic analyses of p.Cys50Tyr effects predicted them to be damaging for both the structure and the function of the edn2 protein, and to create a new site of splicing that may also affect the pattern of pre-mRNA splicing and exon definition. Sanger sequencing of this variant on the rest of the sample set confirmed the segregation pattern obtained with whole genome re-sequencing. The identification of the causative variant and the development of a diagnostic genetic test enable the efficient design of matings to keep the effective population size as high as possible, as well as providing insights into the first *EDN2*-associated hereditary disease in cattle or other species.

KEYWORDS

Bos taurus, juvenile lethality, respiratory alterations, severe growth retardation, whole-genome sequencing

INTRODUCTION

Inherited developmental diseases are less often observed in local bovine breeds, where use of artificial insemination is scarce in comparison with dairy and cosmopolitan beef breeds (Charlier et al., 2008). However, the situation of the Spanish Lidia breed, highly fragmented (Cortés et al., 2014) into lineages called '*castas*' and into herds within lineages based on reproductive isolation, can also lead to a high risk of homozygosity and the outbreak of recessive genetic defects, as reflected by an increase in the inbreeding rate of >1% in all the lineages analysed by Cortés et al. (2014).

In 2004 a new bovine syndrome characterised by growth retardation, respiratory alterations and juvenile lethality broke out in a Lidia cattle herd. These signs, which started at birth, were stated as compatible with syncytial virus infections and the affected calves displayed respiratory difficulties, severe growth retardation, impaired mobility, weakness, prostration, arched posture, dramatic convex frontonasal profile, notorious snoring, swallowing difficulty and dehydration, with a fatal end before 1 year of age in all cases. After analysing the herd book data, we identified the possible carriers, which showed the risks of producing affected offspring with frequencies of 0.05 and 0.16 for sires and dams respectively, and traced them back to a common ancestor born in 1986. Our genetic counselling to the breeder was to limit the use of putative carrier parents (non-affected but producing at least one affected calf) and their close relatives under the initial hypothesis of a hereditary syndrome. As a result, the last affected calves were born in 2017. As far as we know, no compatible inherited disease has been described in the literature nor in OMIA (https://omia.org/) or OMIM (https://www.omim.org).

Towards confirming the hypothesis of a classic Mendelian recessive mode of inheritance for this syndrome, we collected blood samples from 13 affected animals and 24 putative non-carrier parents, and performed a genome-wide SNP scan using the 67 K Affymetrix Bovine Axiom Array, mapping this disorder to a wide 6 cM region on chromosome BTA3. To narrow down the associated region, whole genome re-sequencing (WGS) data were generated from three affected calves and three putative non-carrier parents.

Here we describe the novel hereditary syndrome in cattle termed **G**rowth and **R**espiratory **L**ethal syndrome (GRL), caused by a single mutation in *EDN2* gene. The identification of the causative variant and the development of a diagnostic genetic test have practical implications in the lineage where it arose, and also provide a model for the study of a lethal growth and respiratory syndrome determined by mutations in *EDN2* and never reported in this or other species.

MATERIALS AND METHODS

Animals

Cases were reported and photographed by the breeder of the herd. Necropsies of five affected calves (four females and one male) were performed in 2009 by a veterinary pathologist, determining a bovine syncytial virus as the most likely cause of the observed symptoms. Our genetic advice was requested in 2018, when the breeder suspected some specific sires and dams as the origin of an increased probability of affected offspring and was recommended to avoid any crossbreeding of these individuals. As a result, no more affected calves were born, the last being from 2017. In this context, from the 54 calves affected by the syndrome, only blood samples from 13 were available for study together with 24 putative carrier parents (non-affected but having at least one affected calf) and 24 putative non-carrier parents (non-affected and without any affected calf) (Table S1). Samples were collected and preserved in Magic Buffer® (Biogen, Diagnostica Spain). Genomic DNA was isolated using the E.Z.N.A.® blood DNA mini kit (Omega Bio-Tek) according to the manufacturer's instructions. DNA concentration was measured with Implen NanoPhotometer™ (BioNova). The samples were kept at −20°C until use.

Herd book information

The herd book analysed included 10 442 animals, 4976 males and 5466 females, born between 1935 and 2017. The base population (animals with one or two unknown parents) included 3962 animals. On average, the sires and dams had 57.8 and 5.6 offspring respectively. The Meuwissen and Luo (1992) algorithm, implemented in the PEDIG software (Boichard, 2002), was used to characterise the completeness level of the pedigree from the reference population and to compute the individual inbreeding coefficients of all animals included in the pedigree as the probability of an individual having two alleles identical by descent. This software was also used to calculate both the effective number of founders (Lacy, 1989), defined as the number of equally contributing founders that would be expected to produce the same genetic diversity as that observed in the population under study, computed as the reciprocal of the probability that two alleles drawn at random came from the same founder, and the effective number of ancestors (Boichard et al., 1997), defined as the minimum number of ancestors, founders or not, accounting for the complete genetic diversity of a population, and calculated as the reciprocal of the probability that two alleles drawn at random came from the same ancestor. Finally, generation lengths were computed for the four possible genetic pathways (sire–son,

sire–daughter, dam–son and dam–daughter) as the average age of parents at the birth of their future breeding offspring.

SNP chip analysis

The Affymetrix Bovine Axiom Array (>67 000 markers, assembly UMD 3.1) was used for genotyping 13 affected and 24 putative non-carrier samples. The final number of SNPs after quality control analysis following standard parameters values recommended in the Axiom Analysis Suite software was 55 090. A case–control association test was conducted using the --assoc option implemented in PLINK 1.07 software (Purcell et al., 2007) with the following quality control parameters: --geno 0.10, --maf 0.05, --mind 0.10 and --hwe 1e-6. The Manhattan plot was created with the R software package QQMAN (http://cran.r-project.org/web/packages/qqman/).

Whole-genome resequencing

The WGS data were generated from three affected calves and three putative non-carrier parents using the BGISEQ-500 platform with an average sequence coverage of 13.6× and following a 150 pb paired-end protocol (available at NCBI accession no. PRJNA838078). After sequencing, the raw reads were filtered, removing adaptor sequences, contamination and low-quality reads from raw reads using the SOAPNUKE software developed by BGI Genomics, and deleting the entire reads if more than 25% matched the adapter sequence, more than 50% bases had a quality value lower than 20 or there were more than 3% N in the read. Clean sequence reads were aligned to the *Bos taurus* (assembly ARS-UCD1.2) reference genome using the BWA-MEM algorithm (Li & Durbin, 2010; version 0.7.17) and sorted with SAMTOOLS (Li et al., 2009; version 1.14). Polymerase chain reaction (PCR) duplicates were marked as secondary reads using PICARDTOOLS (http://github.com/broadinstitute/picard; version 2.26.1). The statistics of clean data are shown in Table S2. FREEBAYES (Garrison & Marth, 2012; version 1.3.5) was used for haplotype-based variant detection with the following parameters: --gvcf -g 500 -C 5. Variations with an estimated probability of not being polymorphic less than phred 20 and a coverage well below 10 were removed using vcffilter in VCFLIB (Dawson et al., 2021; version 1.0.2).

Identification of private candidate variants from the WGS data

Private variants were identified using the vcf-contrast tool included in the VCFTOOLS software (Danecek et al., 2011; version 0.1.16). Only those variations homozygous for the alternative allele in the affected samples and homozygous for the reference allele in the putative non-carrier individuals were further investigated. Variants were annotated and their effect predicted with SNPEFF TOOL (Cingolani et al., 2012; version 5.1). The pathogenicity of the *EDN2* c.149G>A (p.Cys50Tyr) variant was evaluated using the web-based bioinformatic prediction tool POLYPHEN-2 with the dataset pairs HumDiv and HumVar (Adzhubei et al., 2010). Their scores range from 0 to 1, and the amino acid substitution is predicted to be damaging if scores are bigger than 0.85. The effect of the p.Cys50Tyr polymorphism on a possible exonic splicing enhancer (ESE) was predicted using the web based tool ESEFINDER 3.0 (Cartegni et al., 2003; Smith et al., 2006).

Sanger sequencing validation

The putative mutation in the bovine *EDN2* gene was validated by Sanger sequencing on the rest of the sample set: 10 affected calves (the other three GRL-affected individuals were WGS), 24 putative carrier parents and 21 putative non-carrier parents (the other three non-carrier parents were WGS) (Table S1). Also, 25 individuals from other cohorts and lineages were genotyped to better determine which dimension this hereditary defect represents in the population. The targeted region was amplified using PCR with DNA AMPLITOOLS HOTSPLIT MASTER MIX (Biotools) and the following primers: foreward, 5′-CCTGAGCACTGAGAATGCCA-3′; reverse, 5′-GTTGTTAGCCCCACACTCCA-3′. The PCR products were purified using the ExoSap-It™ reaction (Affymetrix), and then sequenced using BIGDYE TERMINATOR v3.1 (Applied Biosystems) on an ABI 3500 Genetic Analyser (Life Technologies S.A.).

RESULTS

Phenotype

The first five affected calves appeared in 2004 in a Lidia herd and were reported to be suffering from a novel bovine syndrome exhibiting severe growth retardation, respiratory alterations and juvenile lethality. Various systems were affected, including muscular, skeletal and respiratory abnormalities. At the muscular level, the calves displayed weakness, muscular atrophy, impaired mobility and prostration. Skeletal alterations include arched posture and convex frontonasal profile (Figure 1a,b). The respiratory tract was affected by malformation of the nasal turbinate, which may be the cause of the notorious snoring described in these calves, and trachea, with a progressive reduction of the endotracheal lumen (Figure 1c,d) and the fusion of tracheal cartilage rings (Figure 1e,f). Finally, feeding difficulties and dehydration were also reported in these calves, probably as

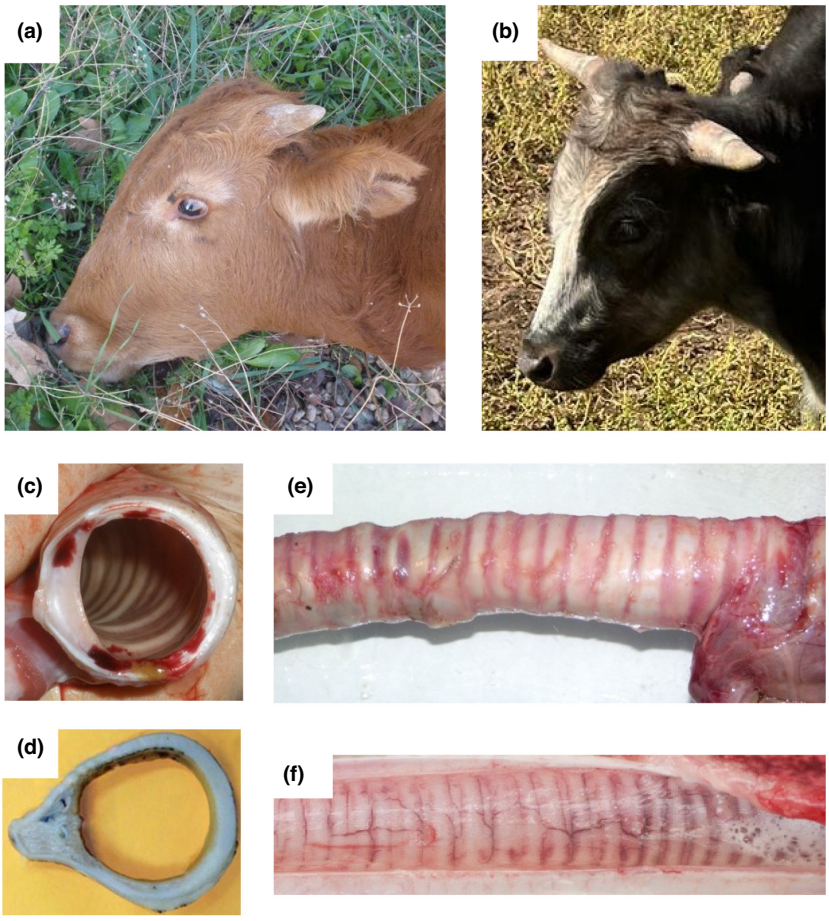

FIGURE 1 Pictures of (a) a convex frontonasal profile of a growth and respiratory lethal syndrome (GRL) affected calf, (b) a normal frontonasal profile of a non-affected calf, (c) the reduction of the endotracheal lumen in a GRL affected calf, (d) a normal bovine tracheal section (Ballarin et al., 2018), (e) the fusion of tracheal cartilage rings in a GRL affected calf, and (f) normal bovine tracheal cartilage rings (source: https://www.agric.wa.gov.au/livestock-biosecurity/ruminant-animal-post-mortem-guide?page=0%2C0)

TABLE 1 Number of affected calves per year since the first cases in 2004

Year	Number of affected calves
2004	5
2005	4
2006	5
2007	3
2008[a]	14
2010	4
2011	1
2014	1
2015	4
2016	4
2017	9

[a]This year the breeder looked for veterinary advice.

consequence of the abnormalities previously described. This GRL syndrome can be diagnosed at birth in most cases, and all affected calves died before 1 year of age. The disease affected both males and females, conceived through natural mounting and born after a normal gestation period. Although the owner of the herd obtained a clinical diagnosis of a viral process, specifically a bovine syncytial virus, this may be a secondary infection, not ruling out the possibility of an underlying genetic aetiology.

Pedigree analysis of the affected individuals

To better understand the genetic inheritance of this syndrome in the affected animals, the pedigree data of the herd were analysed. The effective number of founders (Lacy, 1989) and ancestors (Boichard et al., 1997) for the affected animals considered as reference population (n = 54) were 47 and 15 respectively. In this cohort, the maximum number of generations traced was 11.8, being 5.2 the equivalent complete generations; the percentage of known ancestors decreased from 100% in the first generation, to 95.4% in the second, and to 89.6% in the third; and the average generation length computed for the four possible genetic pathways was 7.6 years.

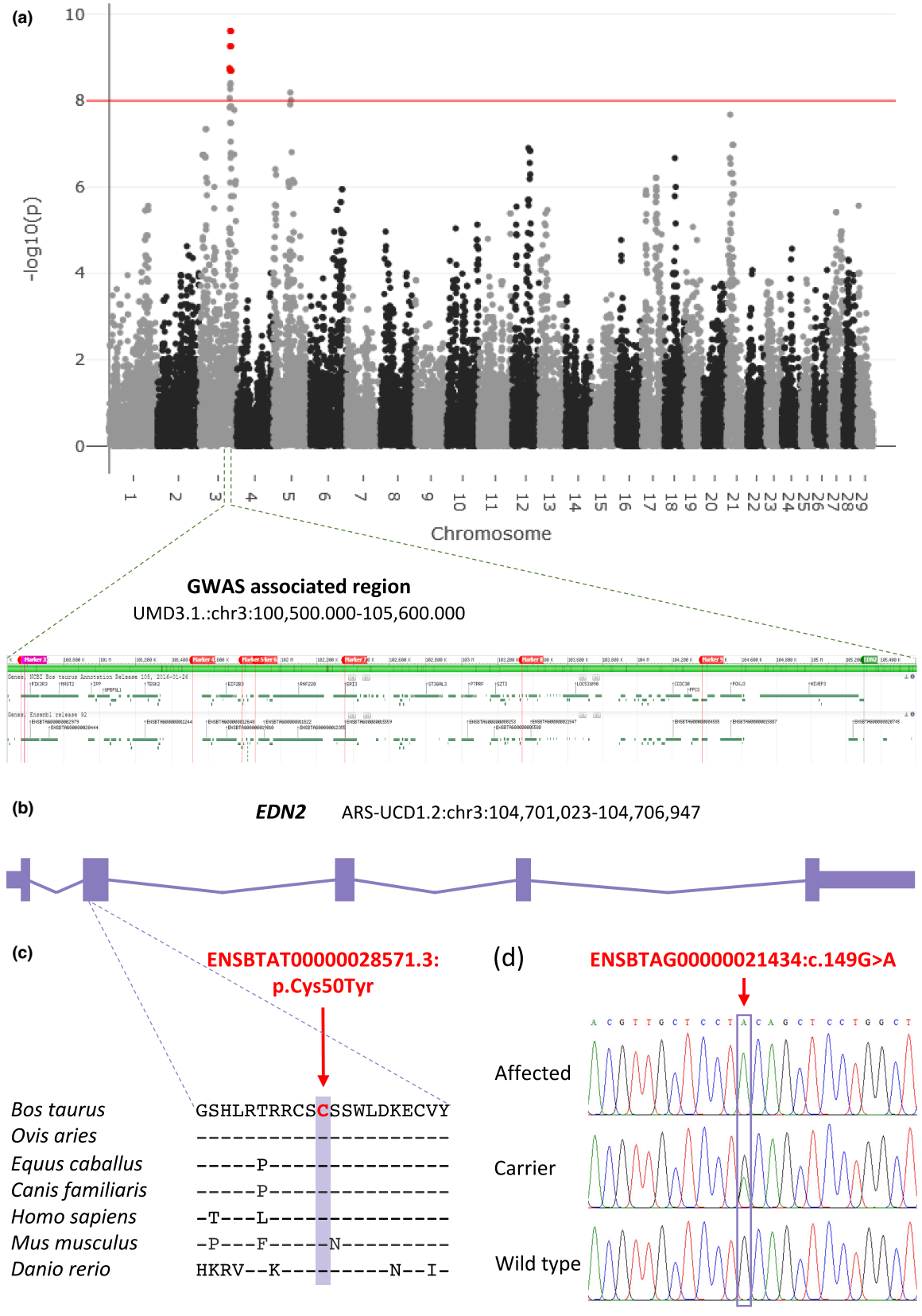

FIGURE 2 Genome-wide association study (GWAS) and whole-genome sequencing (WGS) results. (a) Manhattan plot of GWAS results. The *x*-axis represents the chromosomal position and *y*-axis shows −log10 *p*-values. The red line indicates suggestive *p*-values of 10^{-8} and the red dots show the nine SNPs with lower *p*-values after Bonferroni correction. The GWAS GRL-associated region in the cow genome assembly UMD 3.1 is detailed, including the most significant SNPs and genes. (b) Schematic representation of the bovine *EDN2* gene (chromosome position in the ARS-UCD1.2 assembly). (c) Multispecies protein alignment of the bovine missense variant ENSBTAT00000028571.3:P.Cys50Tyr in exon 2 (for a more detailed multispecies comparison, see Figure S3). (d) Sanger sequencing showing the wild-type, carrier and affected genotypes for the *EDN2* ENSBTAG00000021434:C.149G>A variant

TABLE 2 Private variants present in a homozygous state for the alternative allele in all three affected calves and absent in three putative non-carrier parents causing an amino-acid substitution or included in a splicing region

Chromosome position	Reference allele	Alternative allele	Gene	Effect	Public databases
3:102575190	G	A	*ENSBTAG00000005583*	Splice region, synonymous (c.588G>A|p.Ala196Ala)	Published
3:104701617	G	A	*EDN2*	Missense (c.149G>A|p.Cys50Tyr)	Not published
11:103753484	C	T	*CCDC187*	Missense, transcript no start codon (c.2257G>A|p.Glu753Lys)	Published
11:103757576	CGGCT	TGGCT	*CCDC187*	Missense, transcript no start codon (c.1915G>A|p.Val639Ile)	Published
18:34295184	C	A	*ENSBTAG00000054627*	Missense (c.79C>A|p.Arg27Ser)	Published
18:36473688	T	C	*PDF*	Missense (c.128A>G|p.Gln43Arg)	Published
X:49829094	T	C	*ENSBTAG00000048265*	Missense (c.698A>G|p.Gln233Arg)	Published
X:51060509	C	T	*CENPI*	Missense (c.143G>A|p.Ser48Asn)	Published
X:116114963	G	A	*ENSBTAG00000052068*	Missense (c.1343C>T|p.Ser448Leu)	Published
X:121187371	C	T	*PHEX*	Splice region, intron (c.1404+7G>A)	Published

TABLE 3 Samples included in the study and genotypes for the *EDN2* c.149G>A (ENSBTAT00000028571.3:P.Cys50Tyr) variants

	Affected	Putative carrier parents	Putative non-carrier parents	Individuals from other cohorts and lineages
n	13	24	24	25
Genotype c.149G>A	AA	GA	GG	GG

A total of 54 animals affected by this syndrome were registered in the pedigree (Table 1). The average coefficient of inbreeding of all 54 GRL affected calves was 0.049 vs. a mean inbreeding value of 0.040 for the rest of animals born in the same cohort. We identified multiple loops in the pedigree with common ancestors in the paternal and maternal lineages, tracing the probable variant back to 1986 to the bull 86053 (Figure S1), which left 90 descendants in the herd book. The 18 years between the birth of 86053 (1986) and the first GRL-affected animals (2004) represents an interval of approximately 2.4 generations (average generation interval length is 7.6 years).

Taking all these data into account, the hypothesis of a variant that severely affects the function of a major gene following a classic Mendelian recessive mode of inheritance was investigated.

GWAS using the SNP chip data

To identify the genomic region associated with GRL syndrome, we conducted a genome-wide association study (GWAS) using the SNP chip Affymetrix Bovine Axiom Array (67 K) on 13 affected and 24 putative non-carrier samples. After quality control, genotypes for 55090 SNPs were used to perform a case–control association test. The most significant SNPs peaked around 102 Mbp on chromosome 3 of the UMD 3.1 assembly (p-value less than -10^{-7}; Table S3; Figure 2a).

Identification of private candidate variants from WGS data

To identify the underlying variant responsible for GRL and given that DNA samples were not available from any of the five affected animals from 2004, we sequenced the whole genome of three affected calves born in 2008 that were also related to the common ancestor 86053, suspected of being responsible for the distribution of the variant in the population (Figure S2), as well as three putative non-carrier parents. Based on the pedigree data, we considered an autosomal recessive mode of inheritance when analysing the WGS data. Filtering for private variants present in a homozygous state for the alternate allele in all three affected calves and absent in three putative non-carrier parents revealed 3958 variants shared across all cases (Table S4). Annotation of these variants allowed further filtering for their predicted effect, focusing on those causing an amino-acid substitution or included in a canonical splicing region (Tables 2 and S4). Among these, only one missense variant (ENSBTAG00000021434:c.149G>A|ENSBTAT00000028571.3:p.Cys50Tyr) in exon 2 of the *endothelin 2* (*EDN2*) gene, located on chromosome 3, was not previously described in any public database (Table 2).

The detected variant (ENSBTAG00000021434:c.149G>A; ARS-UCD1.2:chr3:104701617) alters the encoded amino acid residue 50 of *EDN2*, replacing a cysteine (Cys) by a tyrosine (Tyr) (Figure 2b–d). Although

many missense variants have been reported for this gene in different species, as far as we know, no substitution of this particular cysteine has been previously described, including the human gnomAD database (Karczewski et al., 2020) or livestock projects such as the 1000 Bull Genomes or NextGen, among others (Figures 2c and S1). Bioinformatic prediction using the POLYPHEN2 web-tool resulted in HumDiv and HumVar scores of 1, the maximum in the range supporting pathogenicity (Figure S4). The effect of this variant on *EDN2* splicing was investigated using the ESEFINDER webtool, which predicted a new exonic splice site not present in the wild-type DNA (CT**G**CAGC→CT**A**CAGC, SRp40 score = 4.8, SRp40 threshold = 2.67).

Sanger sequencing validation of the *EDN2* c.149G>A|p.Cys50Tyr mutation on the rest of the sample set confirmed the segregation pattern obtained with WGS, with all GRL affected calves displaying the AA mutated genotype, all putative carrier parents the GA genotype and all putative non-carrier parents and the individuals from other cohorts and lineages the GG wild-type genotype (Tables 3 and S1).

DISCUSSION

Here, we report a new hereditary syndrome in bovines, within a Lidia herd, and identify its causal variant in the *EDN2* gene. Its complex phenotype involves severe growth retardation, respiratory alterations and juvenile mortality. As far as we know, this bovine disease has never been described in any species and was termed the growth and respiratory lethal syndrome. Although most respiratory diseases in bovine are produced by virus infections that often impair resistance to bacterial infections in the lower respiratory tract (N'jai et al., 2013), the possibility of an underlying genetic aetiology was supported by pedigree data pointing towards a Mendelian autosomal recessive mode of inheritance. The analysis of the deep pedigree available for this herd allowed the identification of the common ancestor of GRL-affected calves, which was traced back to a male born in 1986 that left 90 descendants in the herd book. The 18 years between the birth of this individual (1986) and the first GRL-affected animals (2004) represents an interval of approximately 2.4 generations (average generation interval length is 7.6 years), this relatively long period until the appearance of the first affected animal probably being a consequence of the relatively low level of inbreeding (0.04). The pedigree data also facilitated an efficient design for matings that prevented the birth of new GRL-affected calves. However, a more accurate tool was needed to keep the effective population size as high as possible, which is especially important in this highly fragmented breed in which the herds (*encastes*) are genetically close and reach high levels of inbreeding.

Based on the hypothesis of an autosomal recessive causal variant, we first performed a GWAS with the 67 K Affymetrix Bovine Axiom Array that mapped the disease to a wide 6 cM region on chromosome 3, not finding evidence for any strong candidate gene. This approach has proved useful in identifying genes associated with economically important traits or diseases in livestock species (Sharma et al., 2015). However, the markers included in the arrays are typically linked to the disease or trait of interest rather than being their direct cause. Even if the causal SNP is included in the array, there is a reasonable likelihood of not detecting even high statistical associations (MacArthur et al., 2014). In such situations, fine mapping methodologies seek to identify the real, not the linked, genetic variant/s responsible for the disease or trait of interest. Therefore, as the cost of high-throughput sequencing technologies has decreased drastically in later years, WGS data of affected and unaffected samples have been obtained to narrow down the associated region. The genomic analysis of private variants in affected calves ultimately revealed the novel missense mutation p.Cys50Tyr in exon 2 of the *EDN2* gene, with a predicted severe impact on protein functionality (Abramowicz & Gos, 2018; Adzhubei et al., 2010; Cartegni et al., 2003; Castro-Chavez, 2010).

EDN2 is a transcriptionally regulated 177 amino acid peptide implicated in vascular homeostasis, but also in female reproduction (Ko et al., 2012), gastrointestinal function (Chang et al., 2013; Takizawa et al., 2005), immune cell function (Grimshaw et al., 2002), heart failure (Brown et al., 2000) and various cancers (Ling et al., 2013). In bovines, *EDN2* expression has been found to be affected in lung epithelial cells during respiratory diseases, specifically in paratuberculosis infection (Çınar et al., 2020; Kiser et al., 2021), and in bronchial epithelial cells after exposure to viral (Herpes virus) and bacterial (*Mannheimia haemolytica*) pathogens (N'jai et al., 2013). Thus, the lesions compatible with a respiratory viral process found in the five necropsied calves could have been facilitated by the alteration of the expression of *EDN2* triggering bovine GRL.

Global loss of edn2 in *EDN2*-null mice has been described as lethal, wherein mice expire around 3 weeks of age from internal starvation, hypothermia and emphysema (Chang et al., 2013), while also displaying ovulatory defects (Cacioppo et al., 2014). These *EDN2*-null mice also exhibited defective energy homeostasis, lung morphology and function alterations, and severe growth retardation (Chang et al., 2013). Starvation is thought to be the primary abnormality caused by the absence of functional edn2, as the animals had normal milk consumption, but blood chemistry indicated that the constitutive *EDN2*-null mice suffered from apparent internal starvation (Chang et al., 2013). Given that nutritional availability is an important determinant of normal lung development, chronic nutrient restriction may induce

structural and functional alterations in the lung (Maritz et al., 2005).

Although the necropsies of five affected calves performed in 2009 focused on respiratory lesions and led to the hypothesis of a bovine syncytial virus as the most plausible cause of death, the muscular, skeletal and respiratory abnormalities exhibited by GRL calves might be secondary effects of edn2 protein deficit, which have been demonstrated to be essential for nutrient availability and energy homeostasis in *EDN2*-null mice. In both cases, the functionality of edn2 is seriously impaired, either by the lack of a portion of exon 1 and all of exon 2 in *EDN2*-null mice or by a missense mutation in bovine exon 2 affecting a cysteine residue (Cys50Tyr) that is highly conserved among multiple animal species (Figure S3) and predicted to be damaging for the structure and function of the protein, as well as to create a new site of splicing also with a harmful effect in GRL calves.

The residue Cys is of special importance in enzyme structure and may be a pathological variant causing disease when it is replaced by any other amino acid (Castro-Chavez, 2010). This agrees with the POLYPHEN-2 results, which predicted a severe impact of the Cys50Tyr variant on the structure and function of edn2 protein based on several features comprising the sequence, and the phylogenetic and structural information characterising this substitution. On the other hand, although the *EDN2* Cys50Tyr variant is in the middle of exon 2, and therefore not affecting canonical splicing sites (located on exon–intron boundaries) or activating cryptic exons (deep intronic mutations), changes in exonic sequences may also affect the pattern of pre-mRNA splicing and exon definition, in this case introducing a new splice site (Abramowicz & Gos, 2018). The bioinformatic analysis performed with ESEFINDER detected a new putative ESE responsive to the human Ser/Arg-rich (SR) protein SRp40, which belongs to a family of conserved splicing factors and promotes exon definition (Cartegni et al., 2003; Graveley, 2000). Further research is needed to disentangle if this amino acid substitution alone or its possible effect on splicing is the cause of the GRL syndrome.

The validation of the p.Cys50Tyr variant on different groups of non-carriers, carriers and affected animals proved that the heterozygous form does not cause any clinically detectable symptoms in cattle, which is in agreement with what was described in mouse. Although it would have been interesting to investigate the mutated *EDN2* transcript and protein, as well as the functional consequences of GRL on physiological parameters and histology, the handling of matings to avoid the birth of affected calves since 2018 prevented us from obtaining these samples.

In conclusion, we reported a new hereditary syndrome in bovine, affecting a Lidia herd, and identified its causal variant in the *EDN2* gene. We showed that the GRL syndrome in cattle shares strikingly similar pathological signs with *EDN2*-null mice. *Endothelin-2* is implicated in many physiological processes, including vascular homeostasis, female reproduction, gastrointestinal, immune cell and heart function, and various cancers, and is also expressed in lung and bronchial epithelial cells. However, studies in *EDN2*-null mice point towards internal starvation as the primary abnormality caused by the absence of functional edn2. Bioinformatic analyses of p.Cys50Tyr effects predicted it to be damaging for both the structure and the function of the protein, as well as to create a new site of splicing also with harmful effects, which together with the validation of its segregation pattern on different groups of non-carriers, carriers and affected animals, as well as the functions of the edn2 protein and the lethal effects of its loss in *EDN2*-null mice, makes this mutation the most plausible and unique cause underlying the development of the GRL syndrome in cattle. The identification of the causative variant and the development of a diagnostic genetic test have practical implications in the lineage where it arose, enabling the efficient design of matings to keep the effective population size as high as possible, but also provide a model for the study of a lethal growth and respiratory syndrome produced by the *EDN2* gene, which has never been reported in this or other species.

ACKNOWLEDGEMENTS

We thank the breeder for providing the blood samples, the pedigree records and the photographs of the necropsy for this study, as well as the Spin-off VELOGEN for funding this research.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY STATEMENT

The SNP array data of 13 GRL-affected animals and 24 putative non-carrier parents are available from the figshare repositories: https://doi.org/10.6084/m9.figshare.19746193.v2 and https://doi.org/10.6084/m9.figshare.19746190.v2. Genomes sequenced for this work are available in NCBI (Bioproject number PRJNA838078). All other data are contained within the article and its supplementary information.

ETHICAL APPROVAL

Approval by an ethics committee for the use of the blood samples was not necessary. We verify that the samples were taken during clinical examinations when the animal was subjected to sampling for medical purposes and the study uses these samples with no supplementary pain for the animal.

ORCID

Paulina G. Eusebi https://orcid.org/0000-0002-0682-038X
Óscar Cortés https://orcid.org/0000-0001-7685-3980
Javier Cañón https://orcid.org/0000-0003-1137-0958
Susana Dunner https://orcid.org/0000-0001-8637-2208
Natalia Sevane https://orcid.org/0000-0003-4766-6291

REFERENCES

Abramowicz, A. & Gos, M. (2018) Splicing mutations in human genetic disorders: examples, detection, and confirmation. *Journal of Applied Genetics*, 59, 253–268.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P. et al. (2010) A method and server for predicting damaging missense mutations. *Nature Methods*, 7, 248–249.

Ballarin, C., Bagnoli, P., Peruffo, A. & Cozzi, B. (2018) Vascularization of the trachea in the bottlenose dolphin: comparison with bovine and evidence for evolutionary adaptations to diving. *Royal Society Open Science*, 5, 171645.

Boichard, D. (2002) PEDIG: a Fortran package for pedigree analysis suited to large populations. In *Proc. 7th world Congr. Genet. Appl. Livest. Prod.* Montpellier, France [CD-ROM]. Paper 28–13.

Boichard, D., Maignel, L. & Verrier, E. (1997) The value of using probabilities of gene origin to measure genetic variability in a population. *Genetics, Selection, Evolution*, 29, 5–23.

Brown, M.J., Sharma, P. & Stevens, P.A. (2000) Association between diastolic blood pressure and variants of the endothelin-1 and endothelin-2 genes. *Journal of Cardiovascular Pharmacology*, 35, S41–S43.

Cacioppo, J.A., Oh, S.W., Kim, H.Y., Cho, J., Lin, P.C., Yanagisawa, M. et al. (2014) Loss of function of endothelin-2 leads to reduced ovulation and CL formation. *PLoS One*, 9, e96115.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q. & Krainer, A.R. (2003) ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Research*, 31, 3568–3571.

Castro-Chavez, F. (2010) The rules of variation: amino acid exchange according to the rotating circular genetic code. *Journal of Theoretical Biology*, 264, 711–721.

Chang, I., Bramall, A.N., Baynash, A.G., Rattner, A., Rakheja, D., Post, M. et al. (2013) Endothelin-2 deficiency causes growth retardation, hypothermia, and emphysema in mice. *Journal of Clinical Investigation*, 123, 2643–2653.

Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J.S., Cambisano, N. et al. (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nature Genetics*, 40, 449–454.

Çınar, M.U., Akyüz, B., Arslan, K., White, S.N., Neibergs, H.L. & Gümüşsoy, K.S. (2020) The EDN2 rs110287192 gene polymorphism is associated with paratuberculosis susceptibility in multibreed cattle population. *PLoS One*, 15, e0238631.

Cingolani, P., Platts, A., Wang, l., Coon, M., Nguyen, T., Wang, L. et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6, 80–92.

Cortés, O., Sevane, N., Baro, J.A. & Cañón, J. (2014) Pedigree analysis of a highly fragmented population, the Lidia cattle breed. *Livestock Science*, 167, 1–8.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A. et al. (2011) The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.

Dawson, E.T., Pedersen, B.S., Garrison, E., Kronenberg, Z.N. & Prins, P. (2021) Vcflib and tools for processing the VCF variant call format. https://doi.org/10.1101/2021.05.21.445151

Garrison, E. & Marth, G. (2012) Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN].

Graveley, B.R. (2000) Sorting out the complexity of SR protein functions. *RNA*, 6, 1197–1211.

Grimshaw, M.J., Wilson, J.L. & Balkwill, F.R. (2002) Endothelin-2 is a macrophage chemoattractant: implications for macrophage distribution in tumors. *European Journal of Immunology*, 32, 2393–2400.

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q. et al. (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581, 434–443.

Kiser, J.N., Wang, Z., Zanella, R., Scraggs, E., Neupane, M., Cantrell, B. et al. (2021) Functional variants surrounding endothelin 2 are associated with *Mycobacterium avium* subspecies paratuberculosis infection. *Frontiers in Veterinary Science*, 8, 625323.

Ko, C., Meidan, R. & Bridges, P.J. (2012) Why two endothelins and two receptors for ovulation and luteal regulation? *Life Sciences*, 91, 501–506.

Lacy, R.C. (1989) Analysis of founder representation in pedigrees: founder equivalent and founder genome equivalents. *Zoo Biology*, 8, 111–124.

Li, H. & Durbin, R. (2010) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 26, 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.

Ling, L., Maguire, J.J. & Davenport, A.P. (2013) Endothelin-2, the forgotten isoform: emerging role in the cardiovascular system, ovarian development, immunology and cancer. *British Journal of Pharmacology*, 168, 283–295.

MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R. et al. (2014) Guidelines for investigating causality of sequence variants in human disease. *Nature*, 508, 469–476.

Maritz, G.S., Morley, C.J. & Harding, R. (2005) Early developmental origins of impaired lung structure and function. *Early Human Development*, 81, 763–771.

Meuwissen, T.I. & Luo, Z. (1992) Computing inbreeding coefficients in large populations. *Genetics, Selection, Evolution*, 24, 305–313.

N'jai, U.A., Rivera, J., Atapattu, D.N., Owusu-Ofori, K. & Czuprynski, C.K. (2013) Gene expression profiling of bovine bronchial epithelial cells exposed in vitro to bovine herpesvirus 1 and Mannheimia haemolytica. *Veterinary Immunology and Immunopathology*, 155, 182–189.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W, Daly M.J. & Sham P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81, 559–575.

Sharma, A., Lee, J.S., Dang, C.G., Sudrajad, P., Kim, H.C., Yeon, S.H. et al. (2015) Stories and challenges of genome wide association studies in livestock – a review. *Asian-Australasian Journal of Animal Sciences*, 28, 1371–1379.

Smith, P.J., Zhang, C., Wang, J., Chew, S.L., Zhang, M.Q. & Krainer, A.R. (2006) An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Human Molecular Genetics*, 15, 2490–2508.

Takizawa, S., Uchide, T., Adur, J., Kozakai, T., Kotake-Nara, E., Quan, J. et al. (2005) Differential expression of endothelin-2 along the mouse intestinal tract. *Journal of Molecular Endocrinology*, 35, 201–209.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Eusebi, P.G., Cortés, Ó., Contreras, E., Cañón, J., Dunner, S. & Sevane, N. (2022) A novel missense variant in *endothelin-2* (*EDN2*) causes a growth and respiratory lethal syndrome in bovine. *Animal Genetics*, 00, 1–9. Available from: https://doi.org/10.1111/age.13246

Table S1. Samples included in the study and genotypes for the *EDN2* c.149G>A (ENSBTAT00000028571.3:p.Cys50Tyr) mutation. In bold, whole-genome sequenced (WGS) samples.

Group	ID	Gender	Year of birth	EDN2 c.149G>A (p.Cys50Tyr) genotype
Affected calves	08310	female	2008	AA
	08312	female	2008	AA
	08314	female	2008	AA
	08334	female	2008	AA
	08352	female	2008	AA
	08363	female	2008	AA
	08003	male	2008	AA
	08012	male	2008	AA
	08040	male	2008	AA
	08052	male	2008	AA
	08057	male	2008	AA
	08070	male	2008	AA
	10491	female	2010	AA
Putative carrier parents	97324	female	1997	GA
	00547	female	2000	GA
	00600	female	2000	GA
	01614	female	2001	GA
	01646	female	2001	GA
	02725	female	2002	GA
	02796	female	2002	GA
	04965	female	2004	GA
	03028	male	2003	GA
	05022	female	2005	GA
	05024	female	2005	GA
	06105	female	2006	GA
	06142	female	2006	GA
	08306	female	2008	GA
	08319	female	2008	GA
	08370	female	2008	GA
	08037	male	2008	GA
	09387	female	2009	GA
	09391	female	2009	GA
	09446	female	2009	GA
	10476	female	2010	GA
	10505	female	2010	GA

	10511	female	2010	GA
	10551	female	2010	GA
Putative non-carrier parents	94159	female	1994	GG
	95197	female	1995	GG
	96261	female	1996	GG
	96309	female	1996	GG
	01691	female	2001	GG
	02015	male	2002	GG
	02794	female	2002	GG
	02810	female	2002	GG
	06046	male	2006	GG
	08030	male	2008	GG
	08045	male	2008	GG
	08055	male	2008	GG
	08064	male	2008	GG
	08315	female	2008	GG
	08346	female	2008	GG
	09003	male	2009	GG
	09026	male	2009	GG
	09027	male	2009	GG
	09030	male	2009	GG
	09055	male	2009	GG
	09057	male	2009	GG
	09064	male	2009	GG
	09381	female	2009	GG
	09422	female	2009	GG
Individuals from other cohorts, lineages	15320	female	2015	GG
	15501	male	2015	GG
	15515	male	2015	GG
	16003	male	2016	GG
	16142	male	2016	GG
	17228	male	2017	GG
	17343	male	2017	GG
	18057	female	2018	GG
	18062	male	2018	GG
	18129	male	2018	GG
	18394	female	2018	GG
	18402	female	2018	GG
	18476	female	2018	GG
	18630	female	2018	GG
	18635	male	2018	GG

19008	female	2019	GG
19092	female	2019	GG
19163	male	2019	GG
19204	female	2019	GG
19553	female	2019	GG
20044	male	2020	GG
20278	male	2020	GG
21121	female	2021	GG
21214	female	2021	GG
21659	male	2021	GG

Table S2. Reads statistics after removing adaptor sequences, contamination and low-quality reads.

Group	Sample Name	Clean Reads	Q20 (%)	GC (%)	% Mapped	% Properly paired	Coverage
Affected calves	**08012**	245317113	97.68%	43.15	99.57%	98.08%	13.76
	08057	245988224	97.88%	42.91	99.27%	97.53%	13.83
	08070	259915939	96.83%	42.82	99.58%	98.17%	13.32
Putative non-carrier parents	**08030**	255575082	96.76%	43.55	99.41%	97.79%	13.35
	09027	248401822	96.97%	42.49	99.56%	98.20%	14.1
	09057	255548101	96.66%	43.25	99.45%	98.09%	13.48

Table S3. Chromosome location (CHR), BP position, SNP identifier (SNP), and *P* values (*P*) of the 9 SNPs with lower *P* values.

CHR	BP	SNP	*P*
3	101513282	AX-117089247	2.43E-10
3	101856941	AX-185119137	2.43E-10
3	102352899	AX-106754329	2.43E-10
3	100584312	AX-115113333	5.49E-10
3	103336589	AX-168428738	5.49E-10
3	98297288	AX-185119781	1.77E-09
3	100563383	AX-106745142	2.00E-09
3	101785952	AX-106723704	2.00E-09
3	104371879	AX-115108067	2.00E-09

Figure S1. Pedigree of the affected family, from the parental generation where the variant arises in the bull 86053 in 1986 to the first five affected calves in 2004. In grey, putative carriers. In black, affected calves (below, inbreeding coefficient).

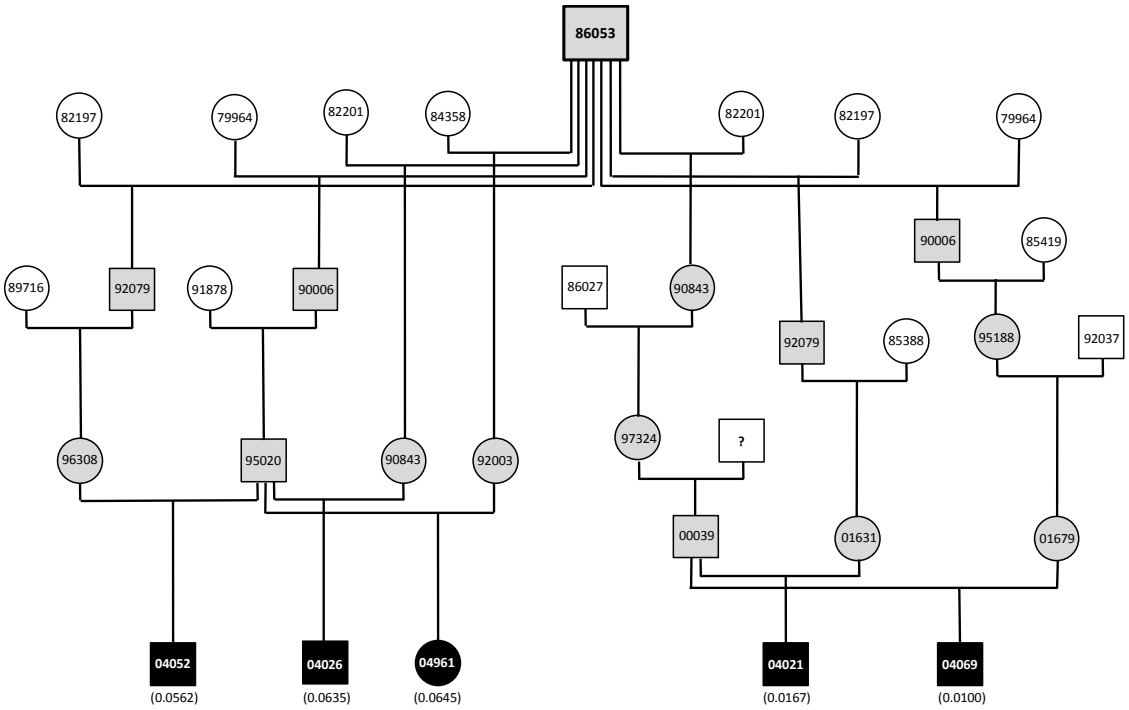

Figure S2. Position of the common ancestor 86053, suspected of being responsible for the distribution of the variant in the population, in the genealogy of the three WGS affected calves. In grey, putative carriers. In black, affected calves (the arrows indicate the sequenced animals). The genotype for the *EDN2* c.149G>A (ENSBTAT00000028571.3:p.Cys50Tyr) variant is indicated for those animals in which a blood sample was available.

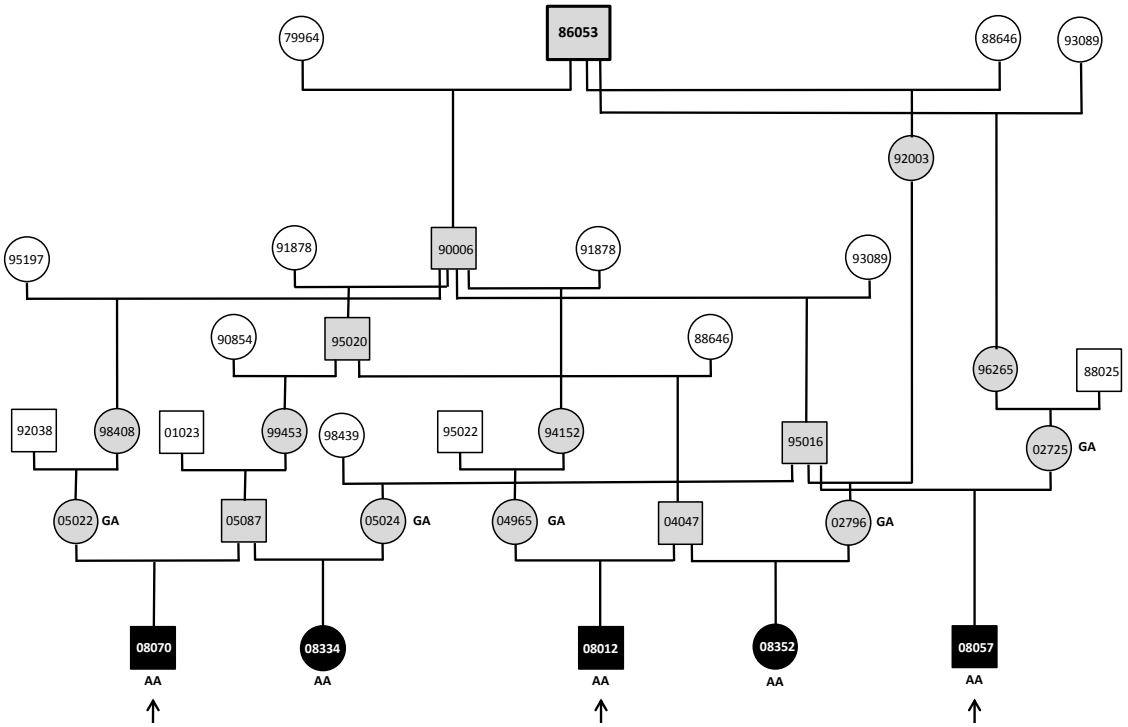

Figure S3. Detailed multispecies protein alignment of the bovine EDN2 ENSBTAT00000028571.3:p.Cys50Tyr missense variant.

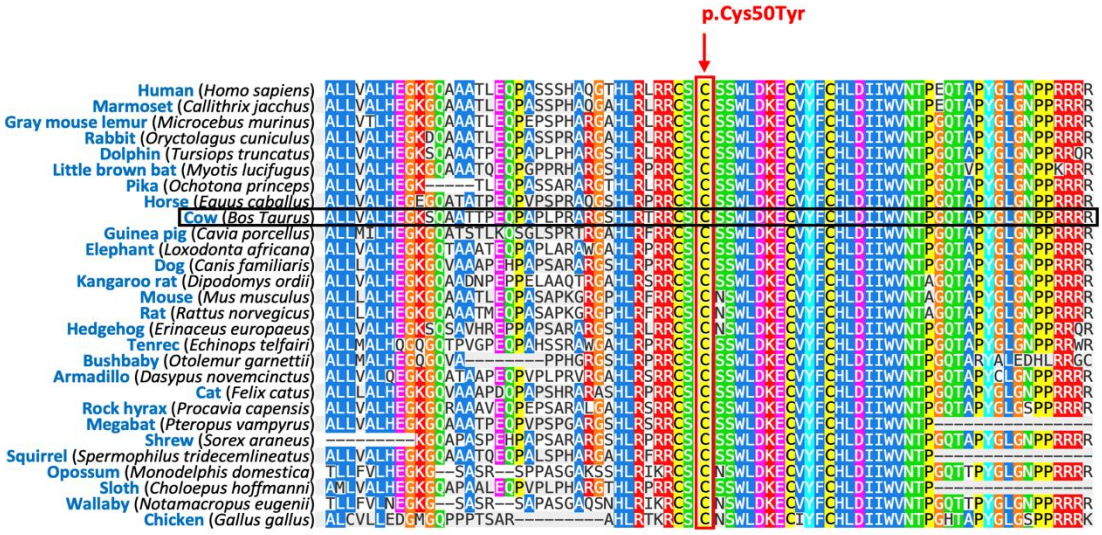

Figure S4. *EDN2* ENSBTAT00000028571.3:p.Cys50Tyr Polyphen2 pathogenicity prediction for the dataset pairs HumDiv and HumVar.