



Truncation of *IFT80* causes early embryonic loss in Holstein cattle associated with Holstein haplotype 2

M. Sofía Ortega,^{1*} Derek M. Bickhart,^{2*} Kelsey N. Lockhart,¹ Daniel J. Null,³ Jana L. Hutchison,³ Jennifer C. McClure,² and John B. Cole^{3†‡}

¹Division of Animal Sciences, College of Agriculture, Food, and Natural Resources, University of Missouri, Columbia 65211

²Cell Wall Biology and Utilization Research Laboratory, U.S. Dairy Forage Research Center, Agricultural Research Service, United States Department of Agriculture, Madison, WI 53706

³Animal Genomics and Improvement Laboratory, Henry A. Wallace Beltsville Agricultural Research Center, Agricultural Research Service, United States Department of Agriculture, Beltsville, MD 20705-2350

ABSTRACT

Recessive alleles represent genetic risk in populations that have undergone bottleneck events. We present a comprehensive framework for identification and validation of these genetic defects, including haplotype-based detection, variant selection from sequence data, and validation using knockout embryos. Holstein haplotype 2 (HH2), which causes embryonic death, was used to demonstrate the approach. Holstein haplotype 2 was identified using a deficiency-of-homozygotes approach and confirmed to negatively affect conception rate and stillbirths. Five carriers were present in a group of 183 sequenced Holstein bulls selected to maximize the coverage of unique haplotypes. Three variants concordant with haplotype calls were found in HH2: a high-priority frameshift mutation resulting, and 2 low-priority variants (1 synonymous variant, 1 premature stop codon). The frameshift in intraflagellar 80 (*IFT80*) was confirmed in a separate group of Holsteins from the 1000 Bull Genomes Project that shared no animals with the discovery set. *IFT80*-null embryos were generated by truncating the *IFT80* transcript at exon 2 or 11 using a CRISPR-Cas9 system. Abattoir-derived oocytes were fertilized in vitro, and zygotes were injected at the one-cell stage either with a guide RNA and CAS9 mRNA complex (n = 100) or Cas9 mRNA (control, n = 100) before return to culture, and replicated 3 times. *IFT80* is activated at the 8-cell stage, and *IFT80*-null embryos arrested at this stage of development, which is consistent with data from mouse hypomorphs and HH2 carrier-to-carrier matings. This frameshift in *IFT80* on

chromosome 1 at 107,172,615 bp (p.Leu381fs) disrupts WNT and hedgehog signaling, and is responsible for the death of homozygous embryos.

Key words: embryonic loss, gene editing, genetic defects, Holstein cattle

INTRODUCTION

Known lethal recessive alleles account for substantial economic losses to cattle breeders (Cole et al., 2016), and 20 such defects (Cole et al., 2020) are routinely tracked in the US population of 9.4 million dairy cows. Economic losses resulting from embryonic death and stillbirths caused by these alleles have been estimated to cost farmers at least \$11 million each year in the United States (Cole et al., 2016). The increased use of artificial insemination on an already reduced genetic pool has exacerbated the problem by increasing the rate at which recessive lethal alleles can be spread in commercial herds. The Holstein haplotype 2 (HH2) allele, which accounts for annual fertility losses of more than \$2 million in the US alone, is present in 1.66% of the domestic Holstein population (VanRaden et al., 2011b) and has a significant negative effect on fertility in the form of early embryonic losses. The patrilineal nature of cattle reproduction and the availability of large numbers of genotyped animals provide an opportunity to identify and validate such recessive mutations and measure their potential effects on other species.

Potential autosomal recessive mutations are routinely tracked during the imputation step of the national genomic evaluations calculated by the Council on Dairy Cattle Breeding (CDCB; Bowie, MD). Genotypes from 48 different arrays are currently stored in the CDCB database (https://queries.uscdcb.com/Genotype/cur_freq.html), with SNP counts ranging from 2,900 to 777,962, and are imputed to a set of 78,964 SNP (80k) for genomic evaluation. In April of 2011 a haplotype affecting fertility was identified on *Bos taurus* auto-

Received January 21, 2022.

Accepted May 31, 2022.

*These authors contributed equally to this work.

†Current address: URUS Group LP, 2418 Crossroads Drive, Suite 3600, Madison, WI 53718.

‡Corresponding author: john.cole@urus.org

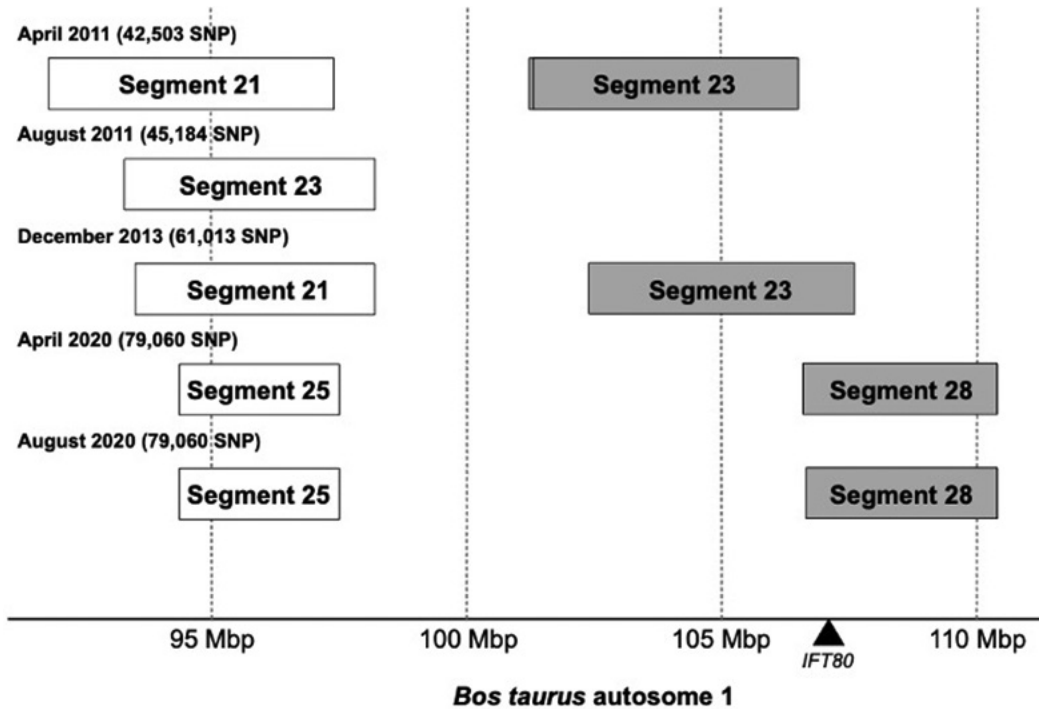


Figure 1. Refinement of candidate Holstein haplotype 2 (HH2) haplotypes over time. The location of the causal variant is indicated by a solid black triangle.

some 1 in the region of 92–97 Mbp and designated as HH2 (Figure 1). At that time, a second haplotype on *Bos taurus* autosome 1 at 101–106 Mbp with a similar carrier list, lack of homozygotes, and negative effects on fertility was identified. When a new SNP list was introduced in August 2011 the haplotype in the 101–106 Mbp region was no longer flagged as a potential lethal. A new 61,000 SNP list was introduced in December 2013, and the haplotype segment upstream from HH2 again appeared on the list of possible lethal haplotypes. In April 2020, the location of HH2 was changed to correspond with the upstream segment based on results from fine-mapping described below. Haplotype blocks are formed using fixed numbers of SNP, rather than linkage disequilibrium boundaries, so it can sometimes be difficult to narrow the candidates down to a single haplotype. In this case, both segments appear to track the causal variant with high concordance.

The goal of this paper is to identify the causal variant underlying the HH2 defect, which causes early embryonic loss in Holstein cattle, and to validate the causal variant in vitro using a gene-editing system.

MATERIALS AND METHODS

No human or animal subjects were used, so this analysis did not require approval by an Institutional

Animal Care and Use Committee or Institutional Review Board.

Routine Assignment of Carrier Status

Single-nucleotide polymorphism genotypes with SNP counts ranging from 2,900 to 777,962 were extracted from the CDCB database and imputed to a common set of 78,964 SNP (80k) using Findhap version 3 (VanRaden et al., 2011a) as part of the routine processing for the August 2020 genomic evaluations in the United States (Wiggans et al., 2017). Carrier status for HH2 was assigned based on the haplotype 28, which spans the interval at 108,398,465–112,161,291 bp on chromosome 1.

Whole-Genome Sequencing, Variant Calling, and Annotation

Three data sets were used to identify putative causal genetic variants by whole-genome sequencing (Figure 2). The first data set consisted of 116 Ayrshire, Holstein, and Jersey samples sequenced by the Animal Genomics and Improvement Laboratory to an average depth of 12 \times . The second data set included 449 Holstein samples from run8 of the 1000 Bull Genomes Project (Hayes and Daetwyler, 2019) sequenced to an

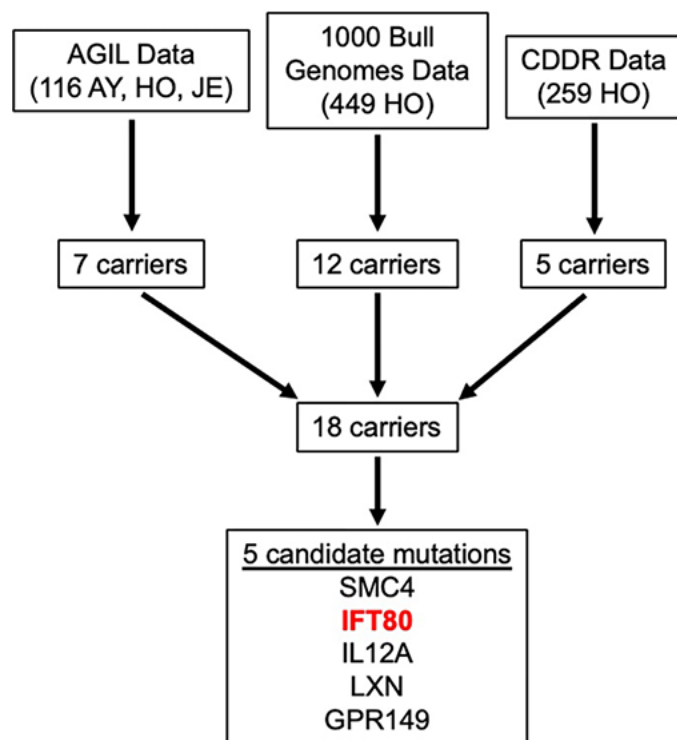


Figure 2. Whole-genome DNA sequence data used in the study included 116 Ayrshire (AY), Holstein (HO), and Jersey (JE) bulls from the Animal Genomics and Improvement Laboratory (AGIL Data), 449 Holstein bulls from Run7 of the 1000 Bull Genomes Project, and 259 Holstein bulls from the Collaborative Dairy DNA Repository (CDDR). Eighteen unique carriers were present after data sets were combined, and candidate causal variants were identified in the *GPR149*, *IFT89*, *IL12A*, *LXN*, and *SMC4* genes.

average depth of 12.7 \times . The third data set included 259 Holstein bulls from the Collaborative Dairy DNA Repository sequenced to an average depth of 17 \times . A total of 824 samples ($n = 688$ unique animals) sequenced using Illumina 150-bp paired-end reads to a target depth of 15 \times were analyzed. Eighteen unique carriers were sequenced: 14 only once, 3 twice, and 1 in all 3 data sets.

Sequence data from the Animal Genomics and Improvement Laboratory and Collaborative Dairy DNA Repository data sets were aligned to the ARS-UCD1.2 reference genome (Rosen et al., 2020) using BWA-MEM as implemented in bwa version 0.7.15-r1140 (<https://doi.org/10.48550/arXiv.1303.3997>) and converted to the binary Sequence Alignment/Map format by samtools version 1.3 (Li et al., 2009). Sequence data variants were identified using the samtools mpileup workflow as described previously (Al-Khudhair et al., 2022). Variants were filtered using version 1.9-259-gb-d769ac of the bcftools “view” command (Li et al., 2009) and the list of filtered variants was annotated

using the SNPEff utility (Cingolani et al., 2012) with the Ensembl annotation of the cattle reference genome (version 101). Variants listed as “High,” “Medium,” and “Low” priority within the haplotype region by SNPEff were selected for further analysis. Effects on predicted protein translation were identified using custom scripts.

Variant calls from the 1000 Bull Genomes run7 data set were used to supplement the genetic background for causal variant concordance analysis (Daetwyler et al., 2014). Data were processed using version 4 of the Genome Analysis Toolkit (McKenna et al., 2010). This included genotyping, variant recalibration for SNP and INDEL, removal of individuals that failed QC, and filtering of monomorphic alleles. The 1000 Bull Genomes VCF files were filtered to remove reads with QS <10 and those identified as “LOWQUAL.” Reads passing edits were aligned against version ARS-UCD1.2 of the bovine genome (Rosen et al., 2020).

Production of *IFT80* Embryos

Guide RNA Design. Guide RNA (gRNA) were designed to target exon 2 or exon 11 (Table 1) of *IFT80*, using the GPP portal from the Broad Institute (<https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design>). To select gRNAs, 2 criteria were used: (1) those with the highest on-target score after design based on the GPP portal and the guide RNA checker portal (Integrated DNA Technologies), and (2) those having the least similarity with other bovine sequences containing a protospacer-adjacent motif. Selected gRNA, as well as the universal 67mer tracrRNA for form guide complexes, were ordered from Integrated DNA Technologies.

The gRNA and tracrRNA were resuspended to a concentration of 500 ng/ μ L. Each guide was then annealed with equal amounts of tracrRNA by heating at 95 $^{\circ}$ C for 5 min and then cooled down from 95 to 25 $^{\circ}$ C using a gradient of -1° C/12 s in a Veriti 96-well gradient thermal cycler (ThermoFisher) to obtain a concentration of 250 ng/ μ L gRNA-tracrRNA complex. Before zygote injection, gRNA-tracrRNA complexes were annealed to CAS9 mRNA (Integrated DNA Technologies) by incubation at room temperature for 15 min. The final injection solution was 50 ng/ μ L each gRNA complex and 20 ng/ μ L Cas9 mRNA. For each exon targeted, guides were injected in pairs, to create a sizable cut in the sequence and facilitate rapid genotyping on an agarose gel. The expected cuts for guides targeting exon 2 were 82 bp for guide pair 1 and 136 bp for guide pair 2, and for those targeting exon 11 were expected to generate a cut of 331 bp.

Table 1. Genomic location, target sequence, and predicted efficiency of guide RNA targeting exons 2 and 11 of the *IFT80* gene

Exon	Guide	Guide name	Position		Sequence – PAM ¹ (5' → 3')	On-target score ²	
			Genomic location	Strand		BI	IDT
2	1	BI283	107,097,118–107,097,140	+	TGTTTTATGTCTCCTTAATA-TGG	33.75	68.0
	2	BI313	107,097,148–107,097,170	+	TCAAGAATTAGTGAGCTGTG-TGG	62.53	65.0
	3	BI423	107,097,269–107,097,291	–	TCTATTGGGTAAATATCATC-AGG	47.26	60.0
11	1	Exon11–1	107,172,343–107,172,365	+	GGAATCAAACTTTATAGGG-TGG	62.05	64.0
	2	Exon 11–3	107,172,670–107,172,692	–	CAGGAAATTTTGGAGAGGAA-AGG	53.38	32.0

¹PAM = protospacer-adjacent motif.

²On-target score: predicted guide efficiency from 2 different design tools: BI (Broad Institute) and IDT (Integrated DNA Technologies).

Production of Embryos In Vitro and Zygote Microinjection. Embryos were produced in vitro with a single sire known to be of high fertility in vitro as previously described (Ortega et al., 2018). At the end of fertilization, putative zygotes (oocytes exposed to sperm) were denuded from the surrounding cumulus cells and split into 2 groups: injected only with Cas9 mRNA and injected with gRNA and Cas9 mRNA against *IFT80*. During the injection procedure zygotes were maintained in manipulation medium (Ortega et al., 2020), and then transferred to 4-well dishes in groups of up to 50 zygotes in 500 μ L of SOF-BE2 (Tribulo et al., 2019), covered with 300 μ L of mineral oil per well at 38.5°C in a humidified atmosphere of 5% (vol/vol) O₂ and 5% (vol/vol) CO₂. Percentage of putative zygotes that cleaved was determined at d 3 of development (d 0 = day of insemination), and blastocyst rate was estimated at d 8 of development. In addition, to determine any developmental arrest, cell stage of all embryos was recorded at d 64.5, 71.5, 120, and 180 h of development (Figure 3B, 3C, and 4A).

Editing Efficiency. To determine editing efficiency of the guides, embryos subjected to microinjection (guides + CAS9 mRNA, or CAS9 mRNA) were collected individually at d 8 of culture, placed in 6 μ L of Embryo Lysis Buffer (Ortega et al., 2020), and subjected to 30 min at 60°C, followed by 10 min at 85°C in a Veriti 96-well gradient thermal cycler (ThermoFisher). Lysed embryo solution was used as template DNA.

For genotyping, the region surrounding the target sequenced of the gRNA was amplified by endpoint PCR using the following primers (5' → 3'): exon 2 F-GGTTTCTTATCCTGCTTCCATTC; R-GAAATTGAGTGTGAACCTTGGG, and for Exon 11 F-CACTGTTTAGGACTCTGCCT; R-CTCTCTGAGTAATGATACCATAGCA. The PCR products were 639 and 620 bp for exon 2 (Figure 4B) and 11 (Figure 4C), respectively. The PCR reaction, amplification (annealing temperature 53.4°C), and visualization of bands were performed as previously described (Ortega et al., 2020). Editing efficiency was calculated from 3 separate in vitro embryo production runs, with

100 embryos per treatment group per replicate. Editing efficiency for exon 2 was 26% biallelic, 32% monoallelic, and 42% nonedited, and for exon 11 was 50% biallelic, 23% monoallelic, and 27% nonedited.

Nucleic Acid Isolation and RT-qPCR Expression Analysis. The DNA and RNA were isolated from individual 8-cell embryos (65–72 h post-insemination) injected against *IFT80* or injected only with Cas9 mRNA. Embryos were collected and lysed individually in 25 μ L of RLT lysis buffer (Catalog No. 79216, Qiagen), lightly vortexed, transferred to a nano-column (PuroSPIN Luna NANOTECH), and centrifuged for 30 s (all centrifugations were performed at 16,000 \times *g* at room temperature). At this point, DNA is trapped in the column, and the flow-through contained the RNA. The column was placed in a new collection tube and stored at room temp for later DNA purification. The flow-through containing RNA was mixed with 1 vol of 70% ethanol, transferred to a new nano-column, and centrifuged for 30 s. Immediately after, 600 μ L of buffer RW1 (Catalog No. 1053394, Qiagen) was added to the column, and centrifuged for 30 s. Flow-through was discarded from the collection tube, and 500 μ L of buffer RPE (Catalog No. 1018013, Qiagen) was added to the column and centrifuged for 30 s. After discarding the flow-through, 500 μ L of 80% ethanol was added to the column and centrifuged for 2 min. To elute RNA, the column was placed in a new microcentrifuge tube, and 15 μ L of RNase-free water was added directly to the membrane of the column, incubated at room temperature for 1 min and finally centrifuged for 1 min. To complete DNA isolation, 500 μ L of 70% ethanol was added to the nano-column containing DNA, followed by a 1-min centrifugation. Then, the column was placed in a new microcentrifuge tube and 15 μ L of elution buffer (Catalog No. 19086, Qiagen) preheated at 70°C was added directly to the membrane of the column, incubated at room temperature for 10 min, followed by 1-min centrifugation to elute DNA. Embryos' edit status was first assessed using PCR across the site of the edit as described above, and the RNA from homozygous edited embryo cells was selected

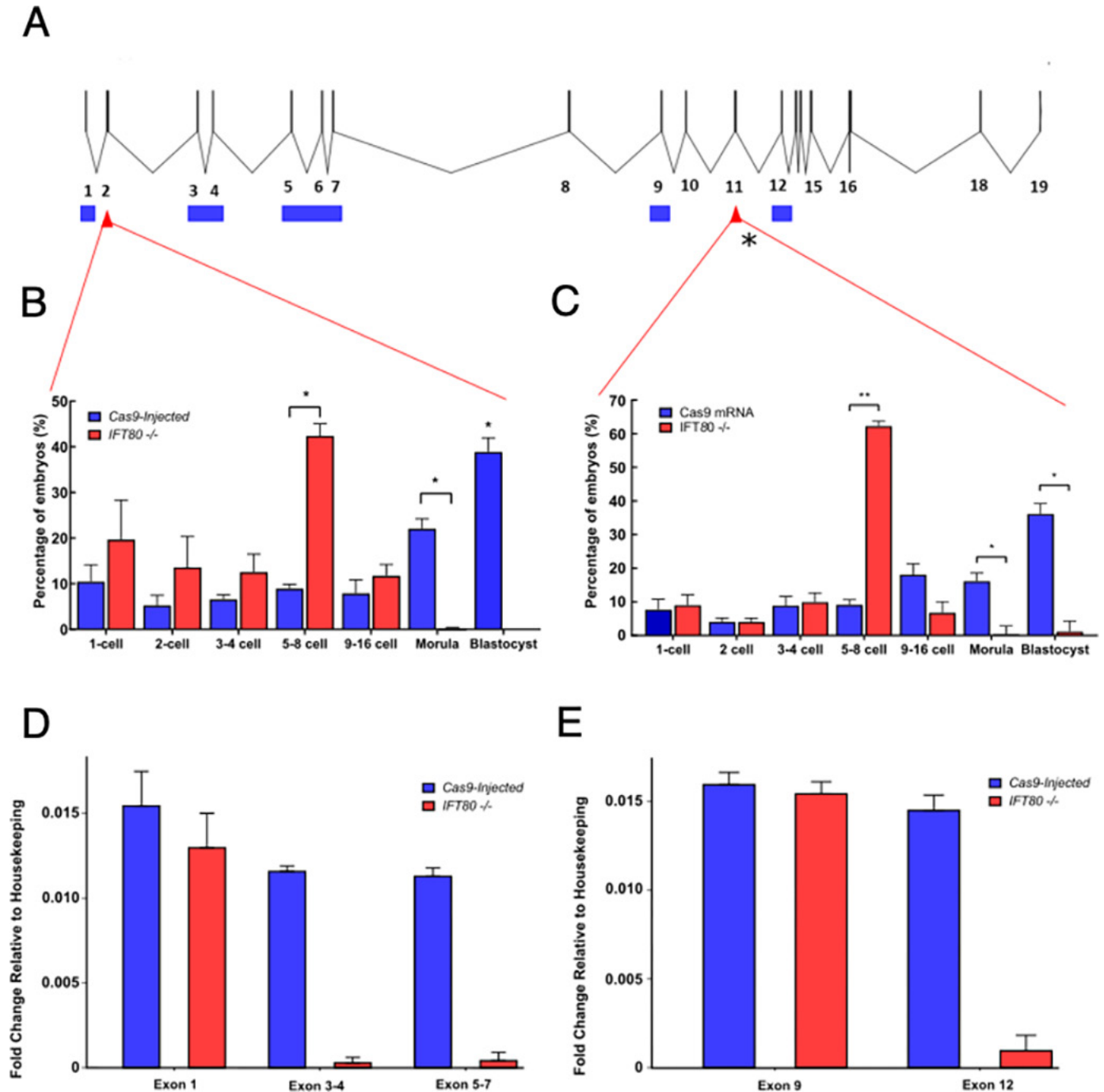


Figure 3. A diagram of the *IFT80* gene (A) shows CRISPR-Cas9 edit sites (red triangles) and quantitative PCR primer design sites (blue boxes). Gene exons are represented by vertical lines, whereas introns are represented by diagonal lines. Embryo counts for each developmental stage for exon 2 (B) and exon 3 (C) demonstrate that homozygous edit embryos (red) do not typically progress past the 8-cell stage compared with vector Cas9 controls (blue). Single (*) and double (**) asterisks indicate Student's *t*-test *P*-values that are less than 0.001 and 0.0001, respectively. Relative fold change of exons downstream of the exon 2 (D) and exon 11 (E) edit sites showed significant decreases in edited embryos (red) compared with the vector Cas9 controls (blue).

for further analysis. Due to the small amount of total RNA in each cell, homozygous embryos were pooled in groups of 4 before cDNA creation. Three pools of edited and control embryos were analyzed in subsequent

analyses. Two-strand cDNA synthesis was conducted using the high-capacity cDNA reverse transcription kit (ThermoFisher) reverse transcriptase following manufacturer instructions. Primer sequences for each locus

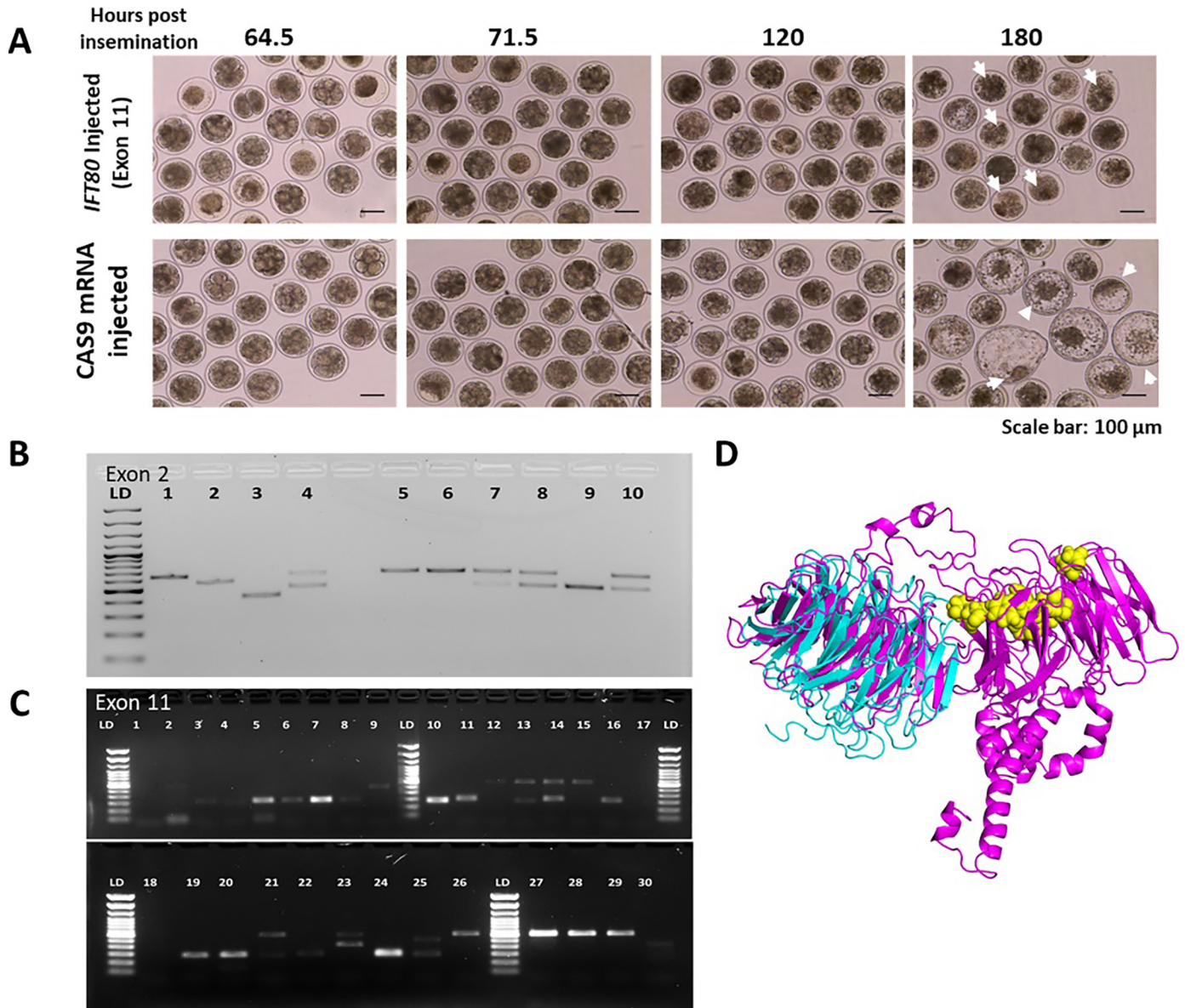


Figure 4. (A) Development of control versus edited embryos. The *IFT80*-injected group have increased numbers of degenerating embryos (white arrows, 120 and 80 h post-insemination) after the 8-cell stage. In control embryos, development progressed normally with formation of blastocysts with compact inner cell mass and defined blastocoel (white arrows, 180 h post-insemination). (B) Embryo genotyping for exon 2 edits. Each numbered lane represents one embryo. Lanes 1, 5, and 6 show nonedited embryos (wild type), lanes 2, 3, and 9 show biallelic edits (absence of wild-type band), and lanes 4, 7, 8, and 10 illustrate monoallelic edits (both wild type and edited band present). (C) Embryo genotyping for exon 11 edits. Each numbered lane represents 1 embryo. Lanes 9, 12, 15, 26, 27, 28, and 29 show nonedited embryos (wild type), lanes 3-8, 10, 11, 16, 19, 20, 22, and 24 show biallelic edits (absence of wild-type band), and lanes 2, 13, 14, 21, and 23 illustrate monoallelic edits (both wild type, and edited band present). (D) Protein model for *IFT80* truncation at exon 11. The pink model represents the 777-amino acid wild-type protein, the blue model represents the 385-amino acid gene-edited protein (after truncation surrounding mutation), and the yellow globular regions indicate putative active binding sites.

were designed using the PrimerQuest (Integrated DNA Technologies). All primers (Table 2) were validated following procedures previously described (Ortega et al., 2017). Quantitative PCR was conducted using the CFX384 Touch Real-Time System (Bio-Rad) using 10 μ L of reaction mix as previously described (Ortega et al., 2020). Given the reduced amount of sample, a

single gene was selected as housekeeping (Steuerwald et al., 1999). The *SDHA* gene was chosen as a suitable housekeeping gene for relative expression calculation in bovine embryos (Goossens et al., 2005). Fold changes were calculated using the method of Livak and Schmittgen (2001) as previously described (Ortega et al., 2017, 2020).

Table 2. Sequence of forward- and reverse-strand primers used for real-time quantitative PCR measurement of *IFT80* gene expression

Exon	Primer forward (5' → 3')	Primer reverse (5' → 3')
1	TGAATTC AATCAGGTGGCATT	AGACATGAGACTGAAGATATCTCTT
3-4	TGGCAGGACGATGGAATTATG	AAAGACGGGAATGCTAAGATCG
5-7	CCGCTTCAACCAAATGCTAAA	GTATGGGATAGTTATGGCCGTC
9-10	GCCTCTTTGAACTATGCACACTTAG	CTACAAAGA AACTGGAATACAC
12	CTCTTTGAGGCATCAACTGGA	TTGGAAATTGCCCTGGATCA

Statistics

Differences in gene expression and development were analyzed by least-squares ANOVA of the Δ CT values using the GLM procedure of the Statistical Analysis System version 9.4 (SAS Institute Inc.). Replicate and treatment were included as main effects in the model.

Availability of Data

Performance and Pedigree Data. The performance and pedigree data used to compute HH2 effects are controlled by the CDCB. Requests to access CDCB data may be sent to João Dürr, CDCB Chief Executive Officer (joao.durr@cdbc.us).

Genotype and Whole-Genome DNA Sequence Data. Access to SNP genotype data and whole-genome DNA sequence data for bulls held by the CDDR must be requested from Jay Weiker, CDDR Administrator (jweiker@naab-css.org).

1000 Bull Genomes Data. Access to whole-genome DNA sequence data from bulls distributed in Run7 is restricted to project members. Inquiries should be directed to Dr. Christy Vander Jagt, Research Scientist, Agriculture Victoria Research (christy.vanderjagt@agriculture.vic.gov.au).

RESULTS AND DISCUSSION

Variant Discovery

We first refined the HH2 genomic locus to narrow the window for variant discovery in subsequent sequencing analysis (Figure 5A). Our analysis was aided by the recent release of a high-quality reference assembly for taurine cattle (Rosen et al., 2020), which had over 100-fold more contiguity than the previously used UMD3.1 reference (Zimin et al., 2009). Using imputed SNP genotypes (VanRaden et al., 2013) derived from more than 3.7 million SNP genotypes of commercial cattle mapped to the new reference assembly, we were able to reduce the size of the locus from 8.9 to 1.5 Mbp. This reduced the number of candidate genes that were previously hypothesized to be causal for the lethality of

the haplotype by 72%, leaving only 7 genes (*PPM1L*, *KPNA4*, *TRIM59*, *SMC4*, *IFT80*, *IL12A*, and *SCHIP1*) in the refined window out of 25 original candidates.

A concordance analysis was used to identify candidate causal variants within the refined haplotype locus, and 1 INDEL within the gene *IFT80* (g.107172616delT; mutation: p.Leu387PhefsTer3) that was >99% concordant with known carrier status was identified (Figure 5B). One carrier animal was predicted to not have a heterozygous copy of g.107172616delTL, but manual review of read pileups on the variant site identified g.107172616delT at a frequency of 25%, which was below the threshold of initial detection (data not shown). We interpreted this result as complete concordance and attributed the lower proportion of variant reads to a bias in sampling the haplotype during sequencing library preparation. The concordant g.107172616delT variant was located on BTA1:107,172,615 and was found to induce a frameshift in the *IFT80* gene's 11th exon that led to early truncation of the protein product (Figure 5A). In silico translation of the expected protein product of the edited transcript suggested an early truncation of the *IFT80* protein, resulting in a protein of 387 amino acids (49.5% of total) compared with the wild type (Figure 4D). The early termination of the gene interrupts one of 7 WD 40-repeat-containing domains (Pfam: SSF50978) and a predicted weak polyampholyte domain (<https://mobidb.bio.unipd.it/A0A4W2EVR7>), both of which likely affect the intraflagellar transport of materials in the cell (Huang et al., 2009). Analysis of potential founders identified the bull Willowholme Mark Anthony (73HO0219, born 1975) as a likely candidate (Figure 6). Pedigree, SNP, and DNA sequence data show that he is the founder with respect to the SNP haplotype, but he is not a carrier of the actual *IFT80* variant, which appears to have arisen in one of his daughters (Elysa Anthony Lea, HOCANF000003628269, born 1981) or granddaughters (Comestar Laurie Sheik-ET, HOCANF000004425038, born 1986). Neither descendant is genotyped or sequenced, and DNA is not available, so the true founder cannot be identified. We note that our concordance analysis did not include structural variation within the haplotype region as these data were missing from our

genetic background information. It is possible that a cryptic structural variant may be the true causal variant for the phenotype and that the frameshift mutation identified within our data set is linked to such a variant. To support our hypothesis that the 1161delT variant causes early embryonic loss in cattle, we performed functional validation experiments on gene-edited cattle zygotes.

Functional Validation

We hypothesized that the early termination of the *IFT80* gene was the causal variant of the phenotype given its perfect concordance and the potential for the mutation to have a high effect on downstream protein function. To induce the phenotype, we used a CRISPR-Cas9 gene-editing system to induce double-strand

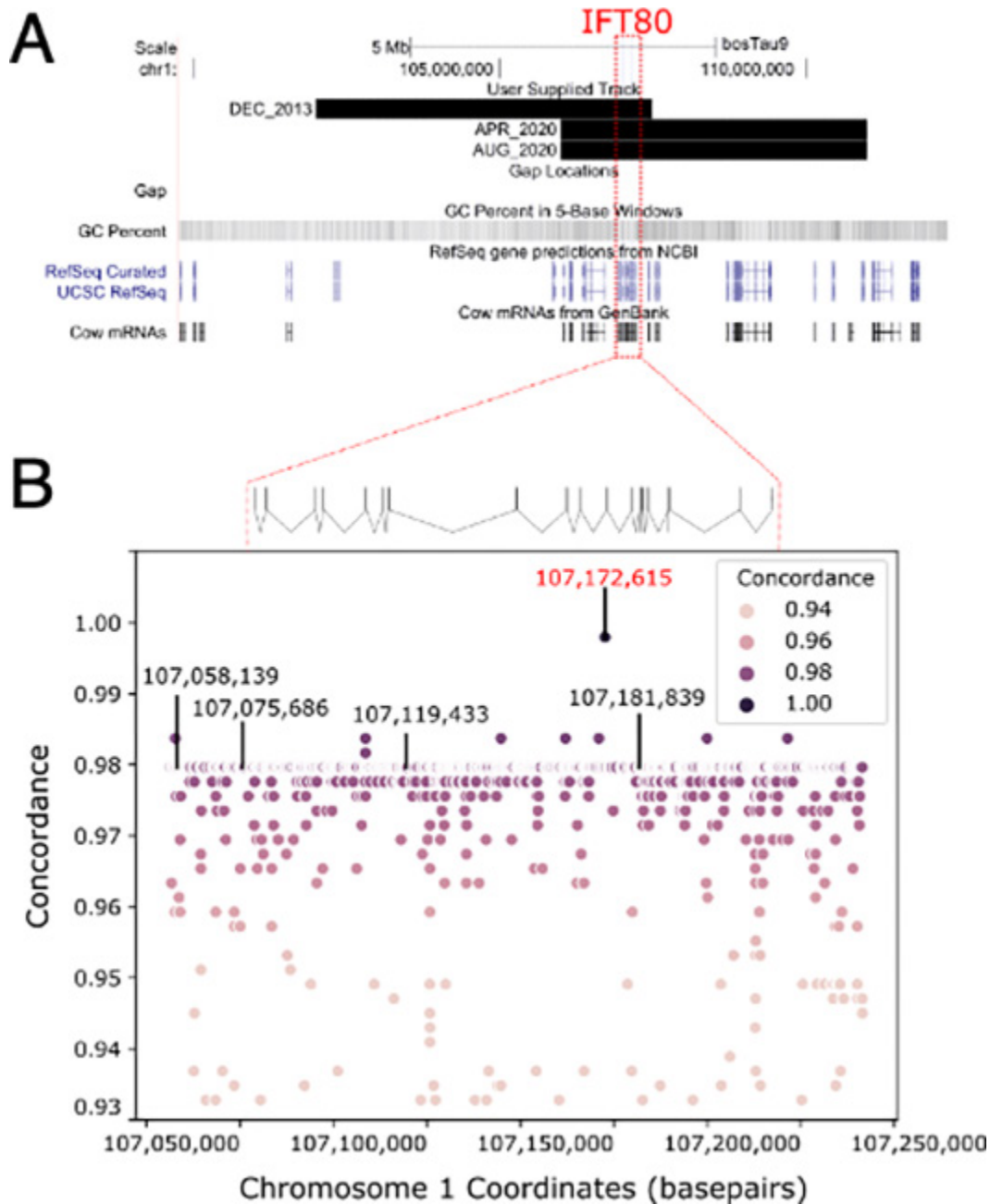


Figure 5. A University of California, Santa Cruz (UCSC) genome browser plot (A) shows the locations of prior Holstein haplotype 2 (HH2) loci as determined in December 2013, April 2020, and August 2020. The *IFT80* gene is highlighted by red dashed lines. Concordance analysis of variant sites against expected carrier status (B) identified only one variant with >99% concordance (highlighted in red). For genomic context, the *IFT80* gene is drawn with exons (straight lines) and introns (diagonal lines) above the concordance plot. Only variants predicted to have high functional effects by SNPEff are highlighted by labels showing coordinate numbers. GC = guanine-cytosine content of DNA.

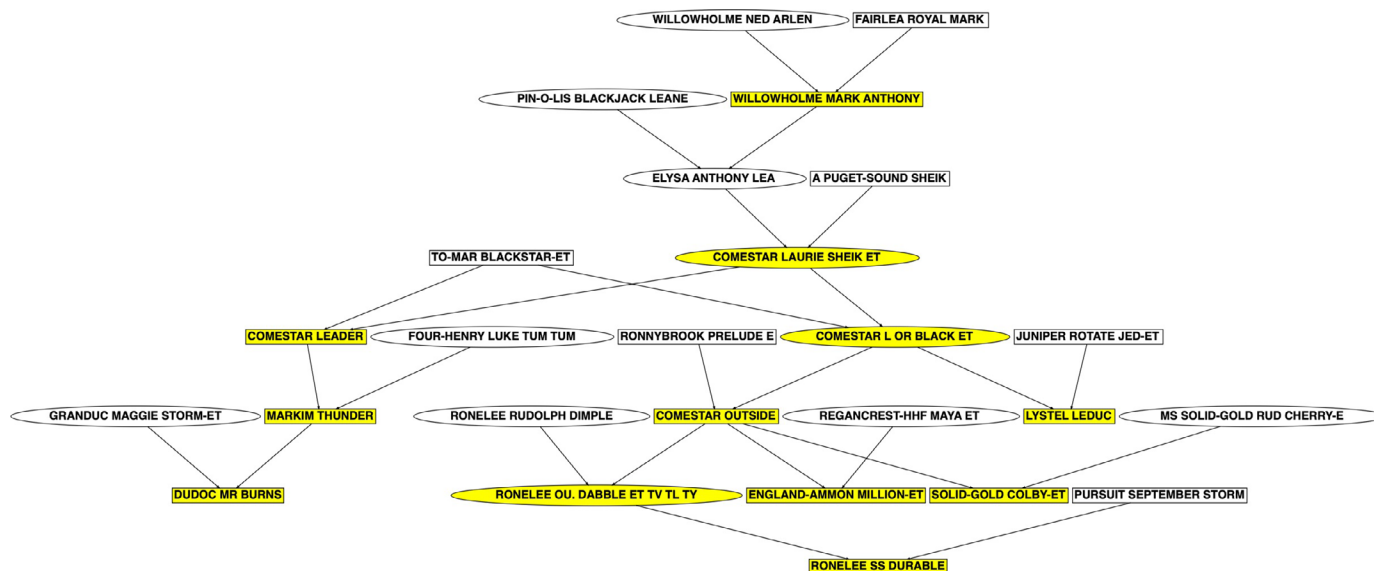


Figure 6. Pedigree of animals carrying the Holstein haplotype 2 (HH2) haplotype (shaded in yellow) show that the bull Willowholme Mark Anthony (73HO0219, born 1975) was the first known carrier of that haplotype. However, he does not carry the causal variant in the *IFT80* gene, which suggests that the original mutation likely arose in his daughter, Elysa Anthony Lea (HOCANF000003628269, born 1981), or granddaughter, Comestar Laurie Sheik-ET (HOCANF000004425038, born 1986). DNA is not available for either cow, so it is not possible to confirm which was the founder animal.

breaks at different exons at or upstream of the candidate causal variant site (Figure 3A). In the first model, *IFT80* was truncated at exon 2, and in a second model, *IFT80* was truncated at exon 11 (variant site). In both cases, treated embryos failed to progress past the 8- to 16-cell transition period (Figure 3A, 3C; $P < 0.0001$), and degenerated after this developmental stage (Figure 4A). As alternative splicing could potentially recover *IFT80* protein function, we quantified the expression at different exons of *IFT80* transcript. For embryos edited at exon 2, expression was measured at exons 1, 3–4, and 5–7 (before and after edit site; Figure 3D). For the embryos edited at exon 11 (exon with causal mutation), *IFT80* expression was determined at exons 9–10 and 12 (before and after the edited exon; Figure 3E) by quantitative PCR assays. Upstream regions of edited sites had no difference in *IFT80* expression (exon 1, $P = 0.44$; exon 9–10, $P = 0.62$), between edited and nonedited embryos. Conversely, in both models (truncation at exon 2 or 11), downstream exons of the edited site had significantly lower *IFT80* expression ($P < 0.0001$) in edited embryos than in nonedited embryos (Figure 3D). Differential expression was most pronounced in exon 12, in which the edited embryos had 62.5% lower fold expression of the exon than the wildtype (Figure 3E). This discrepancy in copy number of the later exons in edited zygotes is contrary to known mechanisms for nonsense-mediated RNA decay, which would result in the post-transcriptional degradation of the entire tran-

script. Some nonsense transcripts have been known to evade nonsense-mediated RNA decay (e.g., Litchfield et al., 2020), though we note that this frameshift occurs earlier in the gene transcript than in previously noted exceptions. Our data suggest that loss of adjacent exons may occur due to this mutation, though the exact mechanism cannot be identified from the results of this study. It is also possible that observed early embryonic loss could also be attributed to off-target edits made by the CRISPR-Cas9 system; however, we note that our gRNA design methods screened designs for off-target activity to minimize the potential for this effect. In addition, BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used to check candidate gRNA to ensure that they were not targeting unexpected regions of the genome.

Mutations to the *IFT80* gene are linked to asphyxiating thoracic dystrophy 2 disease in humans (Beales et al., 2007), and hypomorphic expression of the gene has been shown to result in low postnatal survival in gene-trap mice (Rix et al., 2011). Previous surveys have speculated as to when *IFT80* gene expression is essential during development (Rix et al., 2011), and our results suggest that expression is needed to progress past the 8- to 16-cell developmental phase. This is consistent with previous RNA surveys of cattle preimplantation embryonic development (Graf et al., 2014), and with the known fertility effects of HH2 in cattle (VanRaden et al., 2011b; Cole et al., 2016). The results of editing

experiments suggest that the carboxy-terminal WD 40-repeat or the polyampholyte domains are essential for embryonic development. Recent studies suggest that deletion of *IFT80* disrupts the FGF2 signaling pathway, affecting cell proliferation, and self-renewal of odontoblasts (Yuan et al., 2019). In the preimplantation mammalian embryo, FGF2 signaling is essential for lineage commitment and blastocyst formation (Lanner and Rossant, 2010; Kuijk et al., 2012), in which disruption of FGF2 signaling coincides with the developmental arrest observed in this study.

Yang et al. (2022) recently published an independent analysis of whole-genome and transcriptome data from 8 Chinese Holsteins (4 carriers and 4 noncarriers) in which they identified 5 putative functional variants of protein-coding genes associated with HH2. They acknowledged that our methodology provides experimental confirmation of the developmental effects of the variant we discovered (Li et al., 2022). Interestingly, they did not observe allele-specific transcriptional bias in the blood of carriers, which supports our observation that nonsense-mediated RNA decay does not act on the mutant allele's transcript (Yang et al., 2022). However, they did not specifically report the absence of adjacent exons in gene transcripts that we report. These results underscore the importance of replication to confirm putative associations of DNA variants with phenotypic effects.

CONCLUSIONS

Using genotypes from millions of Holstein cattle and CRISPR-Cas9 gene editing, we were able to track and validate a recessive lethal allele in the *IFT80* gene, respectively. This approach conclusively identified the target mutation in such a way that it provided further insight into the function of the gene in development of other mammalian species. This discovery has further economic benefits, as identification of the exact causal mutation underlying this haplotype will benefit dairy producers by allowing them to avoid carrier-to-carrier matings, which result in pregnancy losses responsible for ~\$437,000/yr in US Holsteins. Concordance analysis and molecular validation on naturally occurring genetic variants represent a potent tool for ongoing efforts to translate genotype to phenotype in the field of quantitative genetics.

ACKNOWLEDGMENTS

Bickhart was supported by appropriated projects 8042-31000-001-00-D, “Enhancing Genetic Merit of Ruminants Through Improved Genome Assembly,





Annotation, and Selection” and 5090-31000-026-00-D, “Investigating Microbial, Digestive, and Animal Factors to Increase Dairy Cow Performance and Nutrient Use Efficiency,” of the Agricultural Research Service (ARS) of the USDA (Washington, DC). Cole, Hutchison, and Null were supported by appropriated project 8042-31000-002-00-D, “Improving Dairy Animals by Increasing Accuracy of Genomic Prediction, Evaluating New Traits, and Redefining Selection Goals,” of ARS, USDA. McClure was supported by appropriated project 5090-31000-026-00-D, “Investigating Microbial, Digestive, and Animal Factors to Increase Dairy Cow Performance and Nutrient Use Efficiency” of ARS, USDA. Clark was supported by USDA NIFA Grant 2019-38420-28972. The authors thank US dairy producers for providing phenotypic, genomic, and pedigree data through the Council on Dairy Cattle Breeding (Bowie, MD) under USDA-ARS Material Transfer Research Agreement 58-8042-8-007. Access to whole-genome sequence data from the Collaborative Dairy DNA Repository (Madison, WI) was provided under USDA-ARS Material Transfer Research Agreement 58-8042-9-0010F. Access to whole-genome sequence data from the 1000 Bull Genomes Project (Melbourne, Victoria, Australia) was provided under USDA-ARS Material Transfer Agreement 14358. The authors also thank Bethany Bauer and Joshua Benne from the University of Missouri (Columbia, MO) for their assistance with microinjections of bovine embryos. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. The USDA is an equal opportunity provider and employer. The authors have not stated any conflicts of interest.

REFERENCES

- Al-Khudhair, A., D. J. Null, J. B. Cole, C. W. Wolfe, D. J. Steffen, and P. M. VanRaden. 2022. Inheritance of a mutation causing neuropathy with splayed forelimbs in Jersey cattle. *J. Dairy Sci.* 105:1338–1345. <https://doi.org/10.3168/jds.2021-20600>.
- Beales, P. L., E. Bland, J. L. Tobin, C. Bacchelli, B. Tuysuz, J. Hill, S. Rix, C. G. Pearson, M. Kai, J. Hartley, C. Johnson, M. Irving, N. Elcioglu, M. Winey, M. Tada, and P. J. Scambler. 2007. *IFT80*, which encodes a conserved intraflagellar transport protein, is mutated in Jeune asphyxiating thoracic dystrophy. *Nat. Genet.* 39:727–729. <https://doi.org/10.1038/ng2038>.
- Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* 6:80–92. <https://doi.org/10.4161/fly.19695>.
- Cole, J. B., D. J. Null, and P. M. VanRaden. 2016. 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>.
- Cole, J. B., P. M. VanRaden, D. J. Null, J. L. Hutchison, and S. M. Hubbard. 2020. AIP research report GENOMIC5: Haplotype tests

- for recessive disorders that affect fertility and other traits. Accessed Jan. 16, 2021. https://www.ars.usda.gov/ARSUserFiles/80420530/Publications/ARR/Haplotype%20tests_ARR-Genomic5.pdf.
- Daetwyler, H. D., A. Capitan, H. Pausch, P. Stothard, R. Van Binsbergen, R. F. Brøndum, X. Liao, A. Djari, S. C. Rodriguez, C. Grohs, D. Esquerre, O. Bouchez, M.-N. Rossignol, C. Klopp, D. Rocha, S. Fritz, A. Eggen, P. J. Bowman, D. Coote, A. J. Chamberlain, C. Anderson, C. P. VanTassell, I. Hulsegge, M. E. Goddard, B. Guldbbrandtsen, M. S. Lund, R. F. Veerkamp, D. A. Boichard, R. Fries, and B. J. Hayes. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* 46:858–865. <https://doi.org/10.1038/ng.3034>.
- Goossens, K., M. Van Poucke, A. Van Soom, J. Vandesompele, A. Van Zeven, and L. J. Peelman. 2005. Selection of reference genes for quantitative real-time PCR in bovine preimplantation embryos. *BMC Dev. Biol.* 5:27. <https://doi.org/10.1186/1471-213X-5-27>.
- Graf, A., S. Krebs, V. Zakhartchenko, B. Schwab, H. Blum, and E. Wolf. 2014. Fine mapping of genome activation in bovine embryos by RNA sequencing. *Proc. Natl. Acad. Sci. USA* 111:4139–4144. <https://doi.org/10.1073/pnas.1321569111>.
- Hayes, B. J., and H. D. Daetwyler. 2019. 1000 Bull Genomes Project to map simple and complex genetic traits in cattle: applications and outcomes. *Annu. Rev. Anim. Biosci.* 7:89–102. <https://doi.org/10.1146/annurev-animal-020518-115024>.
- Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2009. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37:1–13. <https://doi.org/10.1093/nar/gkn923>.
- Kuijk, E. W., L. T. A. van Tol, H. Van de Velde, R. Wubbolts, M. Welling, N. Geijsen, and B. A. J. Roelen. 2012. The roles of FGF and MAP kinase signaling in the segregation of the epiblast and hypoblast cell lineages in bovine and human embryos. *Development* 139:871–882. <https://doi.org/10.1242/dev.071688>.
- Lanner, F., and J. Rossant. 2010. The role of FGF/Erk signaling in pluripotent cells. *Development* 137:3351–3360. <https://doi.org/10.1242/dev.050146>.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Li, S., Y. Gao, O. Canela-Xandri, S. Wang, Y. Yu, W. Cai, B. Li, R. Xiang, A. J. Chamberlain, and E. Pairo-Castineira. K. D. Mellow, K. Rawlik, C. Xia, Y. Yao, P. Navarro, D. Rocha, X. Li, Z. Yan, C. Li, B. D. Rosen, C. P. Van Tassell, P. M. Vanraden, S. Zhang, L. Ma, J. B. Cole, G. E. Liu, A. Tenesa, and L. Fang. 2022. A multi-tissue atlas of regulatory variants in cattle. *Nat. Genet.* <https://doi.org/10.1038/s41588-022-01153-5>.
- Litchfield, K., J. L. Reading, E. L. Lim, H. Xu, P. Liu, M. Al-Bakir, Y. N. S. Wong, A. Rowan, S. A. Funt, T. Merghoub, D. Perkins, M. Lauss, I. M. Svane, G. Jönsson, J. Herrero, J. Larkin, S. A. Quezada, M. D. Hellmann, S. Turajlic, and C. Swanton. 2020. Escape from nonsense-mediated decay associates with anti-tumor immunogenicity. *Nat. Commun.* 11:3800. <https://doi.org/10.1038/s41467-020-17526-5>.
- Livak, K. J., and T. D. Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative pCR and $2^{-\Delta\Delta CT}$ method. *Methods* 25:402–408. <https://doi.org/10.1006/meth.2001.1262>.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytzky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. 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>.
- Ortega, M. S., A. M. Kelleher, E. O’Neil, J. Benne, R. Cecil, and T. E. Spencer. 2020. NANOG is required to form the epiblast and maintain pluripotency in the bovine embryo. *Mol. Reprod. Dev.* 87:152–160. <https://doi.org/10.1002/mrd.23304>.
- Ortega, M. S., J. J. Kurian, R. McKenna, and P. J. Hansen. 2017. Characteristics of candidate genes associated with embryonic development in the cow: Evidence for a role for WBP1 in development to the blastocyst stage. *PLoS One* 12:e0178041. <https://doi.org/10.1371/journal.pone.0178041>.
- Ortega, M. S., J. G. N. Moraes, D. J. Patterson, M. F. Smith, S. K. Behura, S. Pooock, and T. E. Spencer. 2018. Influences of sire conception rate on pregnancy establishment in dairy cattle. *Biol. Reprod.* 99:1244–1254. <https://doi.org/10.1093/biolre/iy141>.
- Rix, S., A. Calmont, P. J. Scambler, and P. L. Beales. 2011. An *Ift80* mouse model of short rib polydactyly syndromes shows defects in hedgehog signalling without loss or malformation of cilia. *Hum. Mol. Genet.* 20:1306–1314. <https://doi.org/10.1093/hmg/ddr013>.
- Rosen, B. D., D. M. Bickhart, R. D. Schnabel, S. Koren, C. G. Elsik, E. Tseng, T. N. Rowan, W. Y. Low, A. Zimin, C. Coudrey, R. Hall, W. Li, A. Rhie, J. Ghurye, S. D. McKay, F. Thibaud-Nissen, J. Hoffman, B. M. Murdoch, W. M. Snelling, T. G. McDaneld, J. A. Hammond, J. C. Schwartz, W. Nandolo, D. E. Hagen, C. Dreischer, S. J. Schultheiss, S. G. Schroeder, A. M. Phillippy, J. B. Cole, C. P. Van Tassell, G. Liu, T. P. L. Smith, and J. F. Medrano. 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* 9:giaa021. <https://doi.org/10.1093/gigascience/gjaa021>.
- Steuerwald, N., J. Cohen, R. J. Herrera, and C. A. Brenner. 1999. Analysis of gene expression in single oocytes and embryos by real-time rapid cycle fluorescence monitored RT-PCR. *Mol. Hum. Reprod.* 5:1034–1039. <https://doi.org/10.1093/molehr/5.11.1034>.
- Tribulo, P., R. M. Rivera, M. S. Ortega, E. A. Jannaman, and P. J. Hansen. 2019. Production and Culture of the Bovine Embryo. Springer.
- VanRaden, P. M., D. J. Null, M. Sargolzaei, G. R. Wiggans, M. E. Tooker, J. B. Cole, T. S. Sonstegard, E. E. Connor, M. Winters, J. B. C. H. M. van Kaam, A. Valentini, B. J. Van Doormaal, M. A. Faust, and G. A. Doak. 2013. Genomic imputation and evaluation using high-density Holstein genotypes. *J. Dairy Sci.* 96:668–678. <https://doi.org/10.3168/jds.2012-5702>.
- VanRaden, P. M., J. R. O’Connell, G. R. Wiggans, and K. A. Weigel. 2011a. Genomic evaluations with many more genotypes. *Genet. Sel. Evol.* 43:10. <https://doi.org/10.1186/1297-9686-43-10>.
- VanRaden, P. M., K. M. Olson, D. J. Null, and J. L. Hutchison. 2011b. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* 94:6153–6161. <https://doi.org/10.3168/jds.2011-4624>.
- Wiggans, G. R., J. B. Cole, S. M. Hubbard, and T. S. Sonstegard. 2017. Genomic selection in dairy cattle: the USDA experience. *Annu. Rev. Anim. Biosci.* 5:309–327. <https://doi.org/10.1146/annurev-animal-021815-111422>.
- Yang, Y., J. Si, X. Lv, D. Dai, L. Liu, S. Tang, Y. Wang, S. Zhang, W. Xiao, and Y. Zhang. 2022. Integrated analysis of whole genome and transcriptome sequencing reveals a frameshift mutation associated with recessive embryonic lethality in Holstein cattle. *Anim. Genet.* 53:137–141. <https://doi.org/10.1111/age.13160>.
- Yuan, X., X. Cao, and S. Yang. 2019. *IFT80* is required for stem cell proliferation, differentiation, and odontoblast polarization during tooth development. *Cell Death Dis.* 10:63. <https://doi.org/10.1038/s41419-018-0951-9>.
- Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10:R42. <https://doi.org/10.1186/gb-2009-10-4-r42>.

ORCID

- M. Sofia Ortega  <https://orcid.org/0000-0003-4027-7314>
 Derek M. Bickhart  <https://orcid.org/0000-0003-2223-9285>
 Daniel J. Null  <https://orcid.org/0000-0003-0012-2115>
 John B. Cole  <https://orcid.org/0000-0003-1242-4401>