Truncation of IFT80 causes early embryonic loss in cattle

M. Sofía Ortega*, Derek M. Bickhart*, Kelsey N. Clark, Daniel J. Null, Jana L. Hutchison, Jennifer C. McClure, John B. Cole†

* = co-first author

† = corresponding author

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

[2]Division of Animal Sciences, College of Agriculture, Food, and Natural Resources, University of Missouri, Columbia, 65211

[3]Animal 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 a substantial risk in populations that have undergone significant bottleneck events. Analyzing millions of genotypes and hundreds of sequenced bulls, we identified one pernicious recessive variant that achieved 1.66% frequency in the US Holstein population and confirmed loss of embryonic development via CRISPR-Cas9 editing. In the process, we identified the developmental significance of two protein domains of the *IFT80* gene in the embryonic development of cattle.

MAIN

Known lethal recessive alleles account for substantial economic losses to cattle breeders[1], and 20 such defects[2] 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 per year[1] in the US and approximately £1.8 million annually in the UK. The advent of artificial insemination techniques has exacerbated the problem by increasing the frequency of recessive lethal alleles in commercial herds. The 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[3] and has a significant effect on cow fertility in the form of early embryonic losses. The patrilineal nature of cattle reproduction and the availability of large numbers of genotyped animals provides an opportunity to identify and validate such recessive mutations and measure their potential impacts on other species.

We first sought to refine the HH2 genomic locus to narrow the window for variant discovery in subsequent sequencing analysis (Fig 1a). Our analysis was aided by the recent release of a high-quality reference assembly for taurine cattle[4] which had over 100-fold more contiguity than the

41 previously used UMD3.1 reference[5]. Using imputed SNP genotypes[6] derived from more than 3.7
42 million SNP genotypes of commercial cattle mapped to the new reference assembly, we were
43 able to reduce the size of the locus from 8.9 Mbp to 1.5 Mbp. This reduced the number of
44 candidate genes that were previously hypothesized to be causal for the lethality of the haplotype
45 by 72%, leaving only seven genes in the refined window out of 25 original candidates.
46
47 Separate, non-overlapping datasets were used for causal variant discovery and validation. The
48 discovery dataset included whole-genome sequence data for 9 carriers and 449 non-carriers from
49 the Animal Genomics and Improvement Laboratory (Beltsville, MD), the Collaborative Dairy
50 DNA Repository (Madison, WI) and the 1000 Bull Genomes Project[7]. A concordance analysis
51 was performed in order to identify candidate causal variants within the haplotype locus, and one
52 INDEL (Fig 1b) that was > 99% concordant with known carrier status was identified. One carrier
53 animal was predicted to not have a heterozygous copy of the INDEL, but manual review of read
54 pileups on the variant site identified the INDEL at a frequency of 25% which was below the
55 threshold of initial detection (data not shown). We interpreted this result as complete
56 concordance and attributed the lower proportion of variant reads to a bias in sampling the variant
57 haplotype during sequencing library preparation. The concordant INDEL was located on
58 btau1:107,172,615 and was found to induce a frameshift in the *IFT80* gene's 11th exon that led
59 to early truncation of the protein product (Supplementary Figure 1). *In silic*o translation of the
60 expected protein product of the edited transcript suggested an early truncation of the IFT80
61 protein, resulting in a protein of 385 amino acids (49.5 % of total) compared to the wild type.
62 The early termination of the gene interrupts one of seven WD 40-repeat containing domains
63 (Pfam: SSF50978) and a predicted weak polyampholyte domain

64 (https://mobidb.bio.unipd.it/A0A4W2EVR7), both of which likely impact the intra-flagellar
65 transport of materials in the cell[8]. Analysis of potential founders identified the bull Willowholme
66 Mark Anthony (73HO0219, born 1975) as a likely candidate. Pedigree, SNP and DNA sequence
67 data show that he is the founder with respect to the SNP haplotype, but he is not a carrier of the
68 actual *IFT80* variant, which appears to have arisen in one of his daughters (Elysa Anthony Lea,
69 HOCANF000003628269, born 1981) or grand-daughters (Comestar Laurie Sheik-ET,
70 HOCANF000004425038, born 1986). Neither descendant is genotyped or sequenced and DNA
71 is not available, so the true founder cannot be identified.
72
73 We hypothesized that the early termination of the *IFT80* gene was the causal variant of the
74 phenotype given its perfect concordance and the potential for the mutation to have a high impact
75 on downstream protein function. To induce the phenotype, we used a CRISPR-Cas9 gene editing
76 system (see Methods) to induce double-strand breaks at different exons at or upstream of the
77 candidate causal variant site (Fig 2a). In the first model, *IFT80* was truncated at exon 2, and in a
78 second model, *IFT80* was truncated at exon 11 (variant site). In both cases, treated embryos
79 failed to progress past the 8- to 16-cell transition period (Fig 2b, 2c; $P < 0.0001$), and
80 degenerated after this developmental stage. As alternative splicing could potentially recover
81 IFT80 protein function, we quantified the expression at different exons of *IFT80* transcript. For
82 embryos edited at exon 2, expression was measured at exons 1, 3–4 and 5–7 (before and after
83 edit site; Fig 2d). For the embryos edited at exon 11 (exon with causal mutation), *IFT80*
84 expression was determined at exons 9–10 and 12 (before and after the edited exon; Fig 2e) by
85 qPCR assays. Upstream regions of edited sites had no difference in *IFT80* expression (Exon 1, P
86 = 0.44; Exon 9-10, P = 0.62), between edited and non-edited embryos. Conversely, in both

87 models (truncation at exon 2 or 11) downstream exons of edited site, had significantly lower
88 *IFT80* expression ($P < 0.0001$) in edited embryos than non-edited embryos, suggesting early
89 termination of the transcript due to the mutation or the altered splicing of these exons from the
90 transcript (Fig 2d). Differential expression was most pronounced in exon 12, in which the edited
91 cell line had 62.5% lower fold expression of the exon than the wildtype.
92
93 Mutations to the *IFT80* gene are linked to asphyxiating thoracic dystrophy 2 (ATD2) disease in
94 human[9] and hypomorphic expression of the gene has been shown to result in low postnatal
95 survival in gene-trap mice[10]. Previous surveys have speculated as to when *IFT80* gene expression
96 is essential during development[10], and our results suggest that expression is needed to progress
97 past the 8- to 16-cell developmental phase. This is consistent with previous RNA surveys of
98 cattle preimplantation embryonic development[11], and with the known fertility effects of HH2 in
99 cattle[1,3]. The results of editing experiments suggest that the carboxy-terminal WD 40-repeat or
100 the polyampholyte domains are essential for embryonic development. Recent studies suggest that
101 deletion of *IFT80* disrupt FGF2 signaling pathway, affecting cell proliferation, and self-renewal
102 of odontoblasts[12]. In the preimplantation mammalian embryo, FGF2 signaling is essential for
103 lineage commitment and blastocyst formation[13,14], in which disruption of FGF2 signaling
104 coincides with the developmental arrest observed in this study.
105
106 Using genotypes from millions of Holstein cattle and CRISPR-Cas9 gene editing, we were able
107 to track and validate a recessive lethal allele in the *IFT80* gene, respectively. This approach
108 conclusively identified the target mutation in such a way that it provided further insight into the
109 function of the gene in development of other mammalian species. There is a further economic

110 benefit to this discovery, as identification of the exact causal mutation underlying this haplotype
111 will benefit dairy producers by allowing them to avoid carrier-to-carrier matings which result in
112 pregnancy losses responsible for ~$437,000/year in US Holsteins. Concordance analysis and
113 molecular validation on naturally occurring genetic variants represent a potent tool for ongoing
114 efforts to translate genotype to phenotype in the field of Quantitative Genetics.

115 **ONLINE METHODS**

116 **Haplotype discovery**

117 Possible autosomal recessive mutations are routinely tracked during the imputation step of the
118 national genomic evaluations calculated by the Council on Dairy Cattle Breeding (CDCB;
119 Bowie, MD). Genotypes from 47 different arrays are currently stored in the CDCB database,
120 with SNP counts ranging from 2,900 to 777,962, and are imputed to a set of 79,276 (80k) for
121 genomic evaluation. No single chip includes all these markers, regardless of density, and
122 genomic calculations require that genotypes are available for all 80k SNP. Chromosomes are
123 phased into maternal and paternal haplotypes and the genome is divided into 100-SNP segments.
124 Inheritance of those segments is traced through the population and missing SNP are filled in
125 using a segment from the pedigree when available, and using the most common segment in the
126 population when pedigree information is not available[3]. Most males have complete pedigrees
127 available for many generations, but many females do not. For this study, haplotyping was
128 performed using the program Findhap version 3[15]. Haplotypes that either never appear in the
129 homozygous state, or which appear with far lower frequency than expected, are flagged as
130 possible recessives[3]. Such haplotypes are then tested for effects on fertility and stillbirth.
131 Founder animals are identified from the pedigree when possible. Currently, 26 such conditions
132 are tracked in the U.S. dairy cattle population

133 (https://www.ars.usda.gov/ARSUserFiles/80420530/Publications/ARR/Haplotype tests_ARR-
134 Genomic5.pdf), and causal mutations are known for all but the HH2 haplotype.
135
136 In April of 2011 a haplotype affecting fertility was identified on *Bos taurus* autosome (BTA) 1 in
137 the region of 92–97 Mbp and designated as HH2 (Figure 1a). At that time, a second haplotype on
138 BTA 1 at 101–106 Mbp with a similar carrier list, lack of homozygotes and negative effects on
139 fertility also was identified. When a new SNP list was introduced in August 2011 the haplotype
140 in the 101–106 Mbp region was no longer flagged as a potential lethal. A new 61,000 SNP list
141 was introduced in December 2013, and the haplotype segment upstream from HH2 again
142 appeared on the list of possible lethal haplotypes. In April 2020 the location of HH2 was
143 changed to correspond with the upstream segment based on results from fine-mapping described
144 below. Haplotype blocks are formed using fixed numbers of SNP, rather than LD boundaries, so
145 it can sometimes be difficult to narrow the candidates down to a single haplotype. In this case,
146 both segments appear to track the causal variant with high concordance.

147 **Whole-genome sequencing, variant calling, and annotation**

148 The dataset used for causal variant discovery and validation included whole-genome sequence
149 data for 9 carriers and 763 non-carriers, including animals from the Ayrshire, Holstein, and
150 Jersey breeds sequenced by the Animal Genomics and Improvement Laboratory (Beltsville,
151 MD), the Collaborative Dairy DNA Repository (Madison, WI), and the 1000 Bull Genomes
152 Project[7]. Sequencing depth for these animals averaged 14.7x. Sequence data were aligned to the
153 ARS-UCD1.2 reference genome[4] using BWA-MEM as implemented in bwa version 0.7.15-
154 r1140[16] and converted to the binary Sequence Alignment/Map format by samtools version 1.3[17].
155 1000 Bull Genomes run8 data were processed using version 4 of the Genome Analysis Toolkit[18],

156 and included genotyping, variant recalibration for SNP and INDEL, removal of individuals that
157 failed QC, and filtering of monomorphic alleles. The resulting VCF file was filtered to remove
158 reads with QS < 10 and those identified as "LOWQUAL". Sequence data variants were
159 identified using the Samtools v1.9 and BCFtools v1.9[19] and reads passing edits were aligned
160 against version ARS-UCD1.2 of the bovine genome[4]. The list of filtered variants was annotated
161 with the SNPEff[20] utility using the Ensembl annotation of the cattle reference genome (version
162 101). Variants listed as "High", "Medium" and "Low" priority within the haplotype region by
163 SNPEff were selected for further analysis. Impacts on predicted protein translation were
164 identified using custom scripts.

165 **Production of IFT80$^{-/-}$ embryos**

166 **Guide RNA design**

167 Guide RNAs (gRNAs) were designed to target exon 2 (Supplementary Table 1), or exon 11
168 (Supplementary Table 2) using the GPP portal from the Broad Institute
169 (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design). To select gRNAs, two
170 criteria were used: (1) those with the highest on-target score after design based on the GPP portal
171 and the guide RNA checker portal (IDT technologies, San José, CA), and (2) those having the
172 least similarity with other bovine sequences containing a protospacer adjacent motif (PAM).
173 Selected gRNAs, as well as the universal 67mer tracrRNA® for form guide complexes were
174 ordered from IDT technologies.

175

176 The gRNAs and tracrRNA were resuspended to a concentration of 500 ng/µl. Each guide was
177 then annealed with equal amounts of tracrRNA by heating at 95 °C for 5 min and then cooled
178 down from 95 to 25 °C using a gradient of -1 °C/12 s in a Veriti 96-well gradient thermal cycler

179 (ThermoFisher, CA) to obtain a concentration of 250 ng/µl gRNA-tracrRNA complex. Prior to
180 zygote injection, gRNA-tracrRNA complexes were annealed to CAS9 mRNA (IDT
181 Technologies) by incubation at room temperature for 15 min. The final injection solution was 50
182 ng/µl each gRNAs complex and 20 ng/µl Cas9 mRNA. For each exon targeted, guides were
183 injected in pairs, to create a sizeable cut in the sequence and facilitate rapid genotyping on an
184 agarose gel. The expected cuts for guides targeting exon 2 was 82 bp and for those targeting
185 exon 11 were expected to generate a cut of 331 bp.

186 **Production of embryos in vitro and zygote microinjection**

187 Embryos were produced *in vitro* with a single sire known to be of high fertility *in vitro* as
188 previously described[21]. At the end of fertilization, putative zygotes (oocytes exposed to sperm)
189 were denuded from the surrounding cumulus cells and split in two groups: injected only with
190 CAS9 mRNA and injected with gRNA/CAS9 mRNA against *IFT80*. During the injection
191 procedure zygotes were maintained manipulation medium[22], and then transferred to four-well
192 dishes in groups of up to 50 zygotes in 500 µl of SOF-BE2[23], covered with 300 µl of mineral oil
193 per well at 38.5 °C in a humidified atmosphere of 5% (v/v) O_2 and 5% (v/v) CO_2. Percentage of
194 putative zygotes that cleaved was determined at day 3 of development (day 0 = day of
195 insemination) and blastocyst rate was estimated at day 8 of development. In addition, to
196 determine any developmental arrest, cell stage of all embryos was recorded at days 3, 5 and 8 of
197 development (Supplementary Figure 2).

198 **Editing efficiency**

199 To determine editing efficiency of the guides, embryos subjected to microinjection were
200 collected individually at day 8 of culture, placed in 6 µL of Embryo Lysis Buffer[22], and

201 subjected to 30 min at 60ºC, followed by 10 min at 85 ºC in a Veriti 96-well gradient thermal
202 cycler (ThermoFisher, CA). Lysed embryo solution was used as template DNA.
203 For genotyping, the region surrounding the target sequenced of the gRNAs was amplified by
204 endpoint PCR using the following primers (5'→3'): exon 2 F-
205 GGTTTCTTATCCTGCTTTCCATTC; R- GAAATTGAGTGTGAACCTTGGG, and for Exon
206 11 F- CACTGTTTAGGACTCTGCCT; R- CTCTCTGAGTAATGATACCATAGCA. The PCR
207 products were 639- and 620 bp for exon 2 (Supplementary Figure 3) and 11 (Supplementary
208 Figure 4), respectively. The PCR reaction, amplification (annealing temperature 53.4 °C), and
209 visualization of bands was performed as previously described[22]. Editing efficiency was
210 calculated from 3 separate *in vitro* embryo production runs, with 100 embryos per treatment
211 group per replicate. Editing efficiency for exon 2 was 26% bi-allelic, 32 % mono-allelic, and
212 42% non-edited, and for exon 11 was 50% bi-allelic, 23% mono-allelic, and 27% non-edited.
213 **Nucleic acids isolation and RT-qPCR expression analysis**
214 DNA and RNA were isolated from individual 8-cell embryos (65–72 h post insemination)
215 injected against *IFT80* or injected only with CAS9 mRNA. Embryos were collected and lysed
216 individually in 25 µl of RLT Lysis buffer (Qiagen, Valencia, CA), lightly vortexed, transferred
217 to a nano column (PuroSPIN™ Luna NANOTECH, Toronto, ON, CA), and centrifuged for 30 s
218 (all centrifugations were performed at 16,000 g). At this point, DNA is trapped in the column
219 and the flow-through contained the RNA. The column was placed in a new collection tube and
220 stored at room temp for later DNA purification. The flow-through containing RNA was mixed
221 with one volume of 70% ethanol, transferred to a new nano column, and centrifuged for 30 s.
222 Immediately after, 600 µl of buffer RW1 (Qiagen) were added to the column, and centrifuged for
223 30 s. Flow-through was discarded from the collection tube and 500 µl of buffer RPE (Qiagen)

224 were added to the column and centrifuged for 30 s. After discarding the flow-through, 500 µl of
225 80% ethanol were added to the column, and centrifuged for 2 min. To elute RNA, the column
226 was placed in a new microcentrifuge tube, and 15 µl of RNAse-free water were added directly to
227 the membrane of the column, incubated at room temperature for one minute and finally
228 centrifuged for 1 min. To complete DNA isolation, 500 µl of 70% ethanol were added to the
229 nano column containing DNA, followed by a 1-min centrifugation. Then, the column was placed
230 in a new microcentrifuge tube and 15 µl of EB buffer (Qiagen) preheated at 70 ºC were added
231 directly to the membrane of the column, incubated at room temperature for 10 min, followed by
232 1 min centrifugation to elute DNA. Embryos edit status was first assessed using PCR across the
233 site of the edit as described above, and the RNA from homozygous edited embryo cells were
234 selected for further analysis. Due to the small amount of total RNA in each cell, homozygous
235 embryos were pooled in groups of four prior to cDNA creation. Three pools of edited and control
236 embryos were analyzed in subsequent analyses. Two strand cDNA synthesis was conducted
237 using the high-capacity cDNA Reverse transcription kit (ThermoFisher) reverse transcriptase
238 following manufacturer instructions. Primers sequences for each locus were designed using the
239 PrimerQuest™ (IDT technologies). All primers were validated following procedures previously
240 described[24]. QPCR was conducted using the CFX384 Touch Real-Time System (Bio-Rad) using
241 10 µl of reaction mix as previously described[22]. The gene *SDHA* was chosen as a suitable
242 housekeeping gene for relative expression calculation in bovine embryos[25]. Fold changes were
243 calculated relative to the housekeeping gene ($2^{\Delta CT}$) as previously described[22,24].

244 **Statistics**

245 Differences in gene expression and development were analyzed by least-squares analysis of
246 variance of the ΔCT values using the GLM procedure of the Statistical Analysis System version

9.4 (SAS Institute Inc., Cary, NC, USA). Replicate and treatment were included as main effects in the model.

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 United States Department of Agriculture (USDA). 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 under USDA Agricultural Research Service (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 was provided under USDA-ARS Material Transfer Agreement 14358. Authors also thank Bethany Redel and Joshua Benne from the University of Missouri 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.

Author Contributions

Cole and Ortega conceived the project. Null and Hutchison identified and annotated variant sites within the haplotype region. Ortega and Clark designed CRISPR-Cas9 gRNAs, generated edited embryos, and performed analysis on resulting embryos. Null and Bickhart ran analysis on predicted variant sites and prepared figures for publication. McClure assisted with image preparation. Ortega and Bickhart wrote the manuscript. All authors read and edited the manuscript.

References

1. 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 (2016).
2. Cole, J. B. *et al.* AIP Research Report Genomic5: Haplotype tests for economically important traits of dairy cattle (update). https://www.ars.usda.gov/ARSUserFiles/80420530/Publications/ARR/Haplotype%20tests_ARR-Genomic5.pdf (2020).
3. VanRaden, P. M., Olson, K. M., Null, D. J. & Hutchison, J. L. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* **94**, 6153–6161 (2011).
4. Rosen, B. D. *et al.* De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* **9**, (2020).
5. Zimin, A. V. *et al.* A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biol.* **10**, R42 (2009).

293 6. VanRaden, P. M. *et al.* Genomic imputation and evaluation using high-density Holstein
294 genotypes. *J. Dairy Sci.* **96**, 668–678 (2013).
295 7. Hayes, B. J. & Daetwyler, H. D. 1000 Bull Genomes Project to map simple and complex
296 genetic traits in cattle: Applications and outcomes. *Ann. Rev. Anim. Biosci.* **7**, 89–102 (2019).
297 8. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths
298 toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13
299 (2009).
300 9. Beales, P. L. *et al.* IFT80, which encodes a conserved intraflagellar transport protein, is
301 mutated in Jeune asphyxiating thoracic dystrophy. *Nat. Genet.* **39**, 727–729 (2007).
302 10. Rix, S., Calmont, A., Scambler, P. J. & Beales, P. L. An Ift80 mouse model of short rib
303 polydactyly syndromes shows defects in hedgehog signalling without loss or malformation
304 of cilia. *Hum. Mol. Genet.* **20**, 1306–1314 (2011).
305 11. Graf, A. *et al.* Fine mapping of genome activation in bovine embryos by RNA sequencing.
306 *Proc. Natl. Acad. Sci. USA* **111**, 4139–4144 (2014).
307 12. Yuan, X., Cao, X. & Yang, S. IFT80 is required for stem cell proliferation, differentiation,
308 and odontoblast polarization during tooth development. *Cell Death & Disease* **10**, 1–11
309 (2019).
310 13. Lanner, F. & Rossant, J. The role of FGF/Erk signaling in pluripotent cells. *Development*
311 **137**, 3351–3360 (2010).
312 14. Kuijk, E. W. *et al.* The roles of FGF and MAP kinase signaling in the segregation of the
313 epiblast and hypoblast cell lineages in bovine and human embryos. *Development* **139**, 871–
314 882 (2012).

315 15. VanRaden, Paul M., O'Connell, Jeffrey R., Wiggans, G. R. & Weigel, K. A. Genomic
316 evaluations with many more genotypes. *Genet. Sel. Evol.* **43**, 10 (2011).
317 16. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
318 *arXiv:1303.3997 [q-bio]* (2013).
319 17. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–
320 2079 (2009).
321 18. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing
322 next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
323 19. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008
324 (2021).
325 20. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide
326 polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-
327 2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
328 21. Ortega, M. S. *et al.* Influences of sire conception rate on pregnancy establishment in dairy
329 cattle. *Biol. Reprod.* **99**, 1244–1254 (2018).
330 22. Ortega, M. S. *et al.* NANOG is required to form the epiblast and maintain pluripotency in the
331 bovine embryo. *Mol. Reprod. Devel.* **87**, 152–160 (2019).
332 23. Tríbulo, P., Rivera, R. M., Ortega Obando, M. S., Jannaman, E. A. & Hansen, P. J.
333 Production and Culture of the Bovine Embryo. in *Comparative Embryo Culture* (ed. Herrick,
334 J. R.) vol. 2006 115–129 (Springer New York, 2019).
335 24. Ortega, M. S., Kurian, J. J., McKenna, R. & Hansen, P. J. Characteristics of candidate genes
336 associated with embryonic development in the cow: Evidence for a role for WBP1 in
337 development to the blastocyst stage. *PLOS ONE* **12**, e0178041 (2017).

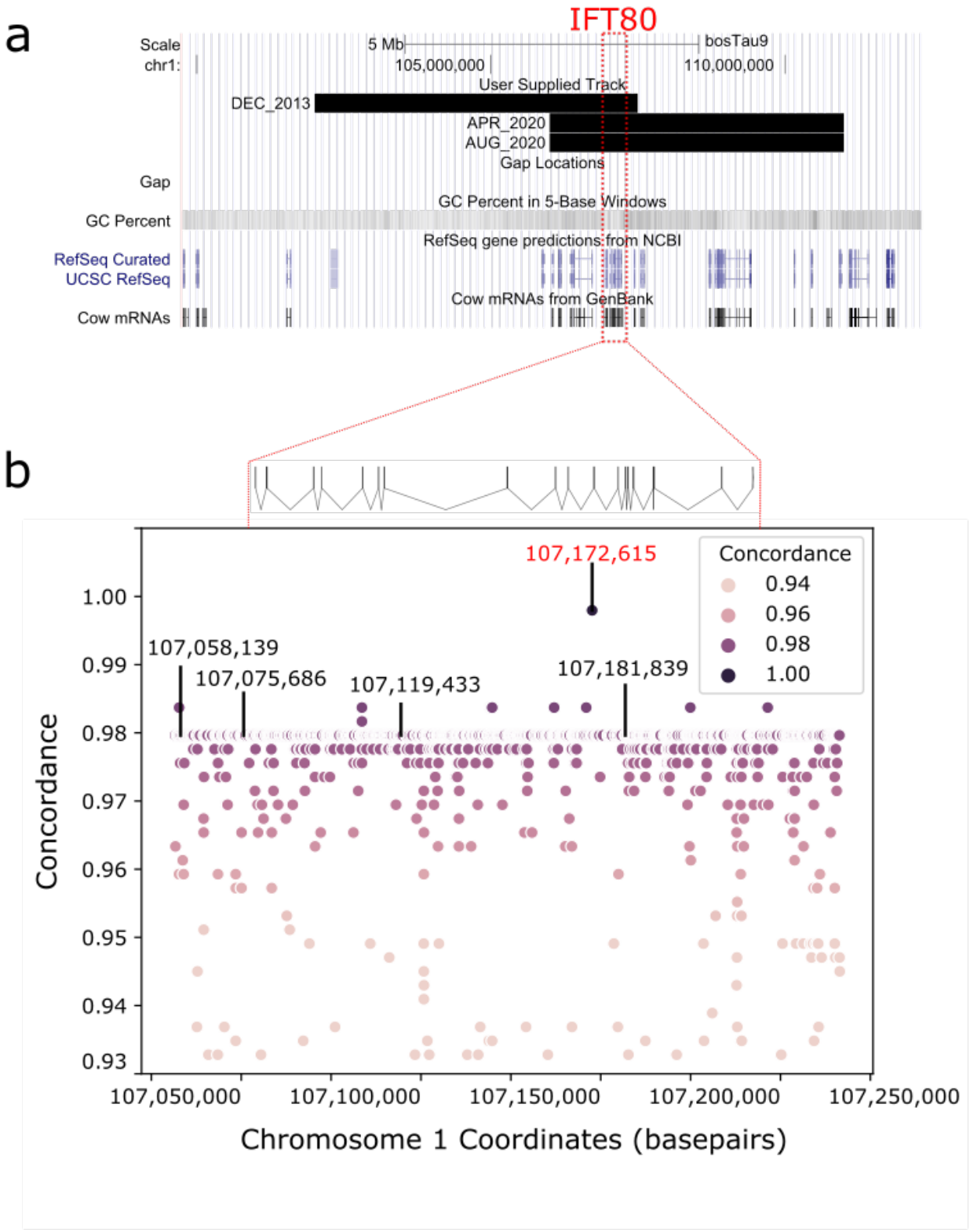
338 25. Goossens, K. *et al.* Selection of reference genes for quantitative real-time PCR in bovine
339 preimplantation embryos. *BMC Devel. Biol.* **5**, 27 (2005).
340

341 **Figure Legends**

342 Figure 1: A UCSC genome browser plot (a) shows the locations of prior Holstein Haplotype 2
343 (HH2) loci as determined in December 2013, April 2020 and August 2020. The *IFT80* gene is
344 highlighted by red dashed lines. Concordance analysis of variant sites against expected carrier
345 status (b) identified only one variant with > 99% concordance (highlighted in red). For genomic
346 context, the *IFT80* gene is drawn with exons (straight lines) and introns (diagonal lines) above
347 the concordance plot. Only variants predicted to have high functional effect by SNPEff are
348 highlighted by labels showing coordinate numbers.
349

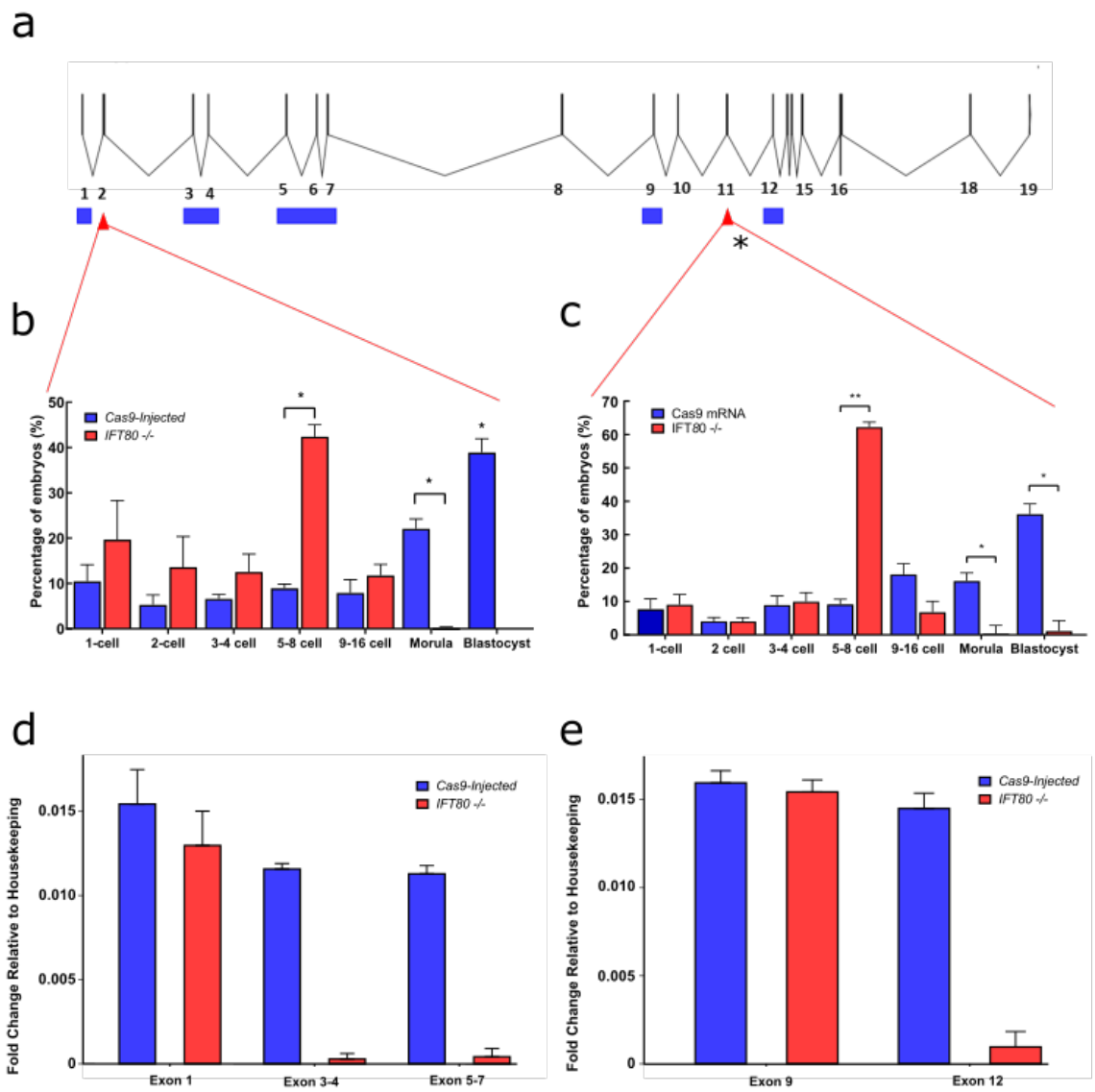
350 Figure 2: A diagram of the *IFT80* gene (a) shows CRISPR-Cas9 edit sites (red triangles) and
351 qPCR primer design sites (blue boxes). Gene exons are represented by vertical lines, whereas
352 introns are represented by diagonal lines. Embryo counts for each developmental stage for exon
353 2 (b) and exon 3 (c) demonstrate that homozygous edit embryos (red) do not typically progress
354 past the 8-cell stage compared to vector Cas9 controls (blue). Single (*) and double (**)
355 asterisks indicate Student's T test p-values that are less than 0.001 and 0.0001, respectively.
356 Relative fold change of exons downstream of the exon 2 (d) and exon 11 (e) edit sites showed
357 significant decreases in edited embryos (red) compared to the vector Cas9 controls (blue).

358 Figure 1



359

Figure 2


a
1 2
3 4
5 6 7
8
9 10
11
12 15 16
18
19
*
b
Cas9-Injected
IFT80 -/-
Percentage of embryos (%)
1-cell
2-cell
3-4 cell
5-8 cell
9-16 cell
Morula
Blastocyst
c
Cas9 mRNA
IFT80 -/-
Percentage of embryos (%)
1-cell
2 cell
3-4 cell
5-8 cell
9-16 cell
Morula
Blastocyst
d
Cas9-Injected
IFT80 -/-
Fold Change Relative to Housekeeping
Exon 1
Exon 3-4
Exon 5-7
e
Cas9-Injected
IFT80 -/-
Fold Change Relative to Housekeeping
Exon 9
Exon 12