RVC OPEN ACCESS REPOSITORY – COPYRIGHT NOTICE

This is the peer-reviewed, manuscript version of the following article:

Adams, H. A., Sonstegard, T. S., VanRaden, P. M., Null, D. J., Van Tassell, C. P., Larkin, D. M. and Lewin, H. A. (2016) 'Identification of a nonsense mutation in APAF1 that is likely causal for a decrease in reproductive efficiency in Holstein dairy cattle', *Journal of Dairy Science*, 99(8), 6693-6701.

The final version is available online: http://dx.doi.org/10.3168/jds.2015-10517.

© 2016. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/.

The full details of the published version of the article are as follows:

TITLE: Identification of a nonsense mutation in APAF1 that is likely causal for a decrease in reproductive efficiency in Holstein dairy cattle

AUTHORS: Adams, H. A., Sonstegard, T. S., VanRaden, P. M., Null, D. J., Van Tassell, C. P., Larkin, D. M. and Lewin, H. A.

JOURNAL: Journal of Dairy Science

PUBLISHER: Elsevier

PUBLICATION DATE: August 2016

DOI: 10.3168/jds.2015-10517

RVC Royal Veterinary College University of London

1 **Identification of a nonsense mutation in *APAF1* that is likely causal for a decrease in**
2 **reproductive efficiency in Holstein dairy cattle**

3 Heather A. Adams[*†], Tad S. Sonstegard[‡], Paul M. VanRaden[§], Daniel J. Null[§], Curt P. Van
4 Tassell[‡], Denis M. Larkin[*], and Harris A. Lewin[*†#1]
5
6 [*]Department of Animal Sciences, and
7 [†]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana 61801
8 [‡]Bovine Functional Genomics Laboratory, and
9 [§]Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville,
10 MD 20705
11 [#]Department of Evolution and Ecology and the UC Davis Genome Center, University of
12 California, Davis, Davis 95616
13 [1]Corresponding author: lewin@ucdavis.edu

14 **ABSTRACT**

15 The HH1 haplotype on chromosome 5 is associated with reduced conception rate and a
16 deficit of homozygotes at the population level in Holstein cattle. The source HH1 haplotype was
17 traced to the bull Pawnee Farm Arlinda Chief (Chief), who was born in 1962 and sired more than
18 16,000 daughters. We identified a nonsense mutation in *APAF1* (*APAF1 p.Q579X)* within HH1
19 using whole-genome resequencing of Chief and three of his sons. This mutation is predicted to
20 truncate 670 amino acids (53.7) percent of the encoded *APAF1* protein that contains a WD40
21 domain critical to protein-protein interactions. Initial screening revealed no homozygous
22 individuals for the mutation in 758 animals previously genotyped, whereas all 497 HH1 carriers
23 possessed one copy of the mutant allele. Subsequent commercial genotyping of 246,773

Holsteins revealed 5,299 *APAF1* heterozygotes and zero homozygotes for the mutation. The causative role of this mutation is also supported by functional data in mice that has demonstrated *Apaf1* to be an essential molecule in the cytochrome-c mediated apoptotic cascade and directly implicated in developmental and neurodegenerative disorders. In addition, most *Apaf1* homozygous knock-outs die by day 16.5 of development. We thus propose that the *APAF1 p.Q579X* nonsense mutation is the functional equivalent of the *Apaf1* knockout. This mutation has caused an estimated 525,000 spontaneous abortions world-wide over the past 35 years, accounting for approximately $420 million in losses. With the mutation identified, selection against the deleterious allele in breeding schemes has aided in eliminating this defect from the population, reducing carrier frequency from 8% in past decades to 2% in 2015.

Keywords: nonsense mutation, *APAF1*, dairy cattle, resequencing

35 # INTRODUCTION

36 Fertility is one of the most important traits determining the sustainability of animal
37 agriculture. For example, the milk produced as a result of a successful pregnancy, and the
38 number of complete gestations in a cow's lifetime, are essential to maintaining the profitability
39 of a modern dairy operation. If reproductive performance decreases, financial losses accrue
40 because of reduced milk production of the herd and the need for replacement animals to maintain
41 herd size. The relative importance of fertility and associated lactation traits to the dairy industry
42 has made reproductive performance an important target in dairy cattle breeding programs (Lucy,
43 2001; Shook, 2006). However, after more than 50 years of using quantitative genetics for
44 genetic improvement, traits related to reproductive performance have been difficult to select for
45 because of low heritability (VanRaden et al., 2004). As an added complexity, inbreeding reduces
46 reproductive efficiency, so intensive efforts have been made to reduce mating between relatives
47 by using pedigree information in breeding decisions. Despite these efforts, the extensive use of
48 artificial insemination (**AI**) in the dairy industry has resulted in an increase in the inbreeding
49 coefficient of the national herd to about 6 percent
50 (https://www/cdcb.us/eval/summary/inbrd.cfm).

51 Inbreeding can rapidly increase the frequency of recessive lethal and sub-lethal alleles in
52 the population, but can also lead to purging of harmful alleles by natural selection. Lethal
53 mutations are the most detrimental economically because no offspring carrying such mutations
54 will survive to reproduce. Some of the top bulls in the history of the dairy industry have been
55 shown to harbor mutations affecting fertility, which spread rapidly in the population through AI.
56 In recent years, a number of these mutations have been identified (see Online Mendelian
57 Inheritance of Animals (http://omia.angis.org.au/). Examples of recessive autosomal lethal or

58 sub-lethal fertility mutations include deficiency of uridine monophosphate synthase (**DUMPS**)
59 (Robinson et al., 1984; Shanks and Robinson, 1989), complex vertebral malformation (**CVM**)
60 (Agerholm et al., 2001; Agerholm et al., 2004), and brachyspina (Charlier et al., 2012). These
61 disorders are caused by point mutations, insertions and deletions that result in aborted fetuses or
62 stillbirths. A recently discovered deletion in Nordic Red cattle (Kadri et al., 2014) causes not
63 only recessive fertility loss but also increased milk yield, maintaining high frequency due to
64 balancing selection. Identification of such mutations has enabled screening programs to avoid
65 matings between carriers from within the population.

66 VanRaden et al. (2011b) used high density SNP genotyping to identify a haplotype on
67 chromosome 5 (BTA5), named HH1, that was associated with a decrease in conception rates and
68 an increase in stillbirths in Holstein cattle. No individuals homozygous for HH1 were found
69 among >78,000 individuals genotyped, despite a haplotype frequency of 2.25%. This led to the
70 hypothesis that a recessive lethal allele located in an 8 Mbp region of BTA5 was circulating in
71 the population.

72 The HH1 haplotype was subsequently traced to a single sire born 50 years ago during the
73 early period of advanced animal breeding (VanRaden et al., 2011a). The present study reports
74 the identification of a stop-gain (nonsense) mutation in the apoptosis peptide activating factor 1
75 (*APAF1*) gene carried on the this bull's HH1 haplotype that is the likely mutation causing
76 reduced conception rate in the Holstein population. This work provides much more complete
77 documentation and validation of the *APAF1* mutation than earlier reports (Adams et al., 2012;
78 Fritz et al. 2013), including pedigree, laboratory, commercial, and across-species bioinformatic
79 validation, sequencing details, fine mapping, economic analysis, and estimates of allele
80 frequency change. Haplotype tests have reduced accuracy with each successive generation due

81 to recombination, whereas causative mutation tests improve breeder confidence in genetic
82 selection and are much simpler to use.
83

84 MATERIALS AND METHODS

85 *Calculations supporting Chief's influence on spontaneous abortions*

86 Using pedigree data, VanRaden et al. (2011a) identified Pawnee Farm Arlinda Chief
87 (Chief), born in 1962 (**Table 1**), as the earliest genotyped ancestor carrying the HH1 haplotype.
88 Chief is one of the most influential sires in the history of the Holstein breed, having produced
89 many sons that became popular sires in addition to >16,000 daughters, >500,000 granddaughters,
90 and >2 million recorded great-granddaughters. We estimated the cumulative number of
91 spontaneous abortions caused by HH1 over the 30 years since Chief alleles became highly
92 frequent to be more than 100,000 nationally and nearly 500,000 worldwide as follows: the
93 estimated cumulative number of spontaneous abortions caused by HH1 in the US Holstein
94 population was calculated as the number of cows (8 million) * .045*(1/2) * .045*(1/2) = 4,050
95 per year, where .045 is the HH1 carrier frequency of one parent, and 1/2 is the probability that
96 the parent will contribute the defect to the offspring. Globally, the estimate is 15,000 per year
97 based on a population size of 30 million. These estimates correspond to approximately 140,000
98 spontaneous abortions in the US and 525,000 world-wide over a 35 year period, the approximate
99 time that Chief alleles appeared on both sides of the pedigree.
100 Carrier frequency in U.S. Holsteins exceeded 0.08 during the 1980's and 1990's, but
101 dropped to about 0.03 in 2010 (VanRaden et al., 2011b). The actual mating pattern for 58,453
102 genotyped Holsteins was used by VanRaden et al. (2011b) who reported 23 expected HH1
103 homozygotes compared to 30 expected when assuming random mating. Use of the actual mating

104 pattern is difficult for national or international populations because many ancestors are either not
105 known or not genotyped. Further direct selection for fertility and against HH1 reduced the
106 frequency to 0.02 in 2015.

107 The economic loss from a mid-term abortion is estimated to be about $800 (Norman et
108 al., 2012), for a total cost of ~$420 million. For comparison, the increased value of milk from
109 using Chief instead of an average bull in 1962 is his genetic contribution to the breed (.143) * the
110 farm price of milk ($0.33/liter) * the increased milk yield (2 liters/day) * 305 days/year * 35
111 years * 30 million cows = $30 billion. Embryonic and fetal loss during gestation was
112 investigated using the national fertility database and occurred mainly from 60 to 200 d of
113 gestation for HH1, but earlier for several other recessive defects (VanRaden et al., 2011a; see
114 also Norman et al., 2012 for more details on the timing of embryo loss, and Fritz et al., 2013 for
115 independent confirmation of fertility effects).
116

117 ***Haplotype detection and crossover analysis***

118 Recombinant haplotypes, defined as a portion but not all of Chief's HH1 source
119 haplotype, were detected within the pedigree of 78,465 animals that had 54,001 SNP genotypes
120 as of 2011 using findhap.f90 as previously described (VanRaden et al., 2011a; Sonstegard et al.,
121 2013). All copies of the 75-marker source haplotype spanning 7.1 Mbp that contained the
122 putative mutation appeared to trace to Chief and to no other prominent ancestors. VanRaden et
123 al. (2011b) studied only the source haplotype, whereas living animals with recombinant
124 haplotypes that are homozygous for only a portion of the source haplotype can rule out that
125 portion of the haplotype as not containing the lethal mutation. This fine mapping method is the
126 mirror image of typical homozygosity mapping, which focuses on the region of homozygosity

127 shared by affected animals instead of ruling out regions of the haplotype homozygous in
128 unaffected animals. After processing all recombinant haplotypes, the area not ruled out was
129 defined as the mutation-critical region, as described by Sonstegard et al. (2013).

130 Recombination events were detected in 78,465 animals genotyped for 43,385 SNPs from
131 the Illumina BovineSNP50 BeadChips (Illumina, San Diego, CA) using edits of Wiggans et al.
132 (2010), and standard output from findhap.f90 (VanRaden et al., 2011a) version 2, which first
133 examined haplotypes of length 600 markers, then 200 markers, and finally output haplotypes of
134 ≤75 markers. The program phases genotypes into haplotypes and detects recombination points
135 between the maternal and paternal haplotype of each genotyped parent. "Recombinant
136 haplotypes contain part of the source haplotype and part of a non-source haplotype, and a
137 descendant's phenotype status may be unknown when crossovers occur. Crossovers were
138 detected from genotypes by directly comparing progeny to parent haplotypes within the
139 pedigree. For each crossover, the last marker known to be from the first parental haplotype and
140 the first marker known to be from the second parental haplotype are output. A gap may remain
141 between those two markers if the parental haplotypes are identical in that region, some genotypes
142 are not called, or both parents were heterozygous and alleles could not be phased leading to an
143 unknown crossover location. Because few dams are genotyped, crossovers occurring in maternal
144 ancestors are often undetected (Sonstegard et al., 2013).

145 Fine mapping was accomplished by checking for animals with both the original HH1
146 haplotype and a recombinant haplotype. Regions homozygous for a section of the source
147 haplotype were removed from consideration of harboring the causative HH1 mutation. For
148 example, if a live animal received the original HH1 haplotype from one parent and the left 20
149 markers of the HH1 haplotype from the other parent, the region containing those 20 markers was

150 removed from consideration, exactly as described in Sonstegard et al. (2013) for Jersey
151 haplotype 1. The frequency of HH1 heterozygotes in all animals genotyped was 3.2% for the
152 source haplotype and 4.5% when recombinant haplotypes were included. Upon identification of
153 the region of the HH1 haplotype including the potentially lethal allele, individuals were selected
154 for study on the basis of their relationship to Chief, allele carrier status, and overall relationship
155 (expected future inbreeding, **EFI**) to the U.S. Holstein breed.

156 Two sons of Chief, Walkway Chief Mark (Mark) and Milu Betty Ivanhoe Chief (Ivanhoe
157 Chief) were identified as carriers of HH1 on the basis of previous genotyping information
158 (VanRaden et al., 2011b). An additional son of Chief, S-W-D Valiant (Valiant), was found not
159 to carry the HH1 haplotype. Chief and Mark had complete genome sequence available from an
160 earlier study (Larkin et al. 2012) and were thus useful for a preliminary screen to identify
161 mutations on the HH1 haplotype. For the present study, additional sequencing was conducted
162 for Chief to increase coverage (see below; **Supplemental Table S1**), thus permitting greater
163 accuracy of base calling in and near the *APAF1* locus. Whole genome sequencing of Ivanhoe
164 Chief and Valiant were conducted as part of an independent project. For Ivanhoe Chief and
165 Valiant, only the sequence data at the *APAF1* locus was used in the present study to assist
166 validation of the inheritance of the *APAF1* mutation (see below). The whole genome sequences
167 of Ivanhoe Chief and Valiant will be reported elsewhere. The DNA sequences of Chief and three
168 of his sons in and near the *APAF1* locus permitted phasing of genotypes into haplotypes, and was
169 useful for designing the 12-marker *APAF1* confirmatory SNP panel (see below).
170

171 ***Sequencing, alignment and mapping***

172 The genomes of Chief, Ivanhoe Chief and Valiant were sequenced using sequencing by
173 synthesis chemistry on an Illumina HiSeq 2000 platform (Illumina Inc., San Diego, CA).
174 Libraries were prepared from 5 μg of genomic DNA purified from semen straws and data was
175 generated using standard sequencing protocols provided by the manufacturer. Previous
176 sequencing results of Mark (12X) and Chief (6X) using 454 Titanium technology were also used
177 (Larkin et al., 2012).

178 Sequence reads were mapped against a whole bovine genome assembly (Btau 4.0) using
179 SOAP2 (Li et al., 2009). Sequence reads were paired-end and 100 bp in length, and in
180 accordance with the mismatch criteria chosen (5 mismatches alloted), only reads with $\geq$ 95%
181 coverage in chromosome sequences were considered mapped. Quality filters included accepting
182 those bases where the quality score (phred-scaled) was > 20 (corresponding to a 1% error rate),
183 and read depth was > 3X coverage per site per individual in each direction. Among mapped
184 reads, those with single matches in chromosome sequences (excluding unassigned contigs), and
185 those identified as a best SOAP2 hit in a chromosome sequence that was better than any other
186 hit, were compiled for subsequent SNP detection.

187

188 ***Detection of SNPs and genes***

189 The SNPs in the suspect region of BTA5 were identified using FreeBayes (Garrison and
190 Marth, 2012). Putative SNPs were accepted if they fit within the following criteria: 4x minimum
191 read coverage with at least two reads aligning in each orientation (forward, reverse), and
192 minimum allele sequencing quality $\geq$ 20. Upon acquiring a list of SNPs in the region, functional
193 annotation of the variants was performed using ANNOVAR (Wang et al., 2010). The
194 ANNOVAR program categorized SNPs by their genic or intergenic locations within the cattle

195 genome. The program reports SNPs located within introns and exons of annotated genes, 5' and
196 3' UTR regions, and those upstream and downstream of gene positions. All coordinates
197 pertaining to SNP and gene positions were converted from Btau4.0 to UMD3.1 genome
198 assemblies using the program LiftOver created by the UCSC Genome Bioinformatics Group
199 (http://genome.ucsc.edu/cgi-bin/hgLiftOver) for consistency with haplotype and genotype
200 datasets.
201
202 ***Selection of animals for APAF1 SNP validation***
203 Animals were selected for validation by querying a large database of 33,415 Holsteins
204 genotyped for 54,001 SNP as constructed previously (Wiggans et al., 2010). Genotype
205 imputation and haplotype frequencies included all 33,415 animals, but the 758 samples selected
206 for further validation were from the Cooperative Dairy DNA Repository, which contains DNA
207 from almost all progeny tested bulls in North America. Haplotype identification was based on
208 the 75 SNP markers designated as the 7.1 Mbp HH1-containing interval on BTA5 (UMD3.1
209 coordinates 58,638,702 to 65,743,920; VanRaden et al., 2011b). An additional query was
210 implemented to select a diverse set of non-carriers that had unique heterozygous haplotype
211 combinations in this interval.
212 A SNP genotyping panel (Sequenom Inc., San Diego, CA) designed for the validation
213 test (Page et al., 2004) was composed of 24 bi-directional assays for 12 putative SNPs in the
214 refined HH1 interval region. This included all SNPs with gene boundaries found within this
215 interval, as well as five additional SNPs observed near adjacent genes in the interval or in distal
216 flanking regions from the *APAF1* stop-gain mutation. A total of 22 of the 24 SNP assays were
217 functional; one SNP locus was monomorphic (**Table 2**). The call rate for all SNP loci was 100%

218 except for UMD3_63107293 (99.3%) and UMD3_62591311 (99.9%). Results from the bi-
219 directional assays for each SNP locus were compared for concordance and integrated into a
220 single marker genotype score for each animal across the 11 SNP loci. Haplotypes of 11
221 informative SNPs were determined by PHASE v2.1.1 (Stephens et al., 2001), and a total of 24
222 probable haplotypes were identified (**Table 3**). These haplotypes are much shorter and different
223 than those originally defined by the 75-marker window (derived from the 54,001 chip) spanning
224 7.1 Mbp that was used to find HH1. Two different numbering systems exist: one for the more
225 than 2,000 different haplotypes in this 7.1 Mbp window, and a second for the 24 haplotypes in
226 the narrow 11 SNP window for validation (**Table 3**).

227 In all, 758 animals were selected for validation genotyping using a 24-SNP (12x2)
228 multiplex panel, and 486 of these were presumed carriers based on the presence of haplotype 12,
229 which was the original designation for the corresponding HH1 haplotype within the 7.1 Mbp
230 interval. Among all animals within the validation set, 246 unique haplotypes existed within the
231 HH1 interval, as well as 323 heterozygous haplotype combinations. Additionally, animals
232 possessing "haplotype 32" (n=11) had a small region of 40 markers in a 1Mb region within the
233 HH1 interval that was expected to be equivalent to haplotype 12 because haplotype 32 was a
234 recombinant haplotype. These individuals were expected to test positive for the causal mutation
235 if the SNP was potentially associated with the recessive lethal effect. After re-genotyping the
236 animals for validation, one animal within the haplotype 12 group was found to have an incorrect
237 genotype (the DNA was actually from a different animal), and was removed from the study. In
238 total, 497 animals (485 with haplotype 12, and 11 with haplotype 32) were expected to be
239 heterozygous for the *APAF1* stop-gain mutation. Following this initial validation, a test for the
240 stop-gain mutation was added to the GeneSeek Genomic Profiler (GGP) BeadChip (GeneSeek-

241 Neogen, Lincoln, NE; Neogen Corp., 2013) and subsequent chips, and genotypes were received

242 for 246,773 Holsteins as part of routine genomic predictions.

243

244 ***Nomenclature***

245 The length of *APAF1* we originally reported in Adams et al. (2012) was 1238 aa, which is

246 the polypeptide length given by the UCSC Genome Browser (https://ucsc.edu/cgi-

247 bin/hgGateway). The UCSC Genome Browser annotation of the APAF1 protein

248 (NP_001178436.1) is based on the NCBI cattle genome annotation. Fritz et al. (2013)

249 apparently used either the human protein, which is 1248 aa, or an alternate annotation of the

250 bovine sequence (e.g., XP_005206703) that is 1248 aa. In UNIPROT, the cattle APAF1 protein

251 (Entry F1MUW4) is 1251 aa, which is caused by addition of 2 aa at the amino terminus and

252 insertion of another internal residue. Upon careful analysis, we have concluded that the

253 UNIPROT annotation is likely to be incorrect. We have also concluded that the 1238 aa

254 polypeptide represents a prediction error or an isoform produced by an alternatively spliced

255 product of the full length *APAF1* mRNA. To adopt the most likely length of the protein, and

256 avoid confusion in the literature, we have chosen 1248 aa for the APAF1 protein. Thus, the

257 nomenclature for the mutation used in this report is Q579X (or Gln579X), which is consistent

258 with the nomenclature described by Fritz et al. (2013).

259

RESULTS AND DISCUSSION

The mutation was fine mapped within the 75-marker HH1 haplotype to a 3.162 Mbp critical region on BTA5 (BTA5:62,435,307 to 65,597,776) bound by 39 SNPs. Fine mapping was based on eight live animals from a group of >78,000 genotyped animals that inherited both the HH1 source haplotype and an HH1-derived recombinant haplotype**.** In addition, 26 animals having one HH1 haplotype with a crossover within the 75-marker interval but outside the fine-mapped region were identified. Animals possessing these recombinant haplotypes were designated as putative carriers and used subsequently for the validation analysis (see below). Within the 39 marker-subinterval, 51 additional animals with crossovers inside the fine-mapped region were detected. The carrier status of these animals could not be fully ascertained at this stage**.**

Chief's genome was sequenced previously, along with his son Mark, because of his influence in the Holstein breed (Larkin et al., 2012). As part of an ongoing study that aims to sequence all influential bulls of the Holstein breed, we added additional sequence coverage to Chief and sequenced two other Chief sons, Ivanhoe Chief and Valiant (**Table 1**). Ivanhoe Chief, Mark and Valiant are also highly influential, having sired many high index bulls and >14,000, >57,000, and >36,000 daughters, respectively. Chief, Mark, Ivanhoe Chief, and Valiant were sequenced to ~31x, 13.5x, ~50x and ~36x coverage, respectively (**Supplemental Table S1**). Having the genome sequence of these bulls allowed us to search within the mapped interval for SNPs and candidate genes for the lethal allele. Normal and carrier progeny of Chief provide information by helping to phase the SNPs. As a preliminary step, alleles possessed by Chief that were inherited only by Valiant, a non-recombinant, non-carrier, were ruled out as causative. If

282 available, sequence for an animal homozygous for Chief's normal haplotype could also rule out
283 those mutations.
284 A list of SNPs within the critical region was acquired by alignment of sequence reads to
285 cattle genome reference assembly Btau4.0 and then converted to coordinates of assembly UMD
286 3.1. The SNPs were then annotated as genic, intergenic, synonymous, nonsynonymous, etc.,
287 using ANNOVAR. Analysis of Chief's DNA within the critical region revealed three unique
288 SNPs in exons while the rest were in intronic, intergenic, 3'UTR and downstream regions (**Table**
289 **4**). Among the exonic SNPs for which Chief was heterozygous, a C→T substitution in exon 11
290 of the gene encoding apoptosis peptide activating factor 1 (*APAF1*) produces a stop-gain
291 mutation at position 579 in the polypeptide (**Table 2; Figure 1**). The *APAF1 p.Q579X* mutation
292 truncates 670 C-terminal amino acids (53.7%) from the 1248 residue full-length *APAF1* protein.
293 Alignments of Mark and Ivanhoe Chief sequence reads, both deemed carriers of HH1, confirmed
294 that these Chief sons were heterozygous for the *APAF1 p.Q579X* mutation. Chief's son Valiant,
295 classified as a non-carrier, was found to be homozygous for the normal allele. The other two
296 exonic SNPs, one in *APAF1* and the other in a predicted gene, both cause synonymous
297 substitutions (**Table 2**), ruling out these mutations as causative. Thus, *APAF1 p.Q579X* was
298 determined to be the putative mutation causing prenatal lethality associated with the HH1
299 haplotype.
300 Validation of *APAF1 p.Q579X* as the likely causative mutation was accomplished by
301 querying a large inventory of archived cattle genomic DNA for all animals carrying the HH1
302 haplotype (see **Materials and Methods**). Among the 758 animals selected for validation
303 genotyping, 497 were presumed carriers of the mutation on the basis of HH1 haplotyping; the
304 rest did not carry the HH1 haplotype and were presumed to be non-carriers. Eleven informative

305 SNPs identified from Chief sequence that are located in the refined HH1 interval, including the
306 SNP producing the stop-gain mutation, were used for the validation studies (**Table 2**). Only one
307 haplotype among the 24 observed 11-SNP haplotypes was associated with the stop-gain mutation
308 and corresponded to the 497 previously identified HH1-positive individuals (**Table 3**).

309 Only one other SNP besides the *APAF1* stop-gain mutation was in high concordance with
310 HH1 (**Table 3**), defined as the absolute value of the percentage of SNP and HH1 genotypes that
311 matched for the 758 validation animals. This SNP is unlikely to be responsible for reduced
312 fertility because it is located in an intron of *SLC25A3* and had a 1.2% false positive detection rate
313 for HH1 haplotypes. In contrast, the stop-gain mutation at position UMD3_63150400 in *APAF1*
314 was 100% concordant with the recessive lethal (**Table 3**). Subsequent test results provided by
315 GeneSeek-Neogen Corp. (Lincoln, NE) for *APAF1 p.Q579X* identified 5,299 heterozygotes and
316 zero homozygotes in 246,773 Holsteins, consistent with the hypothesis that this is the lethal HH1
317 mutation.

318 Additional support for *APAF1 p.Q579X* as the causative mutation derives from functional
319 studies. The protein encoded by *APAF1* is a central component of the cytochrome-c-mediated
320 apoptotic cascade (Apweiler et al., 2004) and has been directly implicated in the etiology of
321 cancer, developmental disorders and neurodegenerative diseases (Honarpour et al., 2001; Blake
322 et al., 2011). The APAF1 protein forms an oligomer that when bound with cytochrome-c and
323 dATP forms the apoptosome, a cytoplasmic structure that binds the *caspase 9* preprotein and
324 cleaves it into its mature active form. The activated form of *caspase 9* initiaties the caspase
325 cascade that ultimately leads to apoptotic cell death. Expression of *Apaf1* during murine
326 development begins between days 7 and 9 in a number of vital tissues and organs, and is crucial
327 for the development of the central nervous system. Homozygous *Apaf1* gene knockout in mice

328 leads to embryonic lethality by day 16.5 or perinatally, and Apaf1-deficient mice exhibit severe
329 abnormalities such as brain overgrowth, persistence of interdigital webs and craniofacial
330 malformations (Cecconi et al., 1998; Yoshida et al., 1998; Honarpour et al., 2001; Muller et al.,
331 2005). Significantly, the deletion of 670 C-terminal amino acids from the APAF1 polypeptide
332 removes 15 WD40 repeats that form a predicted functional WD40 domain in the cattle protein.
333 WD40 domains are found in many proteins involved in signal transduction, transcriptional
334 regulation and apoptosis, and are essential for protein-protein interactions (Acehan et al., 2002).
335 Deletion of the WD40 domain would likely result in failure to form apoptosomes, which are
336 essential for binding of caspase 9 and initiating the apoptosis pathway (Acehan et al., 2002;
337 Riedl and Salvesen, 2007). The severely truncated *APAF1* peptide in homozygous cattle is likely
338 the functional equivalent of the homozygous *Apaf1* knockout in mice. These data strongly
339 support the *APAF1* stop-gain mutation *p.Q579X* as the causative mutation for embryonic, fetal
340 and perinatal loss of cattle homozygous for the HH1 haplotype.

341 An undetected insertion, deletion, or copy number variant carried on HH1 is another
342 possibility, but seems less likely than this obvious candidate. The validation animals were mostly
343 5 to 12 generations removed from Chief, whereas newborn calves may be >14 generations
344 removed, making haplotype detection more difficult, especially with lower density chips.
345 However, the stop-gain mutation is now included in routine genomic evaluation to improve the
346 accuracy of detecting both carrier status and genomic prediction of fertility.

347 Screening for loss of homozygosity is a powerful approach for the identification of
348 chromosomal segments associated with prenatal mortality (VanRaden et al., 2011b; Fritz et al.,
349 2013). We have shown here that detection of loss of homozygosity in combination with
350 moderate coverage whole genome resequencing can be used to rapidly identify causative

351 mutations for prenatal mortality, particularly if there is an available database of DNA sequences
352 of key individuals in animal pedigrees. Such databases are now being widely developed for
353 several cattle breeds (Fritz et al., 2013; Daetwyler et al., 2014; McClure et al., 2014), and the list
354 of haplotypes associated with loss of homozygosity is growing (for an updated list see
355 http://aipl.arsusda.gov/reference/recessive_haplotypes_ARR-G3.html). Identification of the
356 causative mutations for prenatal and perinatal mortality can be translated into genetic screens to
357 rapidly eliminate the unwanted alleles from the breeding population. In the case of *APAF1*, it
358 will now be possible to eliminate a mutation that is estimated to be causal for more than 500,000
359 abortions in Holstein cattle world-wide. Alternatively, these diagnostic tests can also be used to
360 avoid mating of carriers to avoid losing the more prevalent beneficial genetic contributions
361 derived from Chief, whose chromosomes contributed 14% of the current Holstein genome and
362 have been attributed to about $30 billion dollars in increased milk production.
363
364

365
ACKNOWLEDGMENTS

366 This work was supported by USDA, ARS CRIS projects 1265-31000-104-00D and 31S.
367 Thanks to A. Beavers, M. McClure for assistance, and GeneSeek-Neogen Corp. (Lincoln, NE)
368 for providing supporting genotype data. Mention of trade names or commercial products in this
369 article is solely for the purpose of providing specific information and does not imply
370 recommendation or endorsement by the US Department of Agriculture. The USDA is an equal
371 opportunity provider and employer.

372

373 **Table 1.** Pedigree information for four Holstein dairy bulls: Pawnee Farm Arlinda Chief (Chief),
374 Milu Betty Ivanhoe Chief (Ivanhoe Chief), S-W-D Valiant (Valiant), and Walkway Chief Mark
375 (Mark).

Bull	ID[1]	Birth Year	Sire	Dam	Pedigree EFI[2]	Ranking[3]	Alleles Shared (%)[4]
Chief	1427381	1962	Pawnee Farm Reflection Admiral (138326)	Pawnee Farm Glenvue Beauty (4546976)	7.1	5	14.3
Ivanhoe Chief	1578139	1969	Chief	Milu Betty Ovation Ivanhoe (5287566)	5.9	-	-
Valiant	1650414	1973	Chief	Allied Admiral Rose Vivian (6781299)	6.7	11	8.7
Mark	1773417	1978	Chief	Walkway Matt Mamie (8309147)	6.6	16	7.8

376 [1]Holstein breed identification number.
377 [2]Expected Future Inbreeding (**EFI**) of bull's daughters
378 (http://aipl.arsusda.gov/eval/summary/inbrd.cfm?R_Menu=HO.k#StartBody).
379 [3]Bull's EFI ranking, which outlines the relationship of the bull to the Holstein cow population
380 (http://aipl.arsusda.gov/eval/summary/inbrd.cfm?R_Menu=HO.k#StartBody).
381 [4]Percentage of alleles in common with top currently available Holstein sires; the value is an
382 indicator of the bull's overall contribution to the Holstein breed, from a list of the top 25 most
383 influential bulls.

384 **Table 2.** SNPs used for validation and concordance with carrier haplotypes.

Gene	SNP Location	UMD3.1 coordinate	Allele 1[1]	Allele 2[2]	Concordance[3]
ENSBTAG00000038223[4]	Intergenic	62591311	T	C	0.61
ENSBTAG00000038223	Intergenic	62756350	T	A	Not Informative
TMPO	UTR3	63051612	A	G	0.53
TMPO	UTR3	63052631	A	G	0.53
SLC25A3	Intronic	63088973	T	A	0.45
SLC25A3	Intronic	63091578	T	A	0.99
IKBIP	Intronic	63107293	T	C	0.45
APAF1	exonic/stop-gain	63150400	C	T	1.00
APAF1	exonic/synonymous	63198664	C	T	0.52
ENSBTAG00000017385	exonic/synonymous	63209396	C	T	0.64
ANKS1B	downstream	63228106	C	T	0.60
ANKS1B	intergenic	63486133	C	T	0.30

385 [1]Allele found in the reference sequence and found on Chief's haplotype carrying the normal
386 allele.
387 [2]Allele found on Chief's HH1 haplotype.
388 [3]Concordance is the comparison of the HH1 state to the alternative allele.
389 [4]Ensembl identification for predicted genes.
390
391

392 **Table 3.** Haplotypes of the 11 informative SNPs in HH1 validation region.

Haplotype[1]	Haplotype Count
000010-0-0011	7
000011-0-0001	5
000011-0-0111	9
000011-0-1000	64
000011-0-1001	243
000110-0-0001	60
011010-0-1001	1
011110-0-0001	4
011110-0-0011	21
011110-0-0111	179
011110-0-1000	1
011110-0-1001	25
011100-0-0111	9
100010-0-0011	7
100010-0-1001	4
100011-0-0111	22
100011-0-1001	82
100110-0-0001	69
111110-0-0001	1
111110-0-0011	34
111110-0-0110	1
111110-0-0111	117
111110-0-1001	53
111100-1-0111	498

393 [1]*APAF1* stop-gain mutation is the 7th marker of this haplotype, and designated as allele 1.

394

Table 4. Results of the functional annotation analysis using ANNOVAR (Wang et al., 2010).

BTA5 region [62,435,307 – 65,597,776][1]		SNPs[2]	Genes
	Downstream[3]	2	ANKS1B[2][4]
	Exonic	3	APAF1[2], ENSB_17385[1][5]
	Intergenic[6]	1,221	ENSB_38223, SP, SLC25A3, IKBIP, ENSB_17385, ANKS1B, ENSB_40364
	Intronic	354	SP[17], SLC25A3[2], IKBIP[9], APAF1[88], ENSB_17385[1]
	UTR3	2	SP[2]

[1]Coordinates for suspect region on BTA5 from UMD 3.1 genome assembly.

[2]SNPs indicate the number of SNPs identified within each region.

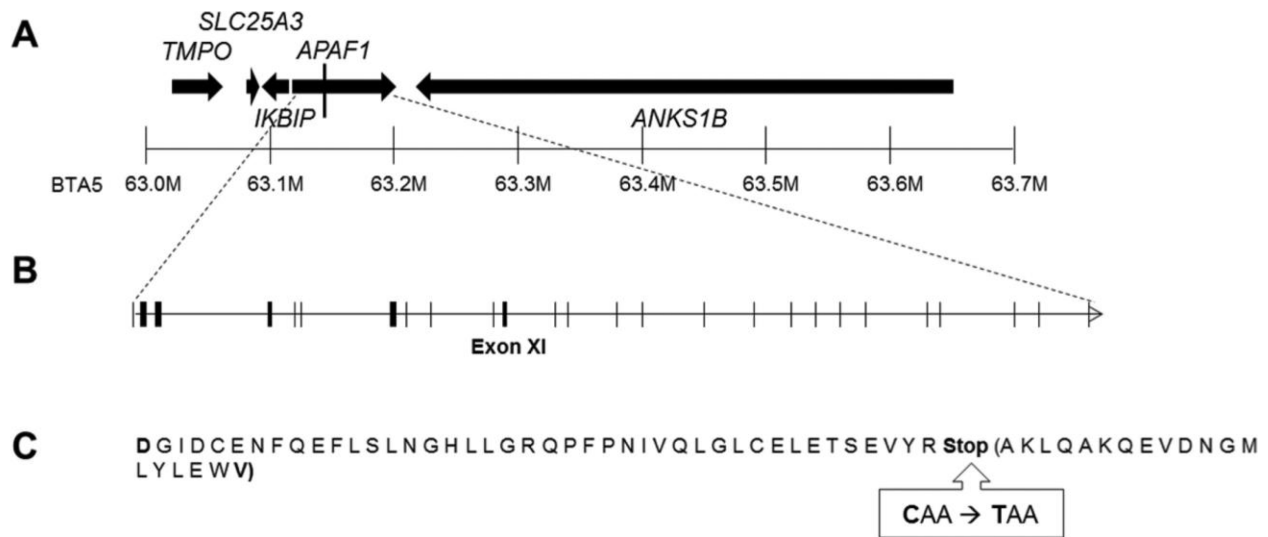
[3]Location of SNP within chromosome: downstream, variant overlaps 1-kb region downstream of transcription end site; exonic, variant overlaps a coding exon; intergenic, variant is in intergenic region; intronic, variant overlaps an intron; UTR3, variant overlaps a 3' untranslated region.

[4]Brackets indicate the number of SNPs associated with the given gene or predicted gene.

[5]'ENSB_' abbreviated from 'ENSBTAG000000' for all predicted genes listed.

[6]Genes listed for SNPs located within intergenic regions are the closest genes that flank the corresponding intergenic SNP.

405 Figure 1. Identification of the APAF1 p.Q579X mutation. (A) The critical region on BTA5 is
406 presented with the SNP-containing genes used in the validation analysis. Genes are shown with
407 arrows indicating their position and orientation. The vertical line within the APAF1 gene denotes
408 the position of the p.Q579X mutation; UMD 3.0 coordinates are included for positional
409 reference. (B) This schematic shows the gene structure for APAF1, including all exons marked
410 by vertical bars. Sequencing revealed a mutation in exon 11. (C) The AA sequence of APAF1
411 exon 11 showing the position where the stop-gain mutation terminates the polypeptide at residue
412 43 of exon 11. Amino acids within the parentheses are those truncated from exon 11. The
413 remainder of the 1,248–AA full-length APAF1 polypeptide is presumed deleted in individuals
414 with the p.Q579X mutation.



415

REFERENCES

Acehan D., X. Jiang, D.G. Morgan, J.E. Heuser, X. Wang and C.W. Akey. 2002. Three-dimensional structure of the apoptosome: implications for assembly, procaspase-9 binding, and activation. Mol. Cell 9: 423-432.

Adams H.A., T. Sonstegard, P.M. VanRaden, D.J. Null, C.P. Van Tassell, D.M. Larkin, and H.A. Lewin. 2012. Identification of a nonsense mutation in APAF1 that is causal for a decrease in reproductive efficiency in dairy cattle. Plant and Animal Genome Meeting, Poster P0555, June 14–18 2012, San Diego.

Agerholm J.S., C. Bendixen, J. Ambjerg, and O. Andersen. 2004. Morphological variation of “complex vertebral malformation” in Holstein calves. J. Vet. Diagn. Invest. 16: 548-553.

Agerholm J.S., C. Bendixen, O. Andersen, and J. Ambjerg. 2001. Complex vertebral malformation in Holstein calves. J. Vet. Diagn. Invest. 13: 283-289.

Apweiler R., A. Bairoch, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, et al. 2004. UniProt: the Universal Protein knowledgebase. Nuleic Acids Res. 32: D115-D119.

Blake J.A., C.J. Bult, J.A. Kadin, J.E. Richardson, and J.T. Eppig; Mouse Genome Database Group. 2011. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. Nucleic Acids Res. 39(suppl 1): D842-D848.

Cecconi F., G. Alvarez-Bolado,B.I. Meyer, K.A. Roth, and P. Gruss. 1998. Apaf1 (CED-4 homolog) regulates programmed cell death in mammalian development. Cell 94(6): 727-737.

446 Charlier, C., J.S. Agerholm, W. Coppieters, P. Karlskov-Mortensen, W. Li, G. de Jong, C.
447 Fasquelle, L. Karim, S. Cirera, N. Cambisano, N. Ahariz, E. Mullaart, M. Georges, and M.
448 Fredholm. 2012. A deletion in the bovine FANCI gene compromises fertility by causing
449 fetal death and brachyspina. PLoS ONE 7:e43085.
450
451 Daetwyler H.D., A. Capitan, H. Pausch, P. Stothard, R. van Binsbergen, R.F. Brondum, X. Liao,
452 A. Djari, S.C. Rodriguez, C. Grohs, D. Esquerre, O. Bouchez, M.N. Rossignol, C. Klopp,
453 D. Rocha, S. Fritz, A. Eggen, P.J. Bowman, D. Coote, A.J. Chamberlain, C. Anderson,
454 C.P. VanTassell, I. Hulsegge, M.E. Goddard, B. Guldbrandtsen, M.S. Lund, R.F.
455 Veerkamp, D.A. Boichard, R. Fries, and B.J. Hayes. 2014. Whole-genome sequencing of
456 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat. Genet. 46(8):
457 858-865.
458
459 Flisikowski K., H. Venhoranta, J. Nowacka-Woszuk, S.D. McKay, A. Flyckt, J. Taponen, R.
460 Schnabel, H. Schwarzenbacher, I. Szczerbal, H. Lohi, et al. 2010. A novel mutation in the
461 maternally imprinted PEG3 domain results in a loss of MIMT1 expression and causes
462 abortions and stillbirths in cattle (Bos taurus). PLoS One 5(11): e15116. doi:
463 10.1371/journal.pone.0015116.
464
465 Fritz S., A. Capitan, A. Djari, S.C. Rodriguez, A. Barbat, A. Baur, C. Grohs, B. Weiss, M.
466 Boussaha, D. Esquerré, C. Klopp, D. Rocha, and D. Boichard. 2013. Detection of
467 haplotypes associated with prenatal death in dairy cattle and identification of deleterious
468 mutations in GART, SHBG and SLC37A2. PLoS One 8(6): e65550. doi:
469 10.1371/journal.pone.0065550.
470
471 Garrison E. and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing.
472 Preprint at arXiv:1207.3907v2 [q-bio.GN].
473
474 Honarpour N., S.L. Gilbert, B.T. Lahn, X. Wang, and J. Herz. 2001. Apaf-1 deficiency and neural
475 tube closure defects are found in fog mice. Proc. Natl. Acad. Sci. USA 98(17): 9683-9687.
476

477 Kadri, N. K., G. Sahana, C. Charlier, T. Iso-Touru, B. Guldbrandtsen, L. Karim, U. S. Nielsen, F.
478 Panitz, G. P. Aamand, N. Schulman, M. Georges, J. Vilkki, M. S. Lund, and T. Druet.
479 2014. A 660-Kb deletion with antagonistic effects on fertility and milk production
480 segregates at high frequency in Nordic Red Cattle: Additional evidence for the common
481 occurrence of balancing selection in livestock. PLoS Genet 10:e1004049.
482
483 Larkin D.M., H.D. Daetwyler, A.G. Hernandez, C.L. Wright, L.A. Hetrick, L. Boucek, S.L.
484 Bachman, M.R. Band, T.V. Akraiko, M. Cohen-Zinder, et al. 2012. Whole-genome
485 resequencing of two elite sires for the detection of haplotypes under selection in dairy
486 cattle. Proc. Natl. Acad. Sci. USA 109(20): 7693-7698.
487
488 Li R., C. Yu, Y. Li, T.W. Lam, S.M.Yiu, K. Kristiansen, and J. Wang. 2009. SOAP2: an improved
489 ultrafast tool for short read alignment. Bioinformatics 25(15): 1966-1967.
490
491 Lucy M.C. 2001. Reproductive loss in high-producing dairy cattle: where will it end? J. Dairy Sci.
492 84: 1277–1293.
493
494 McClure M.C., D. Bickhart, D. Null, P. VanRaden, L. Xu, G. Wiggans, G. Liu, S. Schroeder, J.
495 Glasscock, J. Armstrong, J.B. Cole, C.P. VanTassell, and T.S. Sonstegard. 2014. Bovine
496 exome sequence analysis and targeted SNP genotyping of recessive fertility defects BH1,
497 HH2, and HH3 reveal a putative causative mutation in SMC2 for HH3. PLoS One 9(3):
498 e92769. doi:10.1371/journal.pone.0092769.
499
500 Muller M., J. Berger, N. Gersdorff, F. Cecconi, R. Herken, and F. Quondamatteo. 2005.
501 Localization of Apaf1 gene expression in the early development of the mouse by means of
502 in situ reverse transcriptase-polymerase chain reaction. Devel. Dynam. 234(1): 215-221.
503
504 Neogen Corporation. 2013. GeneSeek Genomic Profiler for Dairy Cattle. Accessed November
505 2013. http://www.neogen.com/Genomics/pdf/Slicks/GGP-LD_Dairy.pdf.
506

507 Norman H.D., R.H. Miller, J.R. Wright, J.L. Hutchison, and K.M. Olson. 2012. Factors associated
508 with frequency of abortions recorded through Dairy Herd Improvement test plans. J. Dairy
509 Sci. 95(7):4074-84.
510
511 Page B.T., E. Casas, R.L. Quaas, R.M. Thallman, T.L. Wheeler, S.D. Shackelford, M.
512 Koohmaraie, S.N. White, G.L. Bennett, J.W. Keele, et al. 2004. Association of markers in
513 the bovine CAPN1 gene with meat tenderness in large crossbred populations that sample
514 influential industry sires. J. Anim. Sci. 82: 3474-3481.
515
516 Riedl S.J. and G.S. Salvesen. 2007. The apoptosome: signaling platform of cell death. Nat. Rev.
517 Mol. Cell Biol. 8: 405-413.
518
519 Robinson J.L., D.B. Dombrowski, G.W. Harpestad, and R.D. Shanks. 1984. Detection and
520 prevalence of UMP synthase deficiency among dairy cattle. J. Hered. 75: 277-280.
521
522 Shanks R.D. and J.L. Robinson . 1989. Embryonic mortality attributed to inherited deficiency of
523 uridine monophosphate synthase. J. Dairy Sci. 72: 3035-3039.
524
525 Shook G.E. 2006. Major advances in determining appropriate selection goals. J. Dairy Sci. 89:
526 1349–1361.
527
528 Stephens M., N.J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype
529 reconstruction from population data. Amer. J. Human Genet. 68: 978-989.
530
531 Sonstegard T.S., J.B. Cole, P.M. VanRaden, C.P. Van Tassell, D.J. Null, S.G. Schroeder, D.
532 Bickhart, and M.C. McClure. 2013. Identification of a nonsense mutation in CWC15
533 associated with decreased reproductive efficiency in Jersey cattle. PLoS One. 8(1):
534 e54872. doi:10.1371/journal.pone.0054872.
535
536 VanRaden P.M., D.J. Null, K.M. Olson, and J.L. Hutchison. 2011a. Reporting of haplotypes with
537 recessive effects on fertility. Interbull Bulletin. 44:117-121.

538

539 VanRaden P.M., K.M. Olson, D.J. Null, and J.L. Hutchison. 2011b. Harmful recessive effects on
540 fertility detected by absence of homozygous haplotypes. J. Dairy Sci. 94(12): 6153-6161.
541

542 VanRaden P.M., A.H. Sanders, M.E. Tooker, R.H. Miller, H.D. Norman, M.T. Kuhn, and G.R.
543 Wiggans. 2004. Development of a national genetic evaluation for cow fertility. J. Dairy
544 Sci. 87: 2285–2292.
545

546 Wang D., W. Wang, P. Dawkins, T. Paterson, N. Kalsheker, J.M. Sallenave, and A.M. Houghton.
547 2011. Deletion of Serpina1a, a murine α1-antitrypsin ortholog, results in embryonic
548 lethality. Exp. Lung Res. 37(5): 291-300.
549

550 Wang K., M. Li, and H. Hakonarson. 2010. ANNOVAR: functional annotation of genetic variants
551 from high-throughput sequencing data. Nucleic Acids Res. 38(16): e164.
552

553 Wiggans G.R., P.M. VanRaden, L.R. BachellerR, M.E. Tooker, J.L. Hutchison, T.A. Cooper, and
554 T.S. Sonstegard. 2010. Selection and management of DNA markers for use in genomic
555 evaluation. J. Dairy Sci. 93(5): 2287-2292.
556

557 Yoshida H., Y.Y. Kong, R. Yoshida, A.J. Elia, A. Hakem, R. Hakem, J.M. Penninger, and T.W.
558 Mak. 1998. Apaf1 is required for mitochondrial pathways of apoptosis and brain
559 development. Cell 94(6): 739-750.