



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

Heather A. Adams,\*†<sup>1</sup> Tad S. Sonstegard,‡<sup>2</sup> Paul M. VanRaden,‡ Daniel J. Null,‡ Curt P. Van Tassell,‡ Denis M. Larkin,\* and Harris A. Lewin\*†§<sup>1</sup>

\*Department of Animal Sciences, and

†Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana 61801

‡Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705

§Department of Evolution and Ecology and the UC Davis Genome Center, University of California, Davis, Davis 95616

### ABSTRACT

The HH1 haplotype on chromosome 5 is associated with a reduced conception rate and a deficit of homozygotes at the population level in Holstein cattle. The source HH1 haplotype was traced to the bull Pawnee Farm Arlinda Chief (Chief), who was born in 1962 and has sired more than 16,000 daughters. We identified a nonsense mutation in *APAF1* (apoptotic protease activating factor 1; *APAF1* p.Q579X) within HH1 using whole-genome resequencing of Chief and 3 of his sons. This mutation is predicted to truncate 670 AA (53.7%) of the encoded APAF1 protein that contains a WD40 domain critical to protein–protein interactions. Initial screening revealed no homozygous individuals for the mutation in 758 animals previously genotyped, whereas all 497 HH1 carriers possessed 1 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 have 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 knockouts 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 worldwide 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.

**Key words:** nonsense mutation, *APAF1*, dairy cattle, resequencing

### INTRODUCTION

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

Inbreeding can rapidly increase the frequency of recessive lethal and sublethal alleles in the population, but it can also lead to purging of harmful alleles by natural selection. Lethal mutations are the most detrimental economically because no offspring carrying such mutations will survive to reproduce. Some of the top bulls in the history of the dairy industry have been

Received October 12, 2015.

Accepted April 23, 2016.

<sup>1</sup>Current address: CRI International Center for Biotechnology, Mt. Horeb, WI 53572.

<sup>2</sup>Current address: Acceligen of Recombinetics Inc., St. Paul, MN 55104.

<sup>3</sup>Corresponding author: [lewin@ucdavis.edu](mailto:lewin@ucdavis.edu)

shown to harbor mutations affecting fertility, which spread rapidly in the population through AI. In recent years, several of these mutations have been identified (see Online Mendelian Inheritance of Animals, <http://omia.angis.org.au/>). Examples of recessive autosomal lethal or sublethal fertility mutations include deficiency of uridine monophosphate synthase (DUMPS; Robinson et al., 1984; Shanks and Robinson, 1989), complex vertebral malformation (CVM; Agerholm et al., 2001, 2004), and brachyspina (Charlier et al., 2012). These disorders are caused by point mutations, insertions, and deletions that result in aborted fetuses or stillbirths. A recently discovered deletion in Nordic Red cattle (Kadri et al., 2014) causes not only recessive fertility loss but also increased milk yield, maintaining high frequency through balancing selection. Identification of such mutations has enabled screening programs to avoid matings between carriers from within the population.

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

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

## MATERIALS AND METHODS

### *Calculations Supporting Chief's Influence on Spontaneous Abortions*

Using pedigree data, VanRaden et al. (2011a) identified Pawnee Farm Arlinda Chief (Chief), born in 1962

(Table 1), as the earliest genotyped ancestor carrying the HH1 haplotype. Chief is one of the most influential sires in the history of the Holstein breed, having produced many sons that became popular sires, in addition to >16,000 daughters, >500,000 granddaughters, and >2 million recorded great-granddaughters. We estimated the cumulative number of spontaneous abortions caused by HH1 over the 30 yr since Chief alleles became highly frequent to be more than 100,000 nationally and almost 500,000 worldwide as follows: the estimated cumulative number of spontaneous abortions caused by HH1 in the US Holstein population was calculated as the number of cows (8 million)  $\times$   $0.045(1/2) \times 0.045(1/2) = 4,050/\text{yr}$ , where 0.045 is the HH1 carrier frequency of one parent, and 1/2 is the probability that the parent will contribute the defect to the offspring. Globally, the estimate is 15,000/yr based on a population size of 30 million. These estimates correspond to approximately 140,000 spontaneous abortions in the United States and 525,000 worldwide over a 35-yr period, the approximate time that Chief alleles appeared on both sides of the pedigree.

Carrier frequency in US Holsteins exceeded 0.08 during the 1980s and 1990s but decreased to about 0.03 in 2010 (VanRaden et al., 2011b). The actual mating pattern for 58,453 genotyped Holsteins was used by VanRaden et al. (2011b), who reported 23 expected HH1 homozygotes compared with 30 expected when assuming random mating. Use of the actual mating pattern is difficult for national or international populations because many ancestors are either not known or not genotyped. Further direct selection for fertility and against HH1 reduced the frequency to 0.02 in 2015.

The economic loss from a midterm abortion is estimated to be about \$800 (Norman et al., 2012), for a total cost of ~\$420 million. For comparison, the increased value of milk from using Chief instead of an average bull in 1962 is his genetic contribution to the breed ( $0.143$ )  $\times$  the farm price of milk ( $\$0.33/\text{L}$ )  $\times$  the increased milk yield ( $2 \text{ L/d}$ )  $\times$   $305 \text{ d/yr}$   $\times$   $35 \text{ yr}$   $\times$   $30$  million cows = \$30 billion. Embryonic and fetal loss during gestation was investigated using the national fertility database and occurred mainly from 60 to 200 d of gestation for HH1, but earlier for several other recessive defects (VanRaden et al., 2011a; see also Norman et al., 2012 for more details on the timing of embryo loss, and Fritz et al., 2013 for independent confirmation of fertility effects).

### *Haplotype Detection and Crossover Analysis*

Recombinant haplotypes, defined as a portion of Chief's HH1 source haplotype but not all of it, were detected within the pedigree of 78,465 animals that

had 54,001 SNP genotypes as of 2011 using findhap.f90 as previously described (VanRaden et al., 2011a; Sonstegard et al., 2013). All copies of the 75-marker source haplotype spanning 7.1 Mbp that contained the putative mutation appeared to trace to Chief and to no other prominent ancestors. VanRaden et al. (2011b) studied only the source haplotype, whereas living animals with recombinant haplotypes that are homozygous for only a portion of the source haplotype can rule out that portion of the haplotype as not containing the lethal mutation. This fine-mapping method is the mirror image of typical homozygosity mapping, which focuses on the region of homozygosity shared by affected animals instead of ruling out regions of the haplotype homozygous in unaffected animals. After processing all recombinant haplotypes, the area not ruled out was defined as the mutation-critical region, as described by Sonstegard et al. (2013).

Recombination events were detected in 78,465 animals genotyped for 43,385 SNP from the Illumina BovineSNP50 BeadChips (Illumina Inc., San Diego, CA) using edits of Wiggans et al. (2010) as well as standard output from findhap.f90 (VanRaden et al., 2011a) version 2, which first examined haplotypes of length 600 markers, then 200 markers, and finally output haplotypes of  $\leq 75$  markers. The program phases genotypes into haplotypes and detects recombination points between the maternal and paternal haplotype of each genotyped parent. The approaches for detecting recombinant haplotypes were described previously (Sonstegard et al., 2013). Briefly, progeny and parent haplotypes were compared directly to identify potential crossover events. For those identified, the marker known to be from the end of the first parental haplotype and that from the beginning of the second parental haplotype

are output. Gaps between these 2 markers denoting the recombinant haplotype region may exist for a number of reasons that include identical parental haplotypes, missing SNP genotypes, or parents with heterozygous alleles that could not be phased. All these phenomena lead to unknown crossover locations. Crossovers derived from maternal ancestors are more difficult to detect, because dam genotypes are not always available.

Fine mapping was accomplished by checking for animals with both the original HH1 haplotype and a recombinant haplotype. Regions homozygous for a section of the source haplotype were removed from consideration of harboring the causative HH1 mutation. For example, if a live animal received the original HH1 haplotype from one parent and the remaining 20 markers of the HH1 haplotype from the other parent, the region containing those 20 markers was removed from consideration, exactly as described in Sonstegard et al. (2013) for Jersey haplotype 1. The frequency of HH1 heterozygotes in all animals genotyped was 3.2% for the source haplotype and 4.5% when recombinant haplotypes were included. Upon identification of the region of the HH1 haplotype including the potentially lethal allele, individuals were selected for study based on their relationship to Chief, allele carrier status, and overall relationship (expected future inbreeding) to the US Holstein breed.

Two sons of Chief, Walkway Chief Mark (Mark) and Milu Betty Ivanhoe Chief (Ivanhoe Chief), were identified as carriers of HH1 based on previous genotyping information (VanRaden et al., 2011b). An additional son of Chief, S-W-D Valiant (Valiant), was found not to carry the HH1 haplotype. Chief and Mark had complete genome sequence available from an earlier study (Larkin et al. 2012) and were thus useful for a preliminary

**Table 1.** Pedigree information for 4 Holstein dairy bulls: Pawnee Farm Arlinda Chief (Chief), Milu Betty Ivanhoe Chief (Ivanhoe Chief), S-W-D Valiant (Valiant), and Walkway Chief Mark (Mark)

Bull	ID <sup>1</sup>	Birth year	Sire	Dam	Pedigree EFI <sup>2</sup>	Ranking <sup>3</sup>	Alleles shared (%) <sup>4</sup>
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

<sup>1</sup>Holstein breed identification number.

<sup>2</sup>Expected future inbreeding (EFI) of bull's daughters ([https://www.cdeb.us/eval/summary/inbrd.cfm?R\\_Menu=HO.k](https://www.cdeb.us/eval/summary/inbrd.cfm?R_Menu=HO.k)).

<sup>3</sup>Bull's EFI ranking, which outlines the relationship of the bull to the Holstein cow population ([https://www.cdeb.us/eval/summary/inbrd.cfm?R\\_Menu=HO.k](https://www.cdeb.us/eval/summary/inbrd.cfm?R_Menu=HO.k)).

<sup>4</sup>Percentage of alleles in common with top currently available Holstein sires; the value is an indicator of the bull's overall contribution to the Holstein breed, from a list of the top 25 most influential bulls.

screen to identify mutations on the HH1 haplotype. For the present study, additional sequencing was conducted for Chief to increase coverage (see below; Supplementary Table S1; <http://dx.doi.org/10.3168/jds.2015-10517>), thus permitting greater accuracy of base calling in and near the *APAF1* locus. Whole genome sequencing of Ivanhoe Chief and Valiant were conducted as part of an independent project. For Ivanhoe Chief and Valiant, only the sequence data at the *APAF1* locus was used in the present study to assist validation of the inheritance of the *APAF1* mutation (see following sections). The whole-genome sequences of Ivanhoe Chief and Valiant will be reported elsewhere. The DNA sequences of Chief and 3 of his sons in and near the *APAF1* locus permitted phasing of genotypes into haplotypes, and was useful for designing the 12-marker *APAF1* confirmatory SNP panel.

### Sequencing, Alignment, and Mapping

The genomes of Chief, Ivanhoe Chief, and Valiant were sequenced using sequencing by synthesis chemistry on an Illumina HiSeq 2000 platform (Illumina Inc.) at the University of Illinois W. M. Keck Center for Comparative and Functional Genomics. Libraries were prepared from 5  $\mu$ g of genomic DNA purified from semen straws, and data were generated using standard sequencing protocols provided by the manufacturer. Previous sequencing results of Mark (12 $\times$ ) and Chief (6 $\times$ ) using 454 Titanium technology were also used (Larkin et al., 2012).

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

### Detection of SNP and Genes

The SNP in the suspect region of BTA5 were identified using FreeBayes [E. Garrison and G. Marth, unpublished data; [arXiv:1207.3907v2](https://arxiv.org/abs/1207.3907v2)(q-bio.GN)]. Putative SNP were accepted if they fit within the following criteria: 4 $\times$  minimum read coverage with at least 2

reads aligning in each orientation (forward, reverse) and minimum allele sequencing quality  $\geq 20$ . Upon acquiring a list of SNP in the region, functional annotation of the variants was performed using ANNOVAR (Wang et al., 2010). The ANNOVAR program categorized SNP by their genic or intergenic locations within the cattle genome. The program reports SNP located within introns and exons of annotated genes, 5' and 3' untranslated regions (UTR), and those upstream and downstream of gene positions. All coordinates pertaining to SNP and gene positions were converted from Btau4.0 to UMD3.1 genome assemblies by using the program LiftOver created by the UCSC Genome Bioinformatics Group (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>) for consistency with haplotype and genotype data sets.

### Selection of Animals for *APAF1* SNP Validation

Animals were selected for validation by querying a large database of 33,415 Holsteins genotyped for 54,001 SNP as constructed previously (Wiggans et al., 2010). Genotype imputation and haplotype frequencies included all 33,415 animals, but the 758 samples selected for further validation were from the Cooperative Dairy DNA Repository (Beltsville, MD), which contains DNA from almost all progeny tested bulls in North America. Haplotype identification was based on the 75 SNP markers designated as the 7.1-Mbp HH1-containing interval on BTA5 (UMD3.1 coordinates 58,638,702 to 65,743,920; VanRaden et al., 2011b). An additional query was implemented to select a diverse set of non-carriers that had unique heterozygous haplotype combinations in this interval.

A SNP genotyping panel (Sequenom Inc., San Diego, CA) designed for the validation test (Page et al., 2004) was composed of 24 bidirectional assays for 12 putative SNP in the refined HH1 interval region. This panel included all SNP with gene boundaries found within this interval, as well as 5 additional SNP observed near adjacent genes in the interval or in distal flanking regions from the *APAF1* stop-gain mutation. A total of 22 of the 24 SNP assays were functional; 1 SNP locus was monomorphic (Table 2). The call rate for all SNP loci was 100% except for UMD3\_63107293 (99.3%) and UMD3\_62591311 (99.9%). Results from the bidirectional assays for each SNP locus were compared for concordance and integrated into a single marker genotype score for each animal across the 11 SNP loci. Haplotypes of 11 informative SNP were determined by PHASE v2.1.1 (Stephens et al., 2001), and 24 probable haplotypes were identified (Table 3). These haplotypes are much shorter and different than those originally defined by the 75-marker window (derived from the 54,001 chip) spanning 7.1 Mbp that was used to find



**Table 2.** Single nucleotide polymorphisms used for validation and concordance with carrier haplotypes

Gene	SNP location	UMD3.1 coordinate	Allele 1 <sup>1</sup>	Allele 2 <sup>2</sup>	Concordance <sup>3</sup>
ENSBTAG00000038223 <sup>4</sup>	Intergenic	62591311	T	C	0.61
ENSBTAG00000038223	Intergenic	62756350	T	A	Not informative
<i>TMPO</i>	UTR3 <sup>5</sup>	63051612	A	G	0.53
<i>TMPO</i>	UTR3	63052631	A	G	0.53
<i>SLC25A3</i>	Intronic	63088973	T	A	0.45
<i>SLC25A3</i>	Intronic	63091578	T	A	0.99
<i>IKBIP</i>	Intronic	63107293	T	C	0.45
<i>APAF1</i>	Exonic/stop-gain	63150400	C	T	1.00
<i>APAF1</i>	Exonic/synonymous	63198664	C	T	0.52
ENSBTAG00000017385	Exonic/synonymous	63209396	C	T	0.64
<i>ANKS1B</i>	Downstream	63228106	C	T	0.60
<i>ANKS1B</i>	Intergenic	63486133	C	T	0.30

<sup>1</sup>Allele found in the reference sequence and found on Chief’s haplotype carrying the normal allele.

<sup>2</sup>Allele found on Chief’s HH1 haplotype.

<sup>3</sup>Concordance is the comparison of the HH1 state to the alternative allele.

<sup>4</sup>Ensembl identification for predicted genes.

<sup>5</sup>Variant overlaps a 3’ untranslated region.

HH1. Two different numbering systems exist: one for the more than 2,000 different haplotypes in this 7.1-Mbp window, and a second for the 24 haplotypes in the narrow 11 SNP window for validation (Table 3).

In all, 758 animals were selected for validation genotyping using a 24-SNP (12 × 2) multiplex panel, and 486 of these were presumed carriers based on the presence of haplotype 12, which was the original des-

ignation for the corresponding HH1 haplotype within the 7.1-Mbp interval. Among all animals within the validation set, 246 unique haplotypes existed within the HH1 interval, as well as 323 heterozygous haplotype combinations. Additionally, animals possessing “haplotype 32” (n = 11) had a small region of 40 markers in a 1-Mb region within the HH1 interval that was expected to be equivalent to haplotype 12 because haplotype 32 is a recombinant haplotype. These individuals were expected to test positive for the causal mutation if the SNP was potentially associated with the recessive lethal effect. After regenotyping the animals for validation, 1 animal within the haplotype 12 group was found to have an incorrect genotype (the DNA was actually from a different animal) and was removed from the study. In total, 497 animals (485 with haplotype 12, and 11 with haplotype 32) were expected to be heterozygous for the *APAF1* stop-gain mutation. Following this initial validation, a test for the stop-gain mutation was added to the GeneSeek Genomic Profiler BeadChip (GeneSeek-Neogen, Lincoln, NE; Neogen Corporation, 2013) and subsequent chips, and genotypes were received for 246,773 Holsteins as part of routine genomic predictions.

**Nomenclature**

The length of *APAF1* originally reported in Adams et al. (2012) was 1,238 AA, which is the polypeptide length given by the UCSC Genome Browser (<https://ucsc.edu/cgi-bin/hgGateway>). The UCSC Genome Browser annotation of the *APAF1* protein (NP\_001178436.1) is based on the NCBI cattle genome annotation. Fritz et al. (2013) apparently used either the human protein,

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

Haplotype <sup>1</sup>	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

<sup>1</sup>*APAF1* stop-gain mutation is the seventh marker of this haplotype and is designated as allele 1.

which is 1,248 AA, or an alternate annotation of the bovine sequence (e.g., XP\_005206703) that is 1,248 AA. In UNIPROT, the cattle APAF1 protein (Entry F1MUW4) is 1,251 AA, which is caused by addition of 2 AA at the amino terminus and insertion of another internal residue. Upon careful analysis, we have concluded that the UNIPROT annotation is likely to be incorrect. We have also concluded that the 1,238-AA polypeptide represents a prediction error or an isoform produced by an alternatively spliced product of the full-length *APAF1* mRNA. To adopt the most likely length of the protein and avoid confusion in the literature, we have chosen 1,248 AA for the APAF1 protein. Thus, the nomenclature for the mutation used in this report is Q579X (or Gln579X), which is consistent with the nomenclature described by Fritz et al. (2013).

## 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 SNP. Fine mapping was based on 8 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 1 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 that of 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 2 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 ~31×, 13.5×, ~50×, and ~36× coverage, respectively (Supplementary Table S1; <http://dx.doi.org/10.3168/jds.2015-10517>). Having the genome sequences of these bulls allowed us to search within the mapped interval for SNP and candidate genes for the lethal allele. Normal and carrier progeny of Chief provided information by helping to phase the SNP. As a preliminary step, alleles possessed by Chief that were inherited only by Valiant, a nonrecombinant

noncarrier, were ruled out as causative. If available, sequence for an animal homozygous for Chief's normal haplotype could also rule out those mutations.

A list of SNP within the critical region was acquired by alignment of sequence reads to cattle genome reference assembly Btau4.0 and then converted to coordinates of assembly UMD 3.1. The SNP were then annotated as genic, intergenic, synonymous, nonsynonymous, and so forth, using ANNOVAR. Analysis of Chief's DNA within the critical region revealed 3 unique SNP in exons, with the rest being intronic, intergenic, 3' UTR, and downstream regions (Table 4). Among the exonic SNP for which Chief was heterozygous, a C→T substitution in exon 11 of the gene encoding apoptosis peptide activating factor 1 (*APAF1*) produces a stop-gain mutation at position 579 in the polypeptide (Table 2; Figure 1). The *APAF1 p.Q579X* mutation truncates 670 C-terminal amino acids (53.7%) from the 1,248-residue full-length *APAF1* protein. Alignments of Mark and Ivanhoe Chief sequence reads, both deemed carriers of HH1, confirmed that these Chief sons were heterozygous for the *APAF1 p.Q579X* mutation. Chief's son Valiant, classified as a noncarrier, was found to be homozygous for the normal allele. The other 2 exonic SNP, one in *APAF1* and the other in a predicted gene, both cause synonymous substitutions (Table 2), ruling out these mutations as causative. Thus, *APAF1 p.Q579X* was determined to be the putative mutation causing prenatal lethality associated with the HH1 haplotype.

Validation of *APAF1 p.Q579X* as the likely causative mutation was accomplished by querying a large inventory of archived cattle genomic DNA for all animals carrying the HH1 haplotype (see Materials and Methods). Among the 758 animals selected for validation genotyping, 497 were presumed carriers of the mutation based on HH1 haplotyping; the rest did not carry the HH1 haplotype and were presumed to be noncarriers. Eleven informative SNP identified from Chief sequence that are located in the refined HH1 interval, including the SNP producing the stop-gain mutation, were used for the validation studies (Table 2). Only 1 haplotype among the 24 observed 11-SNP haplotypes was associated with the stop-gain mutation and corresponded to the 497 previously identified HH1-positive individuals (Table 3).

Only one other SNP besides the *APAF1* stop-gain mutation was in high concordance with HH1 (Table 3), defined as the absolute value of the percentage of SNP and HH1 genotypes that matched for the 758 validation animals. This SNP is unlikely to be responsible for reduced fertility because it is located in an intron of *SLC25A3* and had a 1.2% false-positive detection rate for HH1 haplotypes. In contrast, the stop-gain mutation at position UMD3\_63150400 in *APAF1* was

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

BTA5 region [62,435,307–65,597,776] <sup>1</sup>	SNP (no.) <sup>2</sup>	Gene <sup>3</sup>
Downstream	2	<i>ANKS1B</i> [2]
Exonic	3	<i>APAF1</i> [2], <i>ENSB_17385</i> [1] <sup>4</sup>
Intergenic <sup>5</sup>	1,221	<i>ENSB_38223</i> , <i>SP</i> , <i>SLC25A3</i> , <i>IKBIP</i> , <i>ENSB_17385</i> , <i>ANKS1B</i> , <i>ENSB_40364</i>
Intronic	354	<i>SP</i> [17], <i>SLC25A3</i> [2], <i>IKBIP</i> [9], <i>APAF1</i> [88], <i>ENSB_17385</i> [1]
UTR3	2	<i>SP</i> [2]

<sup>1</sup>Coordinates for suspect region on BTA5 from UMD 3.1 genome assembly. 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.

<sup>2</sup>Number of SNP identified within each region.

<sup>3</sup>Number in brackets indicates the number of SNP associated with the given gene or predicted gene.

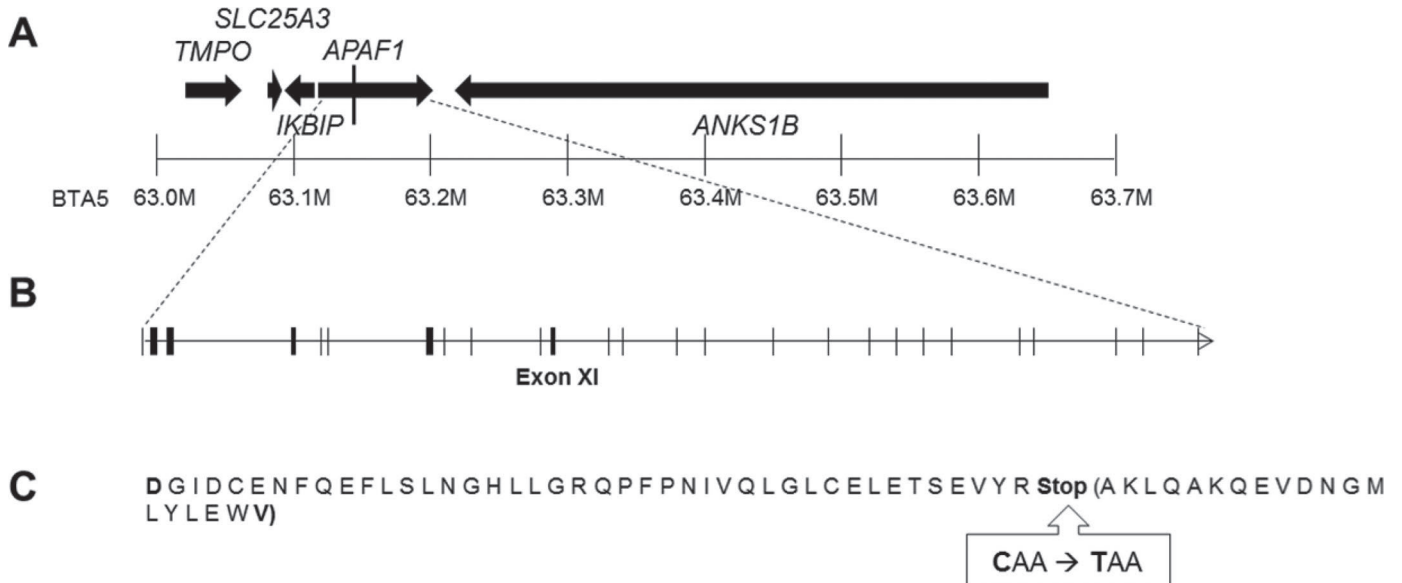
<sup>4</sup>ENSB\_ abbreviated from ENSBTAG000000 for all predicted genes listed.

<sup>5</sup>Genes listed for SNP located within intergenic regions are the closest genes that flank the corresponding intergenic SNP.

100% concordant with the recessive lethal (Table 3). Subsequent test results provided by GeneSeek-Neogen Corporation for *APAF1 p.Q579X* identified 5,299 heterozygotes and zero homozygotes in 246,773 Holsteins, consistent with the hypothesis that this mutation is the lethal HH1 mutation.

Additional support for *APAF1 p.Q579X* as the causative mutation derives from functional studies. The protein encoded by *APAF1* is a central component of the cytochrome-c-mediated apoptotic cascade (Apweiler et al., 2004) and has been directly implicated in the etiology of cancer, developmental disorders, and neurodegenerative diseases (Honarpour et al., 2001; Blake

et al., 2011). The *APAF1* protein forms an oligomer that when bound with cytochrome-c and dATP forms the apoptosome, a cytoplasmic structure that binds the caspase-9 preprotein and cleaves it into its mature active form. The activated form of caspase-9 initiates the caspase cascade that ultimately leads to apoptotic cell death. Expression of *Apaf1* during murine development begins between days 7 and 9 in several vital tissues and organs and is crucial for the development of the central nervous system. Homozygous *Apaf1* gene knock-out in mice leads to embryonic lethality by day 16.5 or perinatally, and *Apaf1*-deficient mice exhibit severe abnormalities such as brain overgrowth, persistence of



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

interdigital webs, and craniofacial malformations (Cecconi et al., 1998; Yoshida et al., 1998; Honarpour et al., 2001; Müller et al., 2005). Significantly, the deletion of 670 C-terminal amino acids from the APAF1 polypeptide removes 15 WD40 repeats that form a predicted functional WD40 domain in the cattle protein. WD40 domains are found in many proteins involved in signal transduction, transcriptional regulation, and apoptosis, and they are essential for protein-protein interactions (Acehan et al., 2002). Deletion of the WD40 domain would likely result in failure to form apoptosomes, which are essential for binding of caspase-9 and initiating the apoptosis pathway (Acehan et al., 2002; Riedl and Salvesen, 2007). The severely truncated *APAF1* peptide in homozygous cattle is likely the functional equivalent of the homozygous *Apaf1* knockout in mice. These data strongly support the *APAF1* stop-gain mutation *p.Q579X* as the causative mutation for embryonic, fetal, and perinatal loss of cattle homozygous for the HH1 haplotype.

An undetected insertion, deletion, or copy number variant carried on HH1 is another possibility, but it seems less likely than the stop-gain mutation in *APAF1*. The validation animals were mostly 5 to 12 generations removed from Chief, whereas newborn calves may be >14 generations removed, making haplotype detection more difficult, especially with lower-density chips. However, the stop-gain mutation is now included in routine genomic evaluation to improve the accuracy of detecting both carrier status and genomic prediction of fertility.

Screening for loss of homozygosity is a powerful approach for the identification of chromosomal segments associated with prenatal mortality (VanRaden et al., 2011b; Fritz et al., 2013). We have shown here that detection of loss of homozygosity in combination with moderate coverage whole genome resequencing can be used to rapidly identify causative mutations for prenatal mortality, particularly if a database of DNA sequences of key individuals in animal pedigrees is available. Such databases are now being widely developed for several cattle breeds (Fritz et al., 2013; Daetwyler et al., 2014; McClure et al., 2014), and the list of haplotypes associated with loss of homozygosity is growing (for an updated list see [http://aipl.arsusda.gov/reference/recessive\\_haplotypes\\_ARR-G3.html](http://aipl.arsusda.gov/reference/recessive_haplotypes_ARR-G3.html)). Identification of the causative mutations for prenatal and perinatal mortality can be translated into genetic screens to rapidly eliminate the unwanted alleles from the breeding population. In the case of *APAF1*, it will now be possible to eliminate a mutation that is estimated to be causal for more than 500,000 abortions in Holstein cattle worldwide. Alternatively, these diagnostic tests can also be used to avoid mating of carriers to avoid losing the

more prevalent beneficial genetic contributions derived from Chief, whose chromosomes contributed 14% of the current Holstein genome and have been attributed to about \$30 billion dollars in increased milk production.

## ACKNOWLEDGMENTS

This work was supported by USDA Agricultural Research Service CRIS projects 1265-31000-104-00D and 31S. The authors thank Gene McCoy (University of Illinois at Urbana-Champaign) and A. Beavers and M. McClure (Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD) for assistance, as well as GeneSeek-Neogen Corp. (Lincoln, NE) for providing supporting genotype data. 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 U.S. Department of Agriculture (Washington, DC).

## 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. Poster P0555 (Abstr.) in Plant and Animal Genome XX Meeting, San Diego, CA. Accessed May 23, 2016. <https://pag.confex.com/pag/xx/webprogram/Paper3932.html>.
- 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, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi, and L. S. Yeh. 2004. UniProt: The Universal Protein knowledgebase. *Nucleic Acids Res.* 32:D115–D119.
- Blake, J. A., C. J. Bult, J. A. Kadin, J. E. Richardson, J. T. Eppig, and the 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* 94:727–737.
- Charlier, C., J. S. Agerholm, W. Coppieters, P. Karlskov-Mortensen, W. Li, G. de Jong, C. Fasquelle, L. Karim, S. Cirera, N. Cambisano, N. Ahariz, E. Mullaart, M. Georges, and M. Fredholm. 2012. A deletion in the bovine *FANCI* gene compromises fertility by causing fetal death and brachyspina. *PLoS One* 7:e43085.
- Daetwyler, H. D., A. Capitan, H. Pausch, P. Stothard, R. van Binsbergen, R. F. Brondum, 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. Hulsege, M. E. Goddard, B. Guldbrandtsen, 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.
- Fritz, S., A. Capitan, A. Djari, S. C. Rodriguez, A. Barbat, A. Baur, C. Grohs, B. Weiss, M. Boussaha, D. Esquerré, C. Klopp, D. Rocha, and D. Boichard. 2013. Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in *GART*, *SHBG* and *SLC37A2*. *PLoS One* 8:e65550. <http://dx.doi.org/10.1371/journal.pone.0065550>.
- Honarpour, N., S. L. Gilbert, B. T. Lahn, X. Wang, and J. Herz. 2001. *Apaf-1* deficiency and neural tube closure defects are found in *fog* mice. *Proc. Natl. Acad. Sci. USA* 98:9683–9687.
- Kadri, N. K., G. Sahana, C. Charlier, T. Iso-Touru, B. Guldbbrandtsen, L. Karim, U. S. Nielsen, F. Panitz, G. P. Aamand, N. Schulman, M. Georges, J. Vilkki, M. S. Lund, and T. Druet. 2014. A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red Cattle: Additional evidence for the common occurrence of balancing selection in livestock. *PLoS Genet.* 10:e1004049.
- Larkin, D. M., H. D. Daetwyler, A. G. Hernandez, C. L. Wright, L. A. Hetrick, L. Boucek, S. L. Bachman, M. R. Band, T. V. Akraiko, M. Cohen-Zinder, J. Thimmapuram, I. M. Macleod, T. T. Harkins, J. E. McCague, M. E. Goddard, B. J. Hayes, and H. A. Lewin. 2012. Whole-genome resequencing of two elite sires for the detection of haplotypes under selection in dairy cattle. *Proc. Natl. Acad. Sci. USA* 109:7693–7698.
- Li, R., C. Yu, Y. Li, T. W. Lam, S. M. Yiu, K. Kristiansen, and J. Wang. 2009. SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics* 25:1966–1967.
- Lucy, M. C. 2001. Reproductive loss in high-producing dairy cattle: Where will it end? *J. Dairy Sci.* 84:1277–1293.
- McClure, M. C., D. Bickhart, D. Null, P. VanRaden, L. Xu, G. Wiggans, G. Liu, S. Schroeder, J. Glasscock, J. Armstrong, J. B. Cole, C. P. VanTassell, and T. S. Sonstegard. 2014. Bovine exome sequence analysis and targeted SNP genotyping of recessive fertility defects BH1, HH2, and HH3 reveal a putative causative mutation in *SMC2* for HH3. *PLoS One* 9:e92769. <http://dx.doi.org/10.1371/journal.pone.0092769>.
- Müller, M., J. Berger, N. Gersdorff, F. Cecconi, R. Herken, and F. Quondamatteo. 2005. Localization of *Apaf1* gene expression in the early development of the mouse by means of in situ reverse transcriptase-polymerase chain reaction. *Dev. Dyn.* 234:215–221.
- Neogen Corporation. 2013. GeneSeek genomic profiler for dairy cattle. Accessed November 2013. [http://www.neogen.com/Genomics/pdf/Slicks/GGP-LD\\_Dairy.pdf](http://www.neogen.com/Genomics/pdf/Slicks/GGP-LD_Dairy.pdf).
- Norman, H. D., R. H. Miller, J. R. Wright, J. L. Hutchison, and K. M. Olson. 2012. Factors associated with frequency of abortions recorded through Dairy Herd Improvement test plans. *J. Dairy Sci.* 95:4074–4084.
- Page, B. T., E. Casas, R. L. Quaas, R. M. Thallman, T. L. Wheeler, S. D. Shackelford, M. Koochmaraie, S. N. White, G. L. Bennett, J. W. Keele, T. E. Dikeman, and T. P. Smith. 2004. Association of markers in the bovine *CAPN1* gene with meat tenderness in large crossbred populations that sample influential industry sires. *J. Anim. Sci.* 82:3474–3481.
- Riedl, S. J., and G. S. Salvesen. 2007. The apoptosome: Signaling platform of cell death. *Nat. Rev. Mol. Cell Biol.* 8:405–413.
- Robinson, J. L., D. B. Dombrowski, G. W. Harpestad, and R. D. Shanks. 1984. Detection and prevalence of UMP synthase deficiency among dairy cattle. *J. Hered.* 75:277–280.
- Shanks, R. D., and J. L. Robinson. 1989. Embryonic mortality attributed to inherited deficiency of uridine monophosphate synthase. *J. Dairy Sci.* 72:3035–3039.
- Shook, G. E. 2006. Major advances in determining appropriate selection goals. *J. Dairy Sci.* 89:1349–1361.
- Sonstegard, T. S., J. B. Cole, P. M. VanRaden, C. P. Van Tassell, D. J. Null, S. G. Schroeder, D. Bickhart, and M. C. McClure. 2013. Identification of a nonsense mutation in *CWC15* associated with decreased reproductive efficiency in Jersey cattle. *PLoS ONE* 8:e54872. <http://dx.doi.org/10.1371/journal.pone.0054872>.
- Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68:978–989.
- VanRaden, P. M., D. J. Null, K. M. Olson, and J. L. Hutchison. 2011a. Reporting of haplotypes with recessive effects on fertility. *Interbull Bull.* 44:117–121.
- 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.
- VanRaden, P. M., A. H. Sanders, M. E. Tooker, R. H. Miller, H. D. Norman, M. T. Kuhn, and G. R. Wiggans. 2004. Development of a national genetic evaluation for cow fertility. *J. Dairy Sci.* 87:2285–2292.
- Wang, K., M. Li, and H. Hakonarson. 2010. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:e164.
- Wiggans, G. R., P. M. Van, L. R. Raden, R. Bacheller, M. E. Tooker, J. L. Hutchison, T. A. Cooper, and T. S. Sonstegard. 2010. Selection and management of DNA markers for use in genomic evaluation. *J. Dairy Sci.* 93:2287–2292.
- Yoshida, H., Y. Y. Kong, R. Yoshida, A. J. Elia, A. Hakem, R. Hakem, J. M. Penninger, and T. W. Mak. 1998. *Apaf1* is required for mitochondrial pathways of apoptosis and brain development. *Cell* 94:739–750.