



RESEARCH ARTICLE

**REVISED** Association of *ARRDC3* and *NFIA* variants with bovine congestive heart failure in feedlot cattle [version 2; peer review: 2 approved, 1 approved with reservations]

Michael P. Heaton <sup>1</sup>, Gregory P. Harhay <sup>1</sup>, Adam S. Bassett<sup>2</sup>, Halden J. Clark<sup>2</sup>, Jaden M. Carlson<sup>2</sup>, Erin E. Jobman<sup>2</sup>, Helen R. Sadd<sup>1</sup>, Madeline C. Pelster<sup>2</sup>, Aspen M. Workman<sup>1</sup>, Larry A. Kuehn<sup>1</sup>, Theodore S. Kalbfleisch <sup>3</sup>, Heather Piscatelli<sup>4</sup>, Michael Carrie<sup>4</sup>, Greta M. Krafur<sup>5</sup>, Dale M. Grotelueschen<sup>2</sup>, Brian L. Vander Ley <sup>2</sup>

<sup>1</sup>USDA, ARS, US Meat Animal Research Center, Clay Center, Nebraska, 68933, USA

<sup>2</sup>University of Nebraska-Lincoln, Great Plains Veterinary Educational Center, Clay Center, Nebraska, 68933, USA

<sup>3</sup>University of Kentucky, Lexington, Kentucky, 40506, USA

<sup>4</sup>MatMaCorp, Lincoln, Nebraska, 68507, USA

<sup>5</sup>Anschutz Medical Campus, University of Colorado Denver, Aurora, Colorado, 80045, USA

**v2** First published: 01 Apr 2022, 11:385  
<https://doi.org/10.12688/f1000research.109488.1>  
 Latest published: 04 Mar 2024, 11:385  
<https://doi.org/10.12688/f1000research.109488.2>

**Abstract**

**Background**

Bovine congestive heart failure (BCHF) has become increasingly prevalent among feedlot cattle in the Western Great Plains of North America with up to 7% mortality in affected herds. BCHF is an untreatable complex condition involving pulmonary hypertension that culminates in right ventricular failure and death. Genes associated with BCHF in feedlot cattle have not been previously identified. Our aim was to search for genomic regions associated with this disease.

**Methods**

A retrospective, matched case-control design with 102 clinical BCHF cases and their unaffected pen mates was used in a genome-wide association study. Paired nominal data from approximately 560,000 filtered single nucleotide polymorphisms (SNPs) were analyzed with McNemar’s test.

**Results**

**Open Peer Review**

**Approval Status** ✓✓?

	1	2	3
<b>version 2</b> (revision) 04 Mar 2024		✓ view	? view
<b>version 1</b> 01 Apr 2022	✓ view	↑ ? view	

1. **Alan L. Archibald** , University of Edinburgh, Midlothian, UK
2. **Cord Drögemüller**, University of Bern, Bern, Switzerland
3. **James Reecy**, Iowa State University, Ames, USA

Any reports and responses or comments on the article can be found at the end of the article.

Two independent genomic regions were identified as having the most significant association with BCHF: the arrestin domain-containing protein 3 gene (*ARRDC3*), and the nuclear factor IA gene (*NFIA*, mid-*p*-values,  $1 \times 10^{-8}$  and  $2 \times 10^{-7}$ , respectively). Animals with two copies of risk alleles at either gene were approximately eight-fold more likely to have BCHF than their matched pen mates with either one or zero risk alleles at both genes (CI 95 = 3-17). Further, animals with two copies of risk alleles at both genes were 28-fold more likely to have BCHF than all others (*p*-value =  $1 \times 10^{-7}$ , CI95 = 4-206). A missense variant in *ARRDC3* (C182Y) represents a potential functional variant since the C182 codon is conserved among all other jawed vertebrate species observed. A two-SNP test with markers in both genes showed 29% of 273 BCHF cases had homozygous risk genotypes in both genes, compared to 2.5% in 198 similar unaffected feedlot cattle. This and other DNA tests may be useful for identifying feedlot animals with the highest risk for BCHF in the environments described here.

### Conclusions

Although pathogenic roles for variants in the *ARRDC3* and *NFIA* genes are unknown, their discovery facilitates classifying animals by genetic risk and allows cattle producers to make informed decisions for selective breeding and animal health management.

### Keywords

Bovine, Congestive Heart Failure, *ARRDC3*, *NFIA*, GWAS



This article is included in the [Agriculture, Food and Nutrition gateway](#).

**Corresponding authors:** Michael P. Heaton ([mike.heaton@usda.gov](mailto:mike.heaton@usda.gov)), Brian L. Vander Ley ([bvanderley2@unl.edu](mailto:bvanderley2@unl.edu))

**Author roles:** **Heaton MP:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Harhay GP:** Data Curation, Formal Analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – Review & Editing; **Bassett AS:** Investigation; **Clark HJ:** Investigation, Writing – Review & Editing; **Carlson JM:** Investigation, Validation, Writing – Review & Editing; **Jobman EE:** Data Curation, Investigation, Writing – Review & Editing; **Sadd HR:** Data Curation, Investigation, Validation, Writing – Review & Editing; **Pelster MC:** Investigation; **Workman AM:** Investigation, Writing – Review & Editing; **Kuehn LA:** Investigation, Writing – Review & Editing; **Kalbfleisch TS:** Data Curation, Investigation, Software; **Piscatelli H:** Investigation, Writing – Review & Editing; **Carrie M:** Investigation; **Krafsur GM:** Investigation, Methodology, Resources, Writing – Review & Editing; **Grotelueschen DM:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; **Vander Ley BL:** Conceptualization, Data Curation, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Writing – Review & Editing

**Competing interests:** The authors declare the following competing interests: co-authors H.P. and M.C. are full-time employees at MatMaCorp, a molecular diagnostics company that builds instruments and develops test kits for human and animal health applications. There are no patents, products in development, or marketed products to declare. These interests do not alter the authors' adherence to all the journal's policies on sharing data and materials. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. The USDA is an equal opportunity provider and employer.

**Grant information:** This research was supported in part by the U.S. Department of Agriculture, Agricultural Research Service (5438-32000-033-00D, MPH), the University of Nebraska-Lincoln School of Veterinary Medicine and Biomedical Sciences/Great Plains Veterinary Education Center (DMG, BLV) and the Nebraska Beef Industry Endowment (BLV).

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2024 Heaton MP *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Heaton MP, Harhay GP, Bassett AS *et al.* **Association of *ARRDC3* and *NFIA* variants with bovine congestive heart failure in feedlot cattle [version 2; peer review: 2 approved, 1 approved with reservations]** F1000Research 2024, 11:385 <https://doi.org/10.12688/f1000research.109488.2>

**First published:** 01 Apr 2022, 11:385 <https://doi.org/10.12688/f1000research.109488.1>

**REVISED Amendments from Version 1**

The major differences from the previous version are that We have added Table 5, Figure 9, and four new supplementary tables (S2-S5):

Table 5. Genome sequence context for *ARRDC3* and *NFIA* SNPs associated with BCHF.

Figure 9. Predicted protein folding differences between bovine *ARRDC3* C182Y variants.

S2. Extended comparison of *ARRDC3* residues near the C182Y position in the Gnathostomes.

S3. *ARRDC3* and *NFIA* genotypes for 204 case-control animals corresponding to 12 SNPs listed in Table 2.

S4. *ARRDC3* and *NFIA* genotypes for two SNPs from the 566 animals shown in Figure 8.

S5. *ARRDC3* and *NFIA* genotypes for two SNPs from the 1203 animals and 46 breeds listed in Table 4.

We have edited Figure 8 so that the *NFIA* genotypes correspond to only one SNP (BCHF32) and not the three SNP haplotype of BCHF31, 32, and 2. This change does not significantly affect the results or conclusions. We have made changes in the Introduction and Discussion section to emphasize our aim was to search for genomic regions associated with this disease, and that the SNPs reported are not known to be pathogenic and cause disease. Rather they are merely linked to variants that are causing the increased disease risk. We have also emphasized at the beginning and end of the discussion that 15× whole genomic sequence (WGS) analyses are underway for the 102 case-control pairs of animals and may shed future light on the identification of variants directly involved with increased risk. However, the WGS analysis is beyond the scope of the present report. We have added new discussion related to the relatively high frequency of the risk alleles and how that may impact selection strategies. In addition to these amendments, we have made numerous minor changes throughout in response to the reviewers' comments and suggestions.

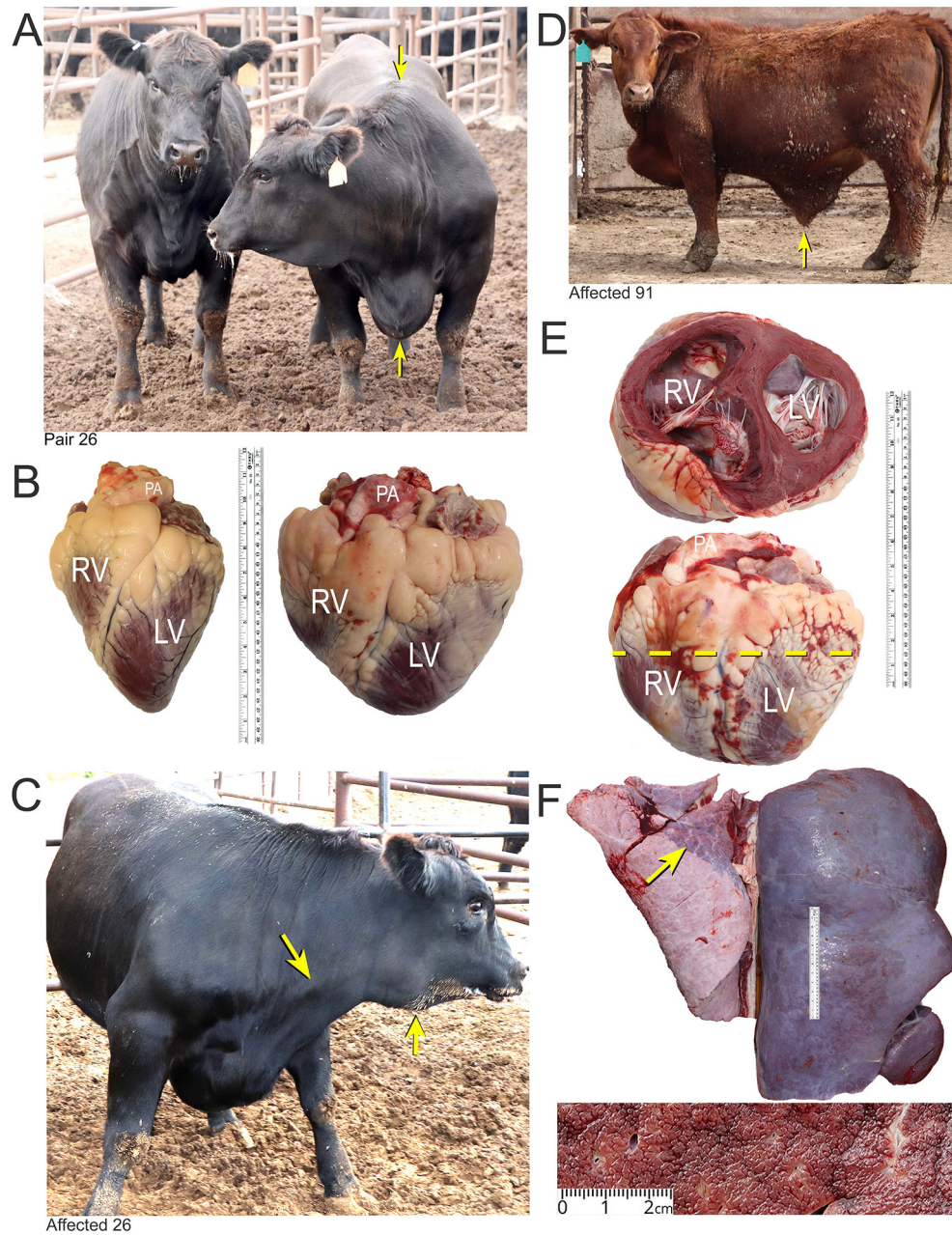
**Any further responses from the reviewers can be found at the end of the article**

## Introduction

Bovine congestive heart failure (BCHF) is a significant cause of death in feedlot cattle at the low to moderate elevations of the Western Great Plains of North America (800 to 1,600 m) (Jensen *et al.*, 1976; Neary *et al.*, 2016). Feedlot cattle in this region are typically a mix of germplasm derived from popular British and Continental European breeds with more than 60% of the animals having a black hide from Angus contributions. Mortality from BCHF has reached 7.5% in severely affected pens of cattle, with annual losses exceeding \$250,000 for a single operation (Heaton *et al.*, 2019). Consequently, reducing the impact of BCHF is a priority for the U.S. cattle industry. The clinical features of BCHF in feedlot cattle are those caused by pulmonary hypertension, right heart failure, and passive liver congestion (Figure 1). BCHF shares some similarities with the high-elevation “brisket disease” disorder that has been known in the Rocky Mountains of Colorado and Utah for more than 100 years, including severe ventral edema of the chest tissues (Glover and Newsom, 1914, 1918). In the high-elevation disorder, the reduced partial pressure of oxygen causes pulmonary hypoxia, vascular resistance, arterial remodeling, and pulmonary hypertension. This condition has been attributed to “cor pulmonale” and eventually causes right ventricular overload and enlargement, ultimately leading to heart failure (Hecht *et al.*, 1962). In BCHF, pulmonary arteries and arterioles also have lesions consistent with hypoxia-induced pulmonary hypertension (Moxley *et al.*, 2019). However, it is unclear whether hypobaric hypoxia is the underlying cause. Some evidence suggests that left heart dysfunction may also play a role in initiating BCHF (Krafsur *et al.*, 2019). Thus, disease pathogenesis of BCHF in feedlot cattle maintained at moderate elevations remains unclear.

Evidence in humans, mice, and cattle suggests that genetic factors influence the risk of pulmonary hypertension and right heart failure. In humans, dozens of inherited DNA sequence variants are known to directly or indirectly affect biological processes involved with cardiopulmonary disease (Morrell *et al.*, 2019; Reza and Owens, 2020). In cattle, specific pathogenic mutations identified in the *PPP1R13L* and *OPA3* genes result in their loss of function (LOF) and cause recessive autosomal cardiomyopathies (Simpson *et al.*, 2009; Owczarek-Lipska *et al.*, 2011). Human and mouse LOF mutations in *PPP1R13L* cause similar cardiomyopathies (Falik-Zaccari, 2017; Herron *et al.*, 2005), as does mouse LOF mutations in *OPA3* (Davies *et al.*, 2008). These examples illustrate the fundamental conservation of mammalian genetics and diseases that can inform the search for optimal cardiovascular health for people and animals.

With regard to the BCHF disorder in Western Great Plains feedlots, there is only indirect evidence of an underlying genetic predisposition. Interviews with feedlot personnel at multiple affected operations indicated that cattle from different ranch sources varied markedly in their incidence of BCHF, despite being from the same breed and raised in the same general region with similar management conditions. Personnel also noted that in most years cattle from the same sources have a predictable prevalence and timing for developing BCHF in their feedlot environment (Heaton, M.P., 2017 unpublished). These observations are consistent with the hypothesis that genomic DNA sequence variants are influencing



**Figure 1. Features of BCHF in feedlot cattle.** Panel A, matched pair number 26 with the clinical case (right) showing ventral edema and a drop in the withers (arrows). The latter occurs as the fluid accumulates in thorax pushing the shoulders outward. Panel B, the affected heart of case 26 (right) compared with a normal heart of a fattened American Angus heifer. Abbreviations: RV, right ventricle; LV, left ventricle; PA, pulmonary artery. Panel C, PA distension and mandibular edema in case 26 (arrows). Panel D and E, clinical case number 91 with ascites (arrow) and enlarged heart with dilated RV. Panel F, Enlarged, encapsulated liver of case 91 with “nutmeg liver” appearance in cross section due to hepatic venous congestion, and atelectasis in lungs (arrow).

BCHF incidence, and each herd has its own frequency of the risk alleles due to founder effects. Based on comparative mammalian genetics, together with these anecdotal observations, we hypothesize that underlying genetic risk factors are influencing BCHF in feedlot cattle.

Candidate gene variants have not been confirmed as risk factors for BCHF in feedlot cattle. A first potential candidate gene variant was reported for cattle with high-altitude pulmonary hypertension (Newman *et al.*, 2015). American Angus cattle affected with pulmonary hypertension at altitudes of 1,478 to 2,618 m had a higher frequency of the endothelial PAS domain-containing protein 1 gene (*EPAS1*) encoding a hypoxia-inducible factor 2 alpha (HIF2 $\alpha$ ) missense variant.

In addition, six other HIF2 $\alpha$  polypeptide sequences were identified in cattle and used to evaluate their potential impact on the adaptive response to chronic hypoxia in U.S. cattle (Heaton *et al.*, 2016). However, a retrospective study with 102 matched case-control pairs of feedlot cattle showed that none of the HIF2 $\alpha$  isoforms encoded by *EPAS1* were associated with BCHF and indicated the need for a wider search (Heaton *et al.*, 2019).

Here we used the same retrospective matched case-control design and the same 102 pairs of cattle from affected feedlots in the Western Great Plains (Table S1, *Extended data*, [Heaton *et al.*, 2019]), together with 778,000 single nucleotide polymorphisms (SNPs), to test the null hypothesis that significant association does not exist between haplotype tagging SNPs and BCHF. Our results rejected the null hypothesis and identified two distinct genomic regions harboring major genetic risk factors associated with BCHF. Cases were matched with contemporary pen mates to reduce confounding association signals that could be caused by population stratification and environmental differences. The strongest association signals were within the arrestin domain-containing protein 3 (*ARRDC3*) gene on chromosome 7, and the gene for nuclear factor IA (*NFIA*) on chromosome 3. The results have implications for deciphering underlying disease mechanisms, selection of breeding animals with reduced risk, and management of animals with high-genetic risk for BCHF.

## Methods

### Ethical statement

The experimental design and procedures used during this research project were reviewed and approved by the Institutional Animal Care and Use Committee of the University of Nebraska-Lincoln as previously described (Experimental Outline numbers 139 and 1172 [Workman *et al.*, 2016; Heaton *et al.* 2019]). The animals were privately owned by commercial feeding operations, and the owners and management approved the use of animals for this study. In every instance, all efforts were undertaken to reduce animal suffering. In addition, animal welfare for these facilities was in accordance with the National Cattlemen's Beef Association's Beef Quality Assurance Feedyard Welfare Assessment program (<http://www.bqa.org/>).

### Animals and study design

Paired samples from 102 affected feedlot cattle and their 102 unaffected matched pen mate controls were identified by monitoring approximately 140,000 total animals in four feedlots located in Nebraska and Wyoming at elevations ranging from 1,163 to 1,280 m. The samples were collected during a 16-month period spanning January 2017 to April 2018 and the 102 pairs originated from 30 different sources with the largest single source contributing 32 pairs (Heaton *et al.*, 2019). The pen sizes ranged from 100 to 300 animals, and control animals were selected based on their willingness to be moved to the sampling area by riders on horseback. Candidate end-stage BCHF cases were identified by feedlot personnel who had experience with clinical signs of BCHF. Researchers were notified and initiated travel to the location (250 to 370 miles). With researchers present, the candidate case animal was euthanized by feedlot personnel and researchers performed necropsy.

The case definition requirement included two or more multiple clinical signs associated with BCHF including: ventral and intermandibular edema ('brisket' and 'bottle jaw'), jugular vein distention and pulsation, ascites/abdominal swelling, and exophthalmia ('bug-eye')<sup>10,11</sup>. The case definition also included two or more non-specific clinical signs: dyspnea, abducted elbows, depression, drooped ears, intermittent watery orange diarrhea, tachycardia, exercise intolerance, open mouth breathing, and weight loss (Heaton *et al.*, 2019). Cases were enrolled in the study only if there was a postmortem presumptive diagnosis of congestive heart failure at necropsy. Of the animals euthanized, the rejection rate for case enrollment was 20% (26/128). Unaffected pen mate controls were matched for source, arrival date, gender, and coat color. The case enrollment rate (102 per 140,000 feedlot cattle monitored, 0.073%) was negatively affected by: 1) removal of candidate case animals for emergency salvage slaughter, 2) removal of matched unaffected pen mates for beef processing, 3) the rapid onset of congestive heart failure and death, 4) reluctance of overburdened personnel to commit extra time despite the \$500 indemnity offer for each matched pair identified early, and 5) exclusion of necropsied candidate animals meeting some, but not all, of the case definition criteria.

None of the 102 unaffected matched pen mate animals developed clinical BCHF signs prior to their processing into beef. Overall, the group of 204 calves was 93% solid black, 70% castrated males, and none had visible horns (Heaton *et al.*, 2019). Most of the 204 calves were from known, well-managed herds that use American Angus genetics. However, breed information was not available for all animals. Tissue samples for genomic DNA isolation included V-shaped ear notches and EDTA whole blood. The ear tissue was desiccated with granular NaCl in the field and stored at -20°C upon return. The plasma and cellular fractions of EDTA whole blood were separated and frozen in liquid nitrogen *en route* and stored at -80°C upon return. Postmortem whole hearts used for gross morphological comparisons representing unaffected cattle were collected during federally-inspected beef processing at the USMARC abattoir from 21 purebred American Angus heifers raised and fattened at an elevation of 578 m.

A panel of 96 unrelated individuals from popular U.S. beef breeds (USMARC Beef Diversity Panel version 2.9 [MBCDPv2.9]) was used for genetic comparisons with case-control animals (Heaton *et al.*, 2016). The panel design was based on a previous set of commercially-available U.S. sires from 19 breeds with minimal pedigree relationships (Heaton *et al.*, 2001). Pedigrees were obtained from leading suppliers of U.S. beef cattle semen and analyzed to identify the least related individuals for inclusion. The number of sires representing each breed (four, five, or six) was based on their numbers of registered progeny circa 2000: Angus (n = 6), Hereford (n = 6), Charolais (n = 6), Simmental (n = 6), Red Angus (n = 6), Limousin (n = 6), Gelbvieh (n = 6), Brangus (n = 5), Beefmaster (n = 5), Salers (n = 5), Shorthorn (n = 5), Maine-Anjou (n = 5), Brahman (n = 5), Chianina (n = 4), Texas Longhorn (n = 4), Santa Gertrudis (n = 4), Braunvieh (n = 4), Corriente (n = 4), and Tarentaise (n = 4). On the basis of the number of registered progeny, the breeds were estimated to represent more than 99% of the germplasm used in the U.S. beef cattle industry. Based on pedigrees, the panel members contained more than 187 unshared haploid genomes (of 192 possible) and allowed a 95% probability of detecting any allele with a frequency greater than 0.016.

### Genotype scoring, SNP filtering, and population substructure evaluation

Unless otherwise indicated, reagents were molecular-biology grade. DNA from ear notches was extracted by standard phenol/chloroform procedures, dissolved in a solution of 10 mM TrisCl, 1 mM EDTA (TE, pH 8.0) and stored at 4°C. (Heaton *et al.*, 2008). SNP genotypes for variants on the BovineHD BeadChip array were scored at GeneSeek Inc. (Lincoln, NE, USA), according to manufacturer's instructions (Illumina, Inc., San Diego, CA, USA). The unfiltered map and ped files from the 102 BCHF case-control animals (Files S1 and S2, *Extended data*) were analyzed for association with BCHF with PLINK v1.90b6.5 64-bit (13 September 2018) software (Chang *et al.*, 2015). The pairs (1 to 102) were designated in the PLINK family ID field. SNPs were removed for having a minor allele frequency less than 0.05 (--maf option). This cutoff was conservative since the frequency of the affected individuals was set at 0.50 *a priori* in this matched case-control design. The missing genotype rate threshold was set at 0.05 (--geno option) and the Hardy-Weinberg equilibrium test with the mid-*p* adjustment threshold was set to 0.0001 (--hwe midp option) (Graffelman and Moreno, 2013). These filters were applied for 204 samples of which 142 were listed as males (steers) and 62 were listed as females (heifers). The animal genders were tested with the --check-sex option and a control animal was identified as being misclassified as a male, leaving 141 steers and 63 heifers, and pair 21 mismatched for sex with the case being a steer and the control being a heifer (Files S3 and S4, *Extended data*).

Since unrecognized population stratification among individuals may obscure association signals and lead to incorrect findings (Deng, 2001; Cardon and Palmer, 2003), multidimensional scaling (MDS) was used with a matrix of pairwise identity-by-state (IBS) distance to visually evaluate the similarity between individuals and pairs. Using PLINK software, unfiltered input map and ped files from the 102 BCHF case-control animals (Files S1 and S2, *Extended data*) were merged with likewise unfiltered input map and ped files from the USMARC Beef Cattle Diversity Panel version 2.9 (MBCDP v2.9) genotyped with BovineHD BeadChip array (Heaton *et al.*, 2016) (Files S5 and S6, *Extended data*). Once merged, the data was filtered as described above to produce a new map and ped files (Files S7 and S8, *Extended data*). MDS data was produced with PLINK software v1.9 and plotted with Microsoft Office 365 Excel software. The IBS/IBD computation was run on the filtered SNP sets with the --genome option. The computation excluded 12,591 variants on non-autosomes from the IBD calculation. Plotting the --cluster output values in the first and second dimensions (i.e., C1 values against C2) gave a scatter plot in which each point is an individual; the two axes correspond to a reduced representation of the data in two dimensions, which facilitates cluster identification. Standard classical (metric) multi-dimensional scaling was used.

### McNemar's test with SNP genotypes

The McNemar's test is a statistical test for paired nominal data (McNemar, 1947). Unlike a traditional case-control GWAS where the scope of inference is over two groups of individuals, a McNemar's test analysis for a case-control GWAS uses only the two matched individuals as the unit of comparison. This approach previously showed significant power to detect the association of *TMEM154* missense variants with ovine progressive pneumonia with 69 matched pairs and a 50k bead array (Heaton *et al.*, 2012). In addition, McNemar's test analysis is not formally available as a convenient option in commonly used, open-source GWAS toolsets that can analyze 560,000 SNPs such as PLINK (v1.90b6.5). The Cochran-Mantel-Haenszel (CMH) test (Mantel and Haenszel, 1959) is available in PLINK and can be run as a specialized case with two clusters; however, it has limited options and does not perform essential McNemar computation with raw data. Thus, a custom workflow was developed with a programming and numeric computing platform (MatLab v.2021.2, MathWorks, Inc., Natick, MA) to produce the critical tables of binomial outputs and McNemar statistics for the filtered SNP genotypes.

In McNemar's test, informative pairs are those which contain exactly one animal with the hypothetical risk allele genotype, i.e., Table 1, quadrants *b* and *c*. Depending upon the scenario tested, the risk allele genotype may include one or two copies of the variant. The ratio of the discordant pairs (*b/c*) provides evidence to reject the null hypothesis. At each

**Table 1. The four binomial outcomes possible in this McNemar's test<sup>a</sup>.**

	Control has risk allele(s)	Control lacks risk allele(s)	Row totals
Case has risk allele(s)	(+,+) <i>a</i>	(+,-) <i>b</i>	<i>a + b</i>
Case lacks risk allele(s)	(-,+) <i>c</i>	(-,-) <i>d</i>	<i>c + d</i>
	<i>a + c</i>	<i>b + d</i>	<i>n</i>

<sup>a</sup>Each binomial outcome is a case-control pair, i.e., (case, control) with the sign indicating the presence or absence of the risk allele. For example, (+,-) indicates an informative outcome where the case has the risk allele and the control does not. The uninformative pairs are (+,+) and (-,-) where each member of the pair has identical risk allele status. Note that this table can be set up for scenarios of 1, 1 or 2, and 2 copies of risk alleles.

biallelic SNP locus, alleles were tested separately in three scenarios as a potential disease risk factor: 1-copy, 1- or 2-copies, and 2-copies, producing six sets of McNemar's test statistics. The presence of a potential risk allele(s) was scored for both animals in every matched pair and used to populate the McNemar's  $2 \times 2$  contingency table for a given SNP. The presence (+) or absence (-) of a possible genetic risk factor was assigned to each animal and the matched pair was then classified by one of four possible binomial outcomes (Table 1, quadrants). A genetic "risk" allele was indicated in the 102 matched pairs when quadrant *b* was greater than quadrant *c* (i.e., odds ratio (OR) > 0). Conversely, a "protective" allele was indicated when quadrant *c* was greater than quadrant *b* (i.e., OR < 0).

McNemar's test analysis for a GWAS imposes statistical constraints at the SNP level that are not present in conventional GWAS. For example, the McNemar's test *p*-values for rejection of the null hypothesis were computed with the binomial distribution. This is different from a conventional retrospective case-control GWAS, where the theoretical distribution of *p*-values may be modeled with Fisher's exact test statistics (or derivative). The binomial distribution in the McNemar's analysis generated discrete *p*-values (e.g., exact *p* and mid *p*) at each SNP site that were dependent on the discordant pairs observed (i.e., *b + c* and *b/c*) for each inheritance model (e.g., 1-copy, 1- or 2-copies, and 2-copies). Thus, *p*-values were computed *de novo* at each locus. To the best of our knowledge, the theoretical genome-wide distribution of McNemar's *p*-values at the SNP level has not been previously described in the GWAS context. Briefly, the input files consisted of PLINK map and ped files with the ped file modified to include the pair identifier in the first column, and the rows sorted in descending pair order from 1 to 102 with the unaffected animal listed first in the pair. The case and control phenotype were coded in the sixth column, with the unaffected control animal coded with 1 (PHENO = 1) and the affected case animal with a 2 (PHENO = 2). The analysis evaluated each SNP for possible risk alleles in the context of McNemar's test as described in README.md file in the GitHub code repository [McNemarsSNPAnalysis](#). Alternatively, the analysis pipeline can be run online within a copy of the [McNemar's SNP Analysis Compute Capsule](#) at [CodeOcean](#). The PLINK map and ped files contained 141 steers and 63 heifers, and SNPs assigned to chromosome 0 through 32 (i.e., included unassigned SNPs, chromosome X, Y, and mitochondrial DNA; Files S9 and S10, *Extended data*). In the first stage of the workflow, biallelic SNP heterozygous animal genotypes were converted to a one-letter International Union of Pure and Applied Chemistry and International Union of Biochemistry and Molecular Biology (IUPAC/IUBMB) ambiguity codes used for nucleotides (i.e., R = a/g, Y = c/t, M = a/c, K = g/t, S = c/g, W = a/t ('Nomenclature Committee for the International Union of Biochemistry [NC-IUB]. [Nomenclature for incompletely specified bases in nucleic acid sequences. Recommendations 1984', 1986](#)) and homozygous animal genotypes were converted to a one-letter code of the same letter (i.e., A, C, T, G).

The McNemar mid-*p* test was used, since it is an improvement over the exact conditional test ([Fagerland et al., 2013](#)). Briefly, the McNemar mid-*p*-value was calculated with [equation 1](#):

$$2*[\text{binomcdf}(n_{12}, b + c, 0.5) - \frac{1}{2}\text{binompdf}(n_{12}, b + c, 0.5)] \quad (1)$$

where  $\text{binomcdf}$  = binomial cumulative distribution function,  $\text{binompdf}$  = binomial probability distribution function, *b* = the number of informative pairs with only the case having the risk factor ("successes"), *c* = the number of informative pairs with only the control having the risk factor ("failures"), *b + c* = "trials" (i.e., the number of discordant pairs),  $n_{12}$  = the smaller value of *b* or *c*, and 0.5 = "probability of success" (i.e., probability of being *b* or *c* by chance). In Excel, this formula returns the McNemar mid-*p*:

$$= 2*((\text{BINOM.DIST}((\text{IF}(b < c, b, c)), b + c, 0.5, \text{TRUE})) - (\frac{1}{2}(\text{BINOM.DIST}((\text{IF}(b < c, b, c)), b + c, 0.5, \text{FALSE}))))$$

where "TRUE" was the option for the cumulative distribution function and "FALSE" was the option for the probability distribution function.



The quantile-quantile (Q-Q) plots of the distribution of the test statistics are challenging to produce because McNemar's test data are discrete for a given number of total pairs and the fraction of discordant pairs  $((b + c)/n, \text{Table 1})$ . This requires the computed theoretical expected distribution of  $p$ -values to mirror the dependency of the observed distribution of  $p$ -values on the fraction of discordant pairs. Thus, for a given informative pair  $(b + c)$  value, theoretical  $p$ -values in Q-Q plots were computed by multiplying the number of observed  $p$ -values by the predicted normalized binomial distribution of all possible  $b$  and  $c$  values, as defined by the  $p$ -value metric used, e.g. exact  $p$  or mid  $p$ . The fraction of discordant pairs is also influenced by minor allele frequency, as the opportunity for the observation of discordant pairs diminishes with frequency.

Systematic use of  $q$ -values in genome-wide tests provides a balance between the number of true and false positives while being automatically calibrated and readily interpreted (Storey and Tibshirani, 2003). The  $q$ -value of a genome-wide data set is the expected proportion of false positives incurred (Storey and Tibshirani, 2003). For instance, a  $q$ -value threshold of 5% results in the false discovery rate (FDR) of 5% of the significant features being truly null. The mafdr function in the Matlab Bioinformatics Toolbox (version 2021.2) was used to compute the  $q$ -values from input  $p$ -values according to procedures previously described (Storey, 2002). The tuning parameter,  $\lambda$ , was used to estimate the *a priori* probability that the null hypothesis is true (equation 2):

$$\left( \pi_0 \equiv \frac{\text{number of true null observations}}{\text{number of all observations}} \right). \quad (2)$$

This parameter was computed both with the 'bootstrap' and 'polynomial' methods in the mafdr function yielding similar results. The polynomial method was a cubic polynomial fit of  $\lambda$  to  $\hat{\pi}_0(\lambda)$  (equation 6 in [Storey, 2002]), as demonstrated for the 2-copy risk allele model (Figure S1A, *Extended data*). In contrast to previous assumptions where the frequency of  $p$ -values tend to remain constant with  $p$ -values  $> 0.5$  (Storey, 2002),  $p$ -values in the present report exhibit an increasing frequency  $> 0.4$  and are not uniformly distributed because of the discrete results generated by the McNemar's test (Figure S1, *Extended data*). Consequently, the polynomial method of computing  $\hat{\pi}_0(\lambda)$  was negatively affected, resulting in poor cubic polynomial fit with  $R^2 = 0.17$ , especially as  $\lambda$  approaches 1. Recalling that when the  $p$ -value equals one, McNemar quadrants  $b$  and  $c$  are also equal, and thus significant SNP associations are not expected; stated another way, it would be extremely unusual for any SNP with a McNemar  $p$ -value of one to be truly associated with disease risk. The computed  $\hat{\pi}_0$  value tends to increase as the frequency of discrete  $p$ -values inflate. This in turn decreases the number of observations that are found to be non-null significant SNPs that pass a given  $q$ -value threshold. For instance, in the scenario where homozygous major alleles are evaluated as genetic risk factors, the number of SNPs passing the  $q$ -value = 0.05 threshold using the polynomial method was 55 with a  $\hat{\pi}_0 = 0.8486$  while the bootstrap methods found 73 significant SNP with a  $\hat{\pi}_0 = 0.5709$ . Therefore, for a given  $q$ -value threshold, the computed SNPs found to be significant in this McNemar analysis were conservative compared to the more conventional GWAS with continuous  $p$ -values distributions that tend to "flatten" with  $p$ -values  $> 0.5$ , reducing  $\hat{\pi}_0$  and increasing the number of non-null significant SNPs found using the canonical Storey analysis as embodied in the mafdr function. To explore the expected genome-wide  $p$ -values for a McNemar-based Q-Q plot, the observed normalized  $p$ -value distribution (parametrically dependent upon McNemar's quadrants  $b$  and  $c$  [informative pairs] for a given  $b + c$ ) was compared to the normalized expected theoretical distribution of  $p$ -values using the same parameters (i.e., observed  $b + c$  values) at each SNP. For example, computing the Q-Q plot's expected distribution of  $p$ -values over all SNP loci required that the expected  $p$ -value be computed relative to the number of observed informative pairs at each locus. In addition, it was necessary to confirm that the approximations used for computing the false discovery rate to correct for multiple test biases were still valid for discrete  $p$ -values.

The six sets of McNemar's pair classifications and test statistics were tallied for each SNP allele and exported to a text file for importation into spreadsheet software (Excel Version 1902, Build 11328.20420). A spreadsheet environment allowed for the SNP set to be used in calculations for McNemar's statistics such as odds ratios (OR),  $p$ -values, and McNemar Chi-squared statistics. After McNemar's testing, SNPs were removed due to their assignment to chromosomes 0, 31, and 32. These SNPs represent markers on unmapped chromosomes, SNPs on chromosome Y, and mitochondrial SNPs. SNPs were also removed if there were no informative pairs resulting from their pairwise analysis (i.e., McNemar quadrants  $b + c = 0$ ).

The upper and lower  $CI_{95}$  for the OR were calculated in equations 3 and 4:

$$e^{(\ln(\text{OR})+1.96(\text{SE of } \ln(\text{OR})))} \quad (3)$$

$$e^{(\ln(\text{OR})-1.96(\text{SE of } \ln(\text{OR})))} \quad (4)$$

respectively, where  $e$  = the natural base,  $\ln$  = natural logarithm, and  $SE$  = the standard error ( $\sqrt{1/b + 1/c}$ ). The McNemar test statistic  $\chi^2 = (b-c)^2/(b+c)$ , and  $\chi^2$  with the continuity correction =  $(|b - c| - 1)^2/(b + c)$ . The effect size (Cohen's  $g$ ) was the larger of  $(b/(b+c))$  or  $c/(b+c)$  - 0.5 (Cohen 1988, Statistical Power Analysis for the Behavioral Sciences).

The spreadsheet format allowed for custom labeling of column headers, row sorting, and color-coded heat mapping. For example, it was useful to sort each of the six sets of data by their best McNemar's  $p$ -value in descending order and color code the rows in blocks of 100 (e.g. purple, red, yellow, dark green, light green). This helped visually identify the top 500 best markers in each data set. When resorted by chromosome and position, the  $p$ -value color-coded rows produced an obvious visual symmetrical pattern due to the biallelic nature of SNPs and the dichotomous phenotype. An exception to the pattern occurred in regions where the major and minor allele frequency were equivalent since their assignment was arbitrary for the set of 204 animals.

The spreadsheet format also allowed for manual production of Manhattan plots from McNemar's data for three inheritance models of the two alleles for each SNP. Briefly, the ARS-UCD1.2 reference assembly SNP coordinates from the PLINK map file (i.e., chr and position) were converted to a linear genome position for the x-axis of the Manhattan plot so that every SNP had a new relative coordinate on a whole genome scale from 1 to 2,651,083,094 bp. This was accomplished by adding the position value of the most terminal SNP on chromosome 1 to each of the SNP position values on chromosome two. This was repeated for each consecutive chromosome (1 to 29 and X). The Bonferroni correction for multiple testing was used at the 0.05 significance level ( $\alpha$ ) as a conservative genome-wide estimate of significance (Johnson *et al.*, 2010) (Equation 5):

$$\text{Bonferroni genome - wide significance threshold} = \alpha / (\text{filtered SNPs}) \quad (5)$$

### Power analysis

Sample size power analysis for the McNemar's test was performed with PASS 2019 Power Analysis and Sample Size Software, version 19.0.2 (NCSS, LLC., Kaysville, Utah, USA). The sample size of 102 matched pairs achieved 95% power to detect an OR of 8.51 and 2.94 with a two-sided McNemar's test and a corrected significance level of 0.01 ( $\alpha$ ) within the range of 25% and 70% discordant pairs, respectively. Power is the probability of rejecting a false null hypothesis, while  $\alpha$  is the probability of rejecting a true null hypothesis.

### CMH test

McNemar's test analysis was compared to that obtained with the CMH association statistic (Mantel and Haenszel, 1959) in PLINK v1.90b6.5. The 102 matched pairing assignments were coded in the fam file in the third column to identify 102 "clusters" (e.g., Pair<sub>1</sub>, Pair<sub>2</sub>, Pair<sub>3</sub>, ... Pair<sub>102</sub>) for use with the CMH association statistic. The --adjust option was also used to produce a sorted list of CMH association results. The Q-Q plot of the distribution of the test statistics was produced with the qqman R package on CRAN: (Turner, SD. 2014. bioRxiv) with 64-bit R version 3.5.1 (2018-07-02).

### Testing BCHF cases and controls for bovine viral diarrhoea virus (BVDV) infection

Blood samples from BCHF cases and controls were tested for infection with BVDV, a possible confounding disease condition of feedlot cattle that is easily detected. However, since these animals were not serially sampled, a positive test result would not distinguish between acute and persistent infection. Blood samples for the first 13 case-control pairs were collected as EDTA whole blood, stored at -80°C, and used for BVDV testing. Blood samples for the remaining 89 case-control pairs were processed by first separating the plasma by centrifugation and storing it separately at -80°C prior to its use in BVDV testing. EDTA blood or plasma was pooled into groups of five or 10, respectively, and total RNA was extracted with a mono-phasic solution of phenol and guanidinium thiocyanate (TRIzol LS, Invitrogen-Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's specifications. BVDV genomic RNA was detected with a previously published, real-time assay that used quantitative reverse transcription polymerase chain reaction (RT-qPCR) primers (5'-GGGNAGTCGTCARTGGTTTCG-3'; 5'-GTGCCATGTACAGCAGAGWTTTT-3') and probe (5'-6-FAM-CCAYGTGGACGAGGGCAYGC-BHQ1-3') (Mahlum *et al.*, 2002). The assay reagents were purchased as a commercial kit (Qiagen OneStep RT-PCR, Germantown, MD) and used according to the manufacturer's instructions. Briefly, cyclic amplification was conducted in a 25  $\mu$ L reaction containing 5  $\mu$ L of extracted RNA, 4.5 mM magnesium chloride, 400  $\mu$ M dNTP, 0.4  $\mu$ M of each forward and reverse primer, 0.2  $\mu$ M probe, and 1  $\mu$ L of an enzyme mix containing reverse transcriptase and a "hot-start" *Taq* polymerase. Cycling conditions included reverse transcription at 50°C for 30 minutes, inactivation of the reverse transcriptase enzyme and activation of the *Taq* polymerase at 95°C for 15 minutes, and 40 cycles of 94°C for 30 seconds, 55°C for 60 seconds, and 72°C for 60 seconds. Positive, negative, no-template, and extraction control samples were included in each group of samples tested.

### SNP genotype assays for *ARRDC3* and *NFIA*

Custom SNP genotype tests were developed on a commercially available genotyping platform that uses “padlock” oligonucleotide probes, combined with fluorescently-labeled probes and isothermal rolling circle amplification technology, for the qualitative detection of alleles (MatMaCorp, Lincoln, NE, USA). Padlock probes are long oligonucleotides whose ends are complementary to adjacent target sequences. The primary target SNPs included *ARRDC3* C182Y (ARS-UCD1.2 chr7:90845941, rs109901274, BovineHD0700027239, BCHF5) and *NFIA* intron 4 (ARS-UCD1.2 chr3:84578325, rs136210820, BovineHD0300024307, BCHF31). However, genotyping tests involving tightly linked SNPs were also used on occasion. The assay steps consisted of sample processing (blood and ear notch), target amplification, and detection. Samples were processed and analyzed on the Solas 8 instrument per the manufacturer’s instructions. Briefly, samples were mixed with a proprietary lysis buffer and heated for three minutes at 95°C. Kit reagents were added to perform an initial PCR amplification with padlock probes. The machine assay conditions were SNP-specific and uploaded to the machine via the internet by the supplier: an initial heating phase at 93° for 3 min, followed by 30 cycles at 93° for 20 s and 57°C for 30 s. Detection of the ligated circular oligonucleotide products in the Solas 8 instrument was accomplished with fluorescently labeled probes and isothermal rolling circle amplification. The total reaction time was approximately two hours, and the genotype assignments were made in real time by the instrument, and the real-time graphical outputs were visually reviewed for quality assurance. Assay performance was validated with known genotypes derived from whole genome sequencing and the BovineHD BeadChip array from a beef cattle diversity panel (Heaton *et al.*, 2016). Ambiguous genotype results from animals of unknown genotypes were resolved by repeat testing with a second assay designed for the opposite strand of *ARRDC3* C182Y (ARS-UCD1.2 chr7:90845941, rs109901274, BovineHD0700027239, BCHF5r) and the *NFIA* intron 4 SNP (ARS-UCD1.2 chr3:84580655, rs133192205, BovineHD0300024308, BCHF32).

### Protein folding prediction of *ARRDC3* structures

The predicted protein structure of the bovine *ARRDC3* was compared with the that from humans. These two proteins were 99% identical across all 414 amino acids (The National Center for Biotechnology Information [NCBI] reference sequences NP\_001069725.1 and NP\_065852.1, respectively). An artificial general intelligence software that uses machine learning was used to fold the *ARRDC3* proteins (ColabFold v1.5.5: AlphaFold2 using MMseqs2; (Mirdita *et al.*, 2022). Each of the three *ARRDC3* sequences was folded separately, and the .pdb structure file with the highest confidence was selected as the best representation of the biologically relevant structure. The two *ARRDC3* variant files were loaded into a molecular graphics system (PyMOL Molecular Graphics System v2.5.3; Schrödinger, LLC, New York, NY), aligned, and rotated to view the variant region.

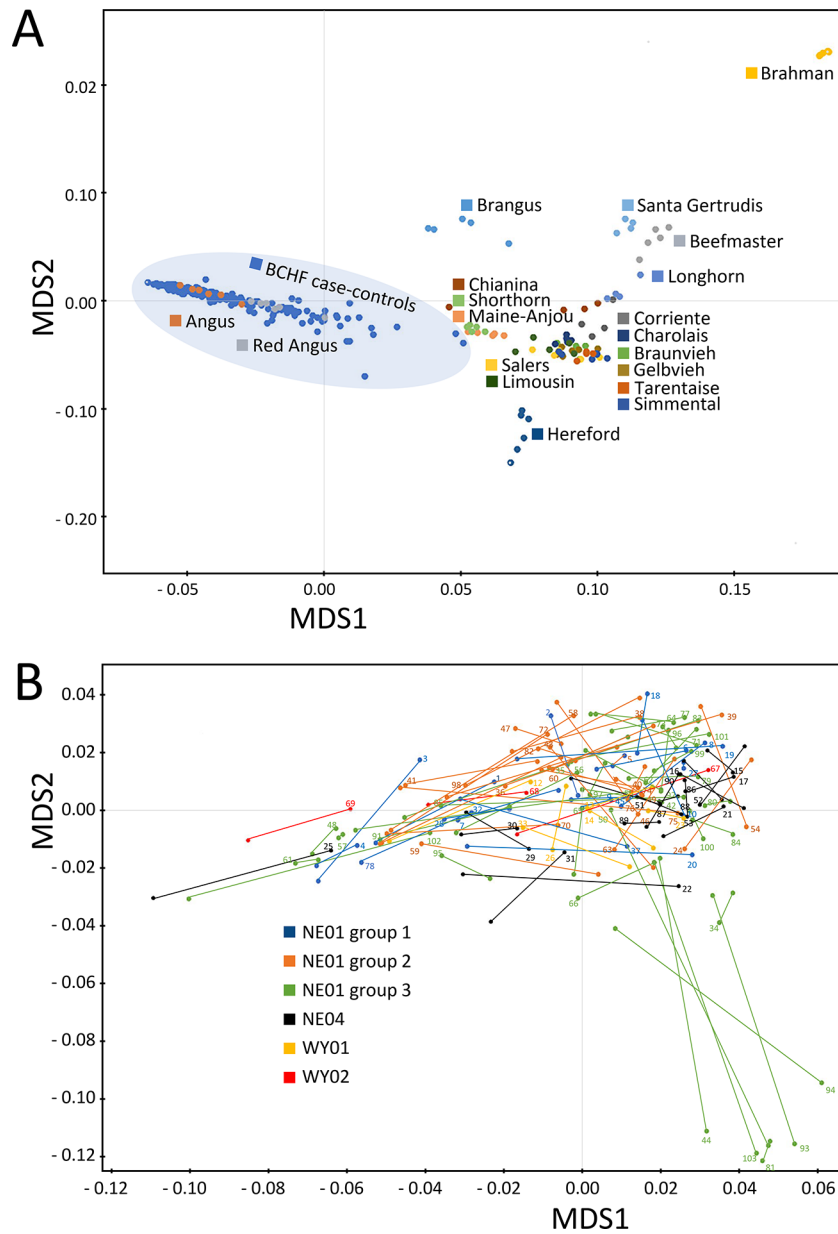
## Results

### Genotype scoring, SNP filtering, and population substructure

Approximately 778,000 SNPs were scored for 204 animals in matched case-control pairs with a total genotyping rate of 0.99 (Files S1 and S2, *Extended data*). Subsequent filtering removed 216,000 SNPs with either low minor allele frequencies, missing genotypes, or deviation from Hardy-Weinberg expectations (173,000, 42,000, and 1,000, respectively). Approximately 560,000 remaining SNPs (Files S3 and S4, *Extended data*) were used for MDS analysis of pairwise IBS distances of the 204 animals from the 102 matched pairs, together with a 19-breed beef diversity panel. The feedlot animals from the 102 matched pairs were tightly clustered with purebred American Angus and Red Angus cattle (Figure 2A). A higher resolution plot with the 102 pairs of BCHF cases and matched controls showed there were seven outliers from one feedlot group, of which five were cases and two were controls (Figure 2B, lower right quadrant). However, many of the matched pairs had short IBS distances, indicating close genetic relationships between random penmates, as would be expected from animals originating from the same ranch. Based on the relatively low number of outliers that were BCHF cases, and the distances of their matched pairs, corrections for population stratification were not made in subsequent analyses.

### Genome-wide McNemar association testing

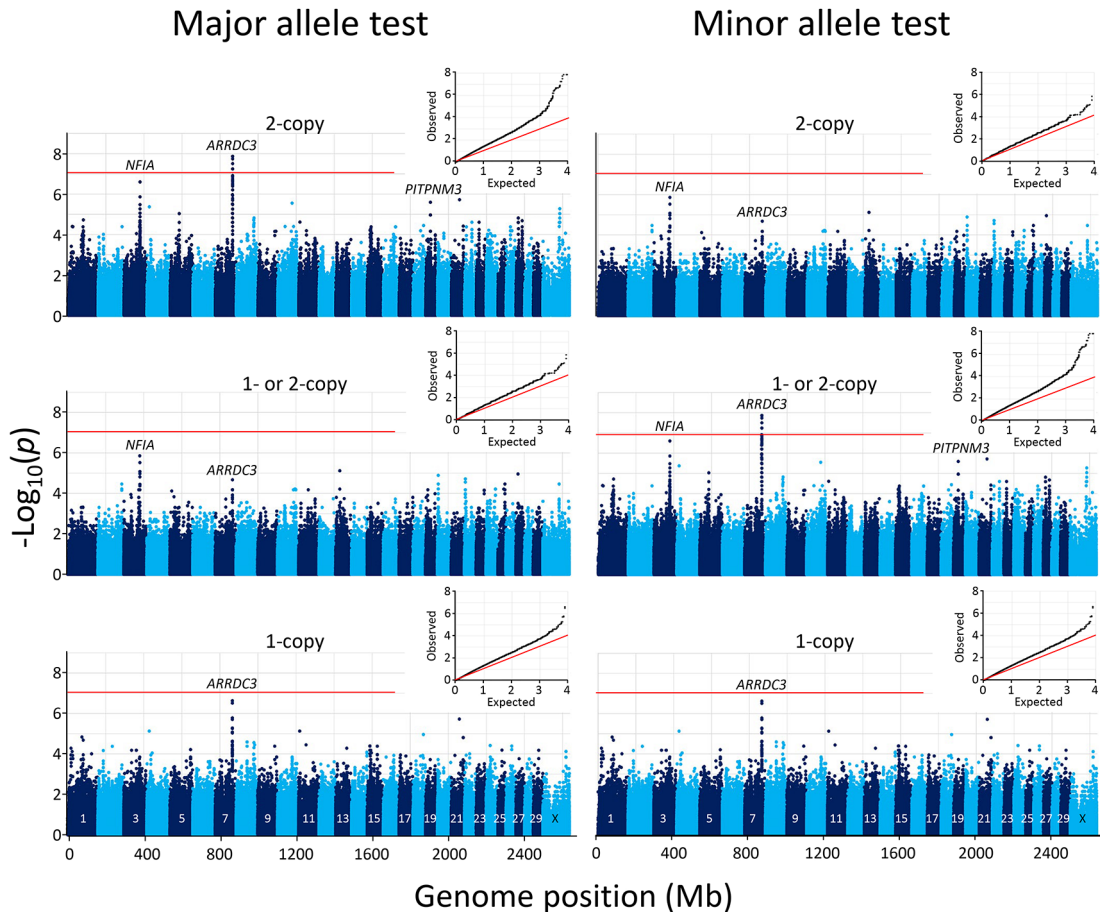
The CMH test was first used to analyze the data set with PLINK software, since it was available and tests for SNP-disease association conditional on clustering of the individuals and is a generalization of McNemar’s test. The CMH test is expected to give the same results as McNemar’s test when each cluster contains a pair of individuals since their test statistics are identical when each stratum shows a pair. However, CMH output in PLINK did not provide easy access to the individual McNemar’s quadrants results for each SNP nor the ability to test 1-copy, 1 or 2-copy, and 2-copy risk allele models. Thus, we developed a custom process on a programming and numeric computing platform as described in the Methods section. After producing the tables of McNemar test statistics (File S11, *Extended data*), SNPs were further removed because they had no informative pairs (n = 704), were unmapped (n = 668), on the chromosome Y (n = 31), or on the mitochondrial genome (n = 1) (File S12, *Extended data*). A Manhattan plot of the resulting SNPs revealed that the most significant genome-wide associations were clustered on chromosome 7 at the *ARRDC3* gene (Figure 3).



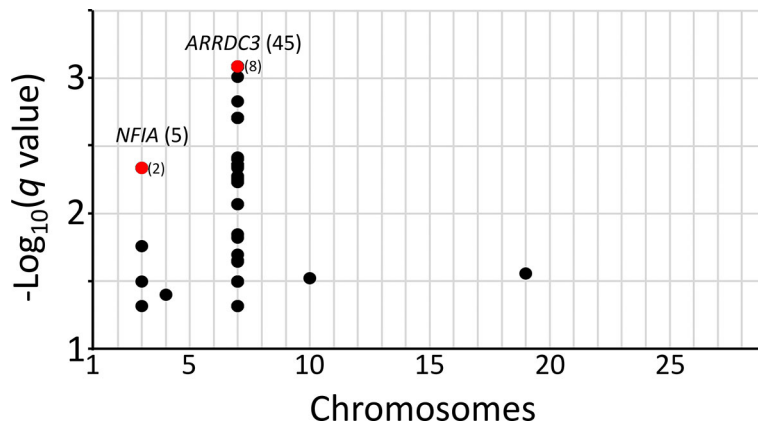
**Figure 2. MDS plot with pairwise IBS distances between breeds and paired individuals.** Panel A, genetic distances between BCHF case-control animals and diverse purebred bulls from 19 U.S. beef breeds. Panel B, genetic distances between matched pair from different feedlot groups and locations in Nebraska (NE) and Wyoming (WY). Lines connecting dots identify each matched pair with numbers corresponding to the pair ID (Table S1) and the clinical case.

No other genomic region contained SNPs that met the Bonferroni threshold at the 0.05 genome-wide significance level ( $-\log_{10} = 7.05$ ).

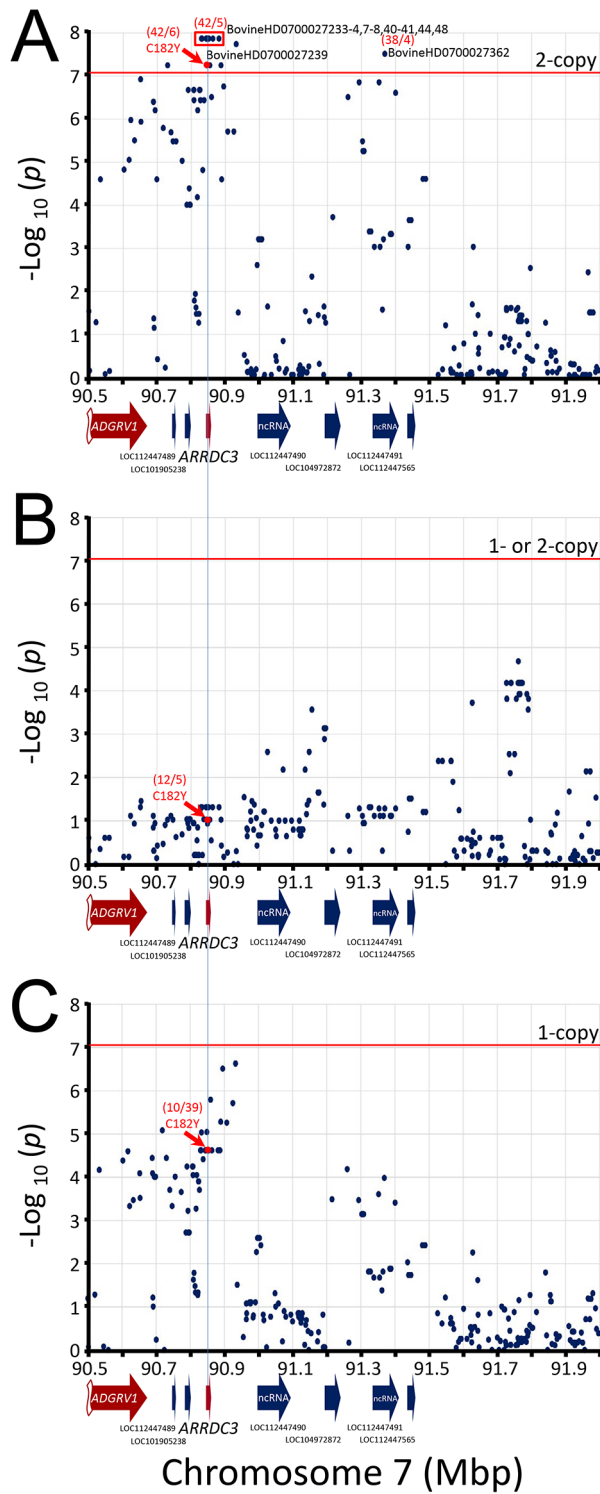
While the Bonferroni correction protects against Type I errors (false positives), it is vulnerable to Type II errors (false negatives). To strike a balance between the number of true and false positives,  $q$ -values were calculated for the McNemar data set to identify regions surpassing a 5% FDR. The  $q$ -value threshold is the expected proportion of significant SNP associations that are false leads. Only models that compared homozygous risk alleles against all other genotypes yielded significant results. There were 53 genome-wide autosomal SNPs passing the 5% FDR: 49 were associated with *ARRDC3* on chromosome 7, five were associated with *NFIA* on chromosome 3, and one SNP each was associated with a region on chromosomes 4, 10, and 19 (Figure 4, File S13 Extended data). Since the three SNPs on chromosomes 4, 10, and



**Figure 3.** Manhattan and Q-Q plots of genome-wide association with 102 BCHF matched case-control pairs. The mid- $p$ -values from the McNemar’s association test were used with major (left) and minor alleles (right). In both panels, the inset shows the Q-Q plot of the distribution of the test statistics and the red horizontal line shows the Bonferroni correction at an  $\alpha$  of 0.05.



**Figure 4.** Genome-wide SNPs with a  $q$ -value threshold less than 0.05 FDR. The  $q$ -values for SNPs were evaluated for the two-copy risk allele model and plotted by chromosome. Numbers in parentheses are linked SNPs with  $q$ -values less than 0.05. Red dots and parentheses represent multiple linked SNPs with equivalent  $q$ -values.



**Figure 5. Manhattan plot of *ARRDC3* gene region.** The mid-*p*-values from the McNemar's association test of the most frequent SNP alleles were used. Panel A, test requiring the presence of homozygous risk alleles. Panel B, test requiring the presence of either 1- or 2-copies of the risk alleles. Panel C, test requiring heterozygosity for the risk alleles. The red horizontal line shows the Bonferroni correction at the 0.05 significance level. The red numbers in parentheses are McNemar's informative pairs from quadrants *b* and *c* (i.e., "*b/c*", Table 1). The ratio of informative pairs indicates the strength of the association, and whether the allele is positively or negatively associated with BCHF (i.e., *b/c* > 0 or < 0).

**Table 2. Homozygous SNPs associated with the highest risk of BCHF in feedlot cattle.**

Genome position <sup>a</sup>		MAF		McNemar statistics							PRSh					
Chr	UMD3.1 (bp)	ARS-UCD1.2 (bp)	SNP ID (rs#)	A <sub>1</sub> /A <sub>2</sub> <sup>b</sup>	case	control	b	c	OR	CI <sub>95</sub>	χ <sup>2c</sup>	-log (p) <sup>d</sup>	-log (q) <sup>e</sup>	Effect size <sup>f</sup>	Proportion of informative pairs <sup>g</sup>	PRSh
7	93229001	90830012	BovineHD0700027233 (rs110586518)	A/G	0.18	0.40	42	5	8.4	3.3-21	27.6	7.86	3.09	0.394	0.461	0.18
7	93231670	90832681	BovineHD0700027234 (rs135607340)	A/G	0.18	0.40	42	5	8.4	3.3-21	27.6	7.86	3.09	0.394	0.461	0.18
7	93243245	90844254	BovineHD0700027237 (rs109439459)	A/G	0.18	0.40	42	5	8.4	3.3-21	27.6	7.86	3.09	0.394	0.461	0.18
7	93243714	90844723	BovineHD0700027238 (rs136675904)	C/T	0.18	0.40	42	5	8.4	3.3-21	27.6	7.86	3.09	0.394	0.461	0.18
7	93244933	90845941	BovineHD0700027239 (rs109901274)	T/C	0.21	0.42	42	6	7.0	3.0-16	25.5	7.24	2.71	0.375	0.471	0.18
7	93247780	90848788	BovineHD0700027240 (rs81180002)	G/T	0.18	0.39	42	5	8.4	3.3-21	27.6	7.86	3.09	0.394	0.461	0.18
7	93251138	90851664	BovineHD0700027241 (rs3423252075)	T/C	0.18	0.40	42	5	8.4	3.3-21	27.6	7.86	3.09	0.394	0.461	0.18
7	93263102	90863630	BovineHD0700027244 (rs3423255708)	C/T	0.18	0.40	42	5	8.4	3.3-21	27.6	7.86	3.09	0.394	0.461	0.18
7	93281534	90882079	BovineHD0700027248 (rs3423252071)	T/C	0.18	0.40	42	5	8.4	3.3-21	27.6	7.86	3.09	0.394	0.461	0.18
3	85123495	84578325	BovineHD0300024307 (rs136210820)	G/A	0.35	0.43	37	5	7.4	2.9-19	22.9	6.60	2.34	0.381	0.412	0.16
3	85125825	84580655	BovineHD0300024308 (rs133192205)	T/C	0.37	0.41	37	5	7.4	2.9-19	22.9	6.60	2.34	0.381	0.412	0.16
3	85254825	84707876	BovineHD0300024366 (rs42583839)	T/C	0.25	0.52	44	11	4.0	2.1-7.7	18.6	5.18	1.32	0.300	0.545	0.16

<sup>a</sup>The SNPs were sorted by chromosome and genomic position. Markers within each genomic region are considered to be statistically equivalent within the margin of error.

<sup>b</sup>A<sub>1</sub> was defined as the most frequent allele in the group of 204 cases and controls and is consistent with the ARS-UCD1.2 top strand (antisense strand in *ARRDC3* and *MFA* genes). The bolded red allele is associated with risk.

<sup>c</sup>McNemar's chi-squared with continuity correction (see Methods section).

<sup>d</sup>Log10 of the mid-p-value.

<sup>e</sup>Log10 of the q-value.

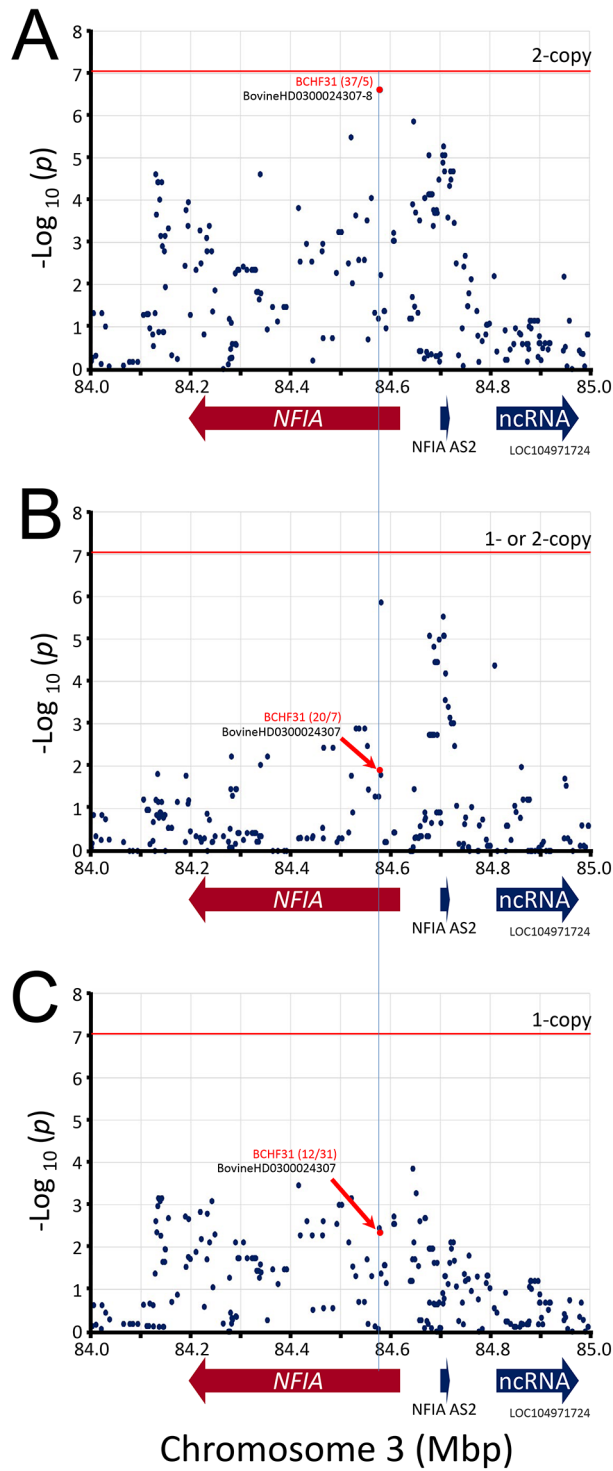
<sup>f</sup>The effect size (Cohen's g) for the McNemar test ranges from 0 to 0.5 where 0.5 is the maximum effect. It is the larger value of: (b/(b+c) or c/(b+c)) - 0.5.

<sup>g</sup>The proportion of informative pairs is (b+c)/n.

<sup>h</sup>The polygenic risk factor value is the (effect size) x (proportion of informative pairs) and represents a value for summing an animal's polygenic risk score.

<sup>i</sup>NP BovineHD0700027239 corresponds to *ARRDC3* C182Y (codon tRt) in exon 4 where tGt encodes C182 and tAt encodes Y182.

<sup>j</sup>NP BovineHD0300024366 was included for comparison, since it was used in early analyses for obtaining frequency data of U.S. cattle breeds in Table 4.



**Figure 6. Manhattan plot of *NFIA* gene region.** The mid- $p$ -values from the McNemar’s association test of the most frequent SNP alleles were used. Panel A, test requiring the presence of homozygous risk alleles. Panel B, test requiring the presence of either 1- or 2-copies of the risk alleles. Panel C, test requiring heterozygosity for the risk alleles. The red horizontal line shows the Bonferroni correction at the 0.05 significance level. The red numbers in parentheses are McNemar’s informative pairs from quadrants  $b$  and  $c$  (i.e., “ $b/c$ ”, Table 1). The ratio of informative pairs indicates the strength of the association, and whether the allele is positively or negatively associated with BCHF (i.e.,  $b/c > 0$  or  $< 0$ ).



19 correspond to the expected proportion of false leads, these regions were dismissed for the purposes of the present study. The remaining genomic regions associated with *ARRDC3* and *NFIA* genes were further characterized.

The best associated region spanned 52 kb and was roughly centered on the *ARRDC3* gene. This region contained eight equivalent linked SNPs with  $-\log_{10}$  mid-*p-values* of 7.86 (Figure 5A) and contained a missense variant in *ARRDC3* codon 182 in exon 4 encoding cysteine (C, tgt) or tyrosine (Y, tat). The homozygous major alleles for the eight best SNPs (and *ARRDC3* Y182) were associated with increased BCHF risk (OR 8.4, CI<sub>95</sub>, 3.3 to 21, Table 2). Conversely, one or two copies of the minor alleles for these eight SNPs (and *ARRDC3* C182) were associated with reduced BCHF risk (OR = 0.12, File S12, Extended data). Heterozygosity (i.e., having one risk allele) was also associated with reduced BCHF risk (0.26). The second best associated region spanned a 2.3 kb region in *NFIA* intron 4, and there were no SNPs present on the BovineHD BeadChip array predicted to affect the *NFIA* coding sequence. The 2.3 kb region contained two equivalent linked SNPs with  $-\log_{10}$  mid-*p-values* of 6.60 (Figure 6). Homozygosity for the major allele of either of these two SNPs was associated with increased BCHF risk (OR 7.4, CI<sub>95</sub>, 2.9 to 19, Table 2). Conversely, one or two copies of the minor alleles for these two SNPs were associated with reduced BCHF risk (OR = 0.14, File S12 Extended data). Heterozygosity was also associated with reduced BCHF risk (0.40). Thus, two copies of the major *ARRDC3* and *NFIA* haplotype alleles were associated with BCHF risk.

### Conservation of *ARRDC3* cysteine codon at position 182

The association of a missense variant in *ARRDC3* with BCHF raised the possibility that the amino acid substitution could affect protein function and thereby increase the risk of disease in feedlot cattle. A comparison with *ARRDC3* codons at position 182 in the jawed-vertebrate species (Gnathostomes) showed the cysteine residue was invariant throughout the 41 representative species examined (Figure 7). Examples of Gnathostome species with alternative residues at the

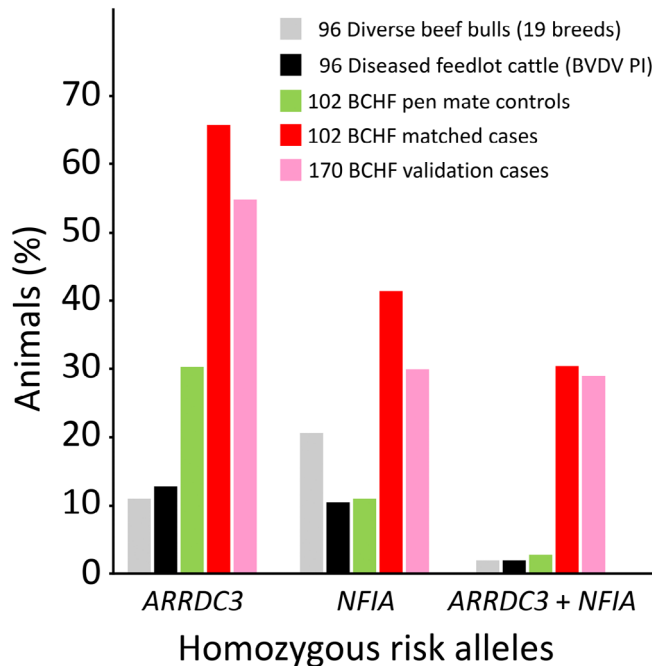
Species types	TMRCA (Ma)	Overall identity (%)	Taxonomic group	ARRDC3 position						
				179	180	181	182	183	184	185
Cattle (Hereford)	0	100.0		K	T	L	Y	C	W	F
Cattle (Brahman)	0	99.8		.	.	.	C	.	.	.
Yak, bison	5	99.8		.	.	.	C	.	.	.
Water buffalo	15	99.8		.	.	.	C	.	.	.
Sheep, goat, chiru	26	99.3		.	.	.	C	.	.	.
Deer, elk	27	99.3		.	.	.	C	.	.	.
Whale, dolphin	56	99.3		.	.	.	C	.	.	.
Swine	62	99.0		.	.	.	C	.	.	.
Camel, alpaca	64	99.0		.	.	.	C	.	.	.
Horse, bear, tiger, fox	78	99.0		.	.	.	C	.	.	.
Bat	79	97.3		.	.	.	C	.	.	.
Shrew	89	98.8		.	.	.	C	.	.	.
Primates	96	98.8		.	.	.	C	.	.	.
Rodents	96	96.9		.	.	.	C	.	.	.
Aardvark, elephant, armadillo	105	98.6		.	.	.	C	.	.	.
Opossum, koala, wombat	164	97.3		.	.	.	C	.	.	.
Eagle, kiwi, quail	310	95.4		.	.	.	C	.	.	.
Alligator, python, turtle	310	97.1		.	.	.	C	.	.	.
Frog	350	90.5		.	.	.	C	.	.	L
Coelacanth	400	91.1		.	.	.	C	.	.	L
Salmon, gar, piranha, tetra	450	88.5		.	.	.	C	.	.	.
Shark	483	84.8		.	.	.	C	.	.	.

**Figure 7. Comparison of *ARRDC3* residues near the C182Y position in the Gnathostomes.** TMRCA is the estimated time to the most recent common ancestor in millions of years (Hedges *et al.*, 2015). The full length *ARRDC3* protein is 414 residues in cattle and most of the Amniota species. The letters are IUPAC/IUBMB codes for amino acids. The dots are amino acid residues identical to those in cattle. A detailed alignment of this region is available in Table S2.

equivalent position were not observed in any genus. The conservation of the C182 residue throughout the Gnathostomes is consistent with the hypothesis that the C182 residue is critical for normal ARRDC3 protein function and that homozygosity of the *B. taurus*-specific Y182 variant may have a deleterious effect.

### Predicting BCHF risk with ARRDC3 and NFIA SNP genotypes

A custom two-SNP test was designed on a commercially available platform to genotype cattle for the major ARRDC3 and NFIA risk alleles. The platform uses “padlock” oligonucleotide probes, combined with fluorescently labeled probes and isothermal rolling circle amplification technology, for the qualitative detection of alleles. Using reference DNA from animals with known genotypes, the call rate and accuracy typically exceeded 97% and 99%, respectively. This included repeated testing of approximately 10% of the samples due to ambiguous allele calls on the first pass. Using custom genotypes to resolve errors in the original BovineHD BeadChip array, the BCHF OR for calves with homozygous risk alleles at both genes increased to 28-fold more ( $b/c = 28/1$ ,  $p\text{-value} = 1.1 \times 10^{-7}$ ,  $CI_{95} = 4\text{-}206$ ). Testing an independent set of 170 BCHF cases from the same feedlots showed approximately 29% were homozygous for both ARRDC3 and NFIA risk alleles, compared to 29% of the BCHF 102 cases used in the GWAS (Figure 8). Conversely, an independent set of feedlot cattle persistently infected with BVDV had about 2% homozygosity for both ARRDC3 and NFIA risk alleles, which was similar to the 3% observed with the 102 unaffected animals used in the GWAS. Assuming a 5% BCHF prevalence in the most severely affected pens of feedlot cattle, the adjusted negative and positive predictive values were 0.99 and 0.10, respectively, with a sensitivity and specificity of 0.91 and 0.58. Thus, testing with two SNPs has significant power to identify which animals have low risk for developing BCHF in feedlots, but has little power to predict which animals will develop disease. ARRDC3 and NFIA genotype configurations and their corresponding expected disease and allele transmission risks, together with their allele frequency estimates by breed, are presented in Tables 3 and 4, respectively. Table 5 provides unambiguous genomic identity and sequence context for ARRDC3 and NFIA SNPs associated with BCHF for future genetic testing designs. Although imperfect, these genetic associations and DNA tests provide a starting point for reducing BCHF risk in the most affected herds.



**Figure 8. Distribution of homozygous ARRDC3 and NFIA risk alleles in cattle cohorts.** Two-SNPs were used for identifying feedlot cattle with homozygous risk alleles (BovineHD0700027239 [ARRDC3] and BovineHD0300024308 [NFIA]). Bovine cohorts: USMARC beef cattle diversity panel v2.9 (19 breeds (Heaton et al., 2016), Kansas feedlot calves persistently infected with bovine viral diarrhoea virus (BVDV PI (Workman et al., 2016), pen matched BCHF controls (this study), BCHF cases (this study), independent BCHF validation cases (this study). The BCHF validation cases were collected by pen riders experienced in BCHF diagnosis, however, they were not confirmed at necropsy by veterinarians and researchers. All 204 of the BCHF case-control cattle tested negative for BVDV, and thus were not persistently infected with BVDV.

**Table 3. *ARRDC3* and *NFIA* genotype configurations, expected disease risk, and relative breeding rank.**

Possible risk allele configurations ( <i>ARRDC3</i> , <i>NFIA</i> ) <sup>a</sup>	Total combined risk alleles	Disease risk (1 to 28 scale) <sup>b</sup>	Probability of risk allele transmission (%)	Breeding rank <sup>c</sup>
0,0	0	1x	0	1
1,0	1	1x	25	2
0,1	1	1x	25	2
1,1	2	1x	50	3
2,0	2	8x	50	3
0,2	2	8x	50	3
2,1	3	8x	75	4
1,2	3	8x	75	4
2,2	4	28x	100	5

<sup>a</sup>Diploid combinations of risk alleles for each gene with markers BovineHD0700027239 (*ARRDC3*) and BovineHD0300024308 (*NFIA*).

<sup>b</sup>The risk of developing BCHF in a feedlot with animals and environments like those described here.

<sup>c</sup>The relative BCHF breeding rank of an animal based on having risk genotypes for *ARRDC3* and *NFIA* from best (1) to worst (5).

**Table 4. *ARRDC3* and *NFIA* risk allele frequencies in 46 breed groups of US cattle.**

Breed group <sup>a</sup>	Animals typed	Risk allele frequency <sup>b</sup>		Animals in each group							
		<i>ARRDC3</i> (BCHF5)	<i>NFIA</i> (BCHF2)	Disease risk <sup>c</sup>			Breeding rank <sup>d</sup>				
				1x	8x	28x	1	2	3	4	5
Angus	30	0.72	0.65	9	16	5	0	3	7	15	5
Ankole-Watusi	20	0.00	0.15	19	1	0	15	4	1	0	0
Ayrshire	24	0.02	0.50	17	7	0	7	9	8	0	0
Beefmaster	29	0.33	0.67	13	16	0	1	6	14	8	0
Belgian Blue	24	0.19	0.85	7	17	0	0	7	8	9	0
Blonde d'Aquitaine	24	0.06	0.42	19	5	0	9	8	6	1	0
Brahman	29	0.00	0.05	29	0	0	26	3	0	0	0
Brahmousin	24	0.00	0.29	22	2	0	12	10	2	0	0
Brangus	29	0.45	0.55	17	10	2	2	8	9	8	2
Braunvieh	28	0.00	0.50	21	7	0	7	14	7	0	0
Brown Swiss	24	0.00	0.50	18	6	0	6	12	6	0	0
Charolais	30	0.05	0.48	24	6	0	7	15	7	1	0
Chianina	28	0.18	0.54	17	10	1	5	10	10	2	1
Corriente	27	0.00	0.59	18	9	0	4	14	9	0	0
Devon	23	0.39	0.26	16	7	0	5	9	6	3	0
Dexter	24	0.00	0.31	23	1	0	10	13	1	0	0
Gelbvieh	29	0.03	0.48	20	9	0	9	11	8	1	0
Guernsey	24	0.00	0.56	19	5	0	2	17	5	0	0
Hereford	30	0.58	0.50	13	17	0	2	2	15	11	0
Highland	24	0.00	0.27	23	1	0	12	11	1	0	0
Holstein	23	0.09	0.63	16	7	0	1	12	9	1	0
Indu-Brazil	24	0.00	0.00	24	0	0	24	0	0	0	0
Jersey	38	0.00	0.58	24	14	0	8	16	14	0	0

**Table 4.** *Continued*

Breed group <sup>a</sup>	Animals typed	Risk allele frequency <sup>b</sup>		Animals in each group							
		ARRDC3 (BCHF5)	NFIA (BCHF2)	Disease risk <sup>c</sup>			Breeding rank <sup>d</sup>				
				1×	8×	28×	1	2	3	4	5
Limousin	30	0.12	0.40	24	5	1	7	17	5	0	1
Maine-Anjou	29	0.24	0.76	13	15	1	0	7	16	5	1
Marchigiana	23	0.00	0.61	11	12	0	7	4	12	0	0
Mini Hereford	24	0.58	0.73	7	12	5	0	5	4	10	5
Mini Zebu	24	0.00	0.04	24	0	0	22	2	0	0	0
Montbeliard	24	0.08	0.42	21	3	0	7	11	5	1	0
Murray Gray	24	0.71	0.42	11	11	2	1	5	7	9	2
Nelore	24	0.00	0.31	23	1	0	10	13	1	0	0
Piedmontese	24	0.00	0.48	20	4	0	5	15	4	0	0
Pinzgauer	24	0.00	0.44	19	5	0	8	11	5	0	0
Red Angus	30	0.73	0.42	12	16	2	0	5	13	10	2
Red Poll	24	0.48	0.48	16	5	3	5	2	10	4	3
Romagnola	24	0.00	0.48	21	3	0	4	17	3	0	0
Salers	29	0.09	0.36	23	6	0	12	8	9	0	0
Santa Gertrudis	28	0.18	0.43	24	3	1	4	18	3	2	1
Senepol	24	0.02	0.65	15	9	0	2	12	10	0	0
Shorthorn	29	0.24	0.57	17	12	0	2	12	10	5	0
Simmental	30	0.13	0.42	23	6	1	9	13	5	2	1
Tarentaise	28	0.09	0.20	27	1	0	14	12	2	0	0
T. Longhorn, MARC	28	0.02	0.41	23	5	0	10	13	4	1	0
T. Longhorn, CTRL	23	0.04	0.61	14	9	0	3	10	10	0	0
Tuli	23	0.00	0.17	23	0	0	15	8	0	0	0
Wagyu	24	0.06	0.73	10	14	0	2	8	12	2	0

<sup>a</sup>Registered cattle chosen for minimal pedigree relationships (Heaton *et al.*, 2016). Breed groups were genotyped from whole genome sequence (WGS) mapped to the bovine reference genome assembly (ARS-UCD1.2).

<sup>b</sup>The risk allele definitions are listed in Table 2. The sites genotyped from WGS data were chr7:90845941 (BCHF5) and chr3: 84707876 (BCHF2).

<sup>c</sup>As described in Table 3.

<sup>d</sup>As described in Table 3.

## Discussion

The present report describes a GWAS with clinical cases of BCHF and their pen matched controls from feedlots in the Western Great Plains. This is a region of the U.S. where outbreaks can be severe and occur annually. The 102 clinical cases enrolled in this study were a subset screened by pen riders from more than 140,000 feedlot cattle over the span of 16 months and were derived from 30 ranch sources. The most significant genome-wide associations were limited to two regions containing *ARRDC3* and *NFIA* genes, respectively. In each region, multiple linked SNPs were associated with BCHF in feedlot cattle. It is unknown whether the associated SNPs are merely linked to variants responsible for the elevated BCHF risk, or whether they are involved directly in pathogenesis. Whole genomic sequence (WGS) analyses are underway for the 102 case-control pairs of animals and may unlock further insights in the future. Despite the lack of resequencing data in the present study, animals having two copies of risk alleles in either gene, or both genes, were eight- and 28-fold more likely to develop BCHF, respectively. Thus, specific DNA sequence variants within *ARRDC3* and *NFIA* are markers for BCHF risk in feedlot cattle in the Western Great Plains, either through linkage to causal variants or direct effects. Neither gene was previously reported to be a direct cause of heart failure in animals. However, both genes play roles in the complex biology of cardiovascular health.

*ARRDC3* is a member of the alpha arrestins family involved in regulating mammalian metabolism. The normal biological functions of *ARRDC3* in cattle have not been studied, however variants have been associated with bovine growth and

**Table 5. Genome sequence context for ARRD3 and NF1A SNPs associated with BCHF.**

Genome position			SNP identifier		Alias	A <sub>1</sub> / A <sub>2</sub>	200 bp flanking genome sequence <sup>c</sup>	Flanking features in 19 beef breeds <sup>d</sup>			
Chr <sup>a</sup>	UMD3.1 (bp)	Gene	Bead Array	rs number				Type	ARS-UCD1.2 (bp)	Diff. (bp)	MAF
7	93229001	ARRD3	BovineHD0700027233	rs110586518	BCHF60	A/G	5'-TACTCCAGAGTATCAGGACCCATGAGGATGATCAAGAAATCAAAAT GGGGAATGGGATTTTTAAATGGGGTGAAGAAGTAAGTAATGGGGTGA [A/G] GATGAAAACCTAGACGAGGAGGTAGCAACGGGAGTGAAGAACAGGGACA GGGACCTGGAGATCTTACGAGAAACCCGAGGAAATTCATACCTACT-3'	SNP	90830000	-12	0.281
7	93231670	ARRD3	BovineHD0700027234	rs135607340	BCHF61	A/G	5'-CTTAGAGAACGAAACGAGTTAAATAGATTTTCCCTCAAATGGAGCAAT TTCCTTCTTTTAATGATCAGGAAAGCTAACTCTCAAAAGGACATAGATG [A/G] CGTAGAATGCCTTAAGGTAACAAAACCTCTATTAATGTTGTTACTAATAGC ACCGAGCGAAGGGTATTTTAAAGGAAATTCACCTGTTGTAATAGGGG-3'	SNP	90832681	-19	0.083
7	93243245	ARRD3	BovineHD0700027237	rs109439459	BCHF33	A/G	5'-TAAGAGACATCCAAATGCTGGGTACAGACTTTTGTATTAAGTATTGACAA TCFAACTTTTAAATCTGTTGTTCAAAAATTAATTAATTTCTCTCTTACCAATGAT [A/G] ATCAAAACCTAATCTGATTTTCAAGAACTCTGCAATATATGCTTAGTAAAG TCAGCCATTTTGTACTTAGGGTGTGATGCTTACTCTGAAATAAGGG-3'	SNP	90844254	0	0.281
7	93243714	ARRD3	BovineHD0700027238	rs136675904	BCHF34	C/T	5'-GGAAATACATTTTAAATCTCAATFAGGGGACACCCCATTCCTCAA TACACAAAATTTTAACTTAATCAATAAATAATCAGAGCAAAAATTAATTTCT [C/T] GCAATACCAGATTAAGCTTAGATAGCCATTTACAAATTAATTAACCTTACGG TCCTTAGGTAGGATAAACCAAGCCAGTTTCATGTTTCATGCTACACATGA-3'	SNP	90844723	0	0.281
7	93244933	ARRD3	BovineHD0700027239 <sup>e</sup>	rs109901274	BCHF5	T/C	5'-TCAGTCTTGTTGGAGTCCACCTCACTAGCTGCTGGGCTATAGCCCTC CTTTCGATCTTGGCCTTAGAGATATGGGCTTAGGTTAAGGACAA [T/C] AGAGAGCTTTTCTTTTTCCTGCTGGGCTGACTGAGATAFAAATTT AGAGRAAAAATCAAGGTATGAAATGAGAGGAAATAAATTTTAAAAA-3'	SNP	90845941	0	0.224
7	93247780	ARRD3	BovineHD0700027240	rs135153741	BCHF35	G/T	5'-ATGACAGACTTGAATGTTCTCTTTTTTTCAGCTAATGCTGATGTTGAAC TATTAAGTGTACTATTTGACTAGCACCTTCAATTAATCAGCTTTAAAT [G/T] AAATACCTTTTAAACCACTACAAAATCCCTGGGAGGAAATTTGAGGTTTCT TTCATAAATTTGGCCCTCCAGATTAATAAATAGCTTTTCTTTTGTAG-3'	SNP	90845997	+56	0.177
7	93251138	ARRD3	BovineHD0700027241	rs110066813	BCHF36	T/C	5'-GTTTGTCCCTTAAAGCTGATTAATTTCTTCTTCAACCCATTTAAACCACTGAC ACCACTCTTTAAAAACCTGTTTCTTCAACCCATTTAAAAAAGCTAC [T/C] AGTCAAAGGAGATCAAAATCAAAAATCCCTGGGCTCTCTTCAAGCCAAATACC AGGACTGTCAATAGGCAAGTCCATTTCTTAATGACTGTCACTGCTTGA-3'	SNP	90848788	0	0.412
7	93263102	ARRD3	BovineHD0700027244	rs109860522	BCHF62	C/T	5'-CAGTGTTTAAATCTGCTGATGAGGATCACTAATAATGTTGTTGTTGAAT ATCTTTTTTTTATTTTCAAGTAGGGCTTCTCTTCCAAAAATAATGTTTGTAT [C/T] AATPAAAAACCTTTTACGTGATTTGAAGACAAATTTTCCACAAAATTTTATAC ATTTTGTCCCCAGATGATTTTAAATATATATATATAGTGTACATATATTGCTA-3'	SNP	90863630	0	0.260
7	93281534	ARRD3	BovineHD0700027248	rs110253449	BCHF63	T/C	5'-TTCCTTTTCCCTGGCTGCAGATCTTGAAGTCACTGCCAAACTAA CTCTCGAGTGGCAACTTCAGAACAAAATAAAGCACTAGACATAGAAATGCT [T/C] GAGCCAAACTTACGTAATAGTCAATTAATGTTTGTAGTCTAGCACCTACC AGCCTTTAGAGAGCTTGTACAGTATACCTTCTCTCTGTGAAGGAAA-3'	SNP	90882079	-86	0.026
								SNP	90882079	0	0.276

**Table 5. Continued**

Genome position		SNP identifier		Alias	A <sub>1</sub> / A <sub>2</sub>	200 bp flanking genome sequence <sup>c</sup>	Flanking features in 19 beef breeds <sup>d</sup>					
Chr <sup>a</sup>	UMD3.1 (bp)	ARS-UCD1.2 (bp)	Gene				Bead Array	rs number	Type	ARS-UCD1.2 (bp)	Diff. (bp)	MAF
3	85123495	84578325	<i>NFIA</i>	BovineHD0300024307	rs136210820	BCHF31	G/A	5'-TTTCCCTGGAGAAATGAACTGCTGCATTCGCTGGAACCTCFATCTTTG CCTCACTTCTCACCCCTGACCGTCAAAACAAGAAATCTGTCTTGAGTTCCAC <b>[G/A]</b> ATGTACCCCATGTCTCATGACTAATCCAAAGCAATTTTATTTACAGTGT CAGAATRAATTTCTATGACAAATCTCATGCCATTTCTCAGCTGCCCTTTGA-3'	SNP SNP	84578236 84578325 84578383	-89 0 +58	0.042 0.448 0.052
3	85125825	84580655	<i>NFIA</i>	BovineHD0300024308	rs133192205	BCHF32	T/C	5'-TTACTAACCATGACCAAGGACACATTTCTCAGCTCTTTGTTCTCA GCAGCATTAAGGCAATTAATAAGCTCCCTCTAATAAAGGTAGATTACA <b>[T/C]</b> GAAGAACTAATAAATAATGTATAAGAAACCAACTTAATAATCATAGAAATG AACAACTAGCTAGCAGATCTGGCTTTCTCTGGCTGCCACTTA-3'	SNP	84580655	0	0.422
3	85254825	84707876	<i>NFIA</i>	BovineHD0300024366 <sup>f</sup>	rs42583839	BCHF2	T/C	5'-tatctgca tgc aatgc ag g ag ac c t g g t t c c a a t c c c t g g g t c a g g a a g a t c c c c t g g g a g g g e a t g g c a a t a c t t c t g c t t g c t t a g a a t a a a a t <b>[T/C]</b> CAGTTCCTCATGCAACCTGGTGGCCCTGAGGCMCCGAACTCCTCGCTGC AGGACCTGACCCCTGTCCTGCCCTGCCCTGCTGACATCTGTGGG-3'	SNP SNP SNP	84707876 84707910 84707917	0 +34 +41	0.464 0.021 0.073

<sup>a</sup>Chromosome.

<sup>b</sup>A1 was defined as the most frequent allele in the group of 204 BCHF cases and controls and the orientation is consistent with the ARS-UCD1.2 top strand (antisense strand in *ARRDC3* and *NFIA* genes). The bolded red allele is associated with risk.

<sup>c</sup>Sequence orientation is consistent with the ARS-UCD1.2 top strand. Each sequence aligned uniquely to the reference genome except BovineHD0300024366 (rs42583839) which contained 75 bp of repetitive sequence denoted in lowercase (blue).

<sup>d</sup>Derived from WGS data with USMARC beef cattle diversity panel version 2.9 (MBCDPV2.9, Heaton et al, 2016), and PLINK software (v1.9, Chang et al, 2015) with filters for MAF greater than or equal to 0.015, and missing genotypes less than or equal to 0.05.

<sup>e</sup>SNP BovineHD0700027239 corresponds to *ARRDC3* C182Y (codon tRt) in exon 4 where tGt encodes C182 and tAt encodes Y182.

<sup>f</sup>SNP BovineHD0300024366 was included for comparison, since it was used in early analyses for obtaining frequency data of U.S. cattle breeds in Table 4.

conformation traits (Saatchi *et al.*, 2014; Bolormaa *et al.*, 2014; Seabury *et al.*, 2017; Abo-Ismael *et al.*, 2017; Jiang *et al.*, 2019; Smith *et al.*, 2019). Studies on humans and mice have shown that *ARRDC3* has numerous functions. It regulates G protein-coupled receptor trafficking and signaling in cancer, suppresses tumors, is differentially regulated by ubiquitin, and functions as a multifaceted adaptor protein to control protein trafficking and cellular signaling (Wedegaertner *et al.*, 2023). The *ARRDC3* protein interacts with at least five membrane receptors or transporters, including:  $\beta$ 2- and  $\beta$ 3-adrenergic receptors ( $\beta$ 2AR and  $\beta$ 3AR), protease activated receptor-1 (PAR1), insulin receptor (IR), and integrin beta 4 (Itg $\beta$ 4) (Wedegaertner *et al.*, 2023). *ARRDC3* has also been linked to regulation of adrenergic signaling through interaction and regulation of ubiquitination and trafficking of the  $\beta$ 2AR (Han *et al.*, 2013; Tian *et al.*, 2016). The  $\beta$ 2AR is present on the cell surface and its stimulation causes smooth muscle relaxation, which may result in peripheral vasodilation bronchodilation with subsequent hypotension and reflex tachycardia.

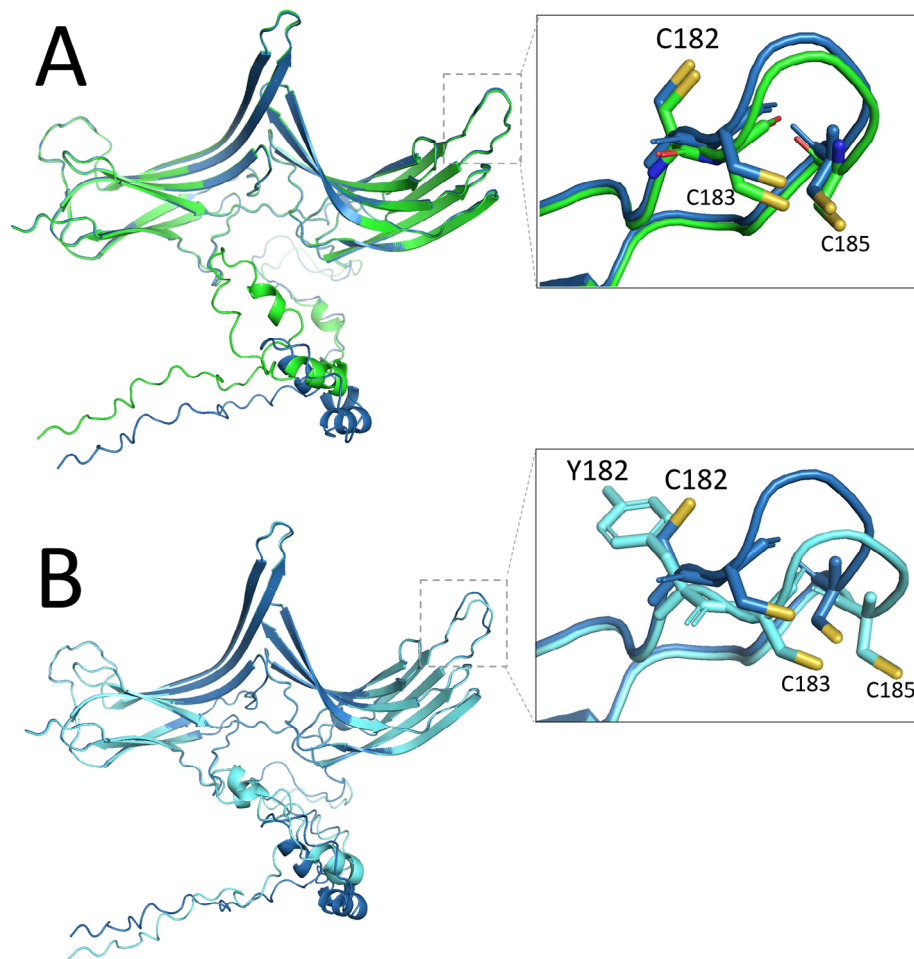
*ARRDC3* can also signal activation of uncoupling protein-1 and the production of heat when thermogenesis is needed during cold stress. Over-activation of human  $\beta$ 2AR promotes cardiac hypertrophy and represents an independent risk factor for cardiovascular morbidity and mortality (Osadchii, 2007). *ARRDC3* also modulates insulin action and glucose metabolism in the liver and binds directly to the insulin receptor in a mouse model for human obesity (Batista *et al.*, 2020). In humans, increased *ARRDC3* expression was associated with an increase in body mass index, while *ARRDC3* deficient mice were resistant to age-related obesity and have increased insulin sensitivity (Patwari *et al.*, 2011). However, 75% of the *ARRDC3* deficient mice did not survive to weaning. Although it is tempting to link human *ARRDC3* obesity mechanisms with the process of fattening cattle, more than half the BCHF cases in our study died in the first, less intensive phase of the 240-day feeding program. This is inconsistent with the hypothesis that fattening is the underlying cause of disease in these cattle (Heaton *et al.*, 2019). Regardless, biological mechanisms by which genetic variation in the *ARRDC3* genomic region affects BCHF are presently unknown.

We hypothesize that the Y182 missense variant negatively alters the normal *ARRDC3* protein functionality but does not eliminate it. The human and bovine *ARRDC3* protein sequences are 99% identical and their aligned tertiary structures from *de novo* structure predictions shows they are nearly superimposable (Figure 9A). The *ARRDC3* C182 residue is part of a three-cysteine loop at the end of two antiparallel beta sheets in the Ig-like beta-sandwich folds at the beginning of the Arrestin-C domain (Figure 9A inset). We speculate that the bovine Y182 variant alters the normal *ARRDC3* function in a way that is deleterious to the individual. This hypothesis is based on the observation that all jawed vertebrates tested have cysteine exclusively at this position. Since the Gnathostomes represent 99% of all living vertebrate species (Brazeau and Friedman, 2015), it is remarkable that only cattle are reported to encode a residue other than cysteine at this position. We also speculate that there may be residual function in bovine *ARRDC3* Y182 based on two observations: 1) *ARRDC3* knockout mice are lowly viable and have a range of perinatal developmental abnormalities including heart and lung defects (La Marca *et al.*, 2023), and 2) homozygous healthy cattle with the non-conserved Y182 genotypes are abundant (this report). However, further research is needed to reject the hypotheses presented here.

*NFIA* is a transcription factor that controls various cellular differentiation programs, including myogenesis and adipogenesis (Hiraike *et al.*, 2020). The only biological function of *NFIA* in cattle was vitamin A effects on *NFIA* expression levels in brown fat depots (Chen *et al.*, 2018), and no prior bovine GWAS has linked *NFIA* to any traits. In humans, *NFIA* was associated with heart failure, reduced ejection fraction, and conduction of the electrical signal throughout the ventricular myocardium (Joseph *et al.*, 2022; Evans *et al.*, 2016). *NFIA* also controls the onset of gliogenesis in the developing spinal cord in mice (Deneen *et al.*, 2006). *NFIA* is best known for activating adipogenesis, as well as actively suppressing myogenesis by binding to brown-fat-specific enhancers before differentiation, and later facilitates the binding of the master transcription factor of adipogenesis: peroxisome proliferator-activated receptor  $\gamma$  (*PPAR* $\gamma$ ) (Hiraike *et al.*, 2017). A reduction in *NFIA* activity shifts the balance towards the myogenesis and white adipocyte gene program, while an increase in *NFIA* activity shifts the balance towards the brown/beige adipocytes program (Hiraike *et al.*, 2020). Thus, it is tempting to speculate that *NFIA* variants in beef cattle have been selected for reduced activity leading to increased muscle and white fat; however, there is no evidence for this.

Still, an intriguing but unproven connection may exist between *ARRDC3* regulation of  $\beta$ -adrenergic signaling and *NFIA* functions relating to cold stress and brown fat thermogenesis. Anecdotally, the incidence of feedlot BCHF increases dramatically each year during the first wave of bitter cold weather, when many animals succumb to BCHF in a short period of time (Personal communication between authors (MPH and BLVL) and Nebraska feedlot owners). In any event, the biological mechanisms explaining how either of these genes may influence BCHF pathogenesis remain to be determined.

A strength of this study was the diagnosis and necropsy of clinical cases at the time of euthanasia. This was accomplished immediately, in the environment and geographic region where the disease is most apparent, and included feedlot pen riders with experience in identifying animals displaying disease signs. The case definition at necropsy was rigorous and



**Figure 9. Predicted protein folding differences between bovine ARRDC3 C182Y variants.** Alignment of ribbon representations of AlphaFold2 predictions of bovine ARRDC3. Panel A, human and bovine ARRDC3 with inset showing the side chains of the C182, C183 and C185 residues. Panel B, bovine ARRDC3 with C182 or Y182 residues with inset showing the sidechains of the variant C182 and Y182 residues, together with invariant C183 and C185 residues.

included multiple specific and non-specific clinical signs to eliminate other potential confounding afflictions. This is in contrast to case definitions based on qualitative cardiomegaly scores obtained during beef processing of animals passing federal antemortem inspection. Cases described in the present study were deemed by pen riders unlikely to survive transport to a processing plant, and even less likely to pass federal inspection, hence their availability for euthanasia and necropsy.

Another strength was that feedlot personnel were incentivized to monitor more than 140,000 head from representative feedlots to find those animals most likely to have terminal BCHF and euthanize them before they died of congestive heart failure. The financial loss of each case enrolled in our study was offset by a \$500 indemnity payment when the animals were identified early, euthanized, and necropsied by the authors. Additional strengths of this study included the use of: diverse ranch sources, matched controls, and custom McNemar allele analysis. Approximately 20% of the collected cases were dismissed due to inadequate pair-matching criteria. Weaknesses of the study included the relatively low number of total pairs (102) that met the qualifications for the study. However, the cost of indemnifying more clinical cases, combined with the time and resources needed for researchers to conduct careful necropsies and sample collections significantly affected the feasibility of going much beyond 100 matched pairs. Ideally, the study would have benefited from 200 matched pairs for discovery, and another 200 matched pairs for validation. Our validation set of 170 clinical cases consisted of unmatched cases that were either animals suspected of having BCHF and shipped early to a salvage beef processor, or those that died in the feedlot.



Genetic testing of young cattle from affected herds can identify animals predisposed to developing BCHF in Western Plains feedlot environments. Using a two-SNP test with markers for *ARRDC3* and *NFIA* risk alleles, BCHF cases are expected to occur in animals with homozygous *ARRDC3* and *NFIA* risk alleles at a rate 28-fold higher than those without homozygous risk alleles in either gene. Although the confidence interval around this OR estimate is wide (four- to 206-fold), the test provides a starting point for applying costly BCHF interventions to a smaller group of animals rather than the whole pen. For example, a subset of high-risk animals could be moved to a lower elevation for fattening, or they could be fattened with less energy-dense feed for a longer interval based on genetic testing. The two-SNP test used here has a remarkably high negative predictive value (0.99) that is powerful for identifying sires without these major risk alleles. Using low-risk sires and either conventional breeding or artificial insemination would have an immediate impact on reducing BCHF in the next generation of calves.

The high frequencies of *ARRDC3* and *NFIA* risk alleles in some affected breeds (> 0.65) raise the question of why the disease prevalence is relatively low even in outbreaks (> 0.07)? This phenomenon may be due to the multifactorial nature of complex disease. The development of BCHF in cattle is presumed to be influenced by exposure to non-genetic factors, such as elevation (hypoxia), extreme weather conditions, feeding schedules, ionophores, feed grade antibiotics, steroid hormone implants, and beta-adrenergic agonist feed additives (Krafsur *et al.*, 2019; Neary *et al.* 2018). These environmental and management factors may play significant but unknown roles in disease development. In addition, most complex diseases are considered polygenic with variation at many genes having small effects (Khera *et al.* 2018). DNA sequence variation identified here in *ARRDC3* and *NFIA* genomic regions may predispose cattle to BCHF under certain conditions, despite the absence of any known causal pathogenic mutations.

The high frequencies of risk alleles observed in some breeds also poses a challenge for selective breeding. Although estimates for the positive predictive value of the *ARRDC3* and *NFIA* markers were low (0.10), the negative predictive value was quite high (0.99). Consequently, identifying breeding stock lacking these risk factors could prove useful for reducing a herd's overall risk. The challenge is finding sires without the *ARRDC3* and *NFIA* risk alleles in breeds with a high prevalence of high-risk alleles. For instance, the largest affected breed in the United States, American Angus, has more than 300,000 bulls registered annually in its association. However, less than 5% (15,000) are expected to possess the optimal low-risk genotypes at both the *ARRDC3* and *NFIA* gene regions (M. Heaton, L. Kuehn, and B. Vander Ley 2022, unpublished). To identify the best bulls for decreasing BCHF risk alleles in their herds, cattle producers may need to obtain genotypes from hundreds of candidate sires. Unfortunately, this information is not widely available for most bulls at present. Despite these challenges, there are a number of reasonable selection strategies that could help reduce the risk of BCHF in the long term, depending on the specific goals and circumstances of the cattle producer.

Identification of the most significant genome-wide markers for BCHF was based on the FDR results presented in Figure 4. Only models that compared homozygous risk alleles against all other genotypes yielded significant results. The choice of best markers for genetic testing was based on results presented in Table 2 and Figure S2. On chromosome 7 in the 52 kb associated genomic region centered on the 14kb *ARRDC3* gene sequence, there were eight SNPs linked to the C182Y missense variant ( $r^2 = 0.87$  to  $0.91$ ). The difference in the OR based on informative McNemar's pairs between the eight linked SNPs and the C182Y variant was not significant ( $b/c = 42/5$  and  $42/6$ , respectively). Thus, any of the eight linked SNPs could be used as a proxy for the C182Y SNP in exon 4 until a variant(s) directly responsible for the elevated BCHF risk is confirmed in this region. In the 2.32 kb associated genomic region of *NFIA* intron 4 there were two linked SNPs ( $r^2 = 0.94$ ) with high statistical association with BCHF and identical McNemar's statistics ( $b/c = 37/5$ ,  $CI_{95} = 2.9-19$ , BovineHD0300024307, BovineHD0300024308). Both *NFIA* intron 4 SNPs were in moderate linkage disequilibrium with a third SNP located 127 kb upstream in *NFIA* antisense (AS) 2 region ( $r^2 = 0.49$  and  $0.45$ , respectively). The *NFIA* AS2 SNP (BovineHD0300024366) had more informative pairs which corresponded to a smaller confidence interval ( $b/c = 44/11$  and  $CI_{95} = 2.1-7.7$ , respectively), but the OR was lower (4.0). Thus, the two *NFIA* intron 4 SNPs are nearly as informative, yet have a better OR (7.4). A minimum two-SNP assay encompassing most of the association would include the *ARRDC3* C182Y and a *NFIA* intron 4 marker (*e.g.*, BovineHD0700027239 [alias BCHF5] and BovineHD0300024308 [alias BCHF32], respectively).

Future prospects for improved BCHF genetic markers are promising. The 560,000 BovineHD BeadChip array SNPs in the present GWAS represent less than 5% of 12 million variants in the Angus genome (Low *et al.*, 2020). Thus, WGS analysis of the 102 case-control pairs would refine the bounds of the genomic region most strongly associated with BCHF and identify the complete set of tagging SNPs appropriate for use in genotyping by sequencing. Any novel variants with increased association would potentially represent candidates for functional variants conferring BCHF risk. A WGS analysis may also identify any newly associated genomic regions previously missed due to ascertainment bias in the SNP selection for the BovineHD BeadChip array. For these reasons, we are endeavoring to produce and analyze 15-fold WGS coverage from these 204 animals. Additional markers from new prospective associated regions could also potentially be

used in genetic tests to calculate polygenic risk scores for individual animals, and further improve disease risk prediction by stratifying populations into risk groups (Knowles and Ashley, 2018; Torkamani *et al.*, 2018; Wise *et al.*, 2019; Claussnitzer *et al.*, 2020). Thus, a two-SNP test for *ARRDC3* and *NFIA* risk alleles presented here represents a first genetic tool for identifying an animal's genetic risk for BCHF and may be helpful in determining whether the *ARRDC3* and *NFIA* markers are also associated with the high-elevation "brisket disease" disorder in the Rocky Mountain region of Colorado and Utah. Regardless of the outcome, additional DNA markers may be needed to increase risk prediction accuracy for BCHF in feedlot cattle.

## Conclusions

DNA sequence variation in *ARRDC3* and *NFIA* was discovered to be associated with BCHF in feedlot cattle at moderate elevations. The role of these variants is unknown and could be either through linkage to causal variants or direct effects on pathogenesis. Animals that were homozygous for risk alleles in both genes were 28-fold more likely to develop heart failure than those without. A custom DNA-based test showed 29% of clinical cases had homozygous risk alleles in both genes, compared to less than 2.5% in 198 similar unaffected feedlot cattle. This type of test may be useful for identifying feedlot animals at the highest risk for BCHF in the Western Great Plains of North America. In herds affected by BCHF, knowledge of which cattle have the highest and lowest genetic risk for disease allows producers to make informed decisions for selective breeding and animal health management.

## Data availability

### Underlying and extended data

Figshare: Additional files for the article entitled, Association of *ARRDC3* and *NFIA* variants with bovine congestive heart failure in feedlot cattle, <https://doi.org/10.6084/m9.figshare.c.4331114.v2> (Heaton, 2022).

This project contains the following data files:

- Table S1. Metadata for 102 BCHF cases and their matched controls
- Table S2. Extended comparison of *ARRDC3* residues near the C182Y position in the Gnathostomes. [https://figshare.com/articles/dataset/Table\\_S2\\_Extended\\_comparison\\_of\\_ARRDC3\\_residues\\_near\\_the\\_C182Y\\_position\\_in\\_the\\_Gnathostomes\\_/22213495](https://figshare.com/articles/dataset/Table_S2_Extended_comparison_of_ARRDC3_residues_near_the_C182Y_position_in_the_Gnathostomes_/22213495)
- Table S3. *ARRDC3* and *NFIA* genotypes for 204 case-control animals corresponding to 12 SNPs listed in Table 2. [https://figshare.com/articles/dataset/Table\\_S3\\_i\\_ARRDC3\\_i\\_and\\_i\\_NFIA\\_i\\_genotypes\\_for\\_204\\_case-control\\_animals\\_corresponding\\_to\\_12\\_SNPs\\_listed\\_in\\_Table\\_2\\_/24986211](https://figshare.com/articles/dataset/Table_S3_i_ARRDC3_i_and_i_NFIA_i_genotypes_for_204_case-control_animals_corresponding_to_12_SNPs_listed_in_Table_2_/24986211)
- Table S4. *ARRDC3* and *NFIA* genotypes for two SNPs from the 566 animals shown in Figure 8. [https://figshare.com/articles/dataset/Table\\_S4\\_i\\_ARRDC3\\_i\\_and\\_i\\_NFIA\\_i\\_genotypes\\_for\\_two\\_SNPs\\_from\\_the\\_566\\_animals\\_shown\\_in\\_Figure\\_8/24986424](https://figshare.com/articles/dataset/Table_S4_i_ARRDC3_i_and_i_NFIA_i_genotypes_for_two_SNPs_from_the_566_animals_shown_in_Figure_8/24986424)
- Table S5. *ARRDC3* and *NFIA* genotypes for two SNPs from the 1203 animals and 46 breeds listed in Table 4. [https://figshare.com/articles/dataset/Table\\_S5\\_i\\_ARRDC3\\_i\\_and\\_i\\_NFIA\\_i\\_genotypes\\_for\\_two\\_SNPs\\_from\\_the\\_1203\\_animals\\_and\\_46\\_breeds\\_listed\\_in\\_Table\\_4\\_/24986442](https://figshare.com/articles/dataset/Table_S5_i_ARRDC3_i_and_i_NFIA_i_genotypes_for_two_SNPs_from_the_1203_animals_and_46_breeds_listed_in_Table_4_/24986442)
- Figure S1. Calculation of expected theoretical  $q$ - and mid  $p$ -values
- Figure S2. Linkage disequilibrium between SNPs associated with BCHF in *ARRDC3* and *NFIA*
- Files S1 and S2. Unfiltered map and ped files (777,962 SNPs) for 102 BCHF matched case-control pairs, filename: BCHF102pairsHD778 (28.9 and 605.4 Mb, respectively.)
- Files S3 and S4. Filtered map and ped files (563,042 SNPs) for 102 BCHF matched case-control pairs, filename: BCHF102pairsHD770Filtered (20.9 and 438.2 Mb, respectively.)
- Files S5 and S6. Unfiltered map and ped files (777,962 SNPs) for 102 BCHF pairs and 96 MBCDPv2.9 bulls, filename: BCHF102\_MBCDP29\_HD778 (27.6 and 890.3 Mb, respectively.)
- Files S7 and S8. Filtered map and ped files (604,746 SNPs) for 102 BCHF pairs and 96 MBCDPv2.9 bulls, filename: BCHF102\_MBCDP29\_HD778Filtered (21.5 and 692.1 Mb, respectively.)

- Files S9 and S10. Filtered map and ped files (563,087 SNPs) for 102 BCHF matched case-control pairs, filename: BCHF102pairsHD778FFFFSortFiltered (20.9 and 438.2 Mb, respectively.)
- File S11. Unformatted McNemar output file (563,087 SNPs) for 102 BCHF matched case-control pairs, filename: BCHF102pairsHD778FFFFSortFiltered.McNemarsResults.csv (468.74)
- File S12. McNemar's test statistics for 561,683 filtered SNPs, filename: BCHF102PairsHD778FFFFSortFiltered.McNemarsResults17.xlsx (450.8 Mb)
- File S13. Genome-wide SNPs passing the 5% FDR. Filename: BCHF102PairsHD778FFFFSortFiltered.ExtractTwoHF\_q\_LT\_0\_05\_MPH3.xlsx (90k)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

### Analysis code

MatLab analysis code available at GitHub: <https://github.com/greg-harhay/McNemarsSNPAnalysis>

Executable MatLab code available at Code Ocean: <https://doi.org/10.24433/CO.0870729.v1>

Data are available under the terms of the [UnLicense](#).

### Acknowledgments

We thank D. George, J. Nierman, D. Ridgway, and S. Schmidt for secretarial and administrative support; S. Johnson for sample collection support; R. Carman, G. Darnall, P. Miller, T. Williams, and their outstanding pen riders for identifying clinical cases, assisting with sample acquisition, and sharing their knowledge of animals with disease; J. Pollak, and M. Boggess for leadership and support during the design and execution of the project. Dr. D. Sjeklocha for assistance with BVDPI calf samples; Dr. A. Oommen and MatMaCorp scientists (Lincoln, NE, USA) for outstanding technical assistance in SNP assay development. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. The USDA is an equal opportunity provider and employer.

### References

- Abo-Ismael MK, Brito LF, Miller SP, *et al.*: **Genome-wide association studies and genomic prediction of breeding values for calving performance and body conformation traits in Holstein cattle.** *Genet. Sel. Evol.* 2017; **49**: 82.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Batista TM, Dagdeviren S, Carroll SH, *et al.*: **Arrestin domain-containing 3 (Arddc3) modulates insulin action and glucose metabolism in liver.** *Proc. Natl. Acad. Sci. U. S. A.* 2020; **117**: 6733–6740.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bolormaa S, Pryce JE, Reverter A, *et al.*: **A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle.** *PLoS Genet.* 2014; **10**: e1004198.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Brazeau MD, Friedman M: **The origin and early phylogenetic history of jawed vertebrates.** *Nature.* 2015; **520**: 490–497.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Cardon LR, Palmer LJ: **Population stratification and spurious allelic association.** *Lancet.* 2003; **361**: 598–604.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chang CC, Chow CC, Tellier LC, *et al.*: **Second-generation PLINK: rising to the challenge of larger and richer datasets.** *Gigascience.* 2015; **4**: 7.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chen H-J, Ihara T, Yoshioka H, *et al.*: **Expression levels of brown/beige adipocyte-related genes in fat depots of vitamin A-restricted fattening cattle.** *J. Anim. Sci.* 2018; **96**: 3884–3896.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Claussnitzer M, Cho JH, Collins R, *et al.*: **A brief history of human disease genetics.** *Nature.* 2020; **577**: 179–189.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Davies VJ, Powell KA, White KE, *et al.*: **A missense mutation in the murine Opa3 gene models human Costeff syndrome.** *Brain.* 2008 Feb; **131**(Pt 2): 368–380.  
[PubMed Abstract](#)
- Deneen B, Ho R, Lukaszewicz A, *et al.*: **The transcription factor NFIA controls the onset of gliogenesis in the developing spinal cord.** *Neuron.* 2006; **52**: 953–968.  
[PubMed Abstract](#)
- Deng HW: **Population admixture may appear to mask, change or reverse genetic effects of genes underlying complex traits.** *Genetics.* 2001; **159**: 1319–1323.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Evans DS, Avery CL, Nalls MA, *et al.*: **Fine-mapping, novel loci identification, and SNP association transferability in a genome-wide association study of QRS duration in African Americans.** *Hum. Mol. Genet.* 2016 Oct 1; **25**(19): 4350–4368.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fagerland MW, Lydersen S, Laake P: **The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional.** *BMC Med. Res. Methodol.* 2013; **13**: 91.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Falik-Zaccai TC, Barsheshet Y, Mandel H, *et al.*: **Sequence variation in PPP1R13L results in a novel form of cardio-cutaneous syndrome.** *EMBO Mol. Med.* 2017 Sep; **9**(9): 1326.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Glover GH, Newsom IE: *Brisket Disease: Dropsy of High Altitudes.* Agricultural Experiment Station of the Agricultural College of Colorado; 1914.
- Glover GH, Newsom IE: **Further studies on brisket disease.** *J. Agric. Res.* 1918; **15**: 409–419.

- Graffelman J, Moreno V: **The mid p-value in exact tests for Hardy-Weinberg equilibrium.** *Stat. Appl. Genet. Mol. Biol.* 2013; **12**: 433–448.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Han S-O, Kommaddi RP, Shenoy SK: **Distinct roles for  $\beta$ -arrestin2 and arrestin-domain-containing proteins in  $\beta$ 2 adrenergic receptor trafficking.** *EMBO Rep.* 2013; **14**: 164–171.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Heaton M: **Additional files for the article entitled, Association of *ARRDC3* and *NFIA* variants with bovine congestive heart failure in feedlot cattle.** *figshare. Collection [dataset].* 2022.  
[Publisher Full Text](#)
- Heaton MP, Bassett AS, Whitman KJ, et al.: **Evaluation of *EPAS1* variants for association with bovine congestive heart failure.** *F1000Research.* 2019; **8**: 1189.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Heaton MP, Chitko-McKown CG, Grosse WM, et al.: **Interleukin-8 haplotype structure from nucleotide sequence variation in commercial populations of U.S. beef cattle.** *Mamm. Genome.* 2001; **12**: 219–226.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Heaton MP, Clawson ML, Chitko-McKown CG, et al.: **Reduced Lentivirus Susceptibility in Sheep with *TMEM154* Mutations.** *PLoS Genet.* 2012; **8**(1): e1002467.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Heaton MP, Keele JW, Harhay GP, et al.: **Prevalence of the prion protein gene *E211K* variant in U.S. cattle.** *BMC Vet. Res.* 2008; **4**: 25.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Heaton MP, Smith TPL, Carnahan JK, et al.: **Using diverse U.S. beef cattle genomes to identify missense mutations in *EPAS1*, a gene associated with pulmonary hypertension.** *F1000Res.* 2016; **5**: 2003.  
[PubMed Abstract](#)
- Hecht HH, Kuida H, Lange RL, et al.: **Brisket disease. III. Clinical features and hemodynamic observations in altitude-dependent right heart failure of cattle.** *Am. J. Med.* 1962; **32**: 171–183.  
[Publisher Full Text](#)
- Hedges SB, Marin J, Suleski M, et al.: **Tree of life reveals clock-like speciation and diversification.** *Mol. Biol. Evol.* 2015; **32**: 835–845.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Herron BJ, Rao C, Liu S, et al.: **A mutation in *NFKB* interacting protein 1 results in cardiomyopathy and abnormal skin development in *wa3* mice.** *Hum. Mol. Genet.* 2005 Mar 1; **14**(5): 667–677.  
[Publisher Full Text](#)
- Hiraike Y, Waki H, Miyake K, et al.: ***NFIA* differentially controls adipogenic and myogenic gene program through distinct pathways to ensure brown and beige adipocyte differentiation.** *PLoS Genet.* 2020; **16**: e1009044.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hiraike Y, Waki H, Yu J, et al.: ***NFIA* co-localizes with *PPAR $\gamma$*  and transcriptionally controls the brown fat gene program.** *Nat. Cell Biol.* 2017; **19**: 1081–1092.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Jensen R, Pierson RE, Braddy PM, et al.: **Brisket disease in yearling feedlot cattle.** *J. Am. Vet. Med. Assoc.* 1976; **169**: 515–517.  
[PubMed Abstract](#)
- Jiang J, Cole JB, Freebern E, et al.: **Functional annotation and Bayesian fine-mapping reveals candidate genes for important agronomic traits in Holstein bulls.** *Communications Biology.* 2019; **2**: 212.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Johnson RC, Nelson GW, Troyer JL, et al.: **Accounting for multiple comparisons in a genome-wide association study (GWAS).** *BMC Genomics.* 2010; **11**: 724.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Joseph J, Liu C, Hui Q, et al.: **Genetic architecture of heart failure with preserved versus reduced ejection fraction.** *Nat. Commun.* 2022; **13**: 7753.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Khera AV, Chaffin M, Aragam KG, et al.: **Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations.** *Nat. Genet.* 2018; **50**: 1219–1224.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Knowles JW, Ashley EA: **Cardiovascular disease: The rise of the genetic risk score.** *PLoS Med.* 2018; **15**: e1002546.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Krafsur GM, Neary JM, Garry F, et al.: **Cardiopulmonary remodeling in fattened beef cattle: a naturally occurring large animal model of obesity-associated pulmonary hypertension with left heart disease.** *Pulm Circ.* 2019; **9**: 2045894018796804.
- La Marca JE, Aubrey BJ, Yang B, et al.: **Genome-wide CRISPR screening identifies a role for *ARRDC3* in TRP53-mediated responses.** *Cell Death Differ.* 2024 Feb; **31**(2): 150–158.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Low WY, Tearle R, Liu R, et al.: **Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle.** *Nat. Commun.* 2020; **11**: 2071.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mahlum CE, Haugerud S, Shivers JL, et al.: **Detection of Bovine Viral Diarrhea Virus by TaqMan® Reverse Transcription Polymerase Chain Reaction.** *J. Vet. Diagn. Investig.* 2002; **14**: 120–125.  
[Publisher Full Text](#)
- Mantel N, Haenszel W: **Statistical aspects of the analysis of data from retrospective studies of disease.** *J. Natl. Cancer Inst.* 1959; **22**: 719–748.  
[PubMed Abstract](#)
- McNemar Q: **Note on the sampling error of the difference between correlated proportions or percentages.** *Psychometrika.* 1947; **12**: 153–157.  
[Publisher Full Text](#)
- Mirdita M, Schütze K, Moriwaki Y, et al.: **ColabFold: making protein folding accessible to all.** *Nat. Methods.* 2022; **19**: 679–682.  
[Publisher Full Text](#)
- Morrell NW, Aldred MA, Chung WK, et al.: **Genetics and genomics of pulmonary arterial hypertension.** *Eur. Respir. J.* 2019; **53**: 1801899.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Moxley RA, Smith DR, Grotelueschen DM, et al.: **Investigation of congestive heart failure in beef cattle in a feedyard at a moderate altitude in western Nebraska.** *J. Vet. Diagn. Investig.* 2019; **31**: 509–522.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Neary JM, Booker CW, Wildman BK, et al.: **Right-Sided Congestive Heart Failure in North American Feedlot Cattle.** *J. Vet. Intern. Med.* 2016; **30**: 326–334.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Neary JM, Garry FB, Gould DH, et al.: **The beta-adrenergic agonist zilpaterol hydrochloride may predispose feedlot cattle to cardiac remodeling and dysfunction [version 1; peer review: 2 approved with reservations].** *F1000Research.* 2018; **7**: 399.  
[Publisher Full Text](#)
- Newman JH, Holt TN, Cogan JD, et al.: **Increased prevalence of *EPAS1* variant in cattle with high-altitude pulmonary hypertension.** *Nat. Commun.* 2015; **6**: 6863.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Nomenclature Committee for the International Union of Biochemistry (NC-IUB): **Nomenclature for incompletely specified bases in nucleic acid sequences. Recommendations 1984.** *Mol. Biol. Evol.* 1986; **3**: 99–108.
- Osadchii OE: **Cardiac hypertrophy induced by sustained  $\beta$ -adrenoreceptor activation: pathophysiological aspects.** *Heart Fail. Rev.* 2007; **12**: 66–86.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Owczarek-Lipska M, Plattet P, Zipperle L, et al.: **A nonsense mutation in the optic atrophy 3 gene (*OPA3*) causes dilated cardiomyopathy in Red Holstein cattle.** *Genomics.* 2011; **97**: 51–57.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Patwari P, Emilsson V, Schadt EE, et al.: **The arrestin domain-containing 3 protein regulates body mass and energy expenditure.** *Cell Metab.* 2011; **14**: 671–683.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Reza N, Owens AT: **Advances in the Genetics and Genomics of Heart Failure.** *Curr. Cardiol. Rep.* 2020; **22**: 132.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Saatchi M, Schnabel RD, Taylor JF, et al.: **Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds.** *BMC Genomics.* 2014; **15**: 724.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Seabury CM, Oldeschulte DL, Saatchi M, et al.: **Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle.** *BMC Genomics.* 2017; **18**: 386.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Simpson MA, Cook RW, Solanki P, et al.: **A mutation in *NFKB* interacting protein 1 causes cardiomyopathy and woolly haircoat syndrome of Poll Hereford cattle.** *Anim. Genet.* 2009; **40**: 42–46.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Smith JL, Wilson ML, Nilson SM, et al.: **Genome-wide association and genotype by environment interactions for growth traits in U.S. Gelbvieh cattle.** *BMC Genomics.* 2019; **20**: 926.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Storey JD: **A direct approach to false discovery rates.** *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 2002; **64**: 479–498.  
[Publisher Full Text](#)
- Storey JD, Tibshirani R: **Statistical significance for genomewide studies.** *Proc. Natl. Acad. Sci.* 2003; **100**: 9440–9445.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tian X, Irannejad R, Bowman SL, et al.: **The  $\alpha$ -Arrestin *ARRDC3* Regulates the Endosomal Residence Time and Intracellular Signaling of the  $\beta$ 2**

**Adrenergic Receptor.** *J. Biol. Chem.* 2016; **291**: 14510–14525.

[PubMed Abstract](#) | [Publisher Full Text](#)

Torkamani A, Wineinger NE, Topol EJ: **The personal and clinical utility of polygenic risk scores.** *Nat. Rev. Genet.* 2018; **19**: 581–590.

[PubMed Abstract](#) | [Publisher Full Text](#)

Wedegaertner H, Bosompra O, Kufareva I, *et al.*: **Divergent regulation of  $\alpha$ -arrestin ARRDC3 function by ubiquitination.** *Mol. Biol. Cell.* 2023 Aug 1; **34**(9): ar93.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Wise AL, Manolio TA, Mensah GA, *et al.*: **Genomic medicine for undiagnosed diseases.** *Lancet.* 2019; **394**: 533–540.

[PubMed Abstract](#) | [Publisher Full Text](#)

Workman AM, Heaton MP, Harhay GP, *et al.*: **Resolving Bovine viral diarrhea virus subtypes from persistently infected U.S. beef calves with complete genome sequence.** *J. Vet. Diagn. Investig.* 2016; **28**: 519–528.

[PubMed Abstract](#) | [Publisher Full Text](#)

# Open Peer Review

Current Peer Review Status:   

Version 2

Reviewer Report 24 July 2024

<https://doi.org/10.5256/f1000research.162702.r252387>

© 2024 Drögemüller C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



## Cord Drögemüller

Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland

What is known about the possible inheritance of BCHF? Have segregation analyses been performed? What is the heritability?

The nomenclature of variants, especially the representation of ARRDC3 alleles, should necessarily follow the current recommendations, see Richards S, et.al.,2015 [Ref 1]

The intronic variant in the NFIA gene remains functionally absolutely unclear, therefore talk about a linked marker rather than a possible causal variant.

The wording "homozygous major alleles" on page 10 is unfortunate. Would it not be better to say homozygous for the major allele. An allele itself cannot be homozygous, but the genotype is either hetero- or homozygous for two or one allele.

In Figure 7 the accession numbers of the listed protein sequences are missing.

Table 4 would be easier to understand if actual genotype distributions were listed. Also missing is a clear table on the genotype distribution of the two highlighted markers for the animals from the discovery and confirmation cohorts to clearly assess the results.

Table 3 is then a further transformation towards possible practical application, but this original publication should include the true genotypes clearly and legibly.

I would be happy to if authors could address these minor revisions.

## References

1. Richards S, Aziz N, Bale S, Bick D, et al.: Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015; **17** (5): 405-24 [PubMed Abstract](#) | [Publisher Full Text](#)

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Animal genetics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 25 April 2024

<https://doi.org/10.5256/f1000research.162702.r263441>

© 2024 Reecy J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



### James Reecy

Department of Animal Science, Iowa State University, Ames, IA, USA

This manuscript reports the search for potential alleles associated with bovine congestive heart failure (BCHF). Using a case-control approach utilizing cattle from the upper plains of western Nebraska, two regions in the genome were identified as potentially containing associated alleles. These regions are centered around ARRDC3 and NFIA. While these alleles can predict risk of disease, they do not fully explain the risk of the disease.

An interesting issue with this disease is the prevalence of exhibiting the disease with associated with the environment, i.e., elevation. This would indicate a genetic by environmental interaction. Furthermore, there is also a dietary component, i.e., energy density of the diet, of the disease, which was not evaluated in this study.

While a genetic variant in each of these genes is presented as being associated with risk of the disease, there is no evidence provided that these are causal mutations. Whole genome sequencing data is not presented, thus it is very possible that the causal mutations have yet to be identified. This is a strong possibility given that other cattle populations have a relatively high frequency of these alleles, but do not exhibit the disease incidence to the extent that the studies populations do. There is variant data that is available in freely available databases, e.g., Ensembl. None of these alleles appear to have been tested.

The authors talk about a validation population of 170 BCHF cases, but nowhere is this population presented in the materials and methods section. Thus, it is unknown if they came from related animals or the same feedlots.

In Figure 2B, the authors claim that the genetic distances between most of the tested population is very small, except for a few animals, which are clearly separated from the other animals. However, how can the reader know that the observed distances are small compared to other populations when no reference information is provided? A pairwise comparison of the 96 breed samples would have this type of reference.

The protein data analysis is not presented in the results section of the manuscript.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Partly

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** I have over 25 years of beef cattle molecular genetics/genomics research experience.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

---

Version 1

Reviewer Report 26 April 2022

<https://doi.org/10.5256/f1000research.120994.r129617>

© 2022 Drögemüller C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Cord Drögemüller**

<sup>1</sup> Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland

<sup>2</sup> Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland

The study aims to identify a possible genetic cause for the occurrence of BCHF in North American feedlot cattle. The case control study using high density SNP genotypes revealed significant



associations to two different genomic regions. This indicates a possible genetic predisposition, at the same time the presented genotype distributions indicate that it is highly likely a complex inheritance, since the genotypes of the markers in these two genomic regions do not alone explain the emergence of the clinical symptoms.

The two protein-coding genes annotated in the two mapped chromosomal regions have not been associated with heart disease in humans or animals. In principle, this could be of great biomedical interest, as it is well known that in animal phenotypes we often have the possibility to assign a (new or further) previously unknown function to more or less well characterized genes. Especially the missense variant in the *ARRDC3* gene seems to be possibly of pathogenic nature due to the presented conservation of the affected amino acid. What speaks against this is the fact that this variant is obviously not rare in cattle in general, otherwise the SNP would not be present on the commercial 777k array. Table 4 shows impressively that the variant occurs similarly frequently in many other breeds or populations, in all of which there has been no mention of a risk for BCHF. Therefore, because no genome sequence data were available, it is rather to assume that in both cases the identified SNPs are only coupled markers and hardly the molecular cause. This should be clearly differentiated in the paper. It remains unclear whether potentially interesting candidate genes for cardiac dysfunction are located in the vicinity of the two association signals and, given the complex development of the disorder, regulatory variants can be assumed as the cause rather than protein-coding variants, which sometimes lead to phenotypes that manifest themselves at birth or at the beginning of life.

Overall, the work is partly difficult to read because too many methodological details are listed and, on the other hand, the text often remains vague in the biological interpretation. Therefore, the broadly presented possibility of implementing these first results in practical breeding work seems too offensive to me. In my opinion, an even broader validation study with many more affected animals is needed before these markers are directly considered in selection. The study represents a very interesting first step, but the results should be considered with caution, especially since there is always a residual risk of coupled unknown pleiotropic variants on traits desired for breeding.

Further comments:

- 'Fedlot cattle' is not a clear term outside of North America, so it would be better to speak of the breed involved (Angus?).
- Were there any other pathological (histological) data on the disease?
- What is known about the possible inheritance of BCHF? Have segregation analyses been performed? What is the heritability?
- If about 5% of feedlot cattle are at risk for BCHF, why have only about 100 affected animals been identified when over 1400000 cattle have been considered? Is the disease rarer than assumed?
- The nomenclature of variants, especially the representation of *ARRDC3* alleles, should necessarily follow the current recommendations, see Richards *et al.* 2015<sup>1</sup>.
- The intronic variant in the *NFIA* gene remains functionally absolutely unclear, therefore talk

about a linked marker rather than a possible causal variant.

- In the methods section, the races in the panel of 96 unrelated controls should be mentioned.
- The actual linkage disequilibrium between markers in the two associated genomic regions remains unclear in places. In particular, the question arises whether the SNPs mentioned at the end of the methods section are really located with a LD of 1 to the NFIA variant in intron 4 of the gene.
- The wording "homozygous major alleles" on page 10 is unfortunate. Would it not be better to say homozygous for the major allele? An allele itself cannot be homozygous, but the genotype is either hetero- or homozygous for two or one allele.
- In Figure 7 the accession numbers of the listed protein sequences are missing.
- On page 15, a section on the evaluation of the pathogenicity of the missense variant in the *ARRDC3* gene is missing next to the presentation of evolutionary conservation. For example, programs such as MutPred2, SIFT or Provean could be used.
- Table 4 would be easier to understand if actual genotype distributions were listed. Also missing is a clear table on the genotype distribution of the two highlighted markers for the animals from the discovery and confirmation cohorts to clearly assess the results. Table 3 is then a further transformation towards possible practical application, but this original publication should include the true genotypes clearly and legibly.
- The second paragraph of the discussion states "an important feature of some BCHF clinical cases..." What exactly does "some" mean here? This should be clarified.

## References

1. Richards S, Aziz N, Bale S, Bick D, et al.: Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015; **17** (5): 405-24 [PubMed Abstract](#) | [Publisher Full Text](#)

### **Is the work clearly and accurately presented and does it cite the current literature?**

Yes

### **Is the study design appropriate and is the work technically sound?**

Yes

### **Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

### **If applicable, is the statistical analysis and its interpretation appropriate?**

I cannot comment. A qualified statistician is required.

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Animal genetics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 13 Feb 2024

**Michael Heaton**

AUTHORS' RESPONSE TO REVIEWER 2

The authors appreciate and are grateful for both reviewers' thoughtful reviews of our article. Below we have provided point-by-point responses to their suggestions and comments:

**Reviewer 2, point 1:** "The study aims to identify a possible genetic cause for the occurrence of BCHF in North American feedlot cattle."

**Authors response 1:** Our aim has been subtly misstated above, but in an important way. Our stated aim in the Abstract was "to search for genomic regions associated with this disease". Our search was designed to test the null hypothesis that no significant association exists between HD770 SNPs tagging genomic regions and BCHF. We have indicated throughout the article that we are reporting the association of SNP alleles and not the pathogenic roles for *ARRDC3* and *NFIA* variants associated with BCHF. Our focus has been on the strength and reproducibility of the detected associations, not their likelihood of being causal. To underscore our stated aim, we have modified the Introduction section final paragraph to include this text: "... to test the null hypothesis that significant association does not exist between haplotype tagging SNPs and BCHF. Our results rejected the null hypothesis and identified two distinct genomic regions harboring major genetic risk factors associated with BCHF."

**Reviewer 2, point 2:** "Therefore, because no genome sequence data were available, it is rather to assume that in both cases the identified SNPs are only coupled markers and hardly the molecular cause. This should be clearly differentiated in the paper."

**Authors response 2:** We agree and have modified the 4<sup>th</sup> and 5<sup>th</sup> sentences of the discussion to state: "The most significant genome-wide associations were limited to two regions containing *ARRDC3* and *NFIA* genes, respectively. In each region, multiple linked SNPs were associated with BCHF in feedlot cattle. It is unknown whether the associated SNPs are merely

*linked to variants responsible for the elevated BCHF risk, or whether they are involved directly in pathogenesis.” As described in our response to Reviewer 1, we underscore this point and make our intentions unambiguous in the last paragraph: “Thus, WGS analysis of the 102 case-control pairs would refine the bounds of the genomic region most strongly associated with BCHF and identify the complete set of tagging SNPs appropriate for use in genotyping by sequencing. Any novel variants with increased association would potentially represent candidates for functional variants conferring BCHF risk. A WGS analysis may also identify any newly associated genomic regions previously missed due to ascertainment bias in the SNP selection for the BovineHD BeadChip array. For these reasons, we are endeavoring to produce and analyze 15-fold WGS coverage from these 204 animals.”*

**Reviewer 2, point 3:** “‘Fedlot cattle’ is not a clear term outside of North America, so it would be better to speak of the breed involved (Angus?).”

**Authors response 3:** We recognize that the details of North American beef production may not be well known around the world. Moreover, it varies greatly within the US. Feedlots may receive all types of beef, dairy, and composite cattle. Overall, Angus-influenced black cattle represent about 70% of the fattened cattle in the study region. In the method section we describe a group of 204 calves as 93% solid black, 70% castrated males, and none had visible horns. Although the cases were defined by clinical presentation, they seemed roughly representative of the feedlots we sampled. We have inserted the following sentence in the third line of the Introduction section: *“Feedlot cattle in this region are typically a mix of germplasm derived from popular British and Continental European breeds with more than 60% of the animals having a black hide from Angus contributions.”*

**Reviewer 2, point 4:** “Were there any other pathological (histological) data on the disease?”

**Authors response 4:** Systematic histological data was not collected on the 102 pairs of cases and controls. Tissues were collected from the clinical cases at necropsy, but the tissues of healthy controls were processed into beef and unavailable.

**Reviewer 2, point 5:** “What is known about the possible inheritance of BCHF? Have segregation analyses been performed? What is the heritability?”

**Authors response 5:** To date, there have been no reports of inheritance, segregation analysis, or heritability of clinical BCHF. Also, no other studies have used a BCHF phenotype to directly obtain an estimate that could correspond to our study. We used a clinically-defined binary phenotype to represent BCHF, specifically, end-stage congestive heart failure mortality in Western Great Plains feedlot cattle. By comparison all matched pen mate control animals were normal and passed their antemortem inspection at beef processing facilities. Thus, with only two phenotypic states, the use of variance as a convenient mathematical quantity was not appropriate for estimating heritability. Moreover, our BCHF study used animals raised in multiple environments and populations, while heritability is a statistical measure defined in the context of a specific environment and a specific population. For these and other reasons, heritability has not been measured for BCHF.

Since our publication, a report used cardiomegaly observed at slaughter as a proxy for "BCHF" (Buchanan, JW et al., *Front. Genet.* 2023; 14:1148301). However, healthy animals with enlarged hearts cattle that have passed antemortem inspection do not meet our criteria for BCHF and would be considered unaffected controls in our study.

**Reviewer 2, point 6:** "If about 5% of feedlot cattle are at risk for BCHF, why have only about 100 affected animals been identified when over 1400000 cattle have been considered? Is the disease rarer than assumed?"

**Authors response 6:** The 5% number is the mortality in severely affected pens of cattle. It was not the percentage of cattle at risk. To clarify reasons for the case enrollment rate, we have added this text to the Methods section, "*The case enrollment rate of 102 cases per 140,000 feedlot cattle monitored (0.073%) was negatively affected by: 1) removal of sick animals for emergency salvage slaughter, 2) removal of matched unaffected pen mates for beef processing, 3) the rapid onset of congestive heart failure and death, 4) reluctance of overburdened personnel to commit valuable time despite the \$500 indemnity offer for each matched pair identified early, and 5) exclusion of necropsied animals meeting some, but not all, of the case definition criteria.*" We also suspect that the disease is underreported due to worker misclassification at necropsy.

**Reviewer 2, point 7:** The nomenclature of variants, especially the representation of ARRDC3 alleles, should necessarily follow the current recommendations, see Richards et al. 20151.

**Authors response 7:** The Richard et al. 2015 article ("Standards and guidelines for the interpretation of sequence variants") provides detailed criteria for classifying pathogenic variants in genes known to cause human diseases. As noted in Authors response 1 above, our stated aim was "to search for genomic regions associated with this disease". We are not reporting the pathogenicity of variants in this report. We have added the rs numbers for the variants highlighted in Table 2 and added context sequence information in a new Table 5.

**Reviewer 2 point 8:** The intronic variant in the NFIA gene remains functionally absolutely unclear, therefore talk about a linked marker rather than a possible causal variant.

**Authors response 8:** Although it is not clear which sentence in the manuscript that Reviewer 2 is referring, we agree. Other than the possibility of the ARRDC3 C182Y missense SNP affecting protein function, all other SNPs associated with BCHF are presumed to be linked variants. Likewise, it is also possible that the ARRDC3 C182Y missense SNP is merely linked to a functional variant. To emphasize this point we have modified this sentence towards the end of the first paragraph of the Discussion section: *Thus, specific DNA sequence variants within ARRDC3 and NFIA are markers for BCHF risk in feedlot cattle in the Western Great Plains, either through linkage to causal variants or direct effects.*

**Reviewer 2, point 9:** In the methods section, the races in the panel of 96 unrelated controls should be mentioned

**Authors response 9:** We agree. This information was in our original submission to

F1000Research. We have restored this content in the Methods section of the revised article.

**Reviewer 2, point 10:** The actual linkage disequilibrium between markers in the two associated genomic regions remains unclear in places. In particular, the question arises whether the SNPs mentioned at the end of the methods section are really located with a LD of 1 to the NFIA variant in intron 4 of the gene.

**Authors response 10:** We agree and have rewritten the penultimate paragraph of the article to clarify and discuss the strength of LD between the BCHF-associated SNPs in the two separate genomic regions. With regards to the two best associated *NFIA* variants spanning 2.32 kb in intron 4, their LD was high ( $r^2 = 0.94$ ). These two SNPs were only moderately linked to a SNP 127 kb upstream in *NFIA AS2* ( $r^2 = 0.49$  and  $0.45$ , Figure S2).

**Reviewer 2, point 11:** The wording "homozygous major alleles" on page 10 is unfortunate. Would it not be better to say homozygous for the major allele? An allele itself cannot be homozygous, but the genotype is either hetero- or homozygous for two or one allele.

**Authors response 11:** We agree this was terribly worded. The sentence on page 10 line 6 now states, "*Homozygosity for the major allele of either of these two SNPs was associated with increased BCHF risk (OR 7.4, CI<sub>95</sub> 2.9 to 19, Table 2).*"

**Reviewer 2, point 12:** In Figure 7 the accession numbers of the listed protein sequences are missing.

**Authors response 12:** Figure 7 summarizes the results of a much larger ARRDC3 peptide alignment with 80 accession numbers. We have added this larger alignment with its accession numbers to the extended and underlying data section (Table S2. Extended comparison of ARRDC3 residues near the C182Y position in the Gnathostomes. <https://doi.org/10.6084/m9.figshare.22213495>)

**Reviewer 2, point 13:** On page 15, a section on the evaluation of the pathogenicity of the missense variant in the ARRDC3 gene is missing next to the presentation of evolutionary conservation. For example, programs such as MutPred2, SIFT or Provean could be used.

**Authors response 13:** We respectfully disagree. As emphasized above in Author response 1, our aim here was not to evaluate the pathogenicity of associated SNPs. In our opinion, such analyses are premature for bead array SNPs, which have a strong selection biased for high minor allele frequencies and represent a only small fraction of the total genetic variation in the region. In silico predictions of pathogenicity of missense variants would be more appropriate once WGS for these regions is available, and the full scope of associated variants is known. However, to increase our understanding of the Y182 variant on the predicted ARRDC3 tertiary structure, we have added alignments from *de novo* AlphaFold2 predicted structure in Figure 9 and introduced a hypothesis for testing (Discussion section).

**Reviewer 2 point 14:** Table 4 would be easier to understand if actual genotype distributions were listed. Also missing is a clear table on the genotype distribution of the two highlighted markers for the animals from the discovery and confirmation cohorts to

clearly assess the results. Table 3 is then a further transformation towards possible practical application, but this original publication should include the true genotypes clearly and legibly.

**Authors response 14:** We are happy to provide additional lists of genotypes for figures. We recognize that extracting these genotypes from the provided 770k SNP bead array PLINK file (Supplementary Files S1 through S9) is not trivial. In addition, the revised article now contains genotypes corresponding to the 12 SNPs represented Table 2 for the 204 animals (Table S3), the two SNPs from 566 animals represented in Figure 8 (Table S4), and two SNPs for the 1203 animals listed in Table 4 (Table S5).

We also note that the homozygous *NFIA* bar chart data in Figure 8 have changed slightly for the 96 Diversity beef bulls and the 96 persistently infected (BVD) feedlot cattle panels. Their bars increased due to inadvertent initial use of a 3-SNP *NFIA* haplotypes instead of just the single *NFIA* SNP rs133192205 (ie., BovineHD0300024308 or BCHF32). Since the 3-SNP *NFIA* haplotypes were not available for the 171 validation cases, we reverted to using only the single *NFIA* SNP for all cohorts.

**Reviewer 2, point 15:** The second paragraph of the discussion states "an important feature of some BCHF clinical cases..." What exactly does "some" mean here? This should be clarified.

**Authors response 15:** We agree. The sentence has been rephased to read: "More than 50% of the BCHF cases in our study died in the first half of the fattening period."

**Competing Interests:** No competing interests were disclosed.

Reviewer Report 20 April 2022

<https://doi.org/10.5256/f1000research.120994.r129614>

© 2022 Archibald A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Alan L. Archibald** 

<sup>1</sup> The Roslin Institute, Royal (Dick) School of Veterinary Studies (R(D)SVS), University of Edinburgh, Midlothian, UK

<sup>2</sup> The Roslin Institute, Royal (Dick) School of Veterinary Studies (R(D)SVS), University of Edinburgh, Midlothian, UK

A thorough analysis of potential associations between genetic factors and susceptibility to bovine congestive heart failure (BCHF) in feedlot cattle is presented. Genetic variants at the *ARRDC3* and *NFIA* loci are shown to be associated with increased risk of BCHF. The SNP variants in these loci do not fully explain the risk of BCHF, nor have causal genetic variants been identified. Genetic tests to

identify animals at risk of BCHF and/or at risk of transmitting increased risk to their offspring are proposed.

The analysis of genetic variation associated with risk of BCHF is limited to SNP variants present on the Illumina HD SNP chip. Given the strong signal associated with the ARRDC3 locus in particular it would have been useful to sequence this locus in multiple individuals of known BCHF phenotype in order to identify candidate causal or risk variants. The authors highlight a missense variant ARRDC3 C182Y for which there is an assay as BovineHD0700027239 on the Illumina HD SNP chip. They note that a cysteine residue is found at this position in the ARRDC3 protein in a wide range of animals separated by millions of years of evolutionary time.

Multiple sequence variants have been reported in the bovine ARRDC3 locus including several other missense mutations and premature stop codons as summarised here:

[Link](#)

Several of these missense and premature stop variants are predicted by SIFT analysis to be deleterious whereas the ARRDC3 C182Y variant is not predicted by SIFT analysis to be deleterious.

Sequence analysis of the ARRDC3 locus in individuals of known BCHF phenotypes might reveal candidate causal or risk variants.

In order to link the SNP information reported in the manuscript more readily to other data sources it would be useful to add rs numbers. For example, the rs number for the ARRDC3 C182Y variant for which there is an assay as BovineHD0700027239 on the Illumina HD SNP chip is rs109901274.

The frequencies of the ARRDC3 risk alleles as summarised in Table 4 are very high in some breeds, including Angus, Red Angus, Murray Grey and i) pose a challenge to selective breeding in these breeds to reduce the risk of BCHF and ii) raises questions about why the incidence of BCHF is not higher.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

I cannot comment. A qualified statistician is required.

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes



**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Farmed animal genetics and genomics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 13 Feb 2024

**Michael Heaton**

AUTHORS' RESPONSE TO REVIEWER 1

The authors appreciate and are grateful for both reviewers' thoughtful reviews of our article. Below we have provided point-by-point responses to their suggestions and comments:

**Reviewer 1, point 1:** "Sequence analysis of the *ARRDC3* locus in individuals of known BCHF phenotypes might reveal candidate causal or risk variants."

**Authors response 1:** We agree. We attempted to indicate this point in the last paragraph of the discussion. However, to further underscore this point and make our intentions unambiguous, we have revised the first paragraph of the discussion: "*It is unknown whether the associated SNPs are merely linked to variants responsible for the elevated BCHF risk, or whether they are involved directly in pathogenesis. Whole genomic sequence (WGS) analyses are underway for the 102 case-control pairs of animals and may unlock further insights in the future.*" The last paragraph of the discussion was also revised to emphasize this point: "*Any novel variants with increased association would potentially represent candidates for functional variants conferring BCHF risk. A WGS analysis may also identify any newly associated genomic regions previously missed due to ascertainment bias in the SNP selection for the BovineHD BeadChip array. For these reasons, we are endeavoring to produce and analyze 15-fold WGS coverage from these 204 animals.*"

**Reviewer 1, point 2:** "In order to link the SNP information reported in the manuscript more readily to other data sources it would be useful to add rs numbers."

**Authors response 2:** We agree and have added rs numbers to Table 2 for the 12 SNPs most associated with BCHF. To provide additional unique molecular information defining these 12 SNP, we also added Table 5 which includes 200 bp of adjacent context nucleotide sequence from the ARS-UCD2.1 bovine reference assembly, together with flanking SNP variants greater than 0.015 frequency as identified in a 96-member beef cattle diversity panel. Together, these details are important for designing accurate assays to score genotypes at SNP sites associated with BCHF in feedlots of the Western Great Plains.

**Reviewer 1, points 3 and 4:** "The frequencies of the *ARRDC3* risk alleles as summarised in Table 4 are very high in some breeds, including Angus, Red Angus, Murray Grey and i) pose a challenge to selective breeding in these breeds to reduce the risk of BCHF and ii) raises questions about why the incidence of BCHF is not higher."

**Authors response 3 and 4:** We agree and have added two sequential paragraphs to the Discussion section: *"The high frequencies of ARRDC3 and NFIA risk alleles in some affected breeds (> 0.65) raise the question of why the disease prevalence is relatively low even in outbreaks (> 0.07)? This phenomenon may be due to the multifactorial nature of complex disease. The development of BCHF in cattle is presumed to be influenced by exposure to non-genetic factors, such as elevation (hypoxia), extreme weather conditions, feeding schedules, ionophores, feed grade antibiotics, steroid hormone implants, and beta-adrenergic agonist feed additives (Krafsur et al, 2019, Neary et al. 2018). These environmental and management factors may play significant but unknown roles in disease development. In addition, most complex diseases are considered polygenic with variation at many genes having small effects (Khera et al. 2018). DNA sequence variation identified here in ARRDC3 and NFIA genomic regions may predispose cattle to BCHF under certain conditions, despite the absence of any known causal pathogenic mutations.*

*The high frequencies of risk alleles observed in some breeds also poses a challenge for selective breeding. Although estimates for the positive predictive value of the ARRDC3 and NFIA markers were low (0.10), the negative predictive value was quite high (0.99). Consequently, identifying breeding stock lacking these risk factors could prove useful for reducing a herd's overall risk. The challenge is finding sires without the ARRDC3 and NFIA risk alleles in breeds with a high prevalence of high-risk alleles. For instance, the largest affected breed in the United States, American Angus, has more than 300,000 bulls registered annually in its association. However, less than 5% (15,000) are expected to possess the optimal low-risk genotypes at both the ARRDC3 and NFIA gene regions (M. Heaton, L. Kuehn, and B. Vander Ley 2022, unpublished). To identify the best bulls for decreasing BCHF risk alleles in their herds, cattle producers may need to obtain genotypes from hundreds of candidate sires. Unfortunately, this information is not widely available for most bulls at present. Despite these challenges, there are a number of reasonable selection strategies that could help reduce the risk of BCHF in the long term, depending on the specific goals and circumstances of the cattle producer."*

**Competing Interests:** No competing interests were disclosed.

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**