PE0326

Association of *ARRDC3* and *NFIA* Genes with Bovine Congestive Heart Failure

MP Heaton[1], Harhay GP[1], Bassett AS[2], Clark HJ[2], Carlson JM[2], Smith HR[1], Pelster MC[2], Workman AM[1], Grotelueschen DM[2] and Vander Ley BL[2]

1) US Meat Animal Research Center, Clay Center, Nebraska, USA; 2) University of Nebraska, Lincoln, Nebraska, USA

Abstract

Background: Bovine congestive heart failure (BCHF) has become increasingly prevalent in feedlot cattle in the Western Great Plains of North America with up to 7% mortality in disease outbreaks. BCHF is an untreatable complex condition involving pulmonary hypertension that culminates in right ventricular failure and death. No candidate genes are presently associated with BCHF, and thus, our aim was to search genome-wide for genetic risk factors in feedlot cattle.

Methods: Samples of 102 clinical BCHF cases and 102 unaffected matched penmates were used in a genome-wide association study (GWAS) with 777,962 single nucleotide polymorphisms (SNPs). The paired nominal data were analyzed with McNemar's test.

Results: Analyses of 563,042 filtered SNPs revealed more than 15 genomic regions highly associated with BCHF. Regions with the strongest association included the arresting domain-containing 3 protein (*ARRDC3*) and nuclear factor IA (*NFIA*) genes. A missense mutation in exon 4 of *ARRDC3* (C182Y), and SNPs in intron 5 of *NFIA* had the best statistical support for association (McNemar's Chi-square > 20). Animals with either or both the *ARRDC3* or *NFIA* risk factors were approximately 7- and 15-fold more likely to have BCHF compared to those without (*p*-value < 10^{-10} for both risks present). A two-SNP genotyping test for *ARRDC3* and *NFIA* risk factors was used to test an independent cohort of feedlot cattle with end-stage heart failure and similar associations with disease were observed.

Conclusions: A matched case-control GWAS identified major genes associated with BCHF in feedlot cattle. Although the roles of these genes in disease pathogenesis are unknown, their discovery facilitates classifying animals by genetic risk for heart failure and will allow producers to make informed decisions for selective breeding and animal health management.

Matched case-control pair

Figure 1. Example of a pen matched case-control pair. Panel A. Steer with end-stage congestive heart failure (left); unaffected penmate (right). Panel B. Heart from an affected feedlot heifer (lower left), and an unaffected fattened Angus heifer (lower right). The view of the sectioned hearts is looking away from the apex and into the valves. The Abbreviations: RV, right ventricle; LV, left ventricle; PA, pulmonary artery.

Outcomes

- *ARRDC3* and *NFIA* variants are major risk factors for heart failure in feedlot cattle.
- A 2-SNP test sorts animals by risk group
- Other loci were also significantly associated

Conclusion

Congestive heart failure in feedlot cattle has major underlying genetic factors.

Testing for risk with two SNPs

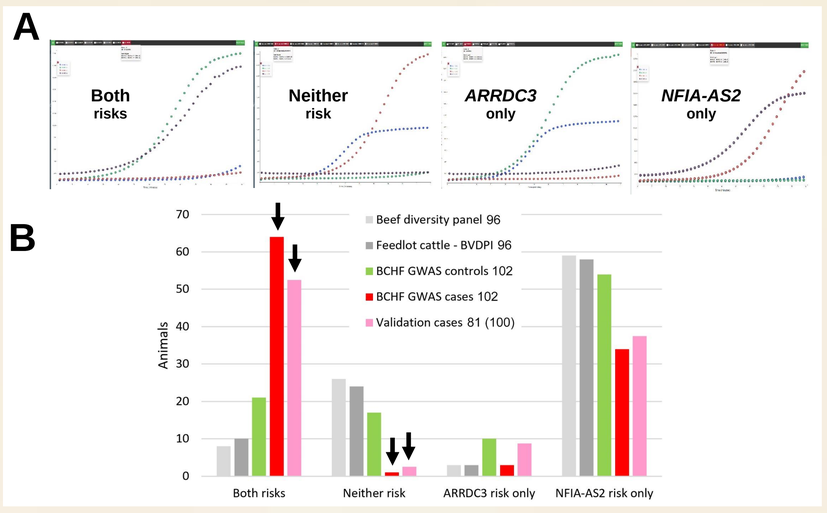

Figure 5. Using a 2-SNP test to identify risk groups in populations. Panel A. Representative animal genotype spectra from "C-SAND" assays based on padlock probe methodology (MatMaCorp, Lincoln, NE). The targeted SNPs were BovineHD0700027239 (*ARRDC3,* BCHF5) and BovineHD0300024366 (*NFIA*, BCHF2). Panel B. Distributions of animals by risk group in four bovine cohorts: USMARC beef cattle diversity panel v2.9 (19 breeds, n = 96, Heaton et al., 2016 **5**:2003), feedlot calves persistently infected with bovine viral diarrhea virus (BVDPI, n = 96, Workman et al., 2016 J. Vet. Diag. Invest. **28**: 519–528), pen-matched BCHF controls (this study, n = 102), pen-matched BCHF cases (this study, n = 102), BCHF independent "validation" feedlot cases (this study, n = 81 [scaled to 100]). The latter samples were collected by experienced pen riders but were not confirmed at necropsy by veterinarians.

McNemar test for association

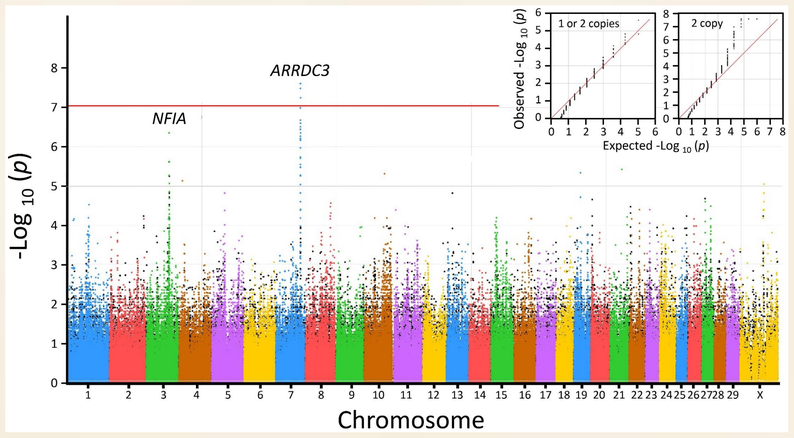

Figure 2. Manhattan and q-q plots of McNemar's exact *p*-values for 563,042 genome-wide SNPs. SNP genotypes on the BovineHD BeadChip array were scored at Neogen Genomics (Lincoln, NE, USA), according to manufacturer's instructions (Illumina, Inc., San Diego). SNPs were filtered with PLINK software v1.9 (Purcell et al., Am J Hum Genet. 2007 81:559-75) and imported into custom software for McNemar test analyses. For each SNP, both alleles were tested for association. Colored dots show results for a "2-copy" model, where disease risk is conferred by having 2 copies of the SNP allele (i.e., recessive). Black dots show results for a "1 or 2-copy" model where disease risk is conferred by having either 1 or 2 copies of the SNP allele (i.e., additive or dominant). The horizontal red line is the Bonferroni correction at 0.05 alpha.

Power to detect association

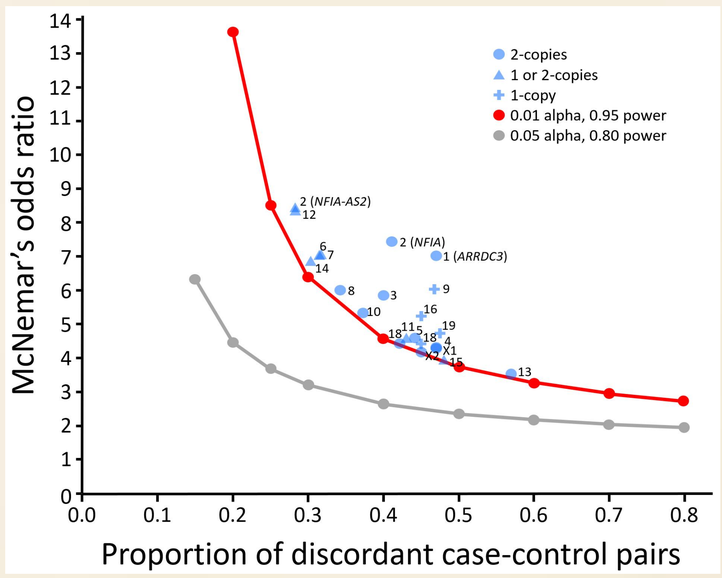

Figure 4. Relationship between effect size of the top 21 associated loci and the power for their detection with 102 matched pairs. SNPs ranked in Table 2 are plotted against two power curves for thresholds of detection. SNPs (562 k) associated with disease were filtered by proportion of discordant pairs (>24) and McNemar's $\chi 2$ (>15).

Genetic distance within pairs

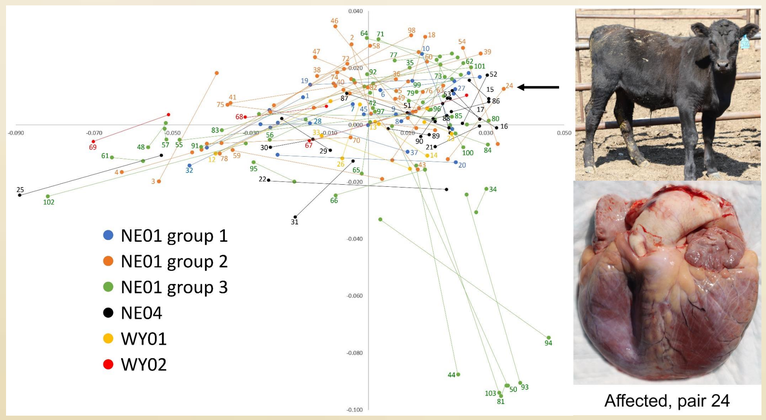

Figure 3. Multidimensional scaling (MDS) plot showing pairwise identity-by-state distances to visualize stratification within and between paired individuals. NE01, NE04, WY01, and WY02 are feedlot locations in Nebraska and Wyoming respectively, Groups 1, 2, and 3 at the NE01 location differ by general source/herd. MDS data was produced with PLINK software v1.9 and plotted with Microsoft Office Excel software.

Table 1. Feedlot sites and sources for case-control pairs.

Site	Altitude m (ft)	Pairs	Sources
NE01	1,242 (4,075)	76	19
NE04	1,163 (3,816)	17	9
WY01	1,263 (4,143)	6	1
WY02	1,280 (4,198)	3	1
Totals		102	30

Heaton et al., 2019 F1000Research **8**:1189

Top 21 SNP associations

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

Rank[a]	Chr	SNP ID	UMD3.1 Chr. pos. (bp)	ARS1.2 Chr. pos. (bp)	Nearest genes	SNP location	A1[b]	A2	A1 Frequency Case	A1 Frequency Control	A1 Frequency Ref.	Risk allele[c]	McNemar pairs 1,1 *a*	1,0 *b*	0,1 *c*	0,0 *d*	*b* + *c*	*n*	(*b* + *c*)/*n*	OR	OR CI95 Lower	OR CI95 Upper	$\chi 2^d$	*p*-value (exact)	-log(*p*)	SNPs in LD (χ2 >15)	Block size[e] (kb)
1	7	BovineHD0700027239	93244933	90845941	ARRDC3	Exon 4 C182Y tRt	A	G	0.794	0.583	0.219	2 copies	A	25	42	6	29	48	102	0.471	7.0	3.0	16.5	25.5	1.01E-07	7.0	56	956
2	3	BovineHD0300024307	85123495	84578325	NFIA	Intron 2	G	A	0.647	0.427	0.448	2 copies	G	7	37	5	53	42	102	0.412	7.4	2.9	18.8	22.9	4.43E-07	6.4	21	596
2	3	ARS-BFGL-NGS-103524	85253155	84706206	NFIA-AS2	Intron3	C	A	0.25	0.525	0.443	1 or 2 copies	A	72	26	3	1	29	102	0.284	8.7	2.6	28.6	16.7	1.52E-05	4.8	21	596
3	10	BovineHD1000021490	75280294	75267920	KCNH5	Intron 4	A	G	0.672	0.476	0.542	2 copies	A	14	35	6	47	41	102	0.402	5.8	2.5	13.9	19.1	4.87E-06	5.3	0	5
4	5	BovineHD4100003664	48418959	48188142	HMGA2_MSRB3	Intergenic	A	G	0.706	0.525	0.203	2 copies	A	15	39	9	39	48	102	0.471	4.3	2.1	8.9	17.5	1.52E-05	4.8	2	5
5	26	BovineHD2600006169	23929257	23685234	CNNM2	Intron 1	A	C	0.884	0.716	0.729	2 copies	A	41	37	8	16	45	102	0.441	4.6	2.2	9.9	17.4	1.54E-05	4.8	0	4
6	19	BovineHD1900007657	25991182	25381310	PITPNM3	Intron 4	A	C	0.799	0.941	0.760	1 or 2 copies	C	7	28	4	63	32	102	0.314	7.0	2.5	20.0	16.5	1.93E-05	4.7	9	30
7	22	BovineHD4100015417	5652914	5605128	GADL1_STT3B	Intergenic	A	G	0.576	0.430	0.552	1 or 2 copies	A	60	28	4	10	32	102	0.314	7.0	2.5	20.0	16.5	1.93E-05	4.7	0	5
8	20	BovineHD2000001515	4762743	4854589	ENIP1	Exon 6 3'UTR	G	A	0.451	0.652	0.859	2 copies	A	3	30	5	64	35	102	0.343	6.0	2.3	15.5	16.5	2.24E-05	4.7	0	8
9	9	BovineHD0900027458	96528831	95086861	SYTL3	Intron 3	A	C	0.490	0.368	0.391	1 copy	A	29	30	5	11	35	75	0.467	6.0	2.3	15.5	16.5	2.24E-05	4.7	0	2
10	4	BovineHD0400003776	12664064	12793811	ASB4	Intron 3	G	A	0.960	0.824	0.688	2 copies	G	61	32	6	3	38	102	0.373	5.3	2.2	12.8	16.4	2.43E-05	4.6	0	22
11	27	BovineHD2700003477	11978259	12942894	AGA_TENM3	Intergenic	A	G	0.451	0.657	0.698	1 or 2 copies	G	48	36	8	10	44	102	0.431	4.5	2.1	9.7	16.6	2.54E-05	4.6	4	48
12	8	BTB-01266056	95400554	93743405	SMC2	Intron 20	C	A	0.855	0.970	0.814	1 or 2 copies	A	3	25	3	71	28	102	0.283	8.3	2.5	27.6	15.8	2.74E-05	4.6	1	21
13	1	BovineHD0100023638	82311685	81722301	MAP3K13	Intron 10	G	A	0.882	0.721	0.792	2 copies	G	36	45	13	8	58	102	0.569	3.5	1.9	6.4	16.6	3.01E-05	4.5	0	4
14	8	Hapmap27238-BTA-163742	93655394	92016089	GRIN3A_CYLC2	Intergenic	A	G	0.825	0.955	0.823	1 or 2 copies	G	5	27	4	66	31	102	0.304	6.8	2.4	19.3	15.6	3.40E-05	4.5	2	6
15	22	BovineHD2200011041	38708159	38567450	CADPS	Intron 3	A	G	0.530	0.690	0.750	1 or 2 copies	G	41	39	10	12	49	102	0.480	4	1.9	7.8	16.0	3.85E-05	4.4	0	2
16	3	ARS-BFGL-NGS-110776	41119556	40971803	COL11A1_OLFM3	Intergenic	C	A	0.667	0.745	0.750	1 copy	A	17	31	6	28	37	82	0.451	5.2	2.2	12.4	15.6	4.13E-05	4.4	0	3
17	22	BovineHD2200002812	9438878	9404132	PDCD6IP_ARPP21	Intergenic	A	G	0.940	0.801	0.662	2 copies	A	53	35	8	6	43	102	0.422	4.4	2.0	9.4	15.7	4.19E-05	4.4	0	8
18	26	BovineHD2600008347	31264198	31000995	SMNDC1_DUSP5	Intergenic	A	G	0.767	0.891	0.651	1 copy	G	6	35	8	46	43	95	0.453	4.4	2.0	9.4	15.7	4.19E-05	4.4	0	12
19	24	BovineHD2400006360	23311439	23021459	NOL4	Intron 4	A	G	0.662	0.799	0.635	1 copy	G	12	33	7	32	40	84	0.476	4.7	2.1	10.7	15.6	4.23E-05	4.4	0	7
1	X	BovineHD3000025651	93114732	87952697	CCNB3	Intron 1	A	C	0.72	0.469	na	Homozygous[f]	A	28	39	9	26	48	102	0.471	4.3	2.1	8.9	17.52	1.52E-05	4.8	10	161
2	X	BovineHD3000025051	91049685	85788201	ZNF41	Intron 2	C	A	0.87	0.662	na	Homozygous	C	48	37	9	8	46	102	0.451	4.1	2.0	8.5	15.85	4.06E-05	4.4	1	1

[a] In the three models tested, the genetic risk factor were defined as having exactly 1, 1 or 2, or exactly 2 copies of the risk allele, respectively.

[b] A1 was defined in the McNemar's test analyses as the most frequent allele in the combined group of 204 cases and controls.

[c] In the three models tested, the genetic risk factor were defined as having exactly 1, 1 or 2, or exactly 2 copies of the risk allele, respectively.

[d] McNemar's chi-squared with continuity correction: $(|b - c| - 1)^2/(b + c)$

[e] Distance between distal SNPs within block of LD where $\chi 2 > 15$. If no additional linked SNPs present, distance between non-linked adjacent SNPs.

[f] In the X-chromosome analysis, pairs of males and pairs of females were analyzed together. Since male X-chromosome genotypes are always homozygous, the risk factor was defined as being homozygous at the position. Both alleles were evaluated for being the risk factor.

C182 conservation in ARRDC3

Table 3. Evolutionary comparison of ARRDC3 residues near the C182Y position in the jawed vertebrates

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

[a] TMRCA is the estimated time to most recent common ancestor in millions of years Hedges SB et. al., Mol Biol Evol. 2015; 32(4): 835–45.

[b] The full length ARRDC3 protein is 414 in cattle and most of the Amniota species

[c] The letters are IUPAC/IUBMB codes for amino acids. The dots are amino acid residues identical to those in cattle.

School of Veterinary Medicine and Biomedical Sciences
GREAT PLAINS VETERINARY EDUCATIONAL CENTER

USDA Agricultural Research Service USMARC

University of Nebraska Lincoln

Acknowledgements

We thank D. George, 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; Drs. G. Krafsur, 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.