Genome-wide association study of beef bull semen attributes

M. L. Butler, A. R. Hartman, J. M. Bormann, R. L. Weaber, D. M. Grieger and M. M. Rolf[*]

Abstract

Background: Cattle production is dependent upon fertility because it results in producing offspring to offset production costs. A number of semen attributes are believed to affect fertility and are frequently measured as part of routine breeding soundness exams or semen collection procedures. The objective of this study was to perform a single-step genome-wide association study (ssGWAS) for beef bull semen attributes. Beef bull fertility phenotypes including volume (VOL), concentration (CONC), number of spermatozoa (NSP), initial motility (IMot), post-thaw motility (PTMot), three-hour post-thaw motility (3HRPTMot), percentage of normal spermatozoa (%NORM), primary abnormalities (PRIM), and secondary abnormalities (SEC) were obtained from two artificial insemination (AI) centers. A total of 1819 Angus bulls with 50,624 collection records were used for ssGWAS. A five-generation pedigree was obtained from the American Angus Association and consisted of 6521 sires and 17,136 dams. Genotypes on 1163 bulls were also obtained from the American Angus Association and utilized in ssGWAS.

Results: A multi-trait animal model was used for the estimation of single nucleotide polymorphism (SNP) effects. Significant SNP were those with a -$\log_{10}$ *P*-value threshold greater than 4.0. Volume, CONC, NSP, IMot, PTMot, 3HRPTMot, %NORM, PRIM, and SEC have five, three, six, seven, two, six, six, and two genome-wide significant SNP, respectively.

Conclusions: Several significant SNP were determined to be near or within quantitative trait loci (QTL) associated with beef bull semen attributes. In addition, genes associated with fertility were found to contain or be near the significant SNP found in the study. The results indicate there are regions of the genome that impact fertility, proving inclusion of genomic information into genetic evaluation should be advantageous for genetic improvement of male fertility traits.

Background

Fertility is a complex trait which is affected by management [1–3], nutrition [4, 5], and genetics [6, 7]. The complexity of fertility may be one reason for the lack information available to beef seedstock producers for making fertility selection decisions. Other reasons may include that fertility data is not currently aggregated by beef breed associations and semen quality data from bull studs is largely proprietary. As genomic testing becomes more affordable and a part of regular management practices for seedstock producers, beef cattle producers become more willing and able to utilize the technology. Utilizing genomic technology in breeding decisions allows cattle producers to more confidently incorporate young, unproven sires into their breeding programs. While using young breeding stock increases the rate of genetic change, the risk of using these bulls is higher because their expected progeny differences (EPD) accuracies are lower. Beef cattle producers do not currently have a selection tool to confidently choose more fertile bulls, nor are current versions of SNP chips enriched in regions where important variants for male fertility exist because they are largely unstudied and unidentified.

*Correspondence: megrolf@ksu.edu
Department of Animal Science Kansas State University, Manhattan, Kansas 66502, USA

Butler *et al. BMC Genomics*   (2022) 23:74

Page 2 of 12

However, thousands of semen records from beef bull semen collection facilities provide an opportunity to identify polymorphisms which may affect spermatogenesis, morphology, and motility of spermatozoa [8]. One way to truly increase the reproductive efficiency of beef cattle production is to build the capacity to identify young beef bulls with sperm abnormalities and low semen quality so that the bulls can be eliminated from the breeding population [8]. Dairy cattle research underscores the ability to capitalize on genomic technologies in these efforts, as researchers have identified multiple QTL regions [9–17] and candidate genes [18–24] associated with male and female fertility. Researchers have identified areas of the genome that are associated with fertility traits in beef bulls [25], but the number of studies is limited. Thus, the need for further validation is a necessity. The objective of this study was to perform a genome-wide association study for beef bull semen attributes and to identify quantitative trait loci (QTL) regions and genes likely associated with fertility traits in beef bulls.

Methods

Data collection

Phenotypic observations for 1819 Angus bulls were obtained from two bull semen collection facilities. The phenotypic observations for beef bull semen attributes included volume (VOL), concentration (CONC), number of spermatozoa (NSP), initial motility (IMot), post-thaw motility (PTMot), three-hour post-thaw motility (3HRPTMot), percentage of normal spermatozoa (%NORM), primary abnormalities (PRIM), and secondary abnormalities (SEC). Trait definitions are listed in Table 1. Data editing procedures were performed to ensure no VOL or CONC phenotypes were 0, all phenotypes recorded as a percentage ranged from 0 to 100, and all bulls had a registration number. After data editing procedures there were 50,624 total collection records utilized for the ssGWAS. Bull stud A contributed 48,131 collection records from 1570 bulls and bull stud B contributed an additional 2493 records from 256 bulls. Summary statistics are shown in Table 1.

The American Angus Association provided pedigree information for 1819 bulls and SNP data for 1163 bulls. The pedigree contained 6521 sires and 17,136 dams. The maximum number of SNP per bull was 54,609. Data were edited to remove SNP with a call rate of < 0.90 ($n = 3921$) and minor allele frequency of < 0.05 ($n = 13,478$). Animals with a call rate of < 0.90 were removed ($n = 3$). Imputation was performed by genotyping providers to the American Angus Association, so that all genotyped bulls had 54,609 SNP. After quality control filtering, a total of 38,515 SNP from 1160 animals were available for analysis. The American Angus Association also provided the SNP positions

Table 1 Trait list, definitions, and summary statistics of beef bull fertility traits collected at an artificial insemination. All post-thaw motility measures are observed the day after the sample is collected. All traits are measured within one hour except for three-hour post-thaw motility

Trait	Units	Definition	N	Mean	Std Dev	Minimum	Maximum
Volume (VOL)	mL	total amount of the ejaculate, measured by milliliters	44,431	7.96	4.22	0.10	74.00
Concentration (CONC)	Million/mL	relative amount of sperm cells per ejaculate, measured by a colorimeter	44,038	1021.16	495.02	10.00	3906.00
Number of spermatozoa (NSP)	Million	calculated by multiplying sperm concentration and semen volume; expressed in millions	44,418	49.18	16.16	0.00	100.00
Initial Motility (IMot)[a]	%	percentage of progressively swimming spermatozoa in the ejaculate immediately after collection	44,038	8004.18	5519.61	64.50	69,795.00
Post-thaw Motility (PTMot)[a]	%	percentage of progressively swimming spermatozoa in the ejaculate, measured within one hour of thawing	29,877	43.53	13.78	0.00	75.00
Three-hour post-thaw motility (3HRPTMot)[a]	%	percentage of progressively swimming spermatozoa in the ejaculate, measured three hours after thawing	8299	15.72	12.48	0.00	60.00
Percentage of Normal Spermatozoa (%NORM)[a]	%	percent morphologically normal spermatozoa	19,455	75.18	8.37	2.00	100.00
Primary Abnormalities (PRIM)[a]	%	percentage of spermatozoa with a defect to the head	19,452	13.00	7.53	0.00	100.00
Secondary Abnormalities (SEC)[a]	%	percentage of spermatozoa with a defect to the tail	19,521	12.09	7.70	0.00	100.00

[a] Trait is measured subjectively by a trained laboratory technician

Butler *et al. BMC Genomics*   (2022) 23:74

Page 3 of 12

and the reference genome used by the American Angus Association was UMD3.1.

Statistical analysis

Beef bull fertility is known to be affected by a variety of environmental factors. A detailed model selection procedure was utilized to select the fixed effects and covariates to include in the analysis and calculate the variance components as described by [26]. In brief, we used a forward selection procedure to identify factors that significantly affected beef bull semen attributes. The factors included in the final model included location where the bull was collected (Location), the class effect of day within year (DayYear), a covariate for age of the bull on the day of collection in days (Age), a covariate for days since previous collection (DaysSince), and a covariate effect of cumulative comprehensive climate index (cumCCI) over the spermatogenesis cycle. The mathematical model used for analysis was as follows:

$$y_{ijk} = S_j + W_{pe} + Location + DayYear + Age + DaysSince + Cum\ CCI + \varepsilon$$

where y is the phenotype being evaluated, S_j is a random animal effect, W_{pe} is the permanent environment effect to account for repeated measures, and ε is the residual.

The mixed model used in the current study is:

$$\begin{bmatrix} X'X & X'Z_a & X'W_{pe} \\ X'Z_a & Z'_a Z_a + \lambda H^{-1} & Z'_a W_{pe} \\ X'W_{pe} & Z'_a W_{pe} & W'_{pe} W_{pe} + I\alpha_{pe} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{pe} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix}$$

where X is an incidence matrix relating the phenotypic observations to the fixed effects in the model (detailed above), Z_a is an incidence matrix relating the phenotypic observations to the additive direct genetic effects, W_{pe} is an incidence matrix relating the phenotypic observations to the additive permanent environment genetic effects, λ is the ratio of residual and additive direct genetic variance, I is an identity matrix, and α is the ratio of residual variance and additive permanent environment variance, $\hat{b}$ is a vector of the fixed effect solutions, $\hat{a}$ is a vector of additive direct genetic effects, $\hat{pe}$ is a vector of additive permanent environment genetic effects, and y is a vector of phenotypic observations. Estimates for specific variance components are found in [26].

All bulls in this study had phenotypic data but because not all phenotyped animals were genotyped, and to include all available information into the analysis, both pedigree and genomic information were utilized in the genetic evaluation of beef bull fertility using a single-step genomic best linear unbiased prediction (ssGBLUP) model. In a ssGBLUP model, the numerator relationship matrix (A) is replaced with H, which augments A with the genotype information. The inverse of the H matrix is generated as follows:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

where A^{-1} is the inverse of the pedigree-based numerator relationship matrix, A^{-1}_{22} is a subset of the numerator relationship matrix for the genotyped individuals, and G is the genomic relationship matrix for the genotyped individuals [27].

The following algorithm was used to back solve for the SNP effects for the ssGWAS [28]:

$$a = Zu$$

where **a** is a vector of breeding values for the genotyped individuals generated from BLUPF90, **Z** is a matrix relating individuals to phenotypes, and **u** is a vector of SNP marker effects.

The variance for the genotyped animal effects is as follows [28]:

$$\text{var}(a) = \text{var}(Zu) = G\sigma_a^2 = ZIZ'\lambda$$

where **G** is the genomic relationship matrix, **I** is an identity matrix, and **λ** is the ratio of the SNP marker effect variance and the breeding value variance.

The SNP effects were predicted utilizing the following equation:

$$\hat{u} = IZ'(ZIZ')^{-1}\hat{a}$$

where **u** is a vector of SNP marker effects, **I** is an identity matrix, **Z** is a matrix relating individuals to phenotypes, and **a** is a vector of breeding values for the genotyped individuals. Each SNP was assumed to have an equal allele substitution effect variance, and it was assumed the SNP affects followed the infinitesimal model. Thus, an unweighted ssGBLUP was performed.

SNP effects were obtained from one trivariate and six bivariate analyses. VOL, CONC, and NSP were included in a trivariate analysis. Due to convergence issues, bivariate analyses were performed among the remaining groups of traits. Initial motility, PTMot, and 3HRPTMot SNP effects were generated from three different bivariate analyses, and %NORM, PRIM, and SEC from an additional three bivariate analyses. For traits where a bivariate analysis was performed, significant SNP from both bivariate analyses are reported. Utilizing multi-trait analyses allows for multiple, genetically correlated traits to be included in the model and improves the predicative capability for each trait. The *P*-values associated with the SNP effects were obtained from the POSTGSF90 program

Butler *et al. BMC Genomics*   (2022) 23:74

Page 4 of 12

within the BLUPF90 software suite as detailed in [29]. The *p*-value for the SNP effect is obtained by [30]:

$$p_i = 2\left(1 - \Phi\left(\left|\frac{\alpha_i}{sd(a_i)}\right|\right)\right)$$

where α_i is the estimate of the marker effects, sd is the standard deviation, and Φ is the cumulative standard normal function.

The *P*-values are generated by back solving for SNP effects from the breeding value estimates. This approach is possible because the fitting of animal as a random effect to generate breeding value estimates is an equivalent model to fitting all SNPs as random effects and solving for these effects directly [31].

After obtaining the SNP effects ($\hat{\boldsymbol{u}}$) and their *P*-values, Manhattan plots for all nine traits were generated utilizing the CMplot package in R [32]. Utilizing the Qvalue R package [33] the -$\log_{10}$ *P*-values were converted to *p*-values so that the false discovery rate (FDR) could be calculated for each *P*-value. While there were no significant SNP identified at an FDR threshold of < 0.0001, < 0.001, < 0.01, < 0.025, or < 0.05, several significant SNP existed with a -$\log_{10}$ *P*-value threshold of 4.0 (Table 2). This less stringent threshold was chosen to allow investigation of the potential biological significance of the QTL regions identified, and is not detrimental to efforts to utilize these results in an unweighted ssGBLUP that does not specifically fit the effect of any single SNP for genomic prediction.

QTL analysis

To account for linkage disequilibrium in the *Bos taurus* genome, QTL regions were formed 250,000 kilobases upstream and downstream from the significant SNP locations [34]. The regions were utilized to identify previously reported QTL which were near the significant SNP (±250,000 kilobases). The QTL regions were identified utilizing the cattle QTL database [35]. For each trait, previously reported QTL were identified close to significant SNP. Where possible, SNP were identified by the corresponding rs number. If no rs number was available, the SNP name was utilized. In addition, the same QTL region boundaries were used to identify genes near the significant SNP using the National Center of Biotechnology Information (NCBI) database. Putative candidate genes were identified near significant SNP and examined for biological meaning. The genes specifically associated with beef bull semen attributes were identified based on the gene functions [36]. The gene list was then used to perform a functional annotation analysis utilizing the Database for Annotation, Visualization, and Integrated

Discovery (DAVID 6.8) [37, 38] for each individual beef bull semen attribute.

Results and discussion

Significant SNP were identified for all beef bull semen attributes. Significant SNP are reported in Table 2 and Manhattan plots are provided in Figs. 1, 2, and 3. For each trait, previously reported QTL were identified close to significant SNP (Table 3). Putative candidate genes specifically associated with beef bull semen attributes are outlined in Table 4.

For VOL, five strongly associated SNP were identified, including BTB-01549373, rs41666488, rs109736826, rs109268478, and rs41575945 (Table 2). Three QTL regions associated with conception rate or non-return rate in dairy cattle [10, 39, 52] were near BTB-01549373 and rs41575945 (Table 3). While conception and non-return rate are generally associated with female fertility, it is interesting that SNP significantly associated with male fertility traits are also near these QTL regions. This is evidence that the ability to successfully reproduce is dependent not only on the female, but also on the semen quantity and quality of the bull [52]. This may also indicate that although the few reported genetic correlations between male and female fertility are low [47, 53, 54], there may be some pleiotropy whereby mutations control aspects of both male and female fertility.

Three significant SNP were identified for CONC, which included rs43067163, rs41623602, and rs29023737 (Table 2). A previously reported mature height QTL region was close to rs41623602 [46]. In addition, a previously reported QTL region associated with weaning weight and mature height was in proximity to rs41623602 (Table 3) [46]. While there are few studies that directly relate size of the bull to the semen quality, weight and height are generally associated with maturity, and the association could be indicative that more mature, larger bulls have greater semen production. In the current study, weight and hip height of the bull at the time of collection did not contribute significantly to concentration [26].

Seven SNP had significant associations with NSP (Table 2). Three different previously reported conception rate QTL regions [39, 40] neighbored significant SNP for NSP. In addition, previously reported QTL regions associated with non-return rate were in the vicinity of two significant SNP from the current study [11, 41]. Like conception rate, non-return rate is usually related to female fertility, but this association provides further validation that male fertility traits are integral in the reproductive success of the herd [52] and/or that male and female fertility share a genetic correlation due to pleiotropy in these regions of the genome.

Butler *et al. BMC Genomics*   (2022) 23:74

Page 5 of 12

Table 2 Genomic regions identified by genome-wide association study contributing significantly to beef bull semen attributes. Significant SNP with a p-value of < 0.00001

Trait	SNP Name	-log$_{10}$ *p*-value	rsID	Chromosome	Position
Volume	BTB-01549373	4.01	No rsID	2	81,679,350
	Hapmap48133-BTA-96243	4.21	rs41666488	3	61,947,686
	ARS-BFGL-NGS-71827	4.26	rs109736826	3	112,997,892
	ARS-BFGL-NGS-25127	4.19	rs109268478	6	102,859,342
	Hapmap49899-BTA-18490	4.44	rs41575945	27	3,495,048
Concentration	BTB-01963898	4.03	rs43067163	1	133,936,071
	BTA-122016-no-rs	4.49	rs41623602	3	38,811,314
	Hapmap55203-rs29023737	4.12	rs29023737	5	3,645,270
Number of Spermatozoa	BTA-110980-no-rs	4.83	rs41618035	1	31,173,269
	ARS-BFGL-NGS-101891	4.28	rs109740774	6	75,137,290
	Hapmap33368-BTA-146079	4.25	rs43567728	8	89,743,053
	Hapmap44146-BTA-83959	4.07	rs41661101	9	50,922,485
	ARS-BFGL-NGS-112914	4.71	rs110005257	11	99,686,154
	ARS-BFGL-NGS-101386	4.55	rs110190516	17	67,017,094
	ARS-BFGL-NGS-85970	4.04	rs108993490	24	16,118,203
Initial Motility	Hapmap48211-BTA-120784	4.69	rs41623436	1	152,412,536
	ARS-BFGL-NGS-74920	4.54	rs109798673	2	827,626
	BTB-00319289	4.13	rs43526428	7	73,777,772
	Hapmap59148-ss46527122	4.10	rs29003479	11	63,430,172
	BTB-01751684	4.04	rs42861585	23	46,295,105
	Hapmap23061-BTC-074055	4.05	rs109512383	25	30,964,321
Post-thaw Motility	Hapmap48211-BTA-120784	4.75	rs41623436	1	152,412,536
	ARS-BFGL-NGS-74920	5.63	rs109798673	2	827,626
	Hapmap45703-BTA-105835	4.10	rs41611445	2	82,083,660
	BTB-01323835	4.22	rs42446055	4	89,708,810
	BTB-01537954	4.32	rs42653645	11	25,036,830
	ARS-BFGL-NGS-41002	4.52	rs110418161	16	36,415,574
	ARS-BFGL-NGS-111058	4.82	rs109719529	17	56,344,620
Three Hour Post-thaw Motility	Hapmap59148-ss46527122	4.50	rs29003479	11	63,430,172
	ARS-BFGL-NGS-111058	4.49	rs109719529	17	56,344,620
Percentage of Normal Spermatozoa	BTA-50285-no-rs	4.41	rs41606310	1	7,669,386
	ARS-BFGL-NGS-43775	5.94	rs110964837	2	73,209,337
	BTA-91078-no-rs	4.85	rs41594758	3	43,031,873
	ARS-BFGL-NGS-25359	4.71	rs109928164	5	70,049,126
	BTA-74480-no-rs	4.66	rs41591913	5	84,578,337
	BTA-98744-no-rs	4.68	rs41666416	10	68,089,016
Primary Abnormalities	BTA-115758-no-rs	4.11	rs41566683	1	36,903,496
	Hapmap57078-ss46526391	4.95	rs41255529	1	60,956,897
	ARS-BFGL-NGS-44433	4.22	rs110487590	1	63,799,085
	BTB-01831345	4.51	rs42940192	4	10,648,384
	ARS-BFGL-NGS-80666	4.05	rs110425782	20	68,324,872
	Hapmap42996-BTA-60916	4.16	rs41646489	22	33,690,494
Secondary Abnormalities	Hapmap42996-BTA-60916	4.28	rs41646489	22	33,690,494

For IMot, 6 SNP on 6 different chromosomes were significant, including rs41623436, rs109798673, rs43526428, rs29003479, rs42861585, and rs109512383. One SNP, rs29003479, on chromosome 11 was significant in both bivariate analyses for initial motility and the other five were only significant in one of the analyses. The SNP significant in both bivariate analyses was near a previously reported scrotal circumference QTL region [46]. Scrotal
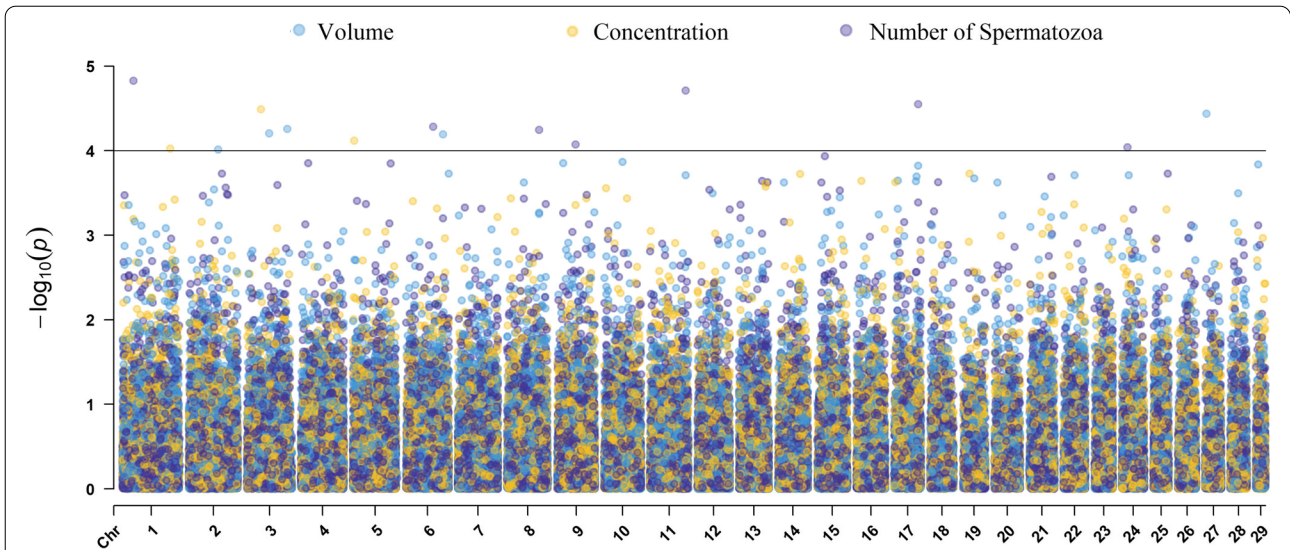
Fig. 1 Manhattan plot showing the result of genome-wide association mapping for volume, concentration, and number of spermatozoa with a significance threshold of 4.0. A trivariate analysis was performed
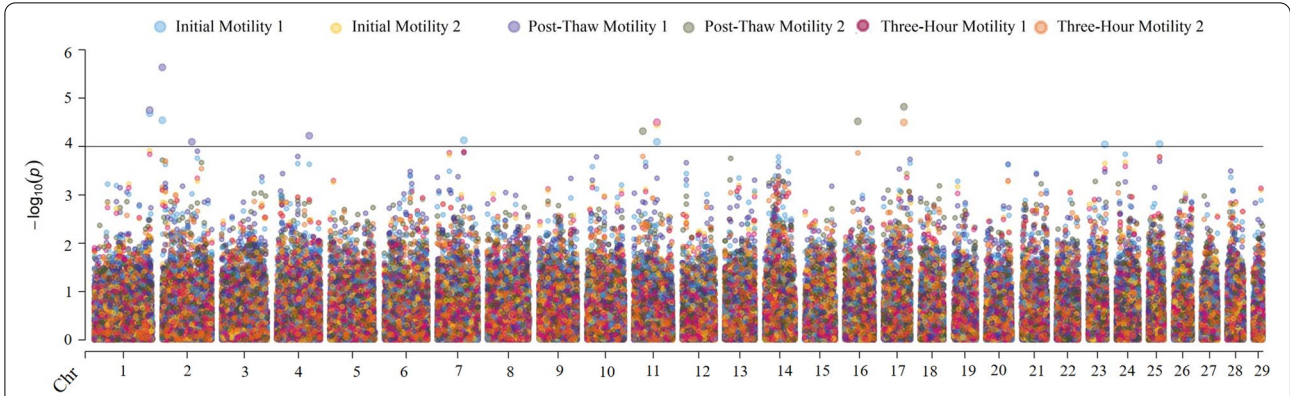

Fig. 2 Manhattan plot showing the result of genome-wide association mapping for motility traits with a significance threshold of 4.0. A bivariate analysis was performed, thus significant SNP from both bivariate analyses are reported
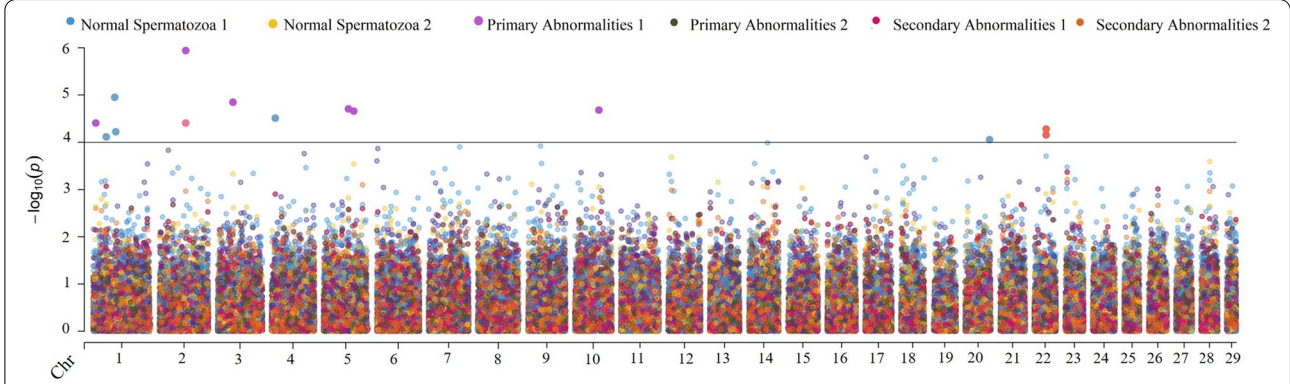

Fig. 3 Manhattan plot showing the result of genome-wide association mapping for percentage of normal spermatozoa and abnormality traits with a significance threshold of 4.0. A bivariate analysis was performed, thus significant SNP from both bivariate analyses are reported

Table 3 Previously reported fertility-associated QTL regions which overlap significant SNP identified for beef bull semen attributes in this study

Trait	SNP Name	rsID	Chr:Position	QTL Trait	Source
Volume	BTB-01549373	No rsID	2: 81679350	Conception rate	[39]
				Non-return rate	[9]
	Hapmap49899-BTA-18490	rs41575945	27:3495048	Non-return rate	[10]
Number of Spermatozoa	BTA-110980-no-rs	rs41618035	1:31173269	Conception rate	[40]
				Inseminations per conception	[11]
				Non-return rate	[11]
	ARS-BFGL-NGS-101891	rs109740774	6:75137290	First service conception	[39]
	Hapmap33368-BTA-146079	rs43567728	8: 89743053	Conception rate	[39]
	Hapmap44146-BTA-83959	rs41661101	9:50922485	Non-return rate	[41]
				Structural soundness of feet, legs, penis, and prepuce	[42]
	ARS-BFGL-NGS-101386	rs110190516	17: 67017094	Non-return rate	[43]
	ARS-BFGL-NGS-85970	rs108993490	24:16118203	First service conception rate	[39]
				Conception rate	[39]
Initial Motility	ARS-BFGL-NGS-74920	rs109798673	2: 827626	Conception rate	[39]
				Fertilization rate	[38]
	BTB-00319289	rs43526428	7:73777772	Conception rate	[39, 44]
				Daughter pregnancy rate	[21, 44]
				Sire conception rate	[45]
				Sexual precocity	[45]
	Hapmap59148-ss46527122	rs29003479	11:63430172	Scrotal circumference	[46]
	Hapmap23061-BTC-074055	rs109512383	25: 30964321	Conception rate	[39]
				Scrotal circumference	[46]
				Sperm average path velocity	[12]
Post-thaw Motility	ARS-BFGL-NGS-74920	rs109798673	2:827626	Conception Rate	[39]
				Fertilization Rate	[38]
	Hapmap45703-BTA-105835	rs41611445	2: 82083660	Conception rate	[39]
				Non-return rate	[9]
	BTB-01537954	rs42653645	11:25036830	Sperm motility	[12, 16]
Three-hour Post-thaw Motility	Hapmap59148-ss46527122	rs29003479	11: 63430172	Scrotal circumference	[46]
Percentage of Normal Spermatozoa	BTA-50285-no-rs	rs41606310	1: 7669386	Daughter pregnancy rate	[22]
	ARS-BFGL-NGS-43775	rs110964837	2:73209337	Conception rate	[12, 17, 39]
				Non-return rate	[9]
	ARS-BFGL-NGS-25359	rs109928164	5: 70049126	Daughter pregnancy rate	[17]
	BTA-74480-no-rs	rs41591913	5:84578337	Conception Rate	[17]
	BTA-98744-no-rs	rs41666416	10: 68089016	Fertility index	[23]
				Scrotal circumference	[47]

Table 3 (continued)

Trait	SNP Name	rsID	Chr:Position	QTL Trait	Source
Primary Abnormalities	BTA-115758-no-rs	rs41566683	1: 36903496	Conception rate	[39, 40]
	Hapmap57078-ss46526391	rs41255529	1:60956897	Conception rate	[39, 40]
				First service conception	[39]
				Non-return rate	[9]
				Scrotal circumference	[46]
	ARS-BFGL-NGS-44433	rs110487590	1: 63799085	Conception rate	[17, 40]
				Daughter pregnancy rate	[17]
				Non-return rate	[9]
				Scrotal circumference	[46]
				Testicular hypoplasia	[48]
	BTB-01831345	rs42940192	4:10648384	Non-return rate	[49, 50]
				Male fertility	[51]

Table 4 Significant SNP with putative candidate genes for semen production traits

SNP	Trait	Chromosome: Pos	Closest gene	Biological Process	Distance from gene (bp)	Relative Location of SNP to gene
rs109798673	Initial Motility	2:827626	HERC2	Causes reduced spermatozoa motility	0	Within
	Initial Motility	2:827626	OCA2	Contributes to spermatid development	196,400	Downstream
	Post-thaw Motility	2:827626	HERC2	Caused reduced spermatozoa motility	0	Within
	Post-thaw Motility	2:827626	OCA2	Contributes to spermatid development	196,400	Downstream
rs110425782	Primary Abnormalities	20:68324872	LOC101902976	Necessity for sperm motility	108,733	Downstream

circumference has been reported to be strongly and positively genetically correlated with motility [6, 55]. A previously reported QTL for sperm average path velocity overlapped the QTL region for rs109512383 [12]. Sperm average path velocity is a measure obtained from a computer-assisted semen analysis and represents the average trajectory of the sperm cell. The trajectory of the spermatozoa is dependent upon the ability of the spermatozoa to move, thus providing further evidence of the relationship between initial motility and sperm average path velocity. A QTL region for sire conception rate [45] was in the vicinity of rs43526428, a SNP significant for IMot (Table 3). In addition, a reported conception rate QTL region [39, 56], and in the current study, a significant SNP for IMot was close to this region. Once the spermatozoa are deposited into the female reproductive tract, it is essential for the sperm cells to progress towards the oocyte for fertilization and, therefore conception [44]. Interestingly, a QTL region found to influence sexual precocity [45] is near a significant SNP associated with IMot

in this study. This is interesting because the average age of bulls in this study was just over two years old, with the youngest bull in the evaluation just under a year old. Bull studs often collect very young bulls and producers demand the semen because these bulls often have more desirable EPD (albeit unproven and lower accuracy) than their older stud mates. Young bulls must have the libido and semen quality to produce viable sperm, so the relationship between sexual precocity and initial motility could indicate young bulls are able to produce a viable ejaculate. Quantitative trait loci associated with milk composition [13, 14] and interval to first estrus after calving [15] were found neighboring rs41623436. A recent study [25] found the same results with crossbred beef bulls, where a significant SNP window was within a QTL for interval to first estrus after calving. Return to estrus after calving is dependent on the hormones gonadotropin releasing hormone (GnRH), luteinizing hormone (LH), and follicle stimulating hormone (FSH). The production of spermatozoa is also dependent on GnRH, signaling

Butler *et al. BMC Genomics*   (2022) 23:74

Page 9 of 12

the release of LH to induce the Leydig cells to produce testosterone. Testosterone paired with FSH signaling to the Sertoli cells causes the production of male gametes. It could be speculated that females which return to estrus more quickly after calving have increased hormone signaling, which could indicate males which have significant SNP near these QTL additionally have stronger hormone signaling; and therefore; produce more and higher quality gametes.

Seven SNP were identified for PTMot (Table 3). A QTL region on chromosome 11 which reportedly influences spermatozoa motility [12, 16] encompassed rs42653645, a SNP significant for PTMot. Two different conception rate QTL regions have been reported [39] near two SNP significant for PTMot in this study. As motile spermatozoa are necessary for conception, this is a logical relationship.

Six significant SNP were identified for %NORM (Table 2). The SNP were rs41606310 on chromosome one, rs110964837 on chromosome two, rs41594758 on chromosome three, rs109928164 on chromosome five, rs41591913 on chromosome five, and rs41666416 on chromosome ten. Notably, rs110964837 was significant in both bivariate models where percentage of normal spermatozoa was evaluated with either PRIM or SEC. The SNP rs110964837 was one of the two SNP near conception-rate QTL regions and a non-return rate QTL region [9, 12, 17, 39]. Previously reported scrotal circumference QTL regions [46], and a significant SNP for %NORM in the current study was near the reported region. Two conception-rate QTL regions and a non-return rate QTL region [9, 12, 17, 39] were close to SNP significantly associated with %NORM. As previously discussed, SNP significantly associated with male fertility traits nearby fertility-related QTL regions is further evidence that the quality of a bull's semen is important for conception or that the same genes underlie attributes affecting conception in both males and females.

Primary abnormalities had six strongly associated SNP: rs41566683, rs41255529, rs110487590, rs42940192, rs110425782, and rs41646489 (Table 2). A previously reported scrotal circumference QTL region found in a population of Angus bulls [46] was near rs41255529, a SNP significant for PRIM. Previously reported genetic correlations between scrotal circumference and PRIM indicate increased scrotal circumference results in fewer PRIM [41, 44]. Three significant SNP for PRIM were in proximity to QTL regions associated with conception rate and non-return rate [39, 40, 46]. Primary abnormalities are abnormalities of the head, and without proper head formation, the sperm cell cannot penetrate the zona pellucida on the oocyte and cause conception [44].

All genes near SNP significant for IMot are outlined in Table 4. A SNP contributing significantly to IMot was within the gene E3 ubiquitin protein ligase 2 (HERC2). In addition, the significant SNP within HERC2 was also downstream of the gene melanosomal transmembrane protein (OCA2). The gene OCA2 is most commonly associated with albinism; however, it also contributes to spermatid development [36]. Reported in Table 4 are additional genes identified within the QTL regions for IMot.

A SNP significant for PTMot was in proximity to collagen like tail subunit of asymmetric acetylcholinesterase (COLQ), which contributes to the structural integrity of the extracellular matrix and heparin binding [57]. It is important to note heparin is an important enhancer of capacitation of the bovine spermatozoa [19]. Furthermore, heparin-binding proteins allow the spermatozoa to undergo the acrosome reaction [20]. In addition, a significant SNP for PTMot was in proximity to the heat shock protein family B (small) member 8 (HSPB8) protein. Elevated heat shock protein levels have been speculated to be associated with immature spermatozoa [58]. This research addressed human infertility and found that an increase in the 70-kDa heat shock protein HspA2 resulted in spermatozoa not completing the phases of changing the plasma membrane during epididymal maturation and having larger amounts of cytoplasmic proteins in mature spermatozoa [58]. As previously discussed, heat shock protein can affect the maturation of the spermatozoa and cause infertility [58]. Reported in Table 4 are additional genes identified near SNP affecting PTMot.

Although not within ±250,000 kilobases of the significant SNP, a pertinent gene nearby was the major facilitator superfamily domain containing 14A (MFSD14A) gene. Also known as HIAT1, it was 300,000 kilobases upstream of the SNP. The MFSD14A gene is important for spermatid nucleus differentiation and sperm mitochondrion organization. It has been reported that mice with a disruption to the MFSD14A gene are infertile due to incomplete acrosome formation and round head defects [59].

Primary abnormalities had a significant SNP in the vicinity of a gene on chromosome 20 (Table 4). Primary abnormalities had a significant SNP within sperm mitochondrial-associated cysteine-rich protein-like (LOC101902976). Sperm-mitochondrial cysteine-rich protein is located in the mitochondrial capsule and plays a role in sperm motility [42]. While the mitochondria are located in the midpiece of the spermatozoa rather than the head, it is still important to note it is associated with male fertility.

Butler *et al. BMC Genomics*   (2022) 23:74

Page 10 of 12

Conclusion

Several QTL associated with fertility traits were identified in this study and validated by previously published literature. In addition, genes within the QTL regions for beef bull semen attributes were discussed. While a few gene functions could be associated with beef bull semen attributes, some of the genes found in the current study do not have any known association with fertility traits reported in previous literature and have unknown related physiological association. Results from the current study, paired with findings in previous literature, validate that bull fertility traits are controlled by genetic factors. Identification of the QTL regions and SNP positions associated with these traits provides the knowledge necessary to enrich these regions in future iterations of SNP chips used within the cattle industries, and thus increase genomic prediction accuracies for these traits should the industry add them to routine genetic evaluation. The increased use of genomic testing in tandem with the number of phenotypes routinely recorded by bull semen collection facilities makes identification of beef bulls with superior genetic merit for fertility traits feasible. This advancement would increase profitability of beef producers by improving fertilization rate, increasing overall herd reproductive rate, and providing producers another genetic tool to use in making selection decisions.

Authors' contributions
MB perfomred the analyses, wrote the main text and prepared the figures. MR, JB, and RW supervised the work and assisted MB with statistical analyses. DG and AH provided reproductive physiology and semen collection procedures expertise. All authors read and approved the final manuscript.

Availability of data and materials
The data that support the findings of this study are available from the American Angus Association and corresponding AI collection centers but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are available from the original data providers and/or the authors upon reasonable request and only with written permission of the original data providers and only in cases where this is consistent with the rules set forth in the data transfer agreement signed by both parties.

Declarations

Ethics approval and consent to participate
Study was undertaken with data owned by the American Angus Association and they consented to the work included in the manuscript.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Bull management to maximize sperm output. In: Proceedings of the Sixth Technical Conference on Artificial Insemination and Reproduction. Milwaukie; 1976. p. 1–10. https://agris.fao.org/agris-search/search.do?recordID=US201303059742. AGRICULTURAL SCIENCE AND TECHNOLOGY INFORMATION. agris.fao.org.
2. Chenoweth PJ. Sexual behavior of the bull: a review. J of Dairy Sci. 1983;66:173–9. https://doi.org/10.3168/jds.S0022-0302(83)81770-6.
3. Fuerst-Waltl B, Schwarzenbacher H, Perner C, Solkner J. Effects of age and environmental factors on semen production and semen quality of Austrian Simmental bulls. Anim. Repro. Sci. 2006;95:27–37. https://doi.org/10.1016/j.anireprosci.2005.09.002.
4. Brito LC, Barth AD, Rawlings NC, Wilde RE, Crews DH Jr, Mire PS, et al. Effect of improved nutrition during calfhood on serum metabolic hormones, gonadotropins, and testosterone concentrations, and on testicular development in bulls. Domestic Anim Endocrinol. 2006;33:460–9. https://doi.org/10.1016/j.domaniend.2006.09.004.
5. Coulter GH, Cook RB, Kastelic JP. Effects of dietary energy on scrotal surface temperature, seminal quality, and sperm production in young beef bulls. J Anim Sci. 1997;75:1048–52. https://doi.org/10.2527/1997.7541048x.
6. Corbet NJ, Burns DM, Johnston DJ, Wolcott ML, Corbet DH, Venus BK, et al. Male traits and herd reproductive capability in tropical beef cattle. 2. Genetic parameters of bull traits. Anim Prod Sci. 2013;53:101–13. https://doi.org/10.1071/AN12163.
7. Huang JZ, Li Q, Hou Q, Wang C, Li J, et al. Solexa sequencing of novel and differentially expressed microRNAs in testicular and ovarian tissues in Holstein cattle. Int J Biol Sci. 2011;7:1016–26. https://doi.org/10.7150/ijbs.7.1016.
8. Taylor JF, Schnabel RD, Sutovsky P. Identification of genomic variants causing sperm abnormalities and reduced male fertility. Anim Repro Sci. 2018;194:57–62.
9. Ben Jemaa S, Fritz S, Guillaume F, Druet T, Denis C, Eggen A, et al. Detection of quantitative trait loci affecting non-return rate in French dairy cattle. J Anim Breed Genet. 2008;125:280–8. https://doi.org/10.1111/j.1439-0388.2008.00744.x.
10. Ron W, Feldmesser E, Golik M, Tager-Cohen I, Kliger D, Reiss V, et al. A complete genome scan of the Israeli Holstein population for quantitative trait loci by a daughter design. J Dairy Sci. 2004;2:476–90. https://doi.org/10.3168/jds.S0022-0302(04)73187-2.
11. Hoglund JK, Buitenhuis AJ, Guldbrandtsen B, Su G, Thomsen B, Lund MS. Overlapping chromosomal regions for fertility traits and production traits in the Danish Holstein population. J Dairy Sci. 2009;92:5712–9. https://doi.org/10.3168/jds.2008-1964.
12. Druet T, Fritz S, Sellem E, Basso B, Gerard O, Salas-Cortes L, et al. Estimation of genetic parameters and genome scan for 15 semen characteristics traits of Holstein bulls. J Anim Breed Genet. 2009;126:269–77. https://doi.org/10.1111/j.1439-0388.2008.00788.x.
13. Costa A, Schwarzenbacher H, Mészáros G, Fuerst-Waltl B, Fuerst C, Sölkner J, et al. On the genomic regions associated with milk lactose in Fleckvieh cattle. J Dairy Sci. 2019. https://doi.org/10.3168/jds.2019-16663.
14. Bouwman AC, Visker MH, van Arendonk JA, Bovenhuis H. Genomic regions associated with bovine milk fatty acids in both summer and winter milk samples. BMC Genet. 2012;13:93.
15. Zhang Z, Kargo M, Liu A, Thomasen JR, Pan Y, Su G. Genotype-by-enviroment interaction of fertility traits in Danish Holstein cattle using a single-step genomic reaction norm model. Heredity. 2019;123:202–14. https://doi.org/10.1038/s41437-019-0192-4.

16. Hering DM, Olenski K, Kaminski S. Genome-wide association study for poor sperm motility in Holstein-Friesian bulls. Anim Reprod Sci. 2014;146:89–97. https://doi.org/10.1016/j.anireprosci.2014.01.012.

17. Jiang J, Ma L, Prakapenka D, VanRaden PM, Cole JB, Da Y. A large-scale genome-wide association study in U.S. Holstein cattle. Front Genet. 2019;10:412. https://doi.org/10.3389/fgene.2019.00412.

18. Westfalewicz B, Dietrich MA, Mostek A, Partyka A, Bielas W, Nizanski W, et al. Identification and functional analysis of bull (Bos taurus) cauda epididymal fluid proteome. J Dairy Sci. 2017;100:6707–19. https://doi.org/10.3168/jds.2016-12526.

19. Handrow RR, Lenz RW, Ax RL. Structural comparisons among glycosaminoglycans to promote an acrosome reaction in bovine spermatozoa. Biochem Biophys Res Commun. 1982;107:1326. https://doi.org/10.1016/S0006-291X(82)80143-5.

20. Lane ME, Therien I, Moreau R, Manjunath P. Heparin and high density lipoprotein mediate bovine sperm capacitation by different mechanisms. Biol Reprod. 1999;60:169–75. https://doi.org/10.1095/biolreprod60.1.169.

21. Frischknecht M, Bapst B, Seefried FR, Signer-Hasler H, Garrick D, Stricker C, et al. Genome-wide association studies of fertility and calving traits in Brown Swiss cattle using imputed whole-genome sequences. BMC Genomics. 2017;18:910. https://doi.org/10.1186/s12864-017-4308-z.

22. Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ Jr, Crooker BA, et al. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics. 2011;12:408. https://doi.org/10.1186/1471-2164-12-408.

23. Schnabel RD, Sonstegard TS, Taylor JF, Ashwell MS. Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. Anim Genet. 2005;36:408–16. https://doi.org/10.1111/j.1365-2052.2005.01337.x.

24. Cai Z, Guldbrandtsen B, Lund MS, Sahana G. Prioritizing candidate genes for fertility in dairy cows using gene-based analysis, functional annotation and differential gene expression. BMC Genomics. 2019;20:255 doi: s12864-019-5638-9.

25. Sweett H, Fonseca PAS, Suárez-Vega A, Livernois A, Miglior F, Canovas A. Genome-wide association study to identify genomic regions and positional candidate genes associated with male fertility in beef cattle. Sci Rep. 2020;10:20102. https://doi.org/10.1038/s41598-020-75758-3.

26. Butler ML, Hartman AR, Bormann JM, Weaber RL, Grieger DM, Rolf MM. Genetic parameter estimation for beef bull semen attributes. J Anim Sci. 2021;99(2):skab013. https://doi.org/10.1093/jas/skab013. https://pubmed.ncbi.nlm.nih.gov/33453111/.

27. Misztal I, Tsuruta S, Laurenco DAL, Masuda Y, Aguilar I, Legarra A, et al. Manual for BLUPF90 family programs: University of Georgia; 2016. http://nce.ads.uga.edu/wiki/doku.php?id=documentation

28. Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet Res. 2012;94:73–83. https://doi.org/10.1017/S0016672312000274.

29. Masuda, Y. Introduction to BLUPF90 suite programs Concise version. University of Georgia. April 2019. Pp. 180–181. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=tutorial_blupf90.pdf. Revision History March 2018 (0.8.0): The first revision prepared for the summer course at UGA. Thanks to Andrés Legarra. April 2019 (0.9.0): Revised to correct some errors and typos. nce.ads.uga.edu.

30. Aguilar I, Legarra A, Cardoso F, Masuda Y, Lourenco D, Misztal I. Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. Genet Sel Evol. 2019511–8.

31. Garrick DJ. Equivalent mixed model equations for genomic selection. J Anim Sci. 2007;85:376.

32. Yin L. Circular manhattan plot. R package version 3.6.0. 2020. https://github.com/YinLiLin/R-CMplot

33. Storey JD, Bass AJ, Dabney A, Robinson D. qvalue: Q-value estimation for false discovery rate control. R package version 2.20.0. 2020. http://github.com/jdstorey/qvalue.

34. McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, et al. Whole genome linkage disequilibrium maps in cattle. BMC Genet. 2007;8:74.

35. Hu Z, Park CA, Reecey JM. Building a livestock genetic and genomic information knowledgebase through integrative developments of animal ATLdb and CorrDB. Nucleic Acids Res. 2019;47:D701–10. https://doi.org/10.1093/nar/gky1084.

36. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47:D506–15.

37. Huang DW, Sherman BT, Lempicki LA. Bioinformatics enrichment tools:paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37:1–13. https://doi.org/10.1093/nar/gkn923.

38. Huang W, Kirkpatrick BW, Rosa GJ, Khatib H. A genome-wide association study using selective DNA pooling identifies candidate markers for fertility in Holstein cattle. Anim Genet. 2009;41:570–8. https://doi.org/10.1111/j.1365-2052.2010.02046.x.

39. Kiser JN, Keuter EM, Seabury CM, Neupane M, Moraes JGN, Dalton J, et al. Validation of 46 loci associated with female fertility traits in cattle. BMC Genomics. 2019;20:576. https://doi.org/10.1186/s12864-019-5935-3.

40. Boichard D, Grohs C, Bourgeois F, Cerqueira F, Faugeras R, Neau A, et al. Detection of genes influencing economic traits in three French dairy cattle breeds. Genetics. 2003;1:77–101. https://doi.org/10.1186/1297-9686-35-1-77.

41. Schrooten C, Bink MCAM, Bovenhuis H. Whole genome scan to detect chromosomal regions affecting multiple traits in dairy cattle. J Dairy Sci. 2004;87:3550–60. https://doi.org/10.3168/jds.S0022-0302(04)73492-X.

42. Hawthorne SK, Goodarzi G, Bagarova J, Gallant KE, Busanelli RR, Olend WJ, et al. Comparative genomics of the sperm mitochondria-associated cysteine-rich protein gene. Genomics. 2006;87:382–91. https://doi.org/10.1016/j.ygeno.2005.09.010.

43. Ashwell MS, Heyen DW, Weller JI, Ron M, Sonstegard TS, Van Tassell CS, et al. Detection of quantitative trait loci influencing conformation traits and calving ease in Holstein-Friesian cattle. J Dairy Sci. 2005;88:4111–9. https://doi.org/10.3168/jds.S0022-0302(05)73095-2.

44. Senger PL. Pathways to pregnancy and parturition. 3rd ed. Redmon, Oregon: Current Conceptions, Inc.; 2012.

45. Melo TP, Fortes MRS, Bresolin T, Mota LFM, Albuquerque LG, Carvalheiro R. Multitrait meta-analysis identified genomic regions associated with sexual precocity in tropical beef cattle. J Anim Sci. 2018;96:4087–99. https://doi.org/10.1093/jas/sky289.

46. McClure MC, Morsci NS, Schnabel RD, Kim JW, Yao P, Rolf MM, et al. A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. Anim Genet. 2010;4:597–607. https://doi.org/10.1111/j.1365-2052.2010.02063.x.

47. Meyer K, Hammond K, Mackinnon MJ, Parnell PF. Estimates of covariances between reproduction and growth in Australian beef cattle. J Anim Sci. 1991;69:3533–43.

48. Neves H, Vargas G, Brito LF, Schenkel FS, Albuquerque LG, Carvalheiro R. Genetics and genomic analyses of testicular hypoplasia in Nellore cattle. PLoS One. 2019;14:e0211159. https://doi.org/10.1371/journal.pone.0211159.

49. Schulman NF, Sahana G, Iso-Touru T, McKay SD, Schnabel RD, Lund MS, et al. Mapping of fertility traits in Finnish Ayrshire by genome-wide association analysis. Anim Genet. 2011;42:263–9. https://doi.org/10.1111/j.1365-2052.2010.02149.x.

50. Pimentel EC, Bauersachs S, Tietze M, Simianer H, Tetens J, Thaller G, et al. Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. Anim Genet. 2011;42:251–62. https://doi.org/10.1111/j.1365-2052.2010.02148.x.

51. Blaschek M, Kaya A, Zwald N, Memili E, Kirkpatrick BW. A whole-genome association analysis of noncompensatory fertility in Holstein bulls. J Dairy Sci. 2011;94:4695–9. https://doi.org/10.3168/jds.2010-3728.

52. DeJarnette JM, Marshal CE, Lenz RW, Monke DR, Ayars WH, Sattler CG. Sustaining the fertility of artificially inseminated dairy cattle: the role of the artificial insemination industry. J Dairy Sci. 2004;87(E. Suppl):E93–E104. https://doi.org/10.3168/jds.S0022-0302(04)70065-X.

53. Hansen M. Genetic investigations on male and female fertility in cattle. Livest Sci. 1979;6:325–34.

54. Syrstad O. Selection for fertility in the basis of AI data. Livest Prod Sci. 1981;8:247–52.

55. Christmas RA, Moser DW, Spire MF, Sargeant JM, Tucker SK. Genetic relationships among breeding soundness traits in yearling bulls. In: Cattlemen's Day. Manhattan; 2001. p. 1–3. https://newprairiepress.org/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1712&context=kaesrr.

56. Zhou Y, Connor EE, Wiggans GR, Lu Y, Tempelman RJ, Schroeder SG, et al. Genome-wide copy number variant analysis reveals variants associated with 10 diverse production traits in Holstein cattle. BMC Genomics. 2018;19:314. https://doi.org/10.1186/s12864-018-4699-5.

57. Deprez P, Inestrosa NC, Krejci E. Two different heparin-binding domains in th triple-helical domain of ColA, the collegen tail subunit of synaptic acetylcholinesterase. J Bio Chem. 2003;278:23233–42. https://doi.org/10.1074/jbc.M301384200.

58. Huszar G, Patrizio P, Vigue L, Willets M, Wilker C, Adhoot D, et al. Cytoplasmic extrusion and the switch from creatine kinase B to M isoform are completed by commencement of epididymal transport in human and stallion spermatozoa. J Androl. 1998;19:11–20. https://doi.org/10.1002/j.1939-4640.1998.tb02465.x.

59. Doran J, Walters C, Kyle V, Wooding P, Hammett-Burke R, College B. Mfsd14a (Hiat1) gene disruption causes globozoospermia and infertility in male mice. Repro. 2016;152:91–9. https://doi.org/10.1530/REP-15-0557.

Publisher's Note