



## A genome-wide association study for clinical mastitis in the dual-purpose German Black Pied cattle breed

Saskia Meier,  Danny Arends,  Paula Korcuć,  Guilherme B. Neumann,  and Gudrun A. Brockmann\* 

Albrecht Daniel Thaer-Institute for Agricultural and Horticultural Sciences, Animal Breeding Biology and Molecular Genetics, Humboldt-Universität zu Berlin, Invalidenstrasse 42, 10115 Berlin, Germany

### ABSTRACT

The dual-purpose German Black Pied Cattle (DSN) has become an endangered breed of approximately 2,550 registered cows in Germany. The breed is genetically related to Holstein-Friesian cattle because the old DSN breed contributed to the selection of the modern Holstein dairy cow. In dairy farms, breeders aim to improve animal health and well-being by reducing the number of mastitis cases, which would also reduce milk losses and treatment costs. On the genomic level, no markers associated with clinical mastitis have been reported in DSN. Therefore, we performed a genome-wide association study on 1,062 DSN cows using a univariate linear mixed model that included a relatedness matrix to correct for population stratification. Although the statistical power was limited by the small population size, 3 markers were significantly associated, and 2 additional markers showed a suggestive association with clinical mastitis. Those markers accounted for 1 to 3% of the variance of clinical mastitis in the examined DSN population. One marker was found in the intragenic region of *NEURL1* on BTA26, and the other 4 markers in intergenic regions on BTA3, BTA6, and BTA9. Further analyses identified 23 positional candidate genes. Among them is *BMPRI1B*, which has been previously associated with clinical mastitis in other dairy cattle breeds. The markers presented here can be used for selection for mastitis-resistant animals in the endangered DSN population, and can broadly contribute to a better understanding of mastitis determinants in dairy cattle breeds.

**Key words:** inflammatory udder disease, bovine clinical mastitis, genomic selection, endangered breed

### INTRODUCTION

The dual-purpose German Black Pied cattle (DSN; Deutsches Schwarzbuntes Niederungs-rind) originated from the Dutch and German North Sea region. The DSN ancestors are considered to have a large genetic contribution to the high-yielding Holstein-Friesian breed (Köppe-Forsthoff, 1967; Grothe, 1993). Because of their advantage in milk production, cattle from that region spread across northern Germany and accompanied emigrants to North America, where the population was further selected for high milk performance. With the reimport of the high-yielding Holstein-Friesians from North America to Europe about a century later in the 1960s, the original dual-purpose DSN breed was almost entirely replaced by Holsteins. For fear of losing the original breed, a genetic reserve of approximately 4,000 herd-book DSN cows was established in the former German Democratic Republic in 1973. Because this genetic reserve was continued without crossbreeding with Jersey or Holstein-Friesian breeds, it captures what is described to be the original North Sea population. After the reunification of Germany in 1990, the genetic reserve was maintained, although the population size started to decline. Today, approximately 2,550 purebred (less than 10% Holstein-Friesian in pedigree) DSN cows are registered, mainly in eastern Germany (Bundesverband Rind und Schwein e.V., 2019).

Albeit the fact that DSN cows produce a lower milk yield (about –2,500 kg/lactation) in comparison to Holstein cows, the breed also possesses some advantageous properties: DSN cows produce higher milk fat and protein contents (4.3% fat and 3.7% protein) and are said to be more fertile, with shorter calving intervals (RBB Rinderproduktion Berlin-Brandenburg GmbH, 2016).

Similar to other dairy cattle breeds, mastitis is a common disease in DSN. Mastitis impairs animal welfare and leads to economic losses as a consequence of lower milk production (about 5% loss per lactation), treatment costs, and losses through premature culling (Seegers et al., 2003). The disease is multifactorial

Received January 16, 2020.

Accepted June 29, 2020.

\*Corresponding author: [gudrun.brockmann@agrar.hu-berlin.de](mailto:gudrun.brockmann@agrar.hu-berlin.de)

and caused by environmental or contagious pathogens. Clinical mastitis (CM) is the inflammation of one or more udder quarters that accompanies clots in the milk. The mastitis incidence in dairy cattle is high. Studies in Canadian herds (93% Canadian Holstein cattle) showed that about 23% of all animals over all lactations had CM (Olde Riekerink et al., 2008), and in Scandinavian herds (mainly Holstein and Red Dairy Cattle), the incidence increased from 12.5% in the first lactation to 26% in the third lactation (Negussie et al., 2010).

Reduction of CM through the improvement of herd management, husbandry, and genetics is a major goal in dairy production. Nevertheless, due to the polygenic characteristic of CM and its low heritability, reducing CM through breeding is a goal, but remains challenging (Sender et al., 2013).

The current breeding program for DSN sires is based on daughter performance data and considers milk recording data and conformation classification. The DSN breeding values include SCS, which has a moderate to high correlation ( $\sim 0.60$ – $0.70$ ) with CM (Rupp and Boichard, 2000; Koeck et al., 2014). Although SCS is correlated with CM, direct selection for genetic markers associated with CM in DSN is desirable to sustainably reduce mastitis incidence in this population. Therefore, the aim of this study was to identify genetic markers linked to CM incidence in DSN. Additionally, the results obtained in this study may be informative for other dairy cattle breeds (e.g., Holstein). However, independent validation in other breeds is necessary.

## MATERIALS AND METHODS

### *Ethics Approval and Consent to Participate*

Ethical review and approval were not required for the animal study because samples were collected based on routine procedures on these farm animals. Ear tags for DNA preparation were taken as part of the required registration procedure.

### *Phenotype Data*

The investigation was carried out on 1,062 purebred DSN cows that were born between 2005 and 2014. Most of them (906 cows) had finished at least 3 lactations. The cows descended from 48 DSN sires. They were kept in the 2 largest DSN farms, which are located in the south of the state Brandenburg, Germany. One farm kept DSN only and had a size of about 800 cows. On the other farm, DSN and Holsteins were housed together with an average herd size of about 300 cows. In both herds, animals were housed in freestall barns. The

1,062 cows represented about two-thirds of the DSN population in Brandenburg. Animals from other farms were not included in this study due to differences in health data recording or small herd size.

The 2 farms recorded individual health data and stored them in the database of the herd management software tool HERDE (version 5.10, dsp-Agrosoft GmbH, <http://www.herde-net.de>) using the ICAR health codes for cattle (ICAR, 2015; Stock and Egger-Danner, 2017). Records for CM from lactations 1 to 3 were analyzed (CM1, CM2, CM3). Each total lactation was additionally split into an early ( $\leq 50$  d postpartum; Lund et al., 1999) and a late ( $> 50$  d postpartum) lactation phase to cope with the specific metabolic status cows undergo at the beginning of a lactation (e.g., CM1\_total, CM1\_early, CM1\_late). All CM cases (inflammation of the mammary gland) were selected from the database. The same diagnosis key recorded within 14 d after the first record is summarized as a single mastitis event (Martin, 2012; Abdel-Shafy et al., 2018). The occurrence of CM was flattened into a binary variable, where 1 meant that mastitis was present, and 0 meant mastitis was absent during the period under consideration.

### *Genotype Data and Quality Control*

Cows were genotyped using the BovineSNP50 version 3 BeadChip (Illumina Inc., San Diego, CA). The call rates for SNP per sample and genotypes per SNP were at least 90%. Genomic positions and the reference genome were obtained from Ensembl Release 94 (Zerbino et al., 2018) based on the UMD3.1 genome assembly (Zimin et al., 2009). We included SNP from all autosomes and the X chromosome. Because only cows were genotyped, the Y chromosome could not be considered. Also, mitochondria were not included. The SNP not segregating in our population were dropped ( $n = 1,082$ ). Per marker, a minimum of 2 genotype groups with a minimum of 10 observations and a minor allele frequency of at least 0.01 were required. All markers that were in 100% linkage ( $n = 195$ ) were removed, keeping only the first one in the running order of the chromosome. After quality control, a total of 38,224 high confident SNP markers were used for further statistical analysis. We used the statistic software R (R Core Team, 2018) for data preparation and plotting.

### *Genome-Wide Association Study*

To perform the genome-wide association study we fitted a univariate linear mixed model with the software GEMMA (version 0.98; Zhou and Stephens, 2014). The univariate linear mixed model was calculated for each

**Table 1.** *P*-values from chi-squared tests for significant effects of environmental influence (covariates) on clinical mastitis<sup>1</sup> for lactations 1 to 3

Covariate	CM1			CM2			CM3		
	Total	Early	Late	Total	Early	Late	Total	Early	Late
Herd	0.10	0.77	0.13	0.16	0.74	0.07 <sup>2</sup>	0.24	0.52	0.48
Calving season	0.05 <sup>2</sup>	0.07 <sup>2</sup>	0.48	0.38	0.02 <sup>2</sup>	0.44	0.00 <sup>2</sup>	0.10 <sup>2</sup>	0.01 <sup>2</sup>
Calving year	0.54	0.13	0.98	0.11	0.72	0.04 <sup>2</sup>	0.67	0.97	0.41

<sup>1</sup>Clinical mastitis in lactation 1 to 3 (CM1, CM2, CM3) is given in total and their early ( $\leq 50$  d postpartum) and late ( $> 50$  d postpartum) phases.

<sup>2</sup>Suggestive ( $P < 0.1$ ) and significant ( $P < 0.05$ ) covariates are used in the univariate linear mixed model.

binary mastitis phenotype (CM 1, 2, 3; total, early, late lactation) separately and checked for significance using Wald test. To account for the population stratification, the standardized relatedness matrix calculated by GEMMA was included. Covariate effects for herd ( $n = 2$ ), calving season ( $n = 4$ ), calving year ( $n = 11$ ), and birth year ( $n = 10$ ) were estimated using a linear mixed model and a chi-squared test for each mastitis phenotype. Only covariates that were at least suggestively significant ( $P$ -value  $< 0.1$ ) were included in the final model (Table 1). Numbers of animals, SNP, and covariates included in the model per phenotype were summarized in Supplemental Table S1 (<https://doi.org/10.3168/jds.2020-18209>). Variance component estimation of proportion of variance in phenotypes explained (PVE; by sparse  $X\beta$  and random effects  $u$ ) with relatedness matrix  $G_S$  for all SNP markers was calculated using Haseman-Elston regression (Sofer, 2017).

### Multiple Testing and Visualization

The number of independent tests was calculated using the simpleM method (Gao et al., 2008) with a block size of 3,000 to account for linkage between markers. After determining the number of independent tests ( $M_{eff} = 14,441$ ), this number was used in Bonferroni correction to adjust the significance threshold. The suggestive level was set to  $\alpha < 0.1$  and the significance level was set to  $\alpha < 0.05$ . The  $P$ -values were converted into  $-\log_{10}(P\text{-value})$ , which led to a threshold of 5.16 when  $\alpha = 0.1$  and 5.46 when  $\alpha = 0.05$ . The ggplot2 (Wickham, 2016) and CMplot (Yin, 2019) packages in R were used for visualization of the resulting association data.

### Positional Candidate Genes

A linkage disequilibrium decay analysis was performed and strong linkage ( $r^2 > 0.6$ ) of markers within 250 kb was observed (Supplemental Figure S1, <https://doi.org/10.3168/jds.2020-18209>). Therefore, genes within  $\pm 250$  kb up- and downstream of significant marker positions were considered as positional candidate genes

and were retrieved using the biomaRt package (Durinck et al., 2005, 2009) based on the *Bos taurus* UMD3.1 assembly. If no gene was located in the 250 kb region, the genes closest to the significant marker were investigated as positional candidate genes.

Positional candidate genes were analyzed for gene ontology (GO) term enrichment using BiNGO 3.0.3 and DAVID 6.8 (Maere et al., 2005; Huang et al., 2009) with their default settings. Bonferroni correction ( $P < 0.05$ ) after hypergeometric testing was used to define significantly overrepresented categories.

## RESULTS AND DISCUSSION

### Mastitis Prevalence

In total, 489 out of 1,062 cows had at least one CM event until the third lactation and 573 cows had never had a CM event recorded within the first 3 lactations. The highest occurrence of CM cases was observed right at the beginning of each of the 3 lactations (Figure 1). This is consistent with observations in Swedish Holstein cattle (Svensson et al., 2006; Nyman et al., 2007; Persson Waller et al., 2009). In our study, the mastitis prevalence was lowest in the early (7.3%) and late (8.3%) phases of the first lactation, increased in lactation 2 (CM2\_early: 7.6%; CM2\_late: 15.9%), and held about the same level in lactation 3 (CM3\_early: 7.9%; CM3\_late: 15.6; Supplemental Table S2, <https://doi.org/10.3168/jds.2020-18209>). On average, DSN cows had a mastitis prevalence of 15.6% in the first lactation (CM1\_total), 25.0% in the second (CM2\_total), and 25.9% in the third (CM3\_total). These frequencies are consistent with reports on other dairy breeds, including the increase of mastitis prevalence from the first to later lactations (Olde Riekerink et al., 2008; Negussie et al., 2010).

### Correlation Between Total Lactations and Their Respective Phases

The SNP effects ( $\beta$ ) from the genome-wide association study for CM in total lactations 1 to 3 (CM1,

**Table 2.** Spearman's correlation coefficients ( $r_s$ ) between SNP effects ( $\beta$ ) from the genome-wide association study for total clinical mastitis in total lactations 1 to 3 (CM1, CM2, CM3) and clinical mastitis cases in their respective early and late lactation phases (early, late)

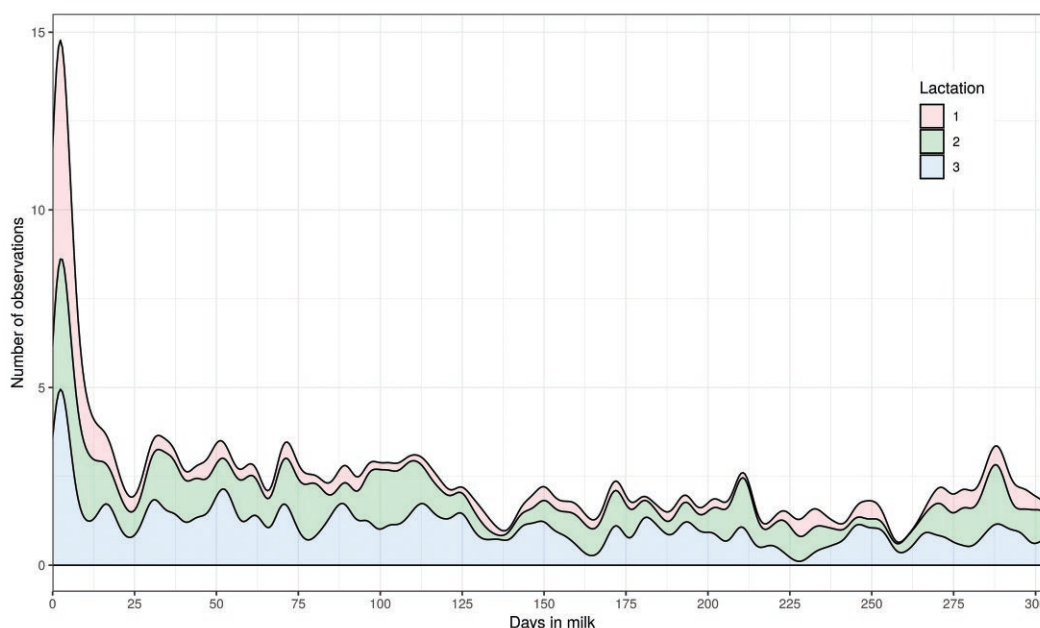
Phenotype	Stage	$r_s$
CM1_total	CM1_early	0.64
	CM1_late	0.76
CM2_total	CM2_early	0.48
	CM2_late	0.81
CM3_total	CM3_early	0.53
	CM3_late	0.83

CM2, CM3) and CM cases in their respective early and late lactation phases correlated (Table 2). In lactation 2, moderate correlation ( $r_s = 0.48$ ) was found between SNP effects for clinical mastitis cases in the total lactation and the early phase (CM2\_total to CM2\_early). Moderate to strong correlation of SNP effects ranging from 0.53 to 0.83 was calculated for the total lactations 1 and 3 and their respective early and late lactation phases. The SNP effects for mastitis in the late lactation phase explained much more of the variance of the total lactation (shown in higher  $r_s$  values of 0.76, 0.81, and 0.83 for CM1\_late, CM2\_late, and CM3\_late, respectively) compared with the early phase (0.64, 0.48, and 0.53 for CM1\_early, CM2\_early, and CM3\_early, respectively). The early phase of lactation represented only 50 d, while the late phase represented all of the time after the first 50 d, which is much longer.

### Markers for CM in DSN

Over all lactations, 3 markers that significantly ( $P < 0.05$ ) correlated with CM were detected on BTA3, BTA9, and BTA26. Two markers on BTA3 and BTA6 were suggestive ( $P < 0.1$ ; Table 3). The widespread location of markers for CM over 4 chromosomes confirmed the polygenic character of the disease (Sahana et al., 2014). The markers were associated with CM in the first and second lactations, in particular with the phenotypes CM1\_total, CM1\_late, and CM2\_early (Figure 2).

The SNP rs42092404 is located in intron 1 of the neutralized E3 ubiquitin protein ligase 1 (*NEURL1*) gene on BTA26. This marker reached the highest significance level for this study [ $-\log_{10}(P) = 6.38$ ] for CM2\_early. The mastitis prevalence increased from cows with genotype T/T (6.6%) to cows with the heterozygous genotype T/C (24.6%). The frequency of the minor and disadvantageous allele C was 0.03. Therefore, the homozygous genotype C/C was missing in our investigated population. Thus far, no association between *NEURL1* and CM or the immune system of mammals has been reported. *NEURL1* has been associated with fat content in Nordic cattle breeds (Iso-Touru et al., 2016). The genes *INA*, *PCGF6*, *TAF5*, *ATP5MD*, *PDCD11*, *CALHM3*, *CALHM2*, *CALHM1*, and *SH3PXD2A* are located in a region of  $\pm 250$  kb around the associated SNP and could poten-



**Figure 1.** Number of first clinical mastitis diagnoses in German Black Pied cattle within 305 DIM in lactations 1 to 3. In all lactations, the highest number of clinical mastitis cases was diagnosed postpartum.



**Table 3.** Marker for clinical mastitis in German Black Pied cattle in the specific lactation (total) or phase (early or late); chromosome (Chr), position, and reference SNP identification number (rs-ID) of the marker are given, as well as the reference (Ref) and alternative (Alt) alleles on the forward strand and the minor allele frequency (MAF)<sup>1</sup>

Chr	Position, bp (rs-ID)	Allele, Ref/Alt (MAF)	Mastitis prevalence, <sup>1</sup> % (n)				Positional candidate gene <sup>2</sup>	-log <sub>10</sub> (P-value)	β <sup>3</sup>	SE (β)	Phenotype
			AA	AB	BB	BB					
3	104,874,970 (rs29026738)	A <sup>1</sup> /G	13.8 (975)	32.1 (84)		<i>GUCA2A, HIVEP3, EDN2</i>	5.21†	0.37	0.08	CM1_total	
3	106,963,577 (rs110237486)	T <sup>1</sup> /G	13.9 (980)	33.8 (80)		<i>MFSD2A, MYCL, TRIT1, PPIE, HPCAL4, NT5C1A, HEYL, PABPC4, MACF1</i>	5.78*	0.40	0.08	CM1_total	
6	31,067,604 (rs29012331)	C/T <sup>4</sup>	21.5 (107)	8.8 (487)	5.9 (460)		5.27†	-0.12	0.03	CM1_late	
9	3,115,242 (rs43581241)	A <sup>1</sup> /C	13.0 (852)	23.0 (191)	53.8 (13)		6.04*	0.25	0.05	CM1_total	
26	24,337,880 (rs42092404)	T <sup>1</sup> /C	6.6 (971)	24.6 (59)		<i>NEURL1, INA, PCGF6, TAF5, ATP5MD, PDCD11, CALHM3, CALHM2, CALHMI, SH3PXD2A</i>	6.38**	0.36	0.05	CM2_early	

<sup>1</sup>Mastitis prevalence for homozygous reference allele genotype (AA), heterozygous genotype (AB), and homozygous alternative allele genotype (BB) are calculated and number of cows within the population with the respective genotype group is given (n).

<sup>2</sup>Positional candidate genes are mainly within 0.25 Mb up- and downstream. The marker position is intronic for *GUCA2A* and *NEURL1*. Genes have been previously identified as immune or mastitis related genes include *HIVEP3*, *EDN2*, *PPIE*, and *BMPR1B*.

<sup>3</sup>β = the effect size of the marker.

<sup>4</sup>The protective allele.

†α < 0.10, \*α < 0.05, \*\*α < 0.01 for -log<sub>10</sub> (P-value).

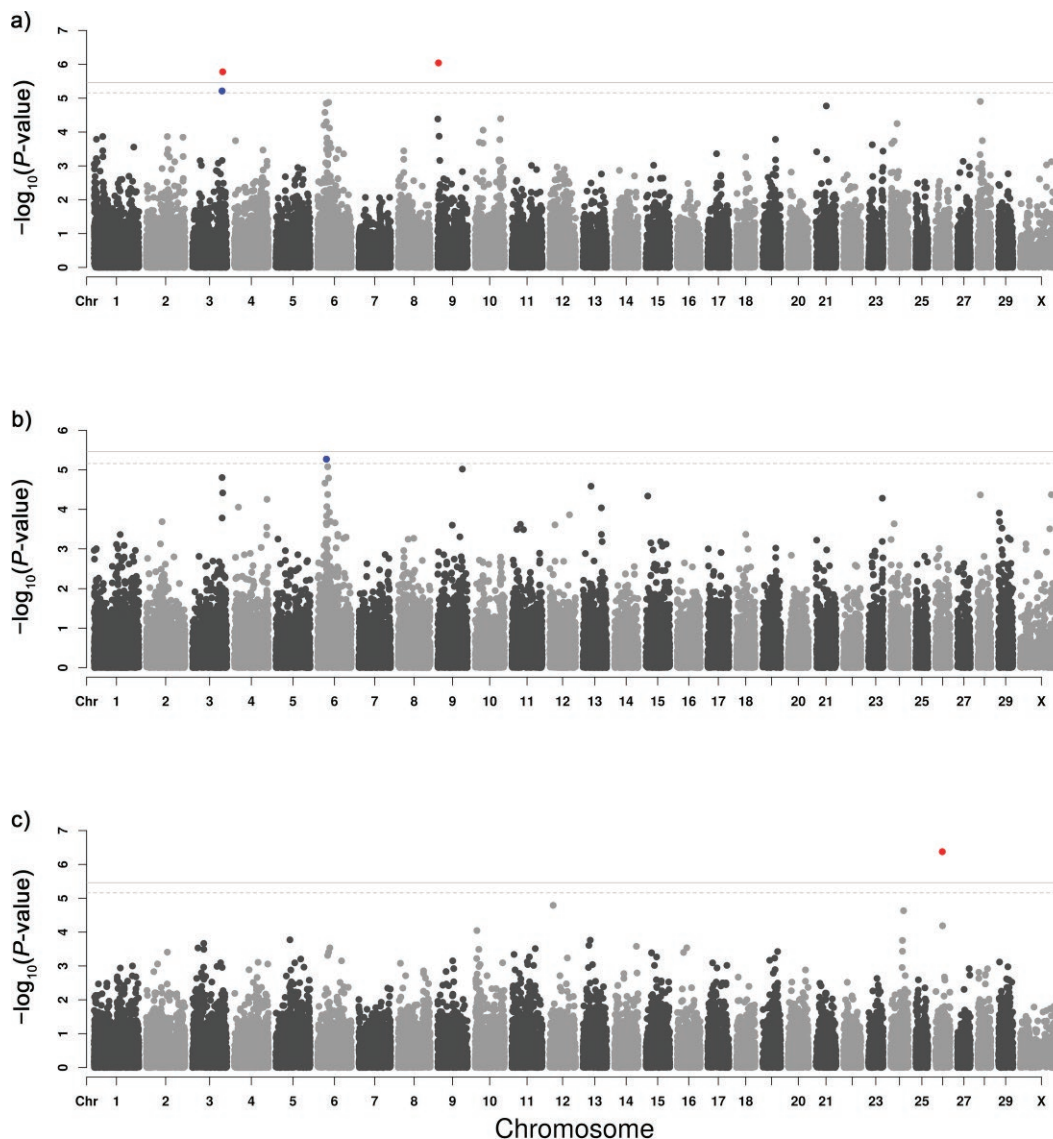
tially be functionally associated with the investigated trait. Interestingly, recently a QTL for CM in Holstein cows has been mapped to the same chromosomal region on BTA26 using selective genotyping. In the Holstein study, the most significant QTL, which was suggestive for CM, was located in a distance of 2 Mb to our DSN peak SNP rs42092404 (Kurz et al., 2019).

The markers rs110237486 and rs29026738 on BTA3 were significantly [-log<sub>10</sub>(P) = 5.78] and suggestively [-log<sub>10</sub>(P) = 5.21], respectively, associated with CM cases during the total first lactation (CM1\_total). These 2 markers are located within 2.1 Mb of each other on the distal end of the chromosome. For the significant marker rs110237486, cows with the genotype T/T (n = 980) showed a mastitis prevalence of 13.9% compared with 33.8% for heterozygous cows (n = 80). The marker is located intergenic between the genes encoding peptidylprolyl isomerase E (*PPIE*) and hippocalcin like 4 (*HPCAL4*). A study found *PPIE* in a rat model as a gene regulated by inflammation (Colitis; Rivera et al., 2006). Gene *PPIE* is suspected to be involved in folding proteins and has shown to have an influence on the adaptive immune system (Nath and Isakov, 2015). As the immune system of mammals has similar mechanisms to regulate inflammation of cells, it is possible that the gene *PPIE* is also responsible for activation of the immune system in cattle. The genes *MFSD2A*, *MYCL*, *TRIT1*, *NT5C1A*, *HEYL*, *PABPC4*, and *MACF1* are also located within a close distance of ±250 kb around this significant SNP.

For the suggestive marker rs29026738, cows with the A/A genotype (n = 975) had a mastitis prevalence of 13.8% compared with cows with the A/G genotype (n = 84) with a prevalence of 32.1%, which is very similar to the significant marker and shows high linkage to each other. The marker is located nearby *GUCA2A*, *EDN2*, and *HIVEP3*, which are reported to be highly associated with resistance to *Mycobacterium avium* ssp. *paratuberculosis* penetration into cells in tissue culture (Settles et al., 2009).

For both markers on BTA3 (rs29026738, rs110237486), homozygous carriers of the alternative allele are missing in our investigated DSN population. The minor allele frequency for both markers was 0.04 (Table 3). Because the minor allele was the disadvantageous allele with respect to mastitis resistance, the low frequency could be due to previous selection against cows with the highest mastitis prevalence.

The SNP rs29012331 [-log<sub>10</sub>(P) = 5.27] was the top marker of a region on BTA6 (6:22,650,675–6:41,969,496, using a -1.5 LOD drop method) associated with CM in the first lactation, particularly in the late phase of lactation 1 (Figure 2a and b). The allele T of the marker rs29012331 was protective, with homozygous T/T and



**Figure 2.** Manhattan plots for (a) the total number of clinical mastitis cases in total lactation 1, (b) the number of clinical mastitis cases in late lactation 1 (d 51 postpartum until end of lactation), and (c) the number of clinical mastitis cases in early lactation 2 (d 0 to 50 postpartum). The Manhattan plots show the distribution of  $-\log_{10}(P\text{-value})$  across all autosomes (BTA1–29) and chromosome X (x-axis). Markers above the significance thresholds are in red (solid line,  $\alpha < 0.05$ ), and markers above the suggestive threshold of significance are blue (dotted line,  $\alpha < 0.1$ ). Markers associated with clinical mastitis were detected on BTA3 (rs29026738, rs110237486), BTA6 (rs29012331), BTA9 (rs43581241), and BTA26 (rs42092404).

heterozygous C/T cows ( $n = 460$  and  $487$ , respectively) showing low mastitis prevalence of 5.9% and 8.8% in the late lactation 1, respectively, while homozygous C/C carriers had a prevalence of 21.5% ( $n = 107$ ). The frequency of the favorite allele T was 0.67. Therefore, further selection for the T allele of this SNP could improve mastitis resistance in DSN cows. The closest gene located to this marker is the bone morphogenetic protein receptor type 1B (*BMPR1B*). The literature has reported *BMPR1B* as a target gene for the immune

response to inflammation of cells infected with *Staphylococcus aureus* (Ju et al., 2018).

The marker rs43581241 on BTA9 (3,115,242) was significantly [ $-\log_{10}(P) = 6.04$ ] associated with CM in the total lactation 1. In the surrounding 250 kb region, no gene was located, but a cluster of copy number variations (CNV) were. Generally, CNV have been associated with changes in many different phenotypes, among which are health- and immune-related traits (Hou et al., 2011; Bickhart et al., 2012; Boussaha et al., 2015;

Keel et al., 2016; Mesbah-Uddin et al., 2018); therefore, CNV in this genomic region could potentially affect the mastitis prevalence. The minor and disadvantageous allele C segregated with a frequency of 0.10 in the DSN population. Even if the frequency of the advantageous allele A is high, this marker is interesting because of its potential to reduce mastitis incidence from 53.8% in cows with the homozygous alternative allele genotype (C/C,  $n = 13$ ) compared with heterozygous cows (A/C = 23.0%,  $n = 191$ ) and to the homozygous reference allele genotype (AA = 13.0%,  $n = 852$ ).

Genetic markers reaching significance across different lactations or lactation phases were not detected. Our initial expectation was that markers significant for CM in lactation 1 would also show significance in other lactations. However, this was not observed.

Estimates of the proportion of variance in CM cases explained by available genotypes, often called chip heritability, ranged in our DSN population between 0.01 and 0.03 (Table 4). Chip heritability estimates are based on all SNP available on the SNP chip and, therefore, will vary when different SNP platforms are used. Furthermore, SNP heritability cannot include unobserved SNP or heritability due to other sources (e.g.,  $G \times G$  interactions) and is, therefore, generally lower than classical broad- or narrow-sense heritability estimates. In general, heritability estimated from a linear model for overall CM was lower (0.04) versus pathogen-specific mastitis in Dutch herds ( $h^2 = 0.02$ – $0.10$ ) on the basis of a threshold model (De Haas et al., 2002). In Swedish Holstein cows, heritability for CM also ranges from 0.01 in the third to 0.03 in the first lactation (Carlén et al., 2004); in Danish Holstein cattle, a heritability for CM of 0.025 could be calculated (Lund et al., 1994). French Holstein cows had a similar heritability for CM as described before (0.024; Rupp and Boichard,

1999). In spite of higher heritability (0.08–0.19) for SCS, it is recommended to select directly against CM than indirectly against high SCS (Heringstad et al., 2000). A pathogen-specific grouping of cows could not be realized with the data presented in this study.

The GO analysis of all 23 positional candidate genes using DAVID (<http://david.abcc.ncifcrf.gov>) showed significant ( $P$ -value = 0.0083 after Bonferroni correction for multiple testing) overrepresentation of GO term 0005261—cation channel activity. Three calcium homeostasis modulator genes (*CALHM1*, *CALHM2*, and *CALHM3*) are found in the list of positional candidate genes contributing to the GO term (Supplemental Table S3, <https://doi.org/10.3168/jds.2020-18209>). The high ranking of these candidate genes is in line with the observation that the calcium concentration in the blood of dairy cows with and without mastitis differs, which could be an indicator for metabolic imbalance (Wegner and Stull, 1978; Hisaeda et al., 2020). No other significant overrepresentation was detected using GO.

### Method Discussion

The small population size, combined with the close genetic relationships among cows in the examined population (Korkuć et al., 2019) and the reliability of health data recorded under production conditions were challenging for this study. Nevertheless, SNP markers reaching the significance threshold could be detected.

The linear mixed model accounted for the fixed effects herd, calving year, and calving season. However, including a herd-year-season interaction as a fixed effect within the model was not possible in terms of small group sizes, which limited the statistical power of the model.

Due to the small size of the DSN population, the cows in this study descended from only 48 sires, which limited the genetic variation in this study. However, because the number of sires was low, linkage groups tended to be larger. Consequently, multiple testing correction was lowered as shown in the simpleM results ( $M_{eff} = 14,444$ ). This could be considered as an advantage in the search for markers reaching significance in more diverse populations. Increasing the number of cows would not increase the number of sires significantly because there was a low number of newly selected DSN sires for performance testing within the DSN population every year.

Because of identified linkage decay in this study, a region of 250 kb surrounding the mastitis-associated markers was chosen for the search of positional candidate genes. In 3 of 5 associated markers, 1 homozygous genotype group was missing in our investigated DSN

**Table 4.** Variance component estimation of PVE<sup>1</sup> with relatedness matrix for all SNP markers using Haseman-Elston regression; estimates are given for total clinical mastitis cases (total) in lactations 1 to 3 (CM1, CM2, CM3) and of clinical mastitis cases in their respective early and late lactation phases (early, late)

Phenotype	Stage	PVE	SE (PVE)
CM1	Total	0.0280	0.0283
	Early	−0.0124	0.0146
	Late	0.0319	0.0323
CM2	Total	0.0187	0.0255
	Early	0.0109	0.0195
	Late	0.0123	0.0238
CM3	Total	0.0207	0.0262
	Early	0.0207	0.0265
	Late	0.0139	0.0257

<sup>1</sup>PVE = proportion of variance in phenotypes explained by sparse  $X\beta$  and random effects  $u$ .



population, which resulted in a skewed distribution of the counted number of animals per genotype group (Table 3). It must be noted that the missing genotypes are the genotypes for which a high mastitis incidence is expected. Therefore, a selection against these genotypes in the past is supposed. Because of the missing genotype groups, it was not possible to determine for these specific markers if the observed genetic effects follow an additive or dominant pattern of inheritance.

For markers on BTA6 and BTA9, all 3 genotype groups were present in the investigated animals, but due to allele frequency distribution, groups with higher mastitis prevalence contained fewer animals than groups of cows with protective alleles. The marker rs43581241 on BTA9 showed the highest difference in mastitis prevalence between genotype classes. The mastitis prevalence was 53.8% for the homozygous reference allele (C/C). However, this genotype group contained only 13 cows, compared with 191 cows for the heterozygous genotype group (A/C = 23.0%) and 852 cows with homozygous reference allele genotype (A/A = 13.0%). For these 2 markers on BTA6 and BTA9, an indirect selection against mastitis could have also occurred. Nevertheless, further enrichment of the advantageous alleles of the markers rs29012331 on BTA6 and rs43581241 on BTA9 that segregate in the current population at a frequency of 0.66 and 0.90, respectively, could improve mastitis resistance in the DSN population. This needs to be further investigated in follow-up studies because these markers could be in weak linkage to the causal allele for CM or might be pleiotropically linked with undesirable traits.

## CONCLUSIONS

In addition to the selection for the protective alleles of CM-associated markers from our investigation, selection against the disadvantageous alleles of the markers that occur at a frequency below 0.04 can be used to reduce CM in DSN. However, the marker effects detected in this study need to be validated, including their influence on other economically important traits such as milk yield. After validation and exclusion of negative effects, these markers can be used for marker-assisted selection of favorable bulls in particular, but also for dams of sires in the existing breeding program. We point out the close genetic relationship between DSN and Holstein, which might cause these results to be more broadly useful for Holstein breeders. Additionally, the methodology presented here might serve as a template for other small endangered breeds. Furthermore, we see a renewed interest in dual-purpose breeds in Europe with the aim of improving animal welfare as well as climate-related reasons.

## ACKNOWLEDGMENTS

The RBB Rinderproduktion Berlin-Brandenburg GmbH (Gross Kreuz, Germany) and DSN farms supported this project by sharing their expertise in animal selection, providing mastitis records, and collecting ear tags. All data required reproducing the analysis, results, and conclusions are contained in the paper and supplemental files. The project was supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the Federal Programme for Ecological Farming and Other Forms of Sustainable Agriculture (Funding number: 2815NA010). S. M., D. A., and G. A. B. designed the study. S. M. prepared mastitis data for statistical analysis. P. K. prepared the 50k SNP Chip data. S. M., D. A., and P. K. performed all computational and statistical analysis and plotted the graphics. S. M. interpreted the data and drafted the manuscript. D. A., P. K., G. B. N., and G. A. B. helped draft the manuscript. All authors read and approved the final manuscript. The first author is an employee of the RBB Rinderproduktion Berlin-Brandenburg GmbH, a cattle breeder's association that produces semen and provides services to DSN and Holstein cattle breeders in the Berlin/Brandenburg area. The authors have not stated any other conflicts of interest.

## REFERENCES

- Abdel-Shafy, H., R. H. Bortfeldt, M. Reissmann, and G. A. Brockmann. 2018. Validating genome-wide associated signals for clinical mastitis in German Holstein cattle. *Anim. Genet.* 49:82–85. <https://doi.org/10.1111/age.12624>.
- Bickhart, D. M., Y. Hou, S. G. Schroeder, C. Alkan, M. F. Cardone, L. K. Matukumalli, J. Song, R. D. Schnabel, M. Ventura, J. F. Taylor, J. F. Garcia, C. P. Van Tassel, T. S. Sonstegard, E. E. Eichler, and G. E. Liu. 2012. Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res.* 22:778–790. <https://doi.org/10.1101/gr.133967.111>.
- Boussaha, M., D. Esquerré, J. Barbieri, A. Djari, A. Pinton, R. Letaief, G. Salin, F. Escudé, A. Roulet, S. Fritz, F. Samson, C. Grohs, M. Bernard, C. Klopp, D. Boichard, and D. Rocha. 2015. Genome-wide study of structural variants in bovine Holstein, Montbéliarde and Normande dairy breeds. *PLoS One* 10:e0135931. <https://doi.org/10.1371/journal.pone.0135931>.
- Bundesverband Rind und Schwein e. V. 2019. *Rinderzucht in Deutschland* 2018. Bonn.
- Carlén, E., E. Strandberg, and A. Roth. 2004. Genetic parameters for clinical mastitis, somatic cell score, and production in the first three lactations of Swedish Holstein cows. *J. Dairy Sci.* 87:3062–3070. [https://doi.org/10.3168/jds.S0022-0302\(04\)73439-6](https://doi.org/10.3168/jds.S0022-0302(04)73439-6).
- De Haas, Y., H. W. Barkema, and R. F. Veerkamp. 2002. Genetic parameters of pathogen-specific incidence of clinical mastitis in dairy cows. *Anim. Sci.* 74:233–242. <https://doi.org/10.1017/S1357729800052401>.
- Durinck, S., Y. Moreau, A. Kasprzyk, S. Davis, B. De Moor, A. Brazma, and W. Huber. 2005. BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* 21:3439–3440. <https://doi.org/10.1093/bioinformatics/bti525>.



- Durinck, S., P. T. Spellman, E. Birney, and W. Huber. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* 4:1184–1191. <https://doi.org/10.1038/nprot.2009.97>.
- Gao, X., J. Starmer, and E. R. Martin. 2008. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.* 32:361–369. <https://doi.org/10.1002/gepi.20310>.
- Grothe, P. O. 1993. *Holstein-Friesian, eine Rasse geht um die Welt*. Landwirtschaftsverlag GmbH, Münster-Hiltrup.
- Heringstad, B., G. Klemetsdal, and J. Ruane. 2000. Selection for mastitis resistance in dairy cattle: A review with focus on the situation in the Nordic countries. *Livest. Prod. Sci.* 64:95–106. [https://doi.org/10.1016/S0301-6226\(99\)00128-1](https://doi.org/10.1016/S0301-6226(99)00128-1).
- Hisaeda, K., T. Koshiishi, A. Sasaki, Y. Shinozuka, N. Isobe, and K. Kawai. 2020. Changes in ionized calcium concentration in the blood of dairy cows with peracute coliform mastitis. *J. Vet. Med. Sci.* 82:457–462. <https://doi.org/10.1292/jvms.19-0678>.
- Hou, Y., G. E. Liu, D. M. Bickhart, M. F. Cardone, K. Wang, E. Kim, L. K. Matukumalli, M. Ventura, J. Song, P. M. VanRaden, T. S. Sonstegard, and C. P. Van Tassel. 2011. Genomic characteristics of cattle copy number variations. *BMC Genomics* 12:127. <https://doi.org/10.1186/1471-2164-12-127>.
- Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4:44–57. <https://doi.org/10.1038/nprot.2008.211>.
- ICAR. 2015. Health codes for bovine. *ICAR Rec. Guidel. Amend.* Accessed May 1, 2019. <https://www.icar.org/index.php/publications-technical-materials/amendments-recording-guidelines/diseases-codes-for-cows/>.
- Iso-Touru, T., G. Sahana, B. Gulbrandtsen, M. S. Lund, and J. Vilkki. 2016. Genome-wide association analysis of milk yield traits in Nordic Red Cattle using imputed whole genome sequence variants. *BMC Genet.* 17:55. <https://doi.org/10.1186/s12863-016-0363-8>.
- Ju, Z., Q. Jiang, G. Liu, X. Wang, G. Luo, Y. Zhang, J. Zhang, J. Zhong, and J. Huang. 2018. Solexa sequencing and custom microRNA chip reveal repertoire of microRNAs in mammary gland of bovine suffering from natural infectious mastitis. *Anim. Genet.* 49:3–18. <https://doi.org/10.1111/age.12628>.
- Keel, B. N., A. K. Lindholm-Perry, and W. M. Snelling. 2016. Evolutionary and functional features of copy number variation in the cattle genome. *Front. Genet.* 7. <https://doi.org/10.3389/fgene.2016.00207>.
- Koeck, A., S. Loker, F. Miglior, D. F. Kelton, J. Jamrozik, and F. S. Schenkel. 2014. Genetic relationships of clinical mastitis, cystic ovaries, and lameness with milk yield and somatic cell score in first-lactation Canadian Holsteins. *J. Dairy Sci.* 97:5806–5813. <https://doi.org/10.3168/jds.2013-7785>.
- Köppe-Forsthoff, J. 1967. *100 Jahre Deutsche Schwarzbuntzucht*. Hiltrup Landwirtschaftsverlag.
- Korkuč, P., D. Arends, and G. A. Brockmann. 2019. Finding the optimal imputation strategy for small cattle populations. *Front. Genet.* 10:52. <https://doi.org/10.3389/fgene.2019.00052>.
- Kurz, J. P., Z. Yang, R. B. Weiss, D. J. Wilson, K. A. Rood, G. E. Liu, and Z. Wang. 2019. A genome-wide association study for mastitis resistance in phenotypically well-characterized Holstein dairy cattle using a selective genotyping approach. *Immunogenetics* 71:35–47. <https://doi.org/10.1007/s00251-018-1088-9>.
- Lund, M. S., J. Jensen, and P. H. Petersen. 1999. Estimation of genetic and phenotypic parameters for clinical mastitis, somatic cell production deviance, and protein yield in dairy cattle using Gibbs sampling. *J. Dairy Sci.* 82:1045–1051. [https://doi.org/10.3168/jds.S0022-0302\(99\)75325-7](https://doi.org/10.3168/jds.S0022-0302(99)75325-7).
- Lund, T., F. Miglior, J. C. M. Dekkers, and E. B. Burnside. 1994. Genetic relationships between clinical mastitis, somatic cell count, and udder conformation in Danish Holsteins. *Livest. Prod. Sci.* 39:243–251. [https://doi.org/10.1016/0301-6226\(94\)90203-8](https://doi.org/10.1016/0301-6226(94)90203-8).
- Maere, S., K. Heymans, and M. Kuiper. 2005. BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21:3448–3449. <https://doi.org/10.1093/bioinformatics/bti551>.
- Martin, G. F. 2012. *Analyse von Behandlungsdaten zur Entwicklung eines Zuchtwertschätzverfahrens für Eutergesundheit beim Milchrind*. PhD thesis. Institut für Agrar- und Ernährungswissenschaften, Martin-Luther-Universität Halle-Wittenberg, Halle, Germany.
- Mesbah-Uddin, M., B. Gulbrandtsen, T. Iso-Touru, J. Vilkki, D. J. De Koning, D. Boichard, M. S. Lund, and G. Sahana. 2018. Genome-wide mapping of large deletions and their population-genetic properties in dairy cattle. *DNA Res.* 25:49–59. <https://doi.org/10.1093/dnares/dsx037>.
- Nath, P. R., and N. Isakov. 2015. Insights into peptidyl-prolyl cis-trans isomerase structure and function in immunocytes. *Immunol. Lett.* 163:120–131. <https://doi.org/10.1016/j.imlet.2014.11.002>.
- Negussie, E., M. H. Lidauer, K. Johansson, U. S. Nielsen, G. P. Aamand, and F. Breeding. 2010. Combining test day SCS with clinical mastitis and udder type traits: A random regression model for joint genetic evaluation of udder health in Denmark, Finland, Sweden. *Interbull Bull.* 42:25–32.
- Nyman, A. K., T. Ekman, U. Emanuelson, A. H. Gustafsson, K. Holtenius, K. P. Waller, and C. H. Sandgren. 2007. Risk factors associated with the incidence of veterinary-treated clinical mastitis in Swedish dairy herds with a high milk yield and a low prevalence of subclinical mastitis. *Prev. Vet. Med.* 78:142–160. <https://doi.org/10.1016/j.prevetmed.2006.10.002>.
- Olde Riekerink, R. G. M., H. W. Barkema, D. F. Kelton, and D. T. Scholl. 2008. Incidence rate of clinical mastitis on Canadian dairy farms. *J. Dairy Sci.* 91:1366–1377. <https://doi.org/10.3168/jds.2007-0757>.
- Persson Waller, K., B. Bengtsson, A. Lindberg, A. Nyman, and H. Ericsson Unnerstad. 2009. Incidence of mastitis and bacterial findings at clinical mastitis in Swedish primiparous cows-Influence of breed and stage of lactation. *Vet. Microbiol.* 134:89–94. <https://doi.org/10.1016/j.vetmic.2008.09.004>.
- RBB Rinderproduktion Berlin-Brandenburg GmbH. 2016. *Deutsches Schwarzbuntes Niederungsring - lebendes Kulturerbe*. Accessed Aug. 4, 2020 <https://www.rinderzucht-bb.de/de/zucht/dsn-genreserve/>.
- R Core Team. 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Au. <https://www.r-project.org/>.
- Rivera, E., I. Flores, E. Rivera, and C. B. Appleyard. 2006. Molecular profiling of a rat model of colitis: Validation of known inflammatory genes and identification of novel disease-associated targets. *Inflamm. Bowel Dis.* 12:950–966. <https://doi.org/10.1097/01.mib.0000231575.11678.Sc>.
- Rupp, R., and D. Boichard. 1999. Genetic parameters for clinical mastitis, somatic cell score, production, udder type traits, and milking ease in first lactation Holsteins. *J. Dairy Sci.* 82:2198–2204. [https://doi.org/10.3168/jds.S0022-0302\(99\)75465-2](https://doi.org/10.3168/jds.S0022-0302(99)75465-2).
- Rupp, R., and D. Boichard. 2000. Relationship of early first lactation somatic cell count with risk of subsequent first clinical mastitis. *Livest. Prod. Sci.* 62:169–180. [https://doi.org/10.1016/S0301-6226\(99\)00056-1](https://doi.org/10.1016/S0301-6226(99)00056-1).
- Sahana, G., B. Gulbrandtsen, B. Thomsen, L.-E. Holm, F. Panitz, R. F. Brøndum, C. Bendixen, and M. S. Lund. 2014. Genome-wide association study using high-density single nucleotide polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle. *J. Dairy Sci.* 97:7258–7275. <https://doi.org/10.3168/jds.2014-8141>.
- Seegers, H., C. Fourichon, and F. Beaudeau. 2003. Production effects related to mastitis and mastitis economics in dairy cattle herds. *Vet. Res.* 34:475–491. <https://doi.org/10.1051/vetres:2003027>.
- Sender, G., A. Korwin-Kossakowska, A. Pawlik, K. G. A. Hameed, and J. Oprządek. 2013. Genetic basis of mastitis resistance in dairy cattle - A review. *Ann. Anim. Sci.* 13:663–673. <https://doi.org/10.2478/aoas-2013-0043>.
- Settles, M., R. Zanella, S. D. McKay, R. D. Schnabel, J. F. Taylor, R. Whitlock, Y. Schukken, J. S. Van Kessel, J. M. Smith, and H. Nebelers. 2009. A whole genome association analysis identifies loci

- associated with *Mycobacterium avium* ssp. *paratuberculosis* infection status in US Holstein cattle. *Anim. Genet.* 40:655–662. <https://doi.org/10.1111/j.1365-2052.2009.01896.x>.
- Sofer, T. 2017. Confidence intervals for heritability via Haseman-Elston regression. *Stat. Appl. Genet. Mol. Biol.* 16:259–273. <https://doi.org/10.1515/sagmb-2016-0076>.
- Stock, K. F., and C. Egger-Danner. 2017. Standardization of health data recording (central health key). Pages 1–10, ICAR Udder health meeting, Edinburgh, UK. ICAR 2017, Rome, Italy.
- Svensson, C., A. K. Nyman, K. P. Waller, and U. Emanuelson. 2006. Effects of housing, management, and health of dairy heifers on first-lactation udder health in southwest Sweden. *J. Dairy Sci.* 89:1990–1999. [https://doi.org/10.3168/jds.S0022-0302\(06\)72266-4](https://doi.org/10.3168/jds.S0022-0302(06)72266-4).
- Wegner, T. N., and J. W. Stull. 1978. Relation between mastitis test score, mineral composition of milk, and blood electrolyte profiles in Holstein cows. *J. Dairy Sci.* 61:1755–1759. [https://doi.org/10.3168/jds.S0022-0302\(78\)83798-9](https://doi.org/10.3168/jds.S0022-0302(78)83798-9).
- Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York, NY. <https://ggplot2.tidyverse.org>.
- Yin, L. 2019. A high-quality drawing tool designed for Manhattan plot of genomic analysis. Huazhong Agricultural University. <https://github.com/YinLiLin/CMplot>.
- Zerbino, D. R., P. Achuthan, W. Akanni, M. R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C. G. Girón, L. Gil, L. Gordon, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, J. K. To, M. R. Laird, I. Lavidas, Z. Liu, J. E. Loveland, T. Maurel, W. McLaren, B. Moore, J. Mudge, D. N. Murphy, V. Newman, M. Nuhn, D. Ogeh, C. K. Ong, A. Parker, M. Patricio, H. S. Riat, H. Schuilenburg, D. Sheppard, H. Sparrow, K. Taylor, A. Thormann, A. Vullo, B. Walts, A. Zadissa, A. Frankish, S. E. Hunt, M. Kostadima, N. Langridge, F. J. Martin, M. Muffato, E. Perry, M. Ruffier, D. M. Staines, S. J. Trevanion, B. L. Aken, F. Cunningham, A. Yates, and P. Flicek. 2018. Ensembl 2018. *Nucleic Acids Res.* 46(D1):D754–D761. <https://doi.org/10.1093/nar/gkx1098>.
- Zhou, X., and M. Stephens. 2014. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* 11:407–409. <https://doi.org/10.1038/nmeth.2848>.
- Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10:R42. <https://doi.org/10.1186/gb-2009-10-4-r42>.

## ORCID

- Saskia Meier  <https://orcid.org/0000-0002-6048-0558>  
Danny Arends  <https://orcid.org/0000-0001-8738-0162>  
Paula Korcuć  <https://orcid.org/0000-0002-3772-9562>  
Guilherme B. Neumann  <https://orcid.org/0000-0001-6581-3019>  
Gudrun A. Brockmann  <https://orcid.org/0000-0002-4387-2947>