



## Genome-wide association study for longevity with whole-genome sequencing in 3 cattle breeds

Qianqian Zhang,\*†<sup>1</sup> Bernt Guldbrandtsen,\* Jørn Rind Thomasen,\*‡ Mogens Sandø Lund,\* and Goutam Sahana\*

\*Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark

†Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, 6700 AH Wageningen, the Netherlands

‡VikingGenetics, Assentoft, DK-8960 Randers, Denmark

### ABSTRACT

Longevity is an important economic trait in dairy production. Improvements in longevity could increase the average number of lactations per cow, thereby affecting the profitability of the dairy cattle industry. Improved longevity for cows reduces the replacement cost of stock and enables animals to achieve the highest production period. Moreover, longevity is an indirect indicator of animal welfare. Using whole-genome sequencing variants in 3 dairy cattle breeds, we carried out an association study and identified 7 genomic regions in Holstein and 5 regions in Red Dairy Cattle that were associated with longevity. Meta-analyses of 3 breeds revealed 2 significant genomic regions, located on chromosomes 6 (META-CHR6–88MB) and 18 (META-CHR18–58MB). META-CHR6–88MB overlaps with 2 known genes: neuropeptide G-protein coupled receptor (*NPFGR2*; 89,052,210–89,059,348 bp) and vitamin D-binding protein precursor (*GC*; 88,695,940–88,739,180 bp). The *NPFGR2* gene was previously identified as a candidate gene for mastitis resistance. META-CHR18–58MB overlaps with zinc finger protein 717 (*ZNF717*; 58,130,465–58,141,877 bp) and zinc finger protein 613 (*ZNF613*; 58,115,782–58,117,110 bp), which have been associated with calving difficulties. Information on longevity-associated genomic regions could be used to find causal genes/variants influencing longevity and exploited to improve the reliability of genomic prediction. **Key words:** genome-wide association study, longevity, cattle, whole-genome sequencing

### INTRODUCTION

Cows that remain healthy and reproduce regularly are of economic importance to breeders. The average

lifespan of cows is far below their biological potential, and disposals due to old age are rare (Essl, 1998). In cows, longevity refers to the period from first parity until disposal, rather than the whole lifespan, although the length of productive life and herd life are more appropriate ways to describe the productive period (Essl, 1998). Longevity is of major importance for the economics of dairy production because involuntary culling of cows is often related to increased costs due to veterinary treatments, poor fertility, or low milk production. Selection for improved longevity is challenging because this trait is expressed late in life (the actual measurement is only available after the cow is culled or dead) and has a relatively low heritability (0.029–0.072; [http://www.nordicebv.info/wp-content/uploads/2015/04/General-description\\_from-old-homepage\\_06052015.pdf](http://www.nordicebv.info/wp-content/uploads/2015/04/General-description_from-old-homepage_06052015.pdf); Pritchard et al., 2013). Identification of genomic regions associated with longevity may help in identifying genes and causal mutations that influence this trait. This information could be included in the genomic selection model for better prediction of its breeding values (Brøndum et al., 2015; Wientjes et al., 2015).

Since 2008, longevity has been included as a trait in joint cattle breeding goals in Denmark, Finland, and Sweden. The Nordic Cattle Genetic Evaluation (NAV; [www.nordicebv.info](http://www.nordicebv.info)) defines the longevity index as the productive longevity of a bull's daughter, without correcting for other traits (e.g., yield or fertility). The economic value of longevity in the Nordic total merit index is 0.38 to 0.51 euros per day, depending on the breed. When this Nordic total merit index is used as a criterion in the selection strategy, 49 to 68% of the maximum progress for longevity can be obtained compared with a strategy that only includes longevity. The genetic trend for longevity has been positive for Holstein (HOL) and Nordic Red Dairy Cattle (RDC), but unchanged for Danish Jersey cattle (JER; [https://www.landbrugsinfo.dk/Kvaeg/Avl/Avlsstatistik/Sider/aarsstat\\_2014.pdf?download=true](https://www.landbrugsinfo.dk/Kvaeg/Avl/Avlsstatistik/Sider/aarsstat_2014.pdf?download=true)).

Received November 30, 2015.

Accepted May 4, 2016.

<sup>1</sup>Corresponding author: [qianqian.zhang@mbg.au.dk](mailto:qianqian.zhang@mbg.au.dk)

Farmers and breeding companies in Nordic countries are taking increasing interest in breeding for longevity. Selection for longevity in cows could increase the number of productive cycles, with an indirect economic effect on the dairy cattle industry. Improved longevity for cows will reduce the replacement cost of stock, enable animals to achieve their highest production period, and improve animal welfare. When the breeding goal in Nordic countries was changed to include both production and functional traits (e.g., fertility, health, and longevity), longevity was improved by about 21 and 18 breeding value units for RDC and HOL, respectively, between the years 1990 and 2010. However, this positive genetic progress for longevity has not been fully reflected in the phenotypic trend of Nordic cows. This discrepancy is because farm-related factors, such as feeding, housing, and management, have a large influence on cattle health (<http://www.nordicebv.info/wp-content/uploads/2015/04/Longevity-of-Nordic-dairy-cows-can-be-improved.pdf>).

With the availability of genome-wide markers, genomic regions associated with longevity in cattle can be identified and studied. Therefore, the objective of this study was to identify QTL associated with longevity in 3 dairy cattle breeds. Within-breed association studies were carried out to detect longevity-associated QTL in 3 cattle breeds, followed by meta-analyses to determine whether additional QTL could be found by combining the breeds.

## MATERIALS AND METHODS

Imputation of 50k genotypes to whole-genome sequencing (WGS) variants for bulls from 3 Nordic cattle breeds was done in 2 steps (Iso-Touru et al., 2016). High-density (HD) genotypes were imputed from 50k genotypes by using HD multibreed reference genotype data, and then the imputed HD genotypes were imputed to WGS variants by using a multibreed WGS reference. Genome-wide association analysis of the imputed WGS variants was also done in 2 steps. First, a genome scan was made separately for each of the 3 cattle breeds, using a modified mixed model approach (Kang et al., 2010). Then, a meta-analysis of the 3 breeds was performed.

### Animals and Phenotypes

Because no animal experiments were performed, approval from the ethics committee was not required. The study included 10,575 genotyped and progeny-tested bulls from 3 Nordic cattle breeds from Denmark, Sweden, and Finland: 5,314 HOL, 4,200 RDC, and 1,061 Danish JER bulls. Population structure of these 3

breeds was reported by Kadri et al. (2015). Deregressed estimated breeding values (DRP) for the longevity trait from routine genetic evaluations were used as phenotypes in the association study. Average reliability of the DRP was 0.748.

In its evaluation of longevity, the NAV considers the number of days from calving to the end of lactation for each lactation up to the fifth, with a maximum of 365 d per lactation. To estimate breeding values of longevity, the NAV uses a multitrait animal model, which includes age at first calving, calving month, and herd-year as fixed effects, heterosis effects as regression effects, and genetic groups, herd  $\times$  year of first calving, and animal effects as random effects. Genetic groups are modeled as phantom parent grouping. Separate genetic evaluations are made for each breed ([http://www.nordicebv.info/wp-content/uploads/2015/04/General-description\\_from-old-homepage\\_06052015.pdf](http://www.nordicebv.info/wp-content/uploads/2015/04/General-description_from-old-homepage_06052015.pdf)). For details regarding recorded phenotypes and models used in routine breeding value estimates, see <https://www.landbrugsinfo.dk/kvaeg/avl/sider/principles.pdf?download=true>.

### Genotypes of Animals

Single nucleotide polymorphism genotyping with a SNP chip and quality control (QC) analyses were performed as described by Sahana et al. (2015). In brief, DNA was extracted from whole blood or semen. All bulls with breeding values for longevity were genotyped with the BovineSNP50 BeadChip (Illumina, San Diego, CA) version 1 or 2. The QC analyses of SNP genotypes were carried out simultaneously for all 3 breeds. The SNP selection parameters were an 85% minimum call rate for individuals and 95% for loci. Marker loci without a known map position, with minor allele frequency <0.05%, or with deviation from Hardy-Weinberg proportions ( $\chi^2$ -test, 1 df,  $P < 0.00001$ ) were excluded. After QC, 43,415 SNP remained in the 50k data set.

A multibreed reference of 3,383 animals (1,222 HOL, 1,326 RDC, and 835 Danish JER) obtained with the Illumina BovineHD Genotyping BeadChip was available in-house and from the EuroGenomics consortium (Lund et al., 2011). The QC parameter set and protocol for HD data were similar to those of the 50k chip described above. Imputation to HD genotypes was done with IMPUTE2 v2.3.1 (Howe et al., 2011). The HD genotyped reference was used to impute the 50k genotypes to the HD array as a bridge to impute to the WGS variants.

### Imputation to WGS Level

Imputed HD genotypes were further imputed to the WGS level by using a multibreed reference of 1,228

animals from run4 of the 1,000 bull genomes project (1,148 cattle, including 288 individuals from the global Holstein-Friesian population, 56 Nordic RDC individuals, 61 JER individuals, and 743 individuals from other breeds; Daetwyler et al., 2014) and private data from Aarhus University (80 cattle, including 23 HOL, 30 Nordic RDC, and 27 Danish JER). Imputation to the WGS level was performed by using Minimac2 (Fuchsberger et al., 2015). A total of 22,751,039 biallelic variants were present in the imputed sequence data.

### Statistical Methods for Association Analysis

Association analysis was carried out by using a modified linear mixed model (LMM) approach for within-breed QTL identification, followed by a meta-analysis across 3 breeds. The efficient mixed-model association expedited (EMMAX) approach was developed previously to correct for population structure and genetic relatedness, and to increase computational speed in association mapping (Kang et al., 2008). Association analysis for each imputed sequence variant was carried out in the EMMAX software tool by using a 2-step variance component-based approach to account for population stratification (Kang et al., 2010). Details about the model are given by Kang et al. (2008, 2010).

In the first step, polygenic and error variances were estimated with the variance component model:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{a} + \mathbf{e},$$

where  $\mathbf{y}$  is a vector of phenotypes (DRP for longevity);  $\mu$  is the intercept;  $\mathbf{a}$  is a vector of random polygenic effects, which are normally distributed as  $\mathbf{a} \sim N(0, \mathbf{G}\sigma_a^2)$ ;  $\mathbf{G}$  is the genomic relationship matrix (GRM), which is built based on imputed HD SNP genotypes;  $\sigma_a^2$  is the additive genetic variance;  $\mathbf{Z}$  is an incidence matrix, relating phenotypes to the corresponding random polygenic effects;  $\mathbf{e}$  is the vector of random individual error terms, assumed to follow  $N(0, \mathbf{I}\sigma_e^2)$ ;  $\mathbf{I}$  is an identity matrix; and  $\sigma_e^2$  is the error variance. The GRM is 2 times the kinship coefficient matrix between each pair of individuals. The kinship matrix is inferred by a simple identical-by-state allele-sharing matrix using “emmax-kin” in EMMAX. Inclusion of a candidate marker in the GRM can lead to loss of power due to double fitting of the candidate marker as both a fixed effect for testing association and a random effect as part of the GRM (Listgarten et al., 2013). Therefore, we followed the leave-one-chromosome-out approach (Yang et al., 2014) to build a kinship matrix specific to each chromosome.

In the second step, each SNP effect was obtained by using a linear regression model:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{x}g + \boldsymbol{\eta},$$

where  $\mu$  is the overall mean;  $\mathbf{x}$  is a vector of imputed allele dosages (expected number of copies of a specified allele, ranging from 0 to 2);  $g$  is the SNP effect; and  $\boldsymbol{\eta}$  is a vector of random residual deviates, with variance  $\mathbf{G}\sigma_a^2 + \mathbf{I}\sigma_e^2$ . This approach was carried out separately for the WGS variants in each of the 3 dairy cattle breeds. A Bonferroni correction was applied to control for false-positive associations.

The variance in the deregressed proofs of longevity explained by a SNP was calculated by  $2p \times (1 - p) \times \alpha^2 / (\text{variance of deregressed proofs of longevity})$  in which  $p$  is the frequency of the allele coded as 1 and  $\alpha$  is the allele substitution effect.

### Meta-Analysis Combining 3 Breeds

**Meta-Analysis.** We carried out a sample size-weighted meta-analysis based on  $P$ -values obtained from the EMMAX analysis for the 3 breeds. The weighted Z-score method was used, as follows:

$$Z = \frac{\sum_i Z_i w_i}{\sqrt{\sum_i w_i^2}},$$

where  $Z_i = \phi^{-1}(p_i/2) \times \text{sign}(\Delta_i)$ ;  $w_i = \sqrt{N_i}$ ;  $N_i$  is the sample size;  $p_i$  is the  $P$ -value;  $\Delta_i$  is the direction of effect for study  $i$ ; and  $\phi$  is the cumulative distribution function of the standard normal distribution. Significant SNPs were detected by using the overall  $P$ -value for the meta-analysis, calculated as  $P = 2\phi(-|Z|)$ , and the Bonferroni-corrected threshold. Analyses were performed in the METAL program (Willer et al., 2010). This approach was carried out for WGS variants across the 3 Nordic dairy cattle breeds.

### Search for Multiple QTL in a Genomic Region

We examined if multiple QTL are segregating in a genomic region on chromosome 6 and 18. First we adjusted the phenotype for the top associated SNP effect (Chr6:88840407 and Chr18:58015050 on chromosomes 6 and 18, respectively) using a LMM followed by a linear regression model to search for additional signal for association. The statistical model is described by the formula:

The statistical model for LMM is described by the following formula:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{x}g + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where  $\mathbf{y}$  is the vector of deregressed proof breeding values for longevity,  $\mathbf{1}$  is a vector of 1 s,  $\mu$  is the general mean,  $g$  is the fixed additive genetic effect of the analyzed SNP,  $\mathbf{x}$  is a vector of allele dosages for the top associated SNP of the chromosome, and  $\mathbf{u}$  is a vector of random polygenic effects, which are normally distributed  $\mathbf{u} \sim N(0, \mathbf{A}\sigma_u^2)$ , where  $\mathbf{A}$  is the pedigree-based additive relationship matrix,  $\sigma_u^2$  is the polygenic variance,  $\mathbf{Z}$  is an incidence matrix relating phenotypes to the corresponding random polygenic effects, and  $\mathbf{e}$  is a vector of residual effects, which are normally distributed  $\mathbf{e} \sim N(0, \mathbf{D}\sigma_e^2)$ , where  $\mathbf{D}$  is a diagonal matrix with elements  $d_{ii} = (1 - r_{DRP}^2)/r_{DRP}^2$  to account for heterogeneous residual variances due to different reliabilities of DRP ( $r_{DRP}^2$ ), and  $\sigma_e^2$  is the residual variance. Analyses were performed using the DMU package (Madsen et al., 2014). The residuals from LMM were used as response variable for the following linear regression model for single variant analysis.

$$\mathbf{y}_r = \mathbf{1}\mu + \mathbf{x}g + \boldsymbol{\varepsilon},$$

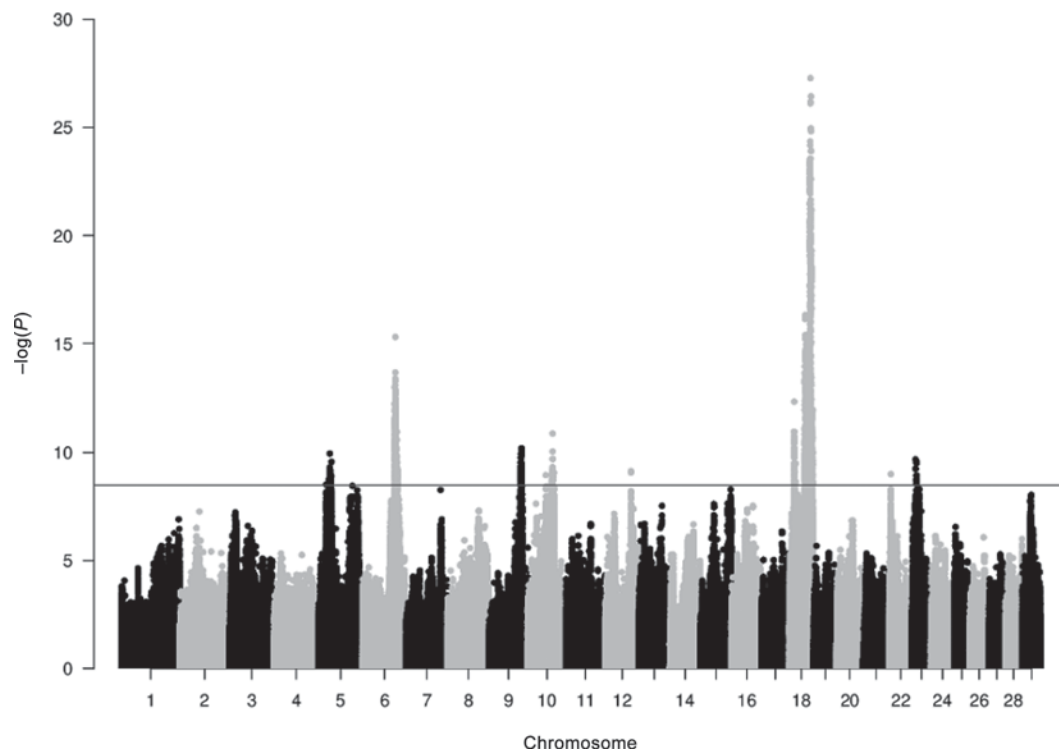
where  $\mathbf{y}_r$  is the vector of residuals from the LMM,  $\boldsymbol{\varepsilon}$  is random error assumed to be distributed  $N(0, \mathbf{I}\sigma_\varepsilon^2)$ , and the rest of the terms in the model are as described above.

## RESULTS AND DISCUSSION

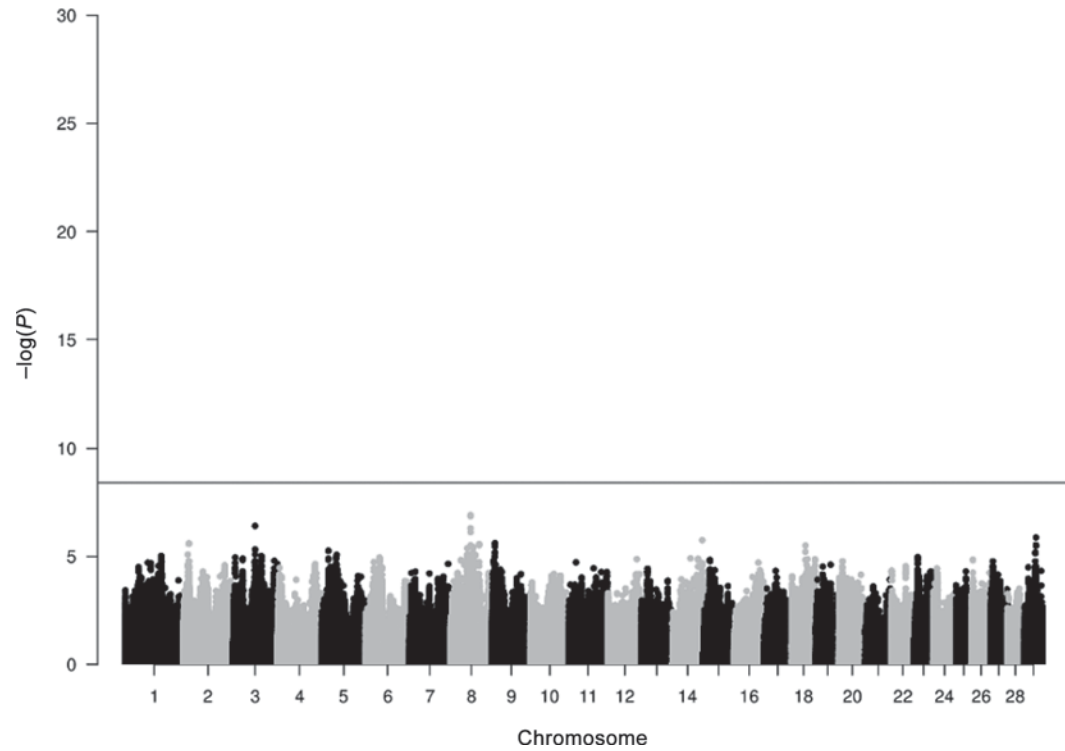
### Within-Breed Association Analyses

Within-breed genome scans for longevity-associated SNP revealed 8 genomic regions on 6 chromosomes in HOL and 5 regions on 4 chromosomes in Nordic RDC (Figures 1–3, Tables S1–S2; <http://dx.doi.org/10.3168/jds.2015-10697>). The 5 most significantly associated genomic regions and representative SNP in each region are presented in Tables S1 and S2 (<http://dx.doi.org/10.3168/jds.2015-10697>) for HOL and RDC. There were no significant SNP associated with longevity in JER (Figure 2).

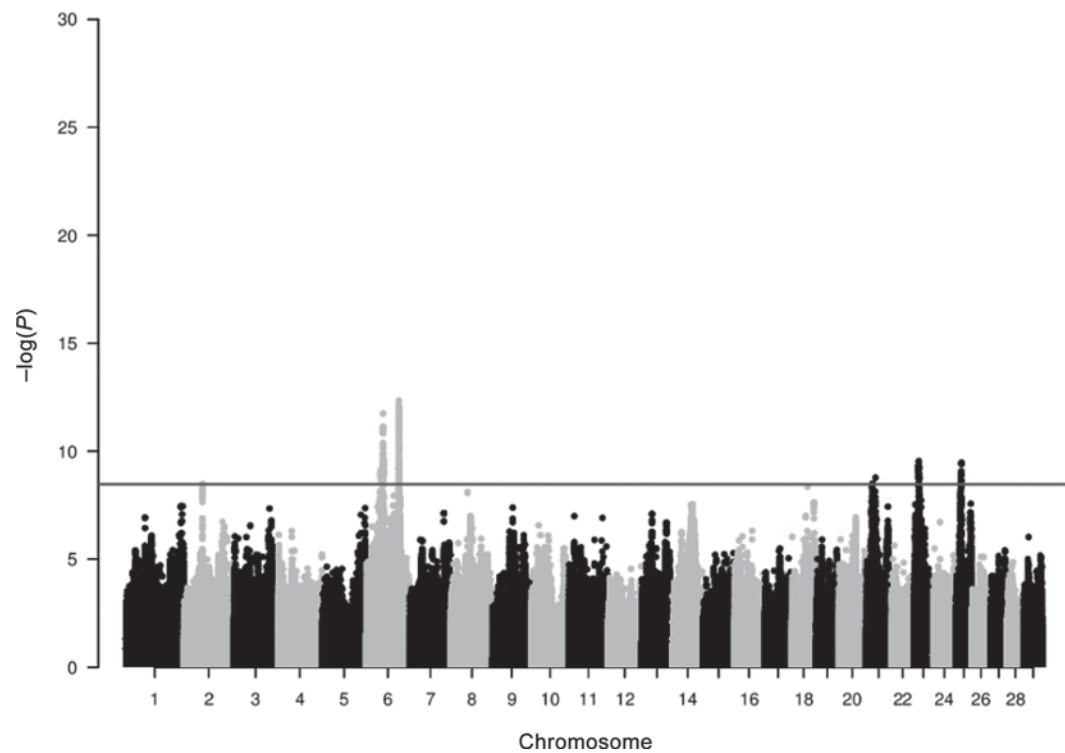
**HOL.** A strong association signal in HOL was found on chromosome 6 around 88 Mbp (HOL-CHR6–88MB). The SNP with the strongest association ( $P = 5.36\text{E-}14$ ) was located at 88,937,771 bp (rs383561794) (Figure 1



**Figure 1.** Manhattan plot for genome-wide association scan for longevity using whole-genome sequence variants in Holstein cattle.



**Figure 2.** Manhattan plot for genome-wide association scan for longevity using whole-genome sequence variants in Jersey cattle.



**Figure 3.** Manhattan plot for genome-wide association scan for longevity using whole-genome sequence variants in Nordic Red Dairy cattle.



and Supplemental Table S1; <http://dx.doi.org/10.3168/jds.2015-10697>), between genes neuropeptide FF receptor 2 (*NPFFR2*; 89,052,210–89,059,348 bp) and group-specific component (vitamin D-binding protein; *GC*; 88,695,940–88,739,180 bp). These genes were reported to be candidate genes associated with mastitis resistance (Sahana et al., 2014). In addition, these longevity QTL (HOL-CHR6–88MB and HOL-CHR6–89MB) also overlaps with the QTL not only associated with quality of udder and quality of feet and legs (Hiendleder et al., 2003; Pausch et al., 2016), but also some productive traits including fat and protein yield and percentage (Prinzenberg et al., 2003; Viitala et al., 2003; Raven et al., 2014). The HOL-CHR5–31MB overlaps with QTL associated with dressing percentage (Stone et al., 1999). The peak associated region around 88 Mbp on chromosome 9 (HOL-CHR9–88MB) was located on gene regulatory (inhibitor) subunit <sup>14</sup>C (*PPP1R14C*; 88,384,683–88,500,749 bp) and overlaps with QTL associated with milk, protein, and fat yield (Zhang et al., 1998; Wiener et al., 2000).

The top associated SNP ( $P = 1.35E-11$ ) in the genomic region around 67 Mbp on chromosome 10 (HOL-CHR10–67MB) was located at 67,632,847 bp (rs381727199), in close proximity to the GTP cyclohydrolase 1 gene (*GCH1*; 67,576,390–67,631,089 bp) and was overlapping with QTL associated with dressing percentage and ovulation rate (MacNeil and Grosz, 2002; Arias and Kirkpatrick, 2004). The HOL-CHR23-MB overlap with QTL associated with milk yield and preweaning ADG (Viitala et al., 2003; Kneeland et al., 2004). The strongest association signal in HOL was found on chromosome 18 around 58 Mbp (HOL-CHR18–58MB). The most significant SNP on this region was located at 58,118,935 bp (rs521076153), very close to the zinc finger protein 613 gene (*ZNF613*; 58,115,782–58,117,110 bp) and zinc finger protein 717 gene (*ZNF717*; 58,130,465–58,141,877 bp), which have been associated with calving difficulties and which have been reported to be associated with calving traits (Mao et al., 2015). It is also very close to the QTL (BTA18: SIGLEC12) reported to be associated with productive life (Ma et al., 2012). This QTL on chromosome 18 also confirms the QTL reported previously for service-sire stillbirth and daughter calving ease (Cole et al., 2011). Meanwhile, this QTL (HOL-CHR18–58MB) overlaps with QTL associated with kidney, pelvic, and heart fat; live weight; and marbling score (MacNeil and Grosz, 2002). This SNP explained 3.45% of the variance in the deregressed proofs of longevity (Supplemental Table S1; <http://dx.doi.org/10.3168/jds.2015-10697>).

**RDC.** In Nordic RDC, 2 significantly associated regions were present on chromosome 6. The most sig-

nificant SNP on RDC-CHR6–88MB had the same locations as SNP on HOL-CHR6–88MB in HOL on chromosome 6 (Figure 3 and Supplemental Table S2; <http://dx.doi.org/10.3168/jds.2015-10697>). However, the QTL on chromosome 6 was not identified in the JER population, which is in line with previous findings for clinic mastitis traits (Sahana et al., 2014). Another region, RDC-CHR6–46MB at 46,099,648 bp (rs133661446) on chromosome 6, was located between 2 genes: coiled-coil domain containing 149 (*CCDC149*; 45,940,147–46,057,441 bp) and leucine-rich repeat LGI family, member 2 (*LGI2*; 46,127,372–46,155,777 bp). RDC-CHR6–46MB also overlaps with QTL associated with longissimus muscle area, hot carcass weight, birth and yearling weight (Casas et al., 2000). RDC-CHR21–19MB overlaps with QTL associated with birth weight (Kneeland et al., 2004). The most significant SNP on chromosome 23 in RDC-CHR23–13MB was located at 13,185,828 bp. This genomic region closely overlaps with the gene potassium channel, 2 pore domain subfamily K, member 16 (*KCNK16*; 13,175,069–13,181,327 bp) and QTL associated with milk yield and preweaning average daily gain (Viitala et al., 2003; Kneeland et al., 2004). The SNP on RDC-CHR6–88MB with highest significance explained 1.4% of the total variance in the deregressed proofs of longevity (Table 1). Mao et al. (2015) reported a major QTL affecting birth index in this HOL population, with the most associated SNP at 57,548,213 bp that explained 4.16% of the total genetic variance for birth index. It is possible that this QTL increases calving difficulties by increasing calf size at birth. Calving difficulties and related complications may lead to culling.

The genomic inflation (lambda) values for HOL, JER, RDC, and meta-analysis were 1.69, 1.30, 1.46, and 1.01, respectively (Supplemental Figures S1–S4; <http://dx.doi.org/10.3168/jds.2015-10697>). In the meta-analysis, the test-statistics were pre-adjusted for the inflation and therefore lambda was close to 1. The highest lambda value was observed for HOL in the within breed analyses. The trait analyzed here is polygenic in nature and phenotypes (DRP) had high reliability; besides, extensive linkage disequilibrium (LD) is present in these breeds due to small effective population sizes. All the markers in LD with the causal factor will show an effect on the trait proportional to their LD ( $r^2$ ) with the causal variant. This is also evident from the fact that we see higher inflation in HOL where we have more power (larger sample sizes) and high LD. The extent of LD is high in HOL compared with the Red Dairy cattle, which are a combination of 3 sub-populations. Jerseys had the smallest sample size, and we observed the lowest inflation. Therefore, we can

**Table 1.** The significantly associated genomic regions with longevity in meta-analysis across 3 cattle breeds

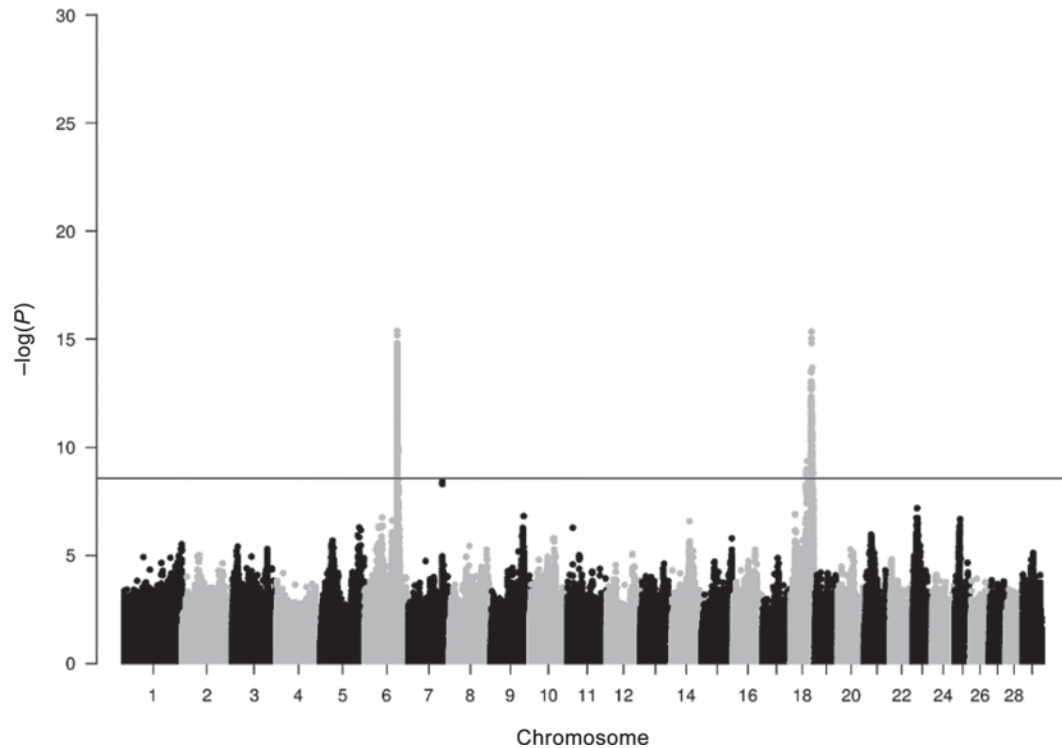
Region	Highest associated SNP position (bp)	No. of significantly associated SNP	Start and end positions of the region (bp)	Gene annotation	SNP reference ID	Sample size	Z-score	P-value
META-CHR6-88MB	88840407			Intergenic variant	rs381295092	9,571	-8.13	4.23E-16
	88899845			Intergenic variant	rs468282389	9,571	8.08	6.73E-16
	88889152	2,565	88025522-88995254	Intergenic variant	rs109723567	9,571	7.97	1.55E-15
	88892105			Intergenic variant	rs379544747	9,571	-7.93	2.16E-15
	88889420			Intergenic variant	rs207890656	9,571	7.93	2.18E-15
META-CHR18-58MB	58015050	596	58014529-58980943	Upstream gene variant	rs445560689	9,571	-8.12	4.55E-16
	58141989			Downstream gene variant	rs381577268	9,571	-8.04	9.35E-16
	58118935			Upstream gene variant	rs521076153	9,571	7.97	1.53E-15
	58405555			Intergenic variant	rs479824187	5,334	7.41	1.31E-13
	58410629			Intergenic variant	rs476912077	5,334	-7.40	1.36E-13

conclude that the inflation we see here is primarily due to the genetic architecture of the trait, sample sizes and extent of LD. According to Yang et al. (2011), “In the absence of population structure and other technical artefacts, but in the presence of polygenic inheritance, substantial genomic inflation is expected. Its magnitude depends on sample size, heritability, LD structure and the number of causal variants.” However, we used the leave-one-chromosome approach for building the relationship matrix. Therefore, we cannot exclude the possibility that some local structure remained unaccounted for, resulting in some inflation in test statistics.

### Meta-Analysis

Meta-analyses of association results from the 3 cattle breeds did not reveal additional QTL for longevity. The 2 genomic regions on chromosomes 6 and 18 that were identified in within-breed analyses were significant in the meta-analyses (Figure 4). The top associated regions are reported in Table 1. The most significantly associated region from the meta-analysis was located on chromosome 6 (META-CHR6-88MB), and the SNP with the highest significance (Chr6:88840407:rs381295092) had an allelic substitution effect of 2.41 for the HOL population (Supplemental Tables S1 and S2; <http://dx.doi.org/10.3168/jds.2015-10697>). META-CHR6-88MB overlaps with 2 genes: neuro-peptide G-protein coupled receptor gene (*NPFFR2*; 89,052,210–89,059,348 bp) and vitamin D-binding protein precursor (*GC*; 88,695,940–88,739,180 bp). This genomic region was reported to harbor mastitis QTL in HOL and RDC (Wu et al., 2015), suggesting that multiple causal variants for mastitis may be located in this region. The second most significantly associated region was META-CHR18-58MB, located very near the zinc finger protein 613 gene (*ZNF613*; 58,115,782–58,117,110 bp) and zinc finger protein 717 gene (*ZNF717*; 58,130,465–58,141,877 bp), which have been associated with calving difficulties.

We studied the LD pattern between the top associated SNP (Chr6:88840407,  $P = 4.23E-16$ ) from the meta-analysis and other significant SNP at META-CHR6-88MB. The most significant SNP was located upstream of *GC* and *NPFFR2*. Association results for significant SNP and the level of LD with the top SNP are presented in Supplemental Figure S5 (<http://dx.doi.org/10.3168/jds.2015-10697>). Results of the search for additional QTL after adjusting phenotypes for the effect of Chr6:88840407 are presented in Supplemental Figure S6 (<http://dx.doi.org/10.3168/jds.2015-10697>). Many of the highly significant SNP within this region were in high LD with Chr6:88840407. Several SNP were located around 88.88 Mbp with low LD with Chr6:



**Figure 4.** Manhattan plot for genome-wide association scan for longevity using whole-genome sequence variants in the meta-analysis of 3 dairy cattle breeds.

88840407 ( $0.20 < r^2 \leq 0.40$ ), as well as around 89.05 Mbp. This result could indicate the presence of additional causal variant(s) in the latter region. However, no significantly associated SNP were observed when an association analysis was carried out for residuals from a model that included the top associated marker (Chr6:88840407) and the polygenic component. The *P*-values of other SNP were under the threshold for significance (Supplemental Figure S6; <http://dx.doi.org/10.3168/jds.2015-10697>). Therefore, it is most likely that only one causal variant affecting longevity is segregating on chromosome 6 between 88 and 90 Mbp.

Analysis of LD between the second most significant region on chromosome 18 and the top associated SNP (Chr18:58015050: rs445560689,  $P = 4.55\text{E-}16$ ) revealed SNP with high and low LD (Supplemental Figure S7; <http://dx.doi.org/10.3168/jds.2015-10697>). There might be multiple causal variants in this region. When the top associated marker (Chr18:58015050) and the polygenic component were included in an association analysis for residuals, the *P*-values of other SNP were below the threshold for significance (Supplemental Figure S8; <http://dx.doi.org/10.3168/jds.2015-10697>). Therefore, it is likely that only one causal variant affecting longevity is segregating on chromosome 18 between 57 and 59 Mbp.

### General Discussion

Only the QTL on chromosome 6 overlapped in HOL and RDC in the within-breed analyses. Meta-analysis did not identify additional QTL. The RDC cattle consist of 3 RDC populations from Denmark, Sweden, and Finland. The population structure of breeds included in the present study has been reported (Kadri et al., 2015; Zhang et al., 2015a,b). Jersey is clearly separated and forms a tight cluster apart from the other populations. The admixture appears to be stronger among the 3 RDC subpopulations. Among the RDC breeds, the population from Denmark seems to be more closely related to HOL, as a result of the import of genes from Red Holstein in the 1970s to avoid inbreeding depression. It is possible that the major QTL segregating in these breeds are of recent origin and unique to the breeds. Some common QTL across these cattle breeds did not show significant association in within-breed analyses, perhaps due to a lack of power because of small effects on the trait or extremely low minor allele frequency. Genetic variants having major effects on longevity are segregating in these breeds, which could be the result of separation for many generations followed by strong artificial selection, as well as random drift due to small effective population sizes.



The QTL on chromosome 6 overlaps with mastitis resistance QTL in the same cattle populations (Sahana et al., 2014). The longevity trait could be influenced by mastitis, given the strong negative genetic correlation between mastitis resistance and risk of being culled ( $-0.40$  to  $-0.53$ ; Neerhof et al., 2000; Roxstrom and Strandberg, 2002). Multiple occurrences of mastitis may cause permanent damage to the udder, leading a cow to be culled from the herd. Alternatively, animals suffering from mastitis may have low immunity to other infections (Sordillo, 2005), which would increase their chances of being culled. The QTL on chromosome 18 overlaps with calf size and calving difficulty QTL in the same populations (Mao et al., 2015). A large size calf can cause damage of the birth canal of a cow during calving, which may lead to a higher risk of culling.

Overall, the most significantly associated SNP were not located within coding regions of genes or were between genes with unknown annotation. Thus, a link between cattle longevity and functions of the identified variants was not obvious. It was not possible to select one candidate gene, even for regions where the associated peaks were located within or near genes (e.g., META-CHR6-88MB and META-CHR18-58MB), due to LD among significant SNP from associated regions. Several other identified genomic regions associated with longevity harbor disease genes, including *KCNK16*, *PPP1R14C*, and *GCH1*. These disease genes and their pathways could be studied further to short-list candidate genes for the longevity trait.

### Data Availability

Data used in this study are from the 1,000 bull genomes project (Daetwyler et al., 2014). Whole-genome sequence data of individual bulls of the 1,000 bull genomes project are already available at the National Center for Biotechnology Information using SRA no. SRP039339 (<http://www.ncbi.nlm.nih.gov/bioproject/PRJNA238491>).

### ACKNOWLEDGMENTS

We are grateful to the Danish Cattle Federation/NAV (Aarhus, Denmark) for providing the phenotypic data used in this study and Viking Genetics, Randers, Denmark, for proving samples for genotyping. Q. Zhang benefited from a joint grant from the European Commission within the framework of the Erasmus-Mundus joint doctorate “EGS-ABG.” This research was supported by the Center for Genomic Selection in Animals and Plants (GenSAP) funded by The Danish Council for Strategic Research (Copenhagen, Denmark). The authors declare that they have no competing inter-

ests. QZ developed and planned the study design, coordinated the study, performed data analyses, and drafted the manuscript; BG participated in the study design, analyses of data, and drafting the manuscript; JRT participated in the study design and drafting the manuscript; MSL participated in the study design and drafting the manuscript; GS participated in the study design and drafting the manuscript. All authors read and approved the final manuscript.

### REFERENCES

- Arias, J., and B. Kirkpatrick. 2004. Mapping of bovine ovulation rate QTL; an analytical approach for three generation pedigrees. *Anim. Genet.* 35:7–13.
- Brøndum, R. F., G. Su, L. Janss, G. Sahana, B. Guldbrandtsen, D. Boichard, and M. S. Lund. 2015. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. *J. Dairy Sci.* 98:4107–4116.
- Casas, E., S. D. Shackleford, J. W. Keele, R. T. Stone, S. M. Kappes, and M. Koohmaraie. 2000. Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms of myostatin. *J. Anim. Sci.* 78:560–569.
- Cole, J. B., G. R. Wiggans, L. Ma, T. S. Sonstegard, T. J. Lawlor, B. A. Crooker, C. P. Van Tassell, J. Yang, S. W. Wang, L. K. Matukumalli, and Y. Da. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. *BMC Genomics* 12:408.
- Daetwyler, H. D., A. Capitan, H. Pausch, P. Stothard, R. Van Binsbergen, R. F. Brøndum, X. P. Liao, A. Djari, S. C. Rodriguez, C. Grohs, D. Esquerre, O. Bouchez, M. N. Rossignol, C. Klopp, D. Rocha, S. Fritz, A. Eggen, P. J. Bowman, D. Coote, A. J. Chatukmalin, C. Anderson, C. P. VanTassell, I. Hulsege, M. E. Goddard, B. Guldbrandtsen, M. S. Lund, R. F. Veerkamp, D. A. Boichard, R. Fries, and B. J. Hayes. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* 46:858–865.
- Essl, A. 1998. Longevity in dairy cattle breeding: A review. *Livest. Prod. Sci.* 57:79–89.
- Fuchsberger, C., G. R. Abecasis, and D. A. Hinds. 2015. minimac2: Faster genotype imputation. *Bioinformatics* 31:782–784.
- Hiendler, S., H. Thomsen, N. Reinsch, J. Bennewitz, B. Leyhe-Horn, C. Looft, N. Xu, I. Medjugorac, I. Russ, C. Kuhn, G. A. Brockmann, J. Blumel, B. Brenig, F. Reinhardt, R. Reents, G. Averdunk, M. Schwerin, M. Forster, E. Kalm, and G. Erhardt. 2003. Mapping of QTL for body conformation and behavior in cattle. *J. Hered.* 94:496–506.
- Howe, B., J. Marchini, and M. Stephens. 2011. Genotype imputation with thousands of genomes. *G3: Genes, Genomes, Genetics* 1:457–470.
- Iso-Touru, T., G. Sahana, B. Guldbrandtsen, M. S. Lund, and J. Vilki. 2016. Genome-wide association analysis of milk yield traits in Nordic Red Cattle using imputed whole genome sequence variants. *BMC Genet.* 17:55.
- Kadri, N. K., B. Guldbrandtsen, M. S. Lund, and G. Sahana. 2015. Genetic dissection of milk yield traits and mastitis resistance quantitative trait loci on chromosome 20 in dairy cattle. *J. Dairy Sci.* 98:9015–9025.
- Kang, H. M., J. H. Sul, S. K. Service, N. A. Zaitlen, S. Y. Kong, N. B. Freimer, C. Sabatti, and E. Eskin. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42:348–354.
- Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, M. J. Daly, and E. Eskin. 2008. Efficient control of population structure in model organism association mapping. *Genetics* 178:1709–1723.

- Kneeland, J., C. Li, J. Basarab, W. M. Snelling, B. Benkel, B. Murdoch, C. Hansen, and S. S. Moore. 2004. Identification and fine mapping of quantitative trait loci for growth traits on bovine chromosomes 2, 6, 14, 19, 21, and 23 within one commercial line of *Bos taurus*. *J. Anim. Sci.* 82:3405–3414.
- Listgarten, J., C. Lippert, E. Y. Kang, J. Xiang, C. M. Kadie, and D. Heckerman. 2013. A powerful and efficient set test for genetic markers that handles confounders. *Bioinformatics* 29:1526–1533.
- Lund, M. S., A. P. W. de Roos, A. G. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. T. Liu, R. Reents, C. Schrooten, F. Seefried, and G. S. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet. Sel. Evol.* 43:43.
- Ma, L., G. R. Wiggans, S. W. Wang, T. S. Sonstegard, J. Yang, B. A. Crooker, J. B. Cole, C. P. Van Tassell, T. J. Lawlor, and Y. Da. 2012. Effect of sample stratification on dairy GWAS results. *BMC Genomics* 13:536.
- MacNeil, M. D., and M. D. Grosz. 2002. Genome-wide scans for QTL affecting carcass traits in Hereford x composite double backcross populations. *J. Anim. Sci.* 80:2316–2324.
- Madsen, P., J. Jensen, R. Labouriau, O. F. Christensen, and G. Sahana. 2014. DMU-A Package for Analyzing Multivariate Mixed Models in quantitative Genetics and Genomics. In *Proc. 10th World Congress of Genetics Applied to Livestock Production*, Vancouver, BC, Canada.
- Mao, X., N. K. Kadri, J. R. Thomasen, D. J. De Koning, G. Sahana, and B. Guldbrandtsen. 2015. Fine mapping of a calving QTL on *Bos taurus* autosome 18 in Holstein cattle. *J. Anim. Breed. Genet.* 133:207–218. <http://dx.doi.org/10.1111/jbg.12187>.
- Neerhof, H. J., P. Madsen, V. P. Ducrocq, A. R. Vollema, J. Jensen, and I. R. Korsgaard. 2000. Relationships between mastitis and functional longevity in Danish black and white dairy cattle estimated using survival analysis. *J. Dairy Sci.* 83:1064–1071.
- Pausch, H., R. Emmerling, H. Schwarzenbacher, and R. Fries. 2016. A multi-trait meta-analysis with imputed sequence variants reveals twelve QTL for mammary gland morphology in Fleckvieh cattle. *Genet. Sel. Evol.* 48:14.
- Prinzenberg, E. M., C. Weimann, H. Brandt, J. Bennewitz, E. Kalm, M. Schwerin, and G. Erhardt. 2003. Polymorphism of the bovine CSN1S1 promoter: Linkage mapping, intragenic haplotypes, and effects on milk production traits. *J. Dairy Sci.* 86:2696–2705.
- Pritchard, T., M. Coffey, R. Mrode, and E. Wall. 2013. Understanding the genetics of survival in dairy cows. *J. Dairy Sci.* 96:3296–3309.
- Raven, L. A., B. G. Cocks, and B. J. Hayes. 2014. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. *BMC Genomics* 15:62.
- Roxstrom, A., and E. Strandberg. 2002. Genetic analysis of functional, fertility-, mastitis-, and production-determined length of productive life in Swedish dairy cattle. *Livest. Prod. Sci.* 74:125–135.
- Sahana, G., B. Guldbrandtsen, B. Thomsen, L. E. Holm, F. Panitz, R. F. Brondum, 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.
- Sahana, G., J. K. Hoglund, B. Guldbrandtsen, and M. S. Lund. 2015. Loci associated with adult stature also affect calf birth survival in cattle. *BMC Genet.* 16:47.
- Sordillo, L. M. 2005. Factors affecting mammary gland immunity and mastitis susceptibility. *Livest. Prod. Sci.* 98:89–99.
- Stone, R. T., J. W. Keele, S. D. Shackelford, S. M. Kappes, and M. Koohmaraie. 1999. A primary screen of the bovine genome for quantitative trait loci affecting carcass and growth traits. *J. Anim. Sci.* 77:1379–1384.
- Viitala, S. M., N. F. Schulman, D. J. de Koning, K. Elo, R. Kinoshita, A. Virta, J. Virta, A. Maki-Tanila, and J. H. Vilkkki. 2003. Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. *J. Dairy Sci.* 86:1828–1836.
- Wiener, P., I. Maclean, J. L. Williams, and J. A. Woolliams. 2000. Testing for the presence of previously identified QTL for milk production traits in new populations. *Anim. Genet.* 31:385–395.
- Wientjes, Y. C. J., M. P. L. Calus, M. E. Goddard, and B. J. Hayes. 2015. Impact of QTL properties on the accuracy of multi-breed genomic prediction. *Genet. Sel. Evol.* 47:42.
- Willer, C. J., Y. Li, and G. R. Abecasis. 2010. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26:2190–2191.
- Wu, X., M. S. Lund, G. Sahana, B. Guldbrandtsen, D. X. Sun, Q. Zhang, and G. S. Su. 2015. Association analysis for udder health based on SNP-panel and sequence data in Danish Holsteins. *Genet. Sel. Evol.* 47:50.
- Yang, J., M. N. Weedon, S. Purcell, G. Lettre, K. Estrada, C. J. Willer, A. V. Smith, E. Ingelsson, J. R. O'Connell, M. Mangino, R. Mägi, P. A. Madden, A. C. Heath, D. R. Nyholt, N. G. Martin, G. W. Montgomery, T. M. Frayling, J. N. Hirschhorn, M. I. McCarthy, M. E. Goddard, and P. M. Visscher. 2011. Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* 19:807–812.
- Yang, J., N. A. Zaitlen, M. E. Goddard, P. M. Visscher, and A. L. Price. 2014. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* 46:100–106.
- Zhang, Q., D. Boichard, I. Hoeschele, C. Ernst, A. Eggen, B. Murkve, M. Pfister-Genskow, L. A. Witte, F. E. Grignola, P. Uimari, G. Thaller, and M. D. Bishop. 1998. Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. *Genetics* 149:1959–1973.
- Zhang, Q., M. P. L. Calus, B. Guldbrandtsen, M. S. Lund, and G. Sahana. 2015a. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genet.* 16:88.
- Zhang, Q., B. Guldbrandtsen, M. Bosse, M. S. Lund, and G. Sahana. 2015b. Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC Genomics* 16:542.

Copyright of Journal of Dairy Science is the property of Elsevier Science and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.