J. Dairy Sci. 100:1–15
https://doi.org/10.3168/jds.2017-12688

Association analysis for young stock survival index with imputed whole-genome sequence variants in Nordic Holstein cattle

Xiaoping Wu,* Bernt Guldbrandtsen,* Ulrik Sander Nielsen,† Mogens Sandø Lund,* and Goutam Sahana*[1]
*Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, 8830 Tjele, Denmark
†Livestock Innovation, SEGES, Danish Agricultural and Food Council F.m.b.A, 8200 Aarhus, Denmark

ABSTRACT

Identification of the genetic variants associated with calf survival in dairy cattle will aid in the elimination of harmful mutations from the cattle population and the reduction of calf and young stock mortality rates. We used de-regressed estimated breeding values for the young stock survival (YSS) index as response variables in a genome-wide association study with imputed whole-genome sequence variants. A total of 4,610 bulls with estimated breeding values were genotyped with the Illumina BovineSNP50 (Illumina, San Diego, CA) single nucleotide polymorphism (SNP) genotyping array. Genotypes were imputed to whole-genome sequence variants. After quality control, 15,419,550 SNP on 29 *Bos taurus* autosomes (BTA) were used for association analysis. A modified mixed-model association analysis was used for a genome scan, followed by a linear mixed-model analysis for selected genetic variants. We identified 498 SNP on BTA5 and BTA18 that were associated with the YSS index in Nordic Holstein. The SNP rs440345507 (Chr5:94721790) on BTA5 was the putative causal mutation affecting YSS. Two haplotype-based models were used to identify haplotypes with the largest detrimental effects on YSS index. For each association signal, 1 haplotype region with harmful effects and the lead associated SNP were identified. Detected haplotypes on BTA5 and BTA18 explained 1.16 and 1.20%, respectively, of genetic variance for the YSS index. We examined whether YSS quantitative trait loci (QTL) on BTA5 and BTA18 were associated with stillbirth. YSS QTL on BTA18 overlapped a QTL region for stillbirth, but most likely 2 different causal variants were responsible for these 2 QTL. Four component traits of the YSS index, defined by sex and age, were analyzed separately by the modified mixed-model approach. The same genomic regions were associated with both bull and heifer calf mortality. Several genes (*EPS8*, *LOC100138951*, and *KLK* family genes) contained a lead associated SNP or were included in haplotypes with large detrimental effects on YSS in Nordic Holstein cattle.

Key words: young stock survival, calf mortality, genome-wide association, quantitative trait loci

INTRODUCTION

Calves and young stock that die during the rearing period result in lost revenue for dairy farmers, fewer heifers for replacement, veterinary costs, and adverse effects on animal welfare. Juvenile death rates among Danish Holstein calves born between 2008 and 2012 were 7.5% in heifer calves and 10% in bull calves (Pedersen et al., 2014). Mortality during the rearing period has greater economic consequences than early embryo loss, abortion, or stillbirth.

Part of the variation in survival has a genetic basis. Fuerst-Waltl and Sørensen (2010) reported heritability values between 0 and 0.08, depending on the age of calves investigated. Norberg et al. (2013) presented heritability for mortality between 0 and 0.03 in the period from 24 h after birth to 180 d in Danish Jersey heifer calves. Some of these deaths may be due to the action of recessive lethal alleles. Due to the widespread use of a limited number of elite dairy cattle bulls, some harmful recessive alleles have spread in the population. These deleterious alleles frequently go unnoticed because the identification of a carrier bull requires mating between its descendants. Mutations causing embryonic lethality and stillbirth have been reported in Nordic dairy cattle by using genomic data (Sahana et al., 2013, 2016; Kadri et al., 2014). However, recessive lethal mutations have not been reported for young stock mortality. An index for young stock survival (**YSS**) in calves was included in the Nordic total merit index by the Nordic Cattle Genetic Evaluation (**NAV**; www.nordicebv.info). Pedersen et al. (2015) reported that correlations between the YSS index and other Nordic total merit indices were generally close to 0. The strongest positive correlations (between 0.1 and 0.2) were observed with 2

Received February 4, 2017.
Accepted April 5, 2017.
[1]Corresponding author: Goutam.sahana@mbg.au.dk

indices, health index and longevity. To improve YSS, it is necessary to select for YSS directly.

The number of dairy cattle with genomic SNP array data has increased rapidly in recent years. The combination of genomic information and breeding values for YSS creates an opportunity to identify genomic variants with harmful effects on YSS. Once these variants have been identified, it will be possible to select against them and prevent at-risk matings between carriers. This selection will improve calf survival and reduce the cost per live cattle produced. Quantitative trait locus information can be used to improve accuracies for genomic prediction (Brøndum et al., 2015). The aim of this study was to identify QTL affecting YSS in Nordic Holstein (**NH**) cattle.

MATERIALS AND METHODS

Phenotype and Genotype Data

We analyzed data from 4,610 NH bulls born in 1998 or later with EBV for YSS. The YSS index was calculated based on 4 survival traits: survival from 2 to 30 d for bull calves (**BP1**) and heifer calves (**HP1**), from 31 to 184 d for young bull calves (**BP2**), and from 31 to 458 d for young heifer calves (**HP2**). Calf death and survival during this period were recorded as 0 and 1, respectively. Calves slaughtered, exported, or with missing records were recorded as missing. The YSS index was calculated by combining EBV for BP1, BP2, HP1, and HP2 by NAV (Denmark), which were weighted by their relative economic values and standardized (Pedersen, 2015).

The breeding value estimation procedure for YSS by NAV was described by Pedersen et al. (2014). Phenotypic data for YSS traits were precorrected for heterosis (by applying regression on the expected total heterosis of all included populations), country (Danish, Swedish, and Finnish data were combined), calf size, calving ease, effect of transfer to another herd, and herd-by-year effect. De-regressed EBV (**DRP**) were derived for animals based on the effective daughter contributions of sire and maternal grandsire (Goddard, 1985; Schaeffer, 1985) by using MiX99 software (Vuori et al., 2006). Supplemental Table S1 (https://doi.org/10.3168/jds.2017-12688) lists descriptive statistics of DRP and reliabilities of the YSS index, its components traits, and stillbirth. Histograms of DRP distributions for BP1, BP2, HP1, HP2, and the YSS index are presented in Supplemental Figure S1 (https://doi.org/10.3168/jds.2017-12688). Heritability estimates for the YSS index and its component traits were reported as being in the range of 0.007 to 0.027 (Pedersen et al., 2014). The phenotypic correlation (DRP) among the YSS index and its components traits was ~0.78 on average (Supplemental Table S2; https://doi.org/10.3168/jds.2017-12688).

An association study for the YSS index was carried out by using imputed whole-genome sequence (**WGS**) data. All bulls were genotyped with the Illumina BovineSNP50 BeadChip (54k; version 1 or 2; Illumina, San Diego, CA). The 54k genotypes were first imputed to the Illumina BovineHD marker set and then to the full WGS level (Iso-Touru et al., 2016; Wu et al., 2016). A total of 22,751,039 biallelic variants (SNP and indels) were imputed from WGS reference data. The SNP with a minor allele frequency less than 1% or those deviating from Hardy-Weinberg proportions ($P < 10^{-6}$) were removed. Eventually, 15,419,550 SNP remained for association analysis. The position of each SNP was defined according to the *Bos taurus* genome assembly UMD3.1 (Zimin et al., 2009). Genes located within or overlapping with top associated SNP were determined by using information from the Variant Effect Predictor tool version 87 of Ensembl (McLaren et al., 2010).

To examine whether YSS QTL were associated with stillbirth, detected QTL regions for the YSS index were analyzed for 5,484 NH bulls with DRP for stillbirth for first parity (4,575 records overlap with YSS). Stillbirth was defined as a calf that was born dead or died within 24 h after birth. Stillbirth might be affected by some of the same genetic factors as mortality at a very early age in life. The EBV of stillbirth was estimated by a multitrait sire model with direct and maternal effects by NAV (http://www.nordicebv.info/dk/).

Association Analyses

A modified linear mixed-model approach (efficient mixed-model association expedited, **EMMAX**) (Kang et al., 2008, 2010) was used to detect associations between imputed sequence variants and YSS index. For details of model description and analysis, see Iso-Touru et al. (2016) and Wu et al. (2016). Significantly associated SNP from the above analysis were reanalyzed using a linear mixed model (**LMM**; Yu et al., 2006), which included a polygenic effect to adjust for familial relatedness and population structures and a fixed effect of the marker. Details of the statistical models and analytical approach were presented by Wu et al. (2016).

Association analyses for BP1, BP2, HP1, and HP2 were only carried out using EMMAX. For each model, a Bonferroni correction was applied to control for false-positive associations that arose due to multiple testing. A SNP was declared significant if its *P*-value was less than 0.05/M, where M is the number of SNP (= 15,419,550). The resulting threshold was $-\log_{10}(P) > 8.49$. For each chromosome, SNP were considered

to be in the same QTL region if the distance between 2 adjacent significant SNP was less than 1 Mb. For each QTL region, the SNP with the lowest *P*-value was designated as the lead SNP.

To check whether the detected lead SNP were recessive lethal mutations, we examined the existence of homozygote animals for the alternative (i.e., nonreference) allele for the lead SNP among the 1,577 sequenced animals, including 450 Holsteins in *run5* of the 1,000 Bull Genomes Project (Daetwyler et al., 2014).

Two haplotype-based association analyses were carried out to identify bulls carrying haplotypes with harmful effects on YSS. Haplotype-based association can fully exploit linkage equilibrium information from multiple markers (Boleckova et al., 2012). In a fixed haplotype model, effect estimates for the low-frequency haplotype could have low estimation accuracy (Becker and Herold, 2009). Therefore, we initially used a random haplotype model (**RHM**), in accordance with Boleckova et al. (2012), followed by the fixed haplotype dosage model (**FHM**) to estimate the effect for a specific haplotype. Imputed high-density 777k genotypes (Wu et al., 2016) for 4,610 animals were used to construct haplotypes by using sliding windows of 15 neighboring SNP covering the associated significant genomic region, as demarcated in the single-marker analysis.

RHM. Each animal carries 2 haplotypes, 1 paternal and 1 maternal in origin. The RHM model of Boleckova et al. (2012) was used:

$$y_i = \mu + h_{i1} + h_{i2} + u_i + e_i,$$

where y_i was the phenotype of bull i; μ was the overall mean; h_{i1} and h_{i2} were random effects of the 2 haplotypes of animal i, assumed to be normally distributed as $N(\mathbf{0}, \mathbf{I}\sigma_h^2)$, where $\mathbf{I}$ was an identity matrix and σ_h^2 was haplotype variance; u_i was the random individual polygenic effect, the vector of u_i, which followed a multi-

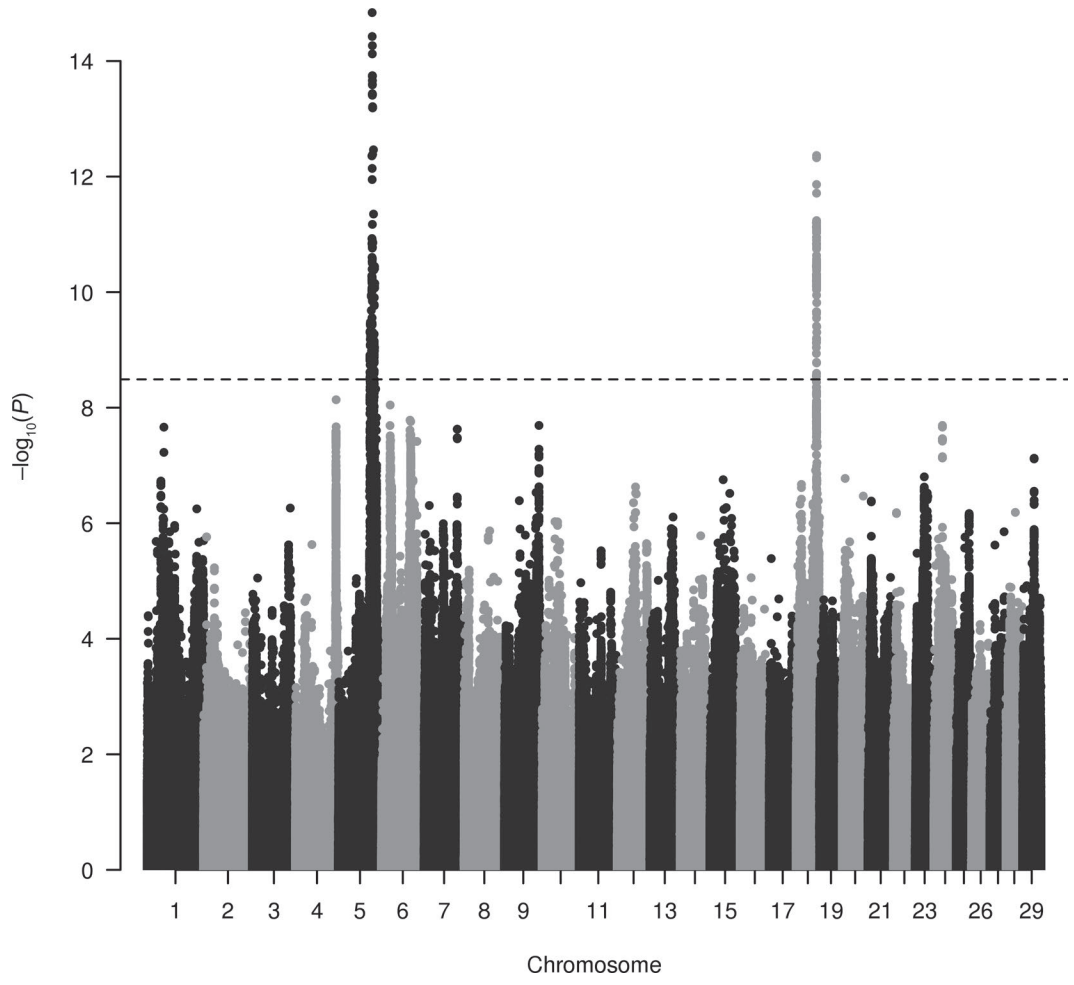

Figure 1. Manhattan plot for association of SNP with young stock survival index for Nordic Holstein cattle, obtained by efficient mixed-model association analysis. Black dashed line indicates genome-wide significance level [$-\log_{10}(P) = 8.49$].

variate normal distribution $\mathbf{u} \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2)$, where $\mathbf{A}$ was the pedigree-based additive relationship matrix and σ_u^2 was polygenic variance; e_i was the random residual effect, the vector of e_i, which followed a multivariate normal distribution, $\mathbf{e} \sim N(\mathbf{0}, \mathbf{D}\sigma_e^2)$, where $\mathbf{D}$ was a diagonal matrix with elements $d_{ii} = (1 - r_{DRP}^2) / r_{DRP}^2$ to account for heterogeneous residual variances due to the different reliabilities of DRP, where d_{ii} is the ith diagonal element of the matrix $\mathbf{D}$ and r_{DRP} is the accuracy for DRP; and σ_e^2 was the residual variance. The significance of the haplotype substitution effect was assessed with a likelihood ratio test comparing the RHM with a null model containing the mean, polygenic effect, and random error terms but no haplotype effects (i.e., $\sigma_h^2 = 0$). Similarly to the single-marker analysis, haplotypes with a P-value less than 3.24×10^{-9} were considered significant. The model predicted the effect of each haplotype, and the haplotype with the smallest P-value was termed the lead haplotype.

FHM. We carried out FHM for the lead haplotype in RHM analysis for each QTL region. Haplotype dosage was modeled as a fixed effect. The FHM was described by

$$y_i = \mu + q_{h_i} + u_i + e_i,$$

where y_i, μ, u_i, and e_i were defined above; and q_{h_i} was the fixed additive genetic effect of the analyzed haplotype, taking values 0, 1, and 2 for animals carrying 0, 1, and 2 copies of the lead haplotype, respectively. The RHM and FHM analyses were performed in the DMU package (Madsen et al., 2014). When haplotypes were tied for largest effect in RHM, these haplotypes were combined in the FHM.

Analysis for Stillbirth. The DRP were derived and an association analysis was conducted for stillbirth by using the same method as for the YSS index, except that for the association analysis only SNP associated with the YSS index from single-marker analysis (EM-

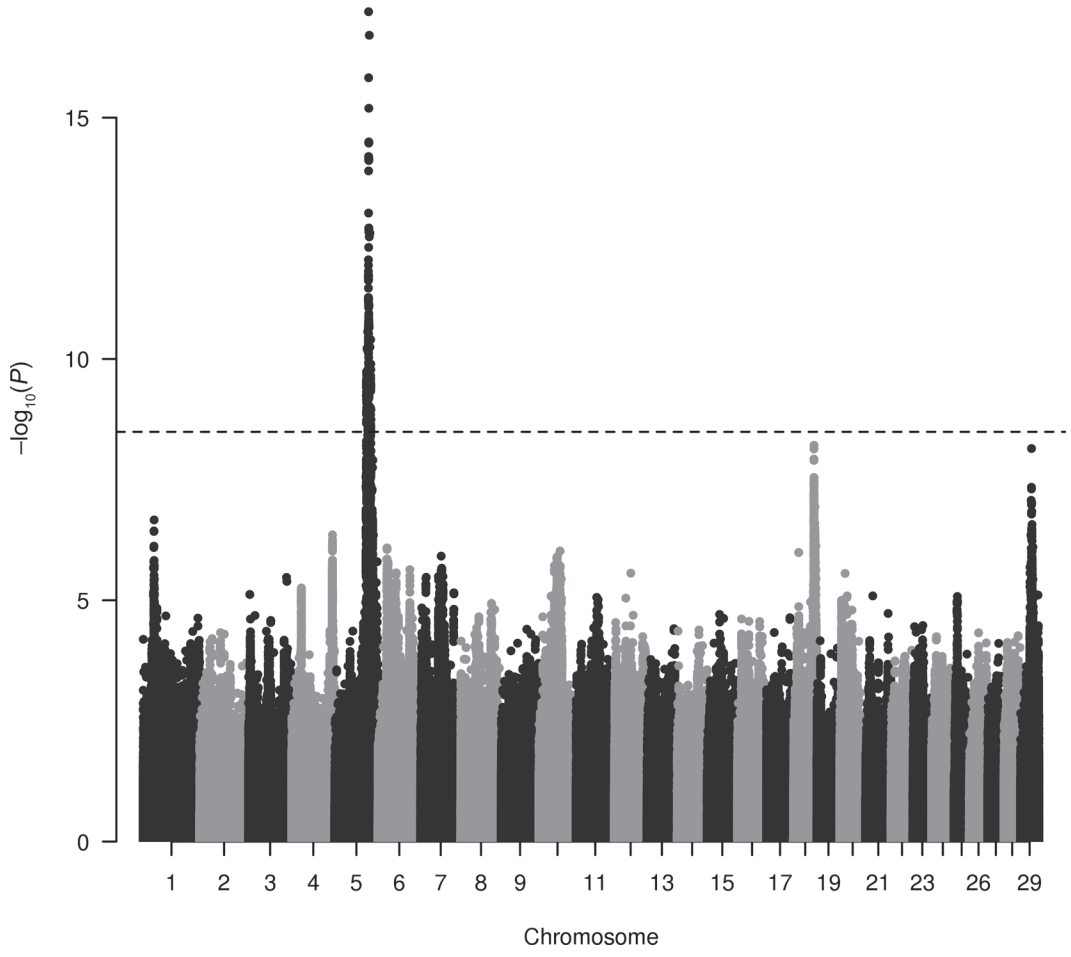

Figure 2. Manhattan plot for association of SNP with young stock survival of bull calves in period 1 for Nordic Holstein cattle, obtained by efficient mixed-model association analysis. Black dashed line indicates genome-wide significance level [$-\log_{10}(P) = 8.49$].

MAX) were used. Lead SNP for the YSS index and stillbirth were analyzed in a bivariate model (Kadri et al., 2015), with pedigree-based polygenic and residual effects as random. The SNP effects for the YSS index and stillbirth were compared between estimates from univariate and bivariate models.

RESULTS

Mixed-Model Analysis for the YSS Index

In the EMMAX analysis, 504 SNP were genome-wide significantly associated with the YSS index in NH (Figure 1). These SNP were located on 2 chromosomes: BTA5 (356 SNP at 88–101 Mb) and BTA18 (148 SNP at 56–58 Mb). Of the 504 SNP that were significant for the YSS index in the EMMAX model, 498 SNP exhibited significant associations in the LMM. The lead SNP on BTA5 (94,721,790; rs440345507) explained 1.74% and on BTA18 (57,577,417; rs477989930) explained 2.95% of the variance of DRP for the YSS index. Average correlation of SNP allele dosage among significant SNP was 0.61 on BTA5 and 0.93 on BTA18. The bull carrying the mutant allele on BTA5 shows lower average DRP for YSS index (Supplemental Figure S2A; https://doi.org/10.3168/jds.2017-12688), and no individual in our analysis show homozygous status for the mutant allele on BTA18 (Supplemental Figure S2B; https://doi.org/10.3168/jds.2017-12688). The top 10 significantly associated SNP on each chromosome are presented in Table 1.

EMMAX Analysis for Component Traits

Component traits of the YSS index were analyzed individually. The following genome-wide-associated SNP were detected: 391 SNP with BP1, 167 SNP with BP2, 337 SNP with HP1, and 57 SNP with HP2. Figures 2–5 are Manhattan plots for association of SNP with BP1, BP2, HP1, and HP2, respectively. The SNP associated

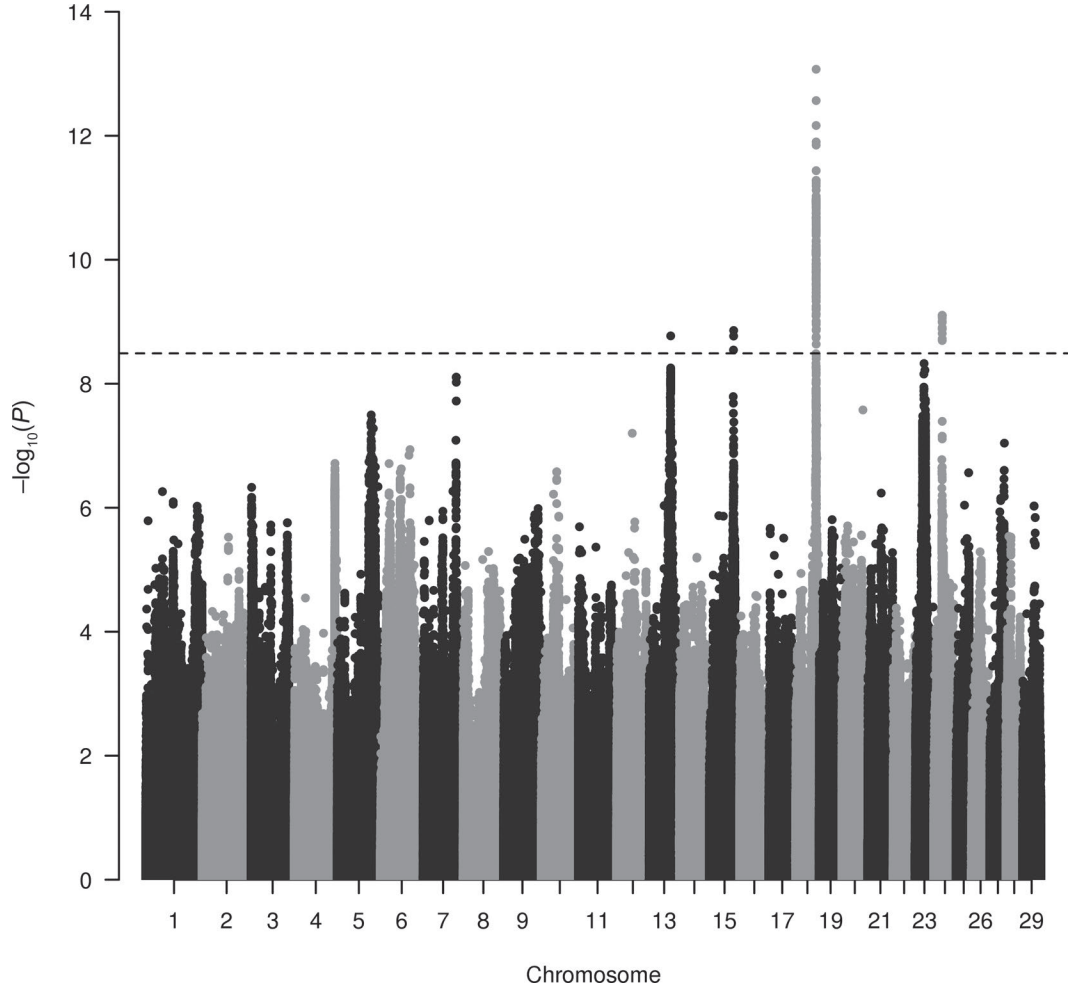

Figure 3. Manhattan plot for association of SNP with young stock survival of bull calves in period 2 for Nordic Holstein cattle, obtained by efficient mixed-model association analysis. Black dashed line indicates genome-wide significance level [$-\log_{10}(P) = 8.49$].

with BP1 were located on BTA5 (391 SNP at 88–100 Mb). The SNP associated with BP2 were located on BTA13 (1 SNP at 60 Mb), BTA15 (3 SNP at 67–68 Mb), BTA18 (148 SNP at 57–58 Mb), and BTA24 (15 SNP at 23 Mb). The SNP associated with HP1 were located on BTA5 (337 SNP at 88–100 Mb). The SNP associated with HP2 were located on BTA18 (57 SNP around 57 Mb).

Of the 498 significant SNP for the YSS index by LMM analysis, 257 SNP were also associated with BP1, 140 SNP with BP2, 250 SNP with HP1, and 57 SNP with HP2 (Figure 6). Additionally, 303 SNP on BTA5 were associated with both BP1 and HP1, and 57 SNP on BTA18 were associated with both BP2 and HP2 (Figure 7). The lead SNP was located at 94,721,790 bp on BTA5 (rs440345507) and 57,577,417 bp on BTA18 (rs477989930). Overlap of associated SNP was observed neither between BP1 and BP2, nor between HP1 and HP2 (Figure 8).

Haplotype Model Analysis for YSS Index

Based on the position of significant SNP in the single-variant analysis, the targeted region for the YSS index included 3,932 SNP on BTA5 (at 88,409,199–101,803,926 bp) and 338 SNP on BTA18 (at 56,855,383–58,141,989 bp) of the imputed high-density 777k chips. Haplotypes were constructed for markers from these 2 regions, and haplotype effects were predicted by using RHM for the YSS index (Figure 9). For the targeted region on BTA5, 411 haplotypes were significant by the likelihood ratio test (Figure 9A). The lead haplotype (HAP1) was located at 94,691,973 to 94,755,948 bp ($P = 1.80 \times 10^{-13}$). The lead SNP from single-variant analysis ($P = 4.56 \times 10^{-19}$) was also located within this haplotype. The frequency of HAP1 was 5.20%. The haplotype effect relative to the population mean was predicted as −0.34 (±1.03) in the DRP scale. For the targeted region on BTA18, 206 haplotypes were significant by the

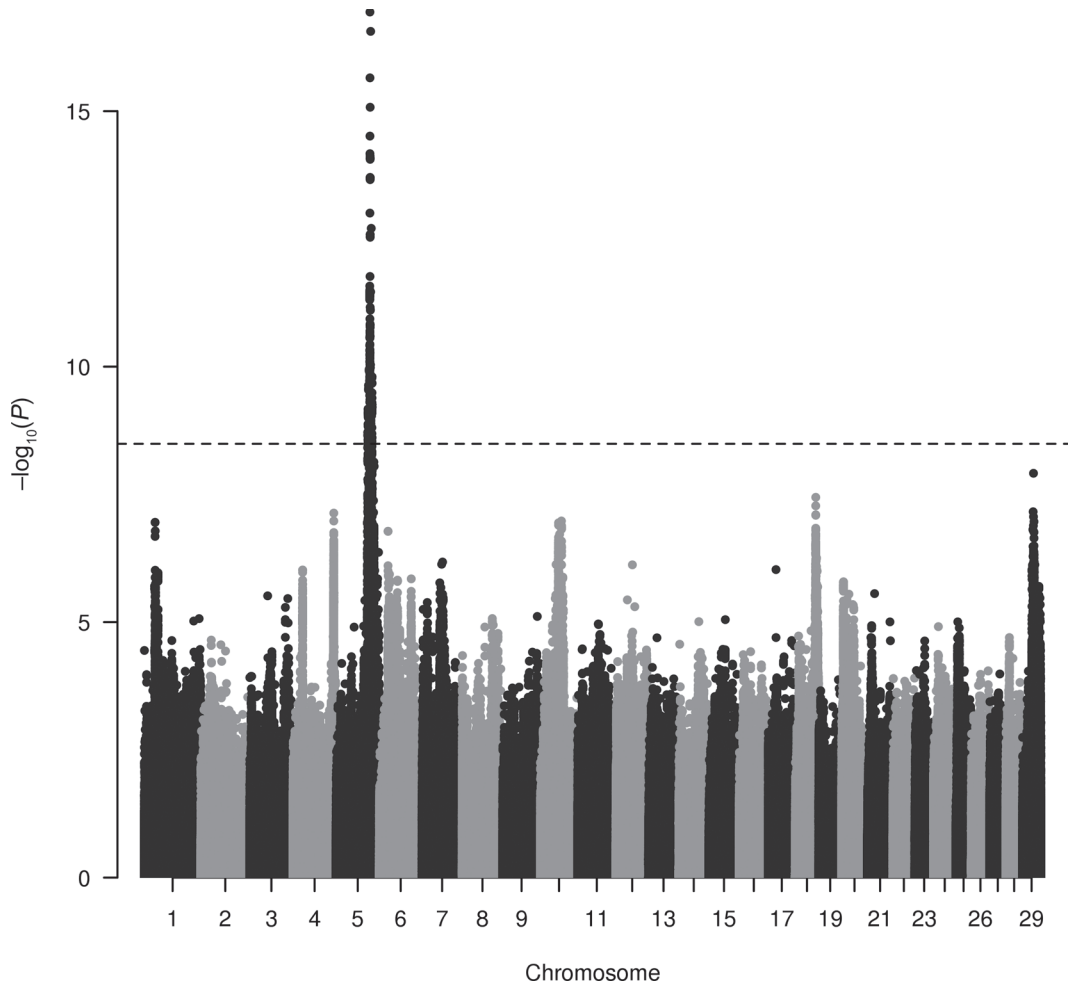

Figure 4. Manhattan plot for association of SNP with young stock survival of heifer calves in period 1 for Nordic Holstein cattle, obtained by efficient mixed-model association analysis. Black dashed line indicates genome-wide significance level [$-\log_{10}(P) = 8.49$].

Table 1. Top 10 SNP detected on chromosome (BTA) 5 and 18 for association with the young stock survival index in Nordic Holstein cattle[1]

BTA	Position (bp)	Reference ID	EMMAX (*P*-value)	LMM (*P*-value)	Annotation	Nearest gene	Distance (bp)
5	94,721,790	rs440345507	1.45E-15	4.65E-19	Stop_gained	*EPS8*	Within
5	94,852,500	rs460885789	3.76E-15	1.05E-18	Intron	*PTPRO*	Within
5	95,208,776	rs461078254	7.52E-15	8.09E-18	Intron	*RERG*	Within
5	95,330,449	rs440042852	3.67E-14	4.03E-17	Intergenic	*PDE6H*	22,678
5	95,341,145	rs465130049	1.84E-14	2.88E-17	Intergenic	*PDE6H*	11,982
5	95,351,189	rs477551269	1.82E-14	2.88E-17	Downstream	*PDE6H*	1,938
5	95,354,456	rs464693471	1.80E-14	2.87E-17	Intron	*PDE6H*	Within
5	95,355,619	rs798756973	5.42E-15	8.64E-18	Intron	*PDE6H*	Within
5	95,356,973	rs447311406	1.79E-14	2.63E-17	Intron	*PDE6H*	Within
5	96,736,654	—	1.39E-15	4.19E-18	Intron	*GRIN2B*	Within
18	57,573,606	rs433342862	5.77E-12	3.04E-16	Upstream	*LOC100138951*	511
18	57,575,616	rs381326326	6.10E-12	3.29E-16	Intron	*LOC100138951*	Within
18	57,576,030	rs378782360	6.07E-12	3.25E-16	Intron	*LOC100138951*	Within
18	57,577,417	rs477989930	4.31E-13	5.45E-17	Intron	*LOC100138951*	Within
18	57,577,596	rs382881665	4.72E-13	7.21E-17	Intron	*LOC100138951*	Within
18	57,577,749	rs380944308	6.27E-12	3.41E-16	Intron	*LOC100138951*	Within
18	57,580,298	rs380618304	6.36E-12	3.51E-16	Intron	*LOC100138951*	Within
18	58,015,050	rs445560689	1.52E-10	1.02E-16	Upstream	*bta-mir-125a*	485
18	58,118,935	rs521076153	4.97E-10	2.14E-16	Upstream	*LOC100138951*	535,298
18	58,141,989	rs381577268	3.97E-10	1.69E-16	Downstream	*ZNF613*	112

[1]EMMAX = efficient mixed-model association expedited; LMM = linear mixed model.

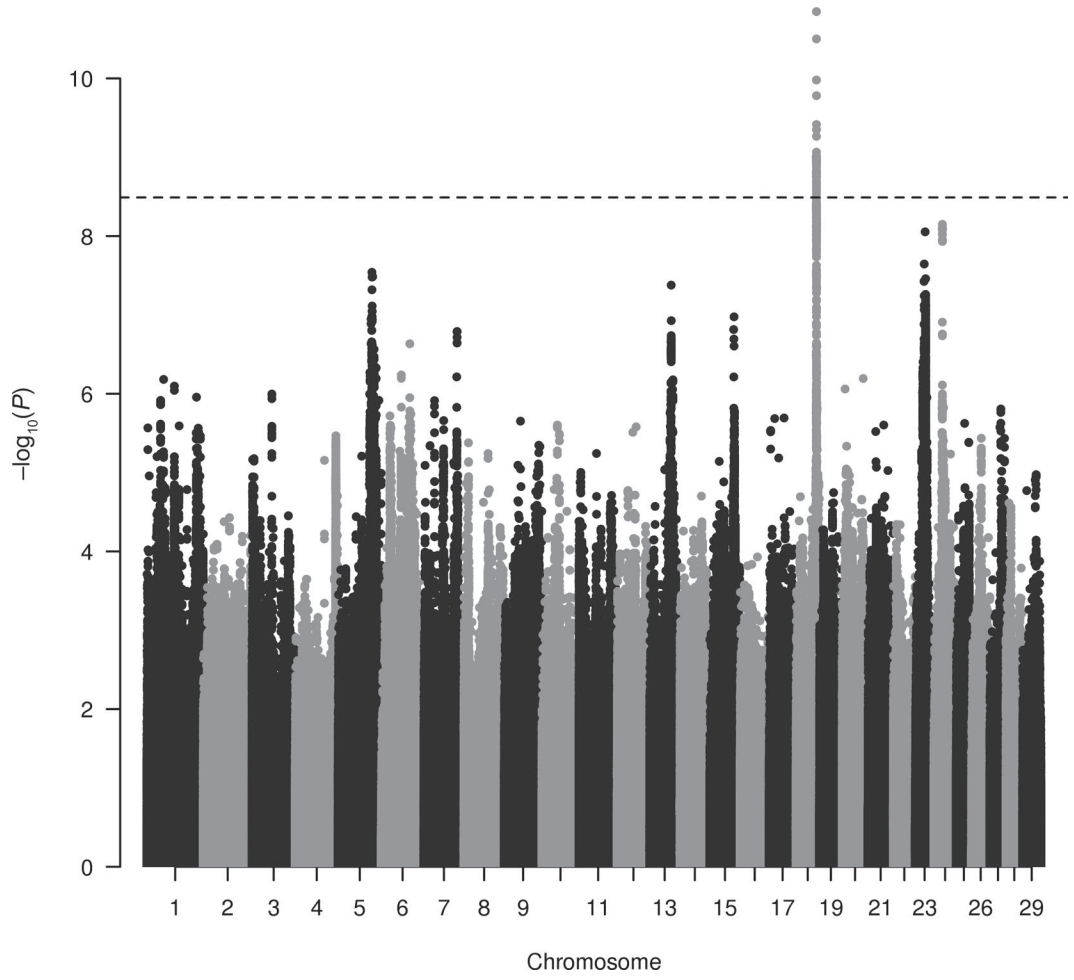

Figure 5. Manhattan plot for association of SNP with young stock survival of heifer calves in period 2 for Nordic Holstein cattle, obtained by the efficient mixed-model association analysis. Black dashed line indicates genome-wide significance level [$-\log_{10}(P) = 8.49$].

likelihood ratio test (Figure 9B). The top 2 haplotypes were located at 57,442,103 to 57,511,637 bp (HAP2a; $P = 2.51 \times 10^{-14}$) and 57,446,310 to 57,516,245 bp (HAP2b; $P = 1.89 \times 10^{-14}$). The HAP2a and HAP2b overlapped and had a frequency of 6.54%. Effects of HAP2a and HAP2b relative to the population mean were estimated as −0.31 (±1.00). The bull carrying lead haplotype, HAP1 and HAP2, shows lower average DRP (Supplemental Figures S2C and S2D; https://doi.org/10.3168/jds.2017-12688).

When the FHM was used, haplotype HAP1 explained 1.16% of the variance, with an estimated effect size of

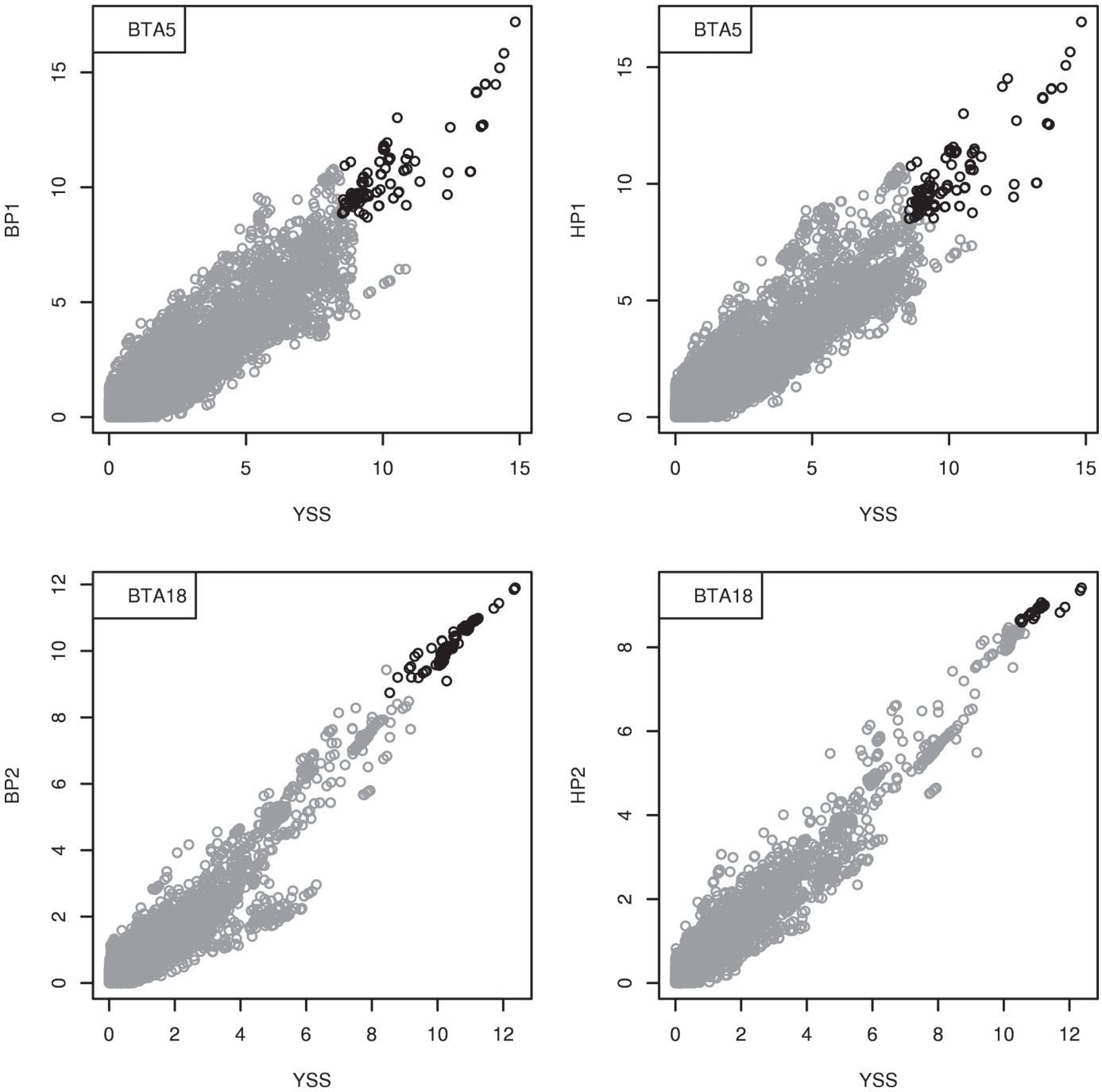

Figure 6. Overlap of associated SNP at 88,000,000 to 102,000,000 bp on BTA5 and 56,000,000 to 59,000,000 bp on BTA18 between young stock survival index (YSS) and its 4 component traits [i.e., YSS in period 1 for bulls (BP1) and heifers (HP1) and YSS in period 2 for bulls (BP2) and heifers (HP2)]. Trait 1 on x-axis, trait 2 on y-axis. The SNP significant for both traits are black, otherwise gray.

−0.38 (±0.82). When HAP2a and HAP2b were fixed separately, they had the same estimated effect size [−0.33 (±0.73)] and explained the same percentage of the variance (1.20%). When these 2 haplotypes were fixed together in the FHM, the estimated effect size was −0.33 (±0.73) for HAP2a and 0 for HAP2b. Therefore, we combined HAP2a and HAP2b as a single QTL region at 57,442,103 to 57,516,245 bp (HAP2).

The lead SNP on BTA5 and 18 are were not present on the BovineSNP50 BeadChip (54k), which is rou-

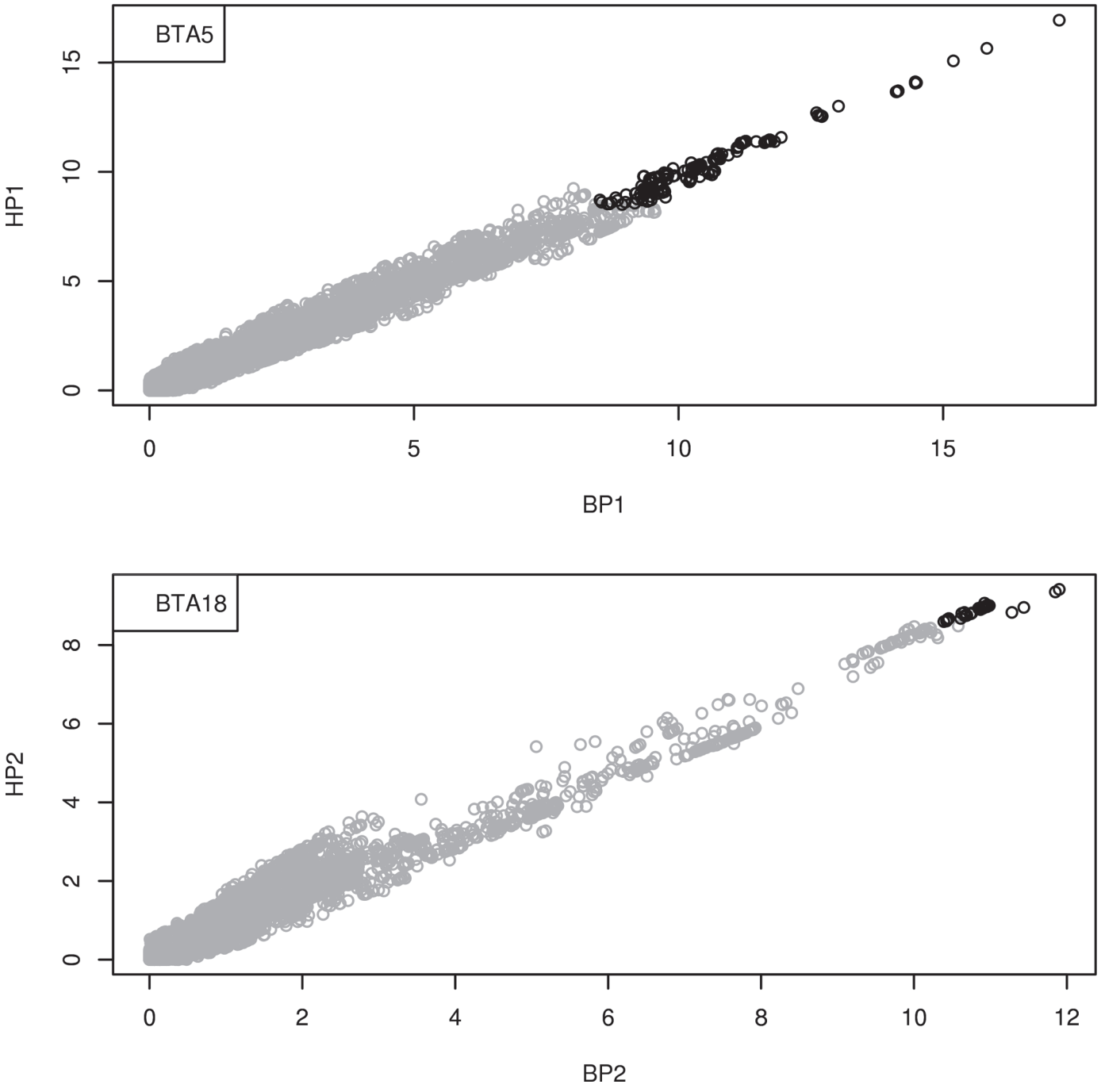

Figure 7. Overlap of associated SNP among 4 component traits of young stock survival (YSS) index. Top: overlapping SNP at 88,000,000 to 102,000,000 bp on BTA5 between YSS index in period 1 for bulls (BP1) and heifers (HP1). Bottom: overlapping SNP at 56,000,000 to 59,000,000 bp on BTA18 between YSS index in period 2 for bulls (BP2) and heifers (HP2). $-\log_{10}(P)$ for trait 1 on x-axis, $-\log_{10}(P)$ for trait 2 on y-axis. The SNP significant for both traits are black, otherwise gray.

tinely used for genomic prediction in cattle. Therefore, we also carried out association analyses using the same RHM constructing haplotypes with the SNP present on the 54k chip for these 2 regions, Chr5:88409199–101803926 and Chr18:56855383–58141989. The lead association haplotypes for these 2 regions were located at Chr5:98693360–99476251 ($P = 8.12 \times 10^{-6}$) and Chr18:57174711–58067310 ($P = 9.70 \times 10^{-9}$; Supplemental Figure S3; https://doi.org/10.3168/jds.2017-12688).

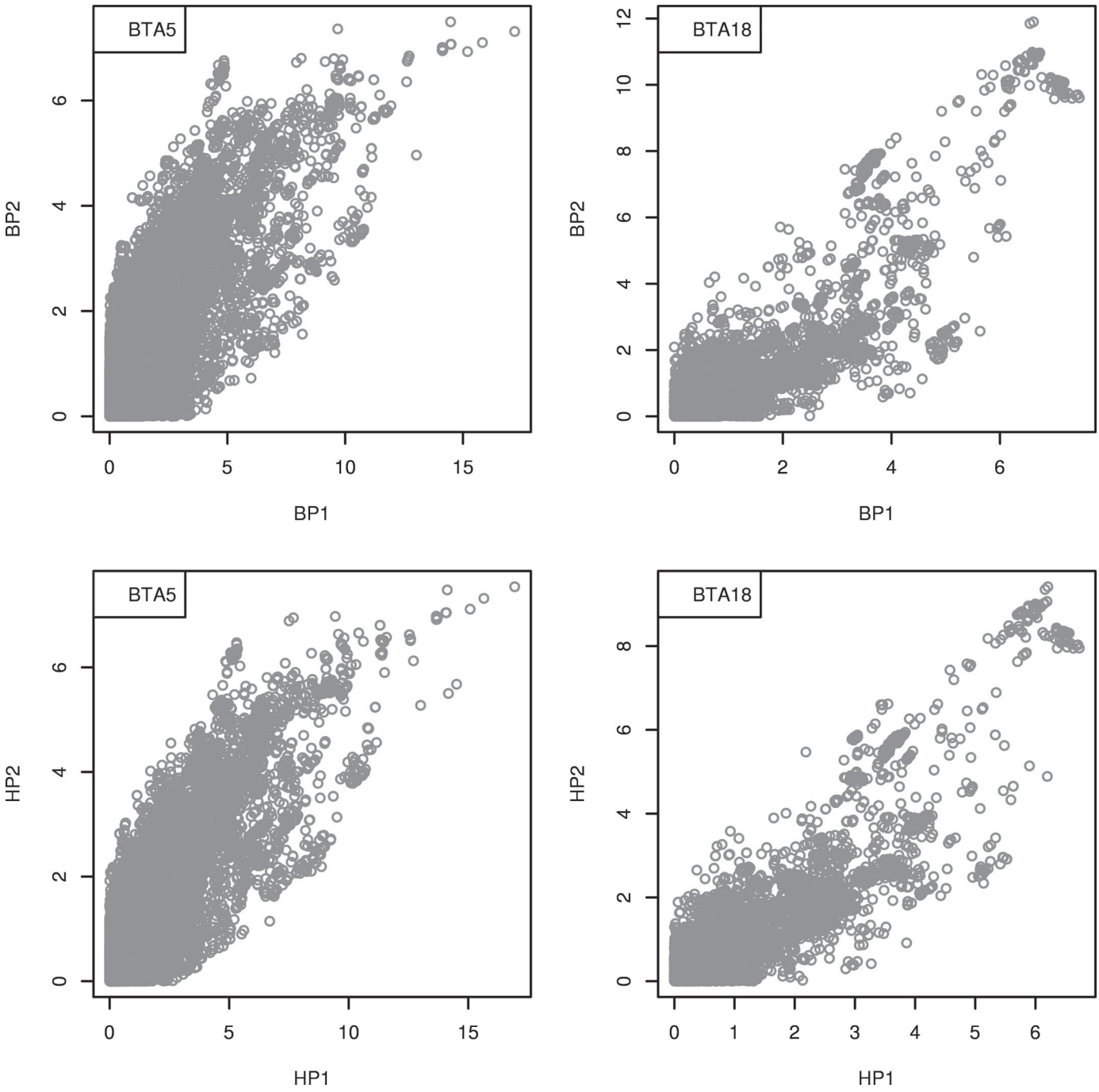

Figure 8. Overlap of associated SNP among 4 component traits of young stock survival (YSS) index on BTA5 (88,000,000–102,000,000 bp) and BTA18 (56,000,000–59,000,000 bp). Top: SNP for YSS of bulls in period 1 (BP1) versus period 2 (BP2). Bottom: SNP for YSS of heifers in period 1 (HP1) versus period 2 (HP2). $-\log_{10}(P)$ for trait 1 on x-axis, $-\log_{10}(P)$ for trait 2 on y-axis. The SNP significant for both traits are black, otherwise gray.

Detection of Lead SNP Homozygous Status in 1,000 Bull Genomes Project Sequence Data

Genotype counts for the putative causative polymorphism in *run5* were obtained from data of the 1,000 Bull Genomes Project (Table 2). For the lead SNP on BTA5, with 19 missing genotypes, 1 homozygote for the alternative allele and 32 heterozygotes were observed in NH. The alternative allele was observed only once (in an Aberdeen Angus individual) among animals from all other breeds. For the lead SNP on BTA18, with 163 missing genotypes, 6 homozygotes for the alternative allele and 130 heterozygotes were observed for Holstein cattle. Three homozygotes and 91 heterozygotes were observed among animals from all other breeds in the 1,000 Bull Genome project.

Association of YSS Index-Related Genomic Regions with Stillbirth

Among the 504 SNP associated with the YSS index in NH, 145 SNP on BTA18 (56,855,383–58,141,989 bp) and 2 SNP on BTA5 (98,559,948 bp and 100,253,082 bp) also had an effect on stillbirth (Figure 10). Only 2 SNP on BTA5 were shared; thus, subsequent analyses for stillbirth were limited to BTA18. On BTA18, the correlation between the lead SNP allele dosage for the YSS index (at 57,577,417, rs477989930) and stillbirth (at 57,020,965, rs476543969) was 0.56. The 2 SNP remained significantly associated with both traits, and the change of the estimated effect size was negligible from the EMMAX to the bivariate model. The SNP rs477989930 explained a greater proportion of the variance for the YSS index, whereas rs476543969 explained a greater proportion of the variance for stillbirth (Table 3).

DISCUSSION

Candidate Genes Underlying the QTL

We performed a genome-wide association study of an NH population and detected 2 QTL regions with large effects on the YSS index on BTA5 and BTA18. The lead SNP rs440345507 on BTA5 (at 94,721,790 bp) was a stop-gained mutation within the *EPS8* gene. In *run5* of the 1,000 Bull Genomes Project sequence data, we found only 1 homozygote in Holstein for the alternative allele and 1 heterozygote in Aberdeen Angus. We detected a haplotype from the 777k genotype data in the targeted region for which no homozygotes existed, indicating recessive lethal gene action. Therefore, the lead SNP constitutes a promising candidate as a recessive semilethal mutation. Haplotype HAP1 (94,691,973–94,755,948 bp) had the strongest association signal in RHM analysis and overlapped with the *EPS8* gene. *EPS8* is critical for activation of the eps8-Abi1-p85-Sos-1 complex, which displays Rac-guanine nucleotide exchange factor activity. The EPS8-related proteins link growth factor stimulation to actin organization, generating functional redundancy in pathways that regulate actin cytoskeletal remodeling (Offenhäuser et al., 2004).

Lead SNP rs477989930 on BTA18 (at 57,577,417 bp) was an intronic variant. The 1,000 Bull Genomes Project sequence data contained 6 Holstein cattle and 3 cattle from other breeds that were homozygous for the alternative allele in *run5*. Therefore, this SNP is not a candidate for recessive lethal mutation. This SNP was located within the *LOC100138951* gene. In addition to the lead SNP, 13 other significantly associated SNP were located within this gene. *LOC100138951* is a *CD33* antigen-like gene. *CD33* participates in the immune system and immune-regulatory interactions between lymphoid and nonlymphoid cells (http://www.reactome.org/content/detail/R-HSA-197719). *CD33*-related siglecs are a major subfamily of proteins that bind sialylated glycans and transmit signals to immune cells (Cao and Crocker, 2011). The most promising haplotype, HAP2, was detected on BTA18 at 57,442,103 to 57,516,245 bp. This region includes a few *KLK* family genes, *KLK9*–*KLK14*, the expressions of which are used as biomarkers for the detection of different human cancers (https://www.ncbi.nlm.nih.gov/gene/).

Previously, a missense mutation in the *TUBD1* gene was found to have a strong association with high perinatal and juvenile mortality in Braunvieh and Fleckvieh cattle (Schwarzenbacher et al., 2016). The *TUBD1* is located on BTA19. We analyzed 6,245 SNP on BTA19 at 10.40 to 11.90 Mb, but did not find any association with the YSS index in NH. Schütz et al. (2016) reported that a 1.3-kb insertion of an endogenous retrovirus into exon 5 of the *APOB* gene at BTA11 (77,959 kb) in German Holstein dairy cattle resulted in cholesterol deficiency. Young calves with this genetic defect only lived for a few months. There are 15,828 SNP located at 74.00 to 77.00 Mb on BTA11 that did not show any association with the YSS index in NH cattle.

Overlap Among Regions Significantly Associated with YSS Traits and Stillbirth

The YSS index is a weighted average of breeding values from 4 component traits. The index shares the QTL region on BTA5 with BP1 and HP1, and the QTL region on BTA18 with BP2 and HP2 (Figure 6). Associated SNP were similar for survival traits for the same period for bull and heifer calves (Figure 7). Thus,

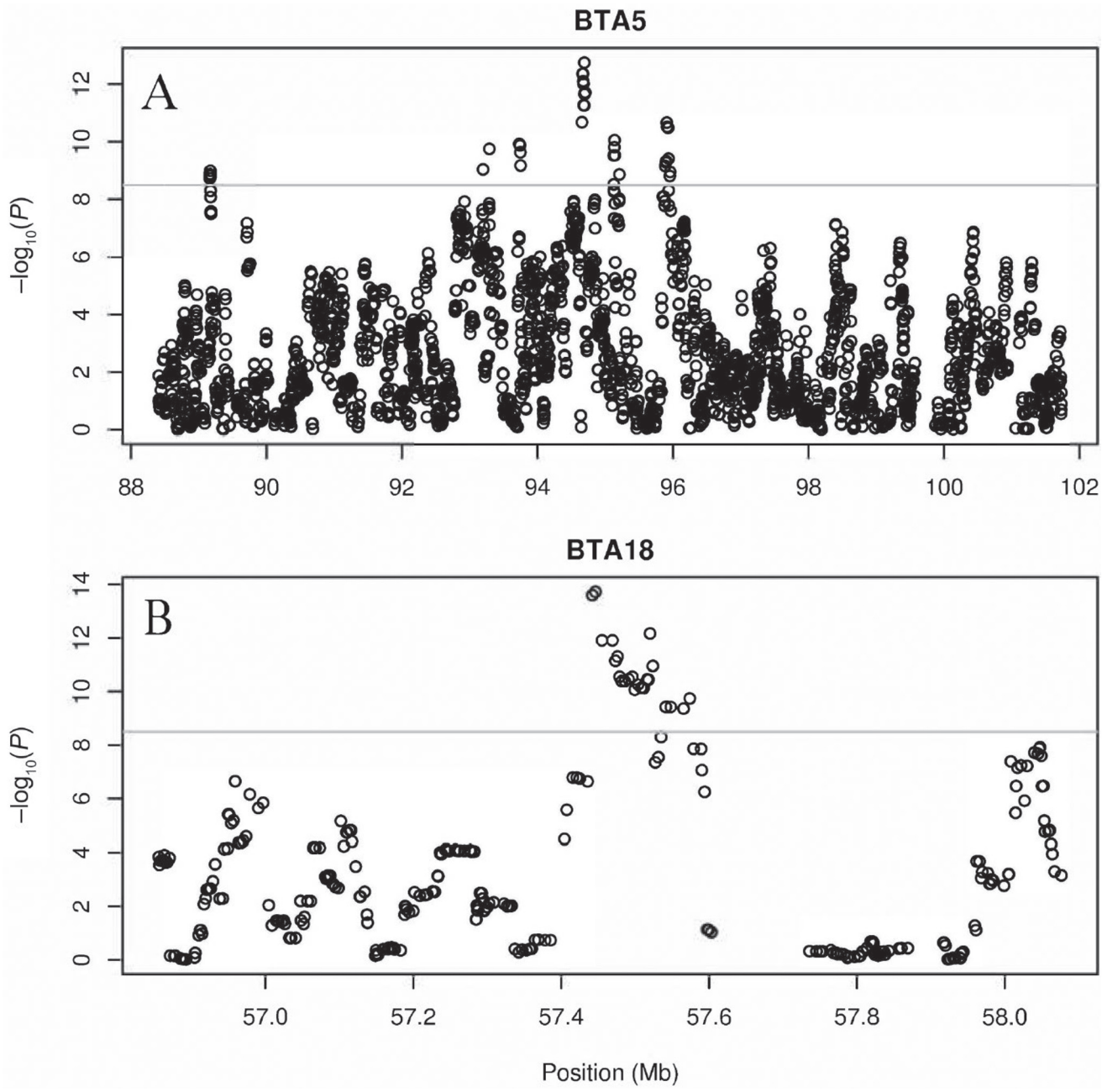

Figure 9. Association signal [$-\log_{10}(P)$] estimated by the random haplotype model for the young stock survival index in Nordic Holstein cattle, plotted against haplotype position on BTA5 (A) and BTA18 (B). Grey line indicates genome-wide significance level [$-\log_{10}(P) = 8.49$].

Table 2. Genotype counts for lead associated SNP, rs440345507 and rs477989930, in the *run5* of the 1,000 Bull Genomes Project whole-genome sequence data

SNP	Breed	Homozygote for the reference allele	Heterozygote	Homozygote for the alternative allele
rs440345507	Holstein	386	30	1
	Red Holstein	24	2	0
	Other	1,114	1	0
rs477989930	Holstein	254	124	4
	Red Holstein	17	6	2
	Other	913	91	3

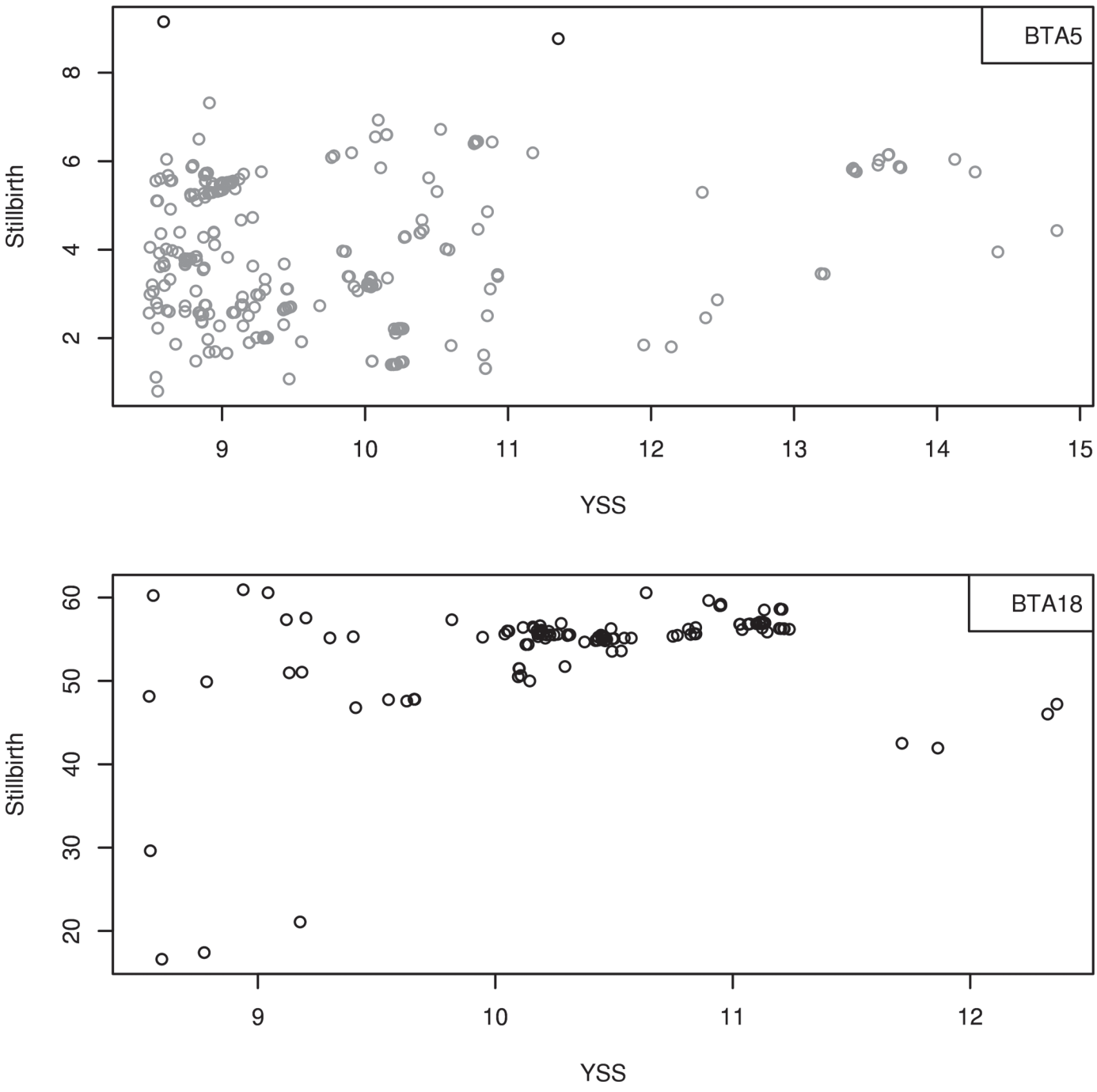

Figure 10. Overlap of associated SNP at 88,000,000 to 102,000,000 bp on BTA5 and 56,000,000 to 59,000,000 bp on BTA18 between young stock survival index (YSS) and stillbirth. Trait 1 on x-axis, trait 2 on y-axis. The SNP significant for both traits are black, otherwise gray.

Table 3. The effect estimates for lead SNP for stillbirth and young stock survival (YSS) located in the associated region of chromosome 18 (Chr18) in the Nordic Holstein cattle

				Effect ± SE (2-trait animal model)		Explained variance (%)	
SNP	Position	Annotation	Gene	YSS index	Stillbirth	YSS index	Stillbirth
rs477989930	Chr18: 57,577,417	Intronic	*CD33*	6.52 ± 0.75	5.26 ± 0.36	3.58	2.71
rs476543969	Chr18: 57,020,965	Intronic	*POLD1*	5.29 ± 0.71	5.67 ± 0.34	1.48	1.99

YSS in males and females seems to share at least some of its genetic basis. In contrast, QTL for early and late YSS did not overlap (Figure 8). This result is consistent with the finding that genetic correlation estimates are high within a period between sexes, but moderate between periods even within a sex (Pedersen et al., 2014). Furthermore, the result supports the notion that different genes control the YSS index over the growing period, but the same genes affect survival in bull and heifer calves for the same development stages. However, a difference could be present in male and female YSS for the genetic factors located on the X-chromosome that was not studied.

A previous study showed that the QTL region on BTA18 exhibits effects on stillbirth, calving ease, and calf size (Mao et al., 2016). The QTL regions on BTA18 for the YSS index and stillbirth overlapped (Figure 10). However, this YSS QTL affected HP2 and BP2 (i.e., calf mortality 30 d after birth). We propose that it is most likely that 2 separate factors are responsible for YSS (after 30 d of birth) and stillbirth (death within 24 h of birth). The top significant associated SNP for stillbirth were not the top associated SNP for YSS.

CONCLUSIONS

Using imputed WGS data, we identified 2 QTL regions, located on BTA5 and BTA18, that were associated with YSS in NH. The SNP rs440345507 (Chr5: 94,721,790) was identified as a putative causal mutation for QTL for the YSS index in NH. The *EPS8*, *LOC100138951*, and *KLK* gene family overlapped with the position of the top SNP or detected haplotypes and, therefore, may be considered as candidate genes for the YSS index in NH. The QTL region on BTA5 was associated with BP1 and HP1, and the QTL region on BTA18 was associated with BP2 and HP2. Thus, the association study for component traits indicated that different genes affected the YSS index over the growing period and that the effects were not sex specific. For future study, the lead SNP has been included in a custom-made low-density SNP chip that is currently used for routine genotyping in 3 Nordic countries.

ACKNOWLEDGMENTS

The authors declare no competing interest.

GS, XW, BG, and MSL conceived and designed the study. XW analyzed the data and wrote the paper. MSL, GS, BG, and USN contributed materials and analysis tools. All authors read, revised, and approved the final manuscript.

We are grateful to the Nordic Cattle Genetic Evaluation (NAV, Aarhus, Denmark) for providing the phenotypic data used in this study and Viking Genetics (Randers, Denmark) for providing samples for genotyping. This work was supported by research projects funded by the Milk Levy Fund (Mælkeafgiftsfonden, Skejby, Aarhus), Green Development and Demonstration Programme (GUDP) project from the Ministry of Environment and Food of Denmark, Viking Genetics, and NAV. The 1,000 Bull Genomes Project is kindly acknowledged for sharing WGS data.

REFERENCES

Becker, T., and C. Herold. 2009. Joint analysis of tightly linked SNPs in screening step of genome-wide association studies leads to increased power. Eur. J. Hum. Genet. 17:1043–1049. https://doi.org/10.1038/ejhg.2009.7.

Boleckova, J., O. F. Christensen, P. Sorensen, and G. Sahana. 2012. Strategies for haplotype-based association mapping in a complex pedigreed population. Czech J. Anim. Sci. 57:1–9.

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. https://doi.org/10.3168/jds.2014-9005.

Cao, H., and P. R. Crocker. 2011. Evolution of CD33-related siglecs: Regulating host immune functions and escaping pathogen exploitation? Immunology 132:18–26. https://doi.org/10.1111/j.1365-2567.2010.03368.x.

Daetwyler, H. D., A. Capitan, H. Pausch, P. Stothard, R. van Binsbergen, R. F. Brøndum, X. Liao, A. Djari, S. C. Rodriguez, C. Grohs, D. Esquerré, O. Bouchez, M.-N. Rossignol, C. Klopp, D. Rocha, S. Fritz, A. Eggen, P. J. Bowman, D. Coote, A. J. Chamberlain, C. Anderson, C. P. VanTassell, I. Hulsegge, 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. https://doi.org/10.1038/ng.3034.

Fuerst-Waltl, B., and M. K. Sørensen. 2010. Genetic analysis of calf and heifer losses in Danish Holstein. J. Dairy Sci. 93:5436–5442. https://doi.org/10.3168/jds.2010-3227.

Goddard, M. 1985. A method of comparing sires evaluated in different countries. Livest. Prod. Sci. 13:321–331. https://doi.org/10.1016/0301-6226(85)90024-7.

Iso-Touru, T., G. Sahana, B. Guldbrandtsen, 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.

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. https://doi.org/10.3168/jds.2015-9599.

Kadri, N. K., B. Guldbrandtsen, P. Sørensen, and G. Sahana. 2014. Comparison of genome-wide association methods in analyses of admixed populations with complex familial relationships. PLoS One 9:e88926–e88926. https://doi.org/10.1371/journal.pone.0088926.

Kang, H., J. Sul, N. Zaitlen, and S. Kong. 2010. Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. 42:348–354. https://doi.org/10.1038/ng.548.

Kang, H., N. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, M. Daly, and E. Eskin. 2008. Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723. https://doi.org/10.1534/genetics.107.080101.

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 Proceedings of 10th World Congress of Genetics Applied to Livestock Production. Vancouver, British Columbia, Canada. http://pure.au.dk/portal/files/82154310/Paper.pdf.

Mao, X., N. K. Kadri, J. R. Thomasen, D. J. De Koning, G. Sahana, and B. Guldbrandtsen. 2016. Fine mapping of a calving QTL on *Bos taurus* autosome 18 in Holstein cattle. J. Anim. Breed. Genet. 133:207–218. https://doi.org/10.1111/jbg.12187.

McLaren, W., B. Pritchard, D. Rios, Y. Chen, P. Flicek, and F. Cunningham. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 26:2069–2070. https://doi.org/10.1093/bioinformatics/btq330.

Norberg, E., J. E. Pryce, and J. Pedersen. 2013. Short communication: A genetic study of mortality in Danish Jersey heifer calves. J. Dairy Sci. 96:4026–4030. https://doi.org/10.3168/jds.2012-5659.

Offenhäuser, N., A. Borgonovo, A. Disanza, P. Romano, I. Ponzanelli, G. Iannolo, P. P. Di Fiore, and G. Scita. 2004. The eps-8-family of proteins links growth factor stimulation to actin reorganization generating functional redundancy in the Ras/Rac pathway. Mol. Biol. Cell 15:91–98. https://doi.org/10.1091/mbc.E03-06-0427.

Pedersen, J. 2015. Economic value of Young Stock Survival index. Pages 1–5. file:///C:/Users/au464965/Downloads/9a_20150113youngstocksurvivaleconomicvalue.pdf.

Pedersen, J., M. Kargo, A. Fogh, J. Pösö, J. Eriksson, U.S. Nielsen, and G. P. Aamand. 2015. Note on economic value of Young Stock Survival. Pages 1–11. http://www.nordicebv.info/wp-content/uploads/2015/10/Economic-value-of-Young-Stock-Survival.pdf.

Pedersen, J., J. Pösö, J. Eriksson, U.S. Nielsen, and G. P. Aamand. 2014. NAV evaluation for calf survival. http://www.nordicebv.info/wp-content/uploads/2015/04/CalfSurvivalFinalReport.pdf.

Sahana, G., T. Iso-Touru, X. Wu, U. S. Nielsen, D.-J. de Koning, M. S. Lund, J. Vilkki, and B. Guldbrandtsen. 2016. A 0.5-Mbp deletion on bovine chromosome 23 is a strong candidate for stillbirth in Nordic Red cattle. Genet. Sel. Evol. 48:35. https://doi.org/10.1186/s12711-016-0215-z.

Sahana, G., U. S. Nielsen, G. P. Aamand, M. S. Lund, and B. Guldbrandtsen. 2013. Novel harmful recessive haplotypes identified for fertility traits in Nordic Holstein cattle. PLoS One 8:e82909. https://doi.org/10.1371/journal.pone.0082909.

Schaeffer, L. R. 1985. Model for international evaluation of dairy sires. Livest. Prod. Sci. 12:105–115. https://doi.org/10.1016/0301-6226(85)90084-3.

Schütz, E., C. Wehrhahn, M. Wanjek, R. Bortfeld, W. E. Wemheuer, J. Beck, and B. Brenig. 2016. The Holstein Friesian Lethal Haplotype 5 (HH5) results from a complete deletion of TBF1M and cholesterol deficiency (CDH) from an ERV-(LTR) insertion into the coding region of APOB. PLoS One 5:1–15. https://doi.org/10.1371/journal.pone.0154602.

Schwarzenbacher, H., J. Burgstaller, F. R. Seefried, C. Wurmser, M. Hilbe, S. Jung, C. Fuerst, N. Dinhopl, H. Weissenböck, B. Fuerst-Waltl, M. Dolezal, R. Winkler, O. Grueter, U. Bleul, T. Wittek, R. Fries, and H. Pausch. 2016. A missense mutation in TUBD1 is associated with high juvenile mortality in Braunvieh and Fleckvieh cattle. BMC Genomics 17:400.

Vuori, K., I. Strandén, M. Lidauer, and E. Mäntysaari. 2006. MiX99–Effective solver for large and complex linear mixed models. Pages 27–33 in Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil.

Wu, X., B. Guldbrandtsen, M. S. Lund, and G. Sahana. 2016. Association analysis for feet and legs disorders with whole-genome sequence variants in 3 dairy cattle breeds. J. Dairy Sci. 99:7221–7231. https://doi.org/10.3168/jds.2015-10705.

Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki, J. F. Doebley, M. D. McMullen, B. S. Gaut, D. M. Nielsen, J. B. Holland, S. Kresovich, and E. S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203–208. https://doi.org/10.1038/ng1702.

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.