Fine-mapping young-stock survival QTL on chromosome 6 in Nordic Red Dairy Cattle

Z. Cai[1], X. Wu[1], B. Thomsen[2], G. Sahana[1]*

[1]Aarhus University, Center for Quantitative Genetics and Genomics, Blichers Allé 20, 8830 Tjele, Denmark. [2]Aarhus University, Department of Molecular Biology and Genetics, C. F. Møllers Allé 3, 8000 Aarhus, Denmark; *goutam.sahana@qgg.au.dk

Abstract
A genome-wide association study was carried out for young-stock survival (YSS) in Nordic Red Dairy cattle. The de-regressed proofs for 3,078 bulls for YSS index and four component traits were analyzed. The 50k genotypes were imputed to whole genome sequence level using *run7* 1000 bull genome sequence data. A major QTL on chromosome 6 was identified with rs133470751 as the lead SNP with $-\log_{10}$(p value) equal to 47.02. The lead SNP is located in a gene desert. The post-GWAS analysis for functional annotations of the variants and presence for structural variants in the QTL peak was evaluated for being causal factor.

Introduction
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 (Wu et al. 2017). Mortality of cattle causes economic loss for dairy farmers. The cost at rearing period is relatively high compare with early embryonic loss, abortion and stillbirth. Calves and young stock that die during the rearing period result in lost revenue, fewer heifers for replacement, higher veterinarian costs and deteriorating animal welfare. Therefore, it is important for both economy and welfare, to breed for reducing mortality of calves and young stock. 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). 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. 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 (Carlen 2016).
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. In this study, we focus to identify the causal variant underlying a major QTL on chromosome 6 in Nordic Red Dairy cattle (RDC).

Materials & methods
Phenotype and genotype data. For genome scan, we used data from 3,078 RDC bulls. The phenotype used was de-regressed proofs for 4 component traits of YSS (i.e. BP1, BP2, HP1 and HP2) and YSS index.
All these RDC bulls were genotyped with Illumina 50k SNP chip; followed by imputed to HD level (777k), and then imputed to whole genome sequence level using *Run7* reference animals

from the 1000 Bull Genomes project (BGP). The whole genome sequencing (WGS) reference panel was phased with Beagle4 (Browning et al. 2018) and recalled by shapeit2 (Delaneau et al. 2011). HD panel was phased using Beagle5 (Browning et al. 2018). Both steps of imputation were performed by Minimac4 (Fuchsberger et al. 2015). The average imputation accuracy, R^2 from Minimac4, was 91.2% across all minor allele frequency classes. The SNP with a minor allele frequency less than 1% or those deviating from Hardy-Weinberg proportions ($p<10^{-6}$) were removed. Association analysis was carried our using a linear-mixed model approach (Kang et al. 2010) using GCTA software (Yang et al. 2011). A SNP was declared significant if its p-value was less than multiple Bonferroni correction threshold (see Wu et al. 2017 for details).

Post-GWAS. To reveal the possible consequence of the variants within the QTL interval, we performed variants annotation using variant effect predictor (VEP) v104 (McLaren et al. 2016). The prediction impact with "HIGH" was retained to check the segregation in the WGS animals from 1000 BGP Run7. The Gene Ontology (GO) (Ashburner et al., 2000) was used to explore the function of identified candidate genes. The precomputed Repeatmasker (Chen 2004) profile was downloaded from NCBI, and checked the presence of the repeat sequence around the lead SNP. Variation in repeats can alter the expression of genes, and changes in the number of repeats have been linked to certain human diseases (de Bustos et al. 2016). The flanking 1 Mb region of was compared to the regulatory elements from previous studies (Kern et al. 2021). We also examined if chromosomal deletions (Mesbah-Uddin et al. 2018) located in or close of the QTL regions could be the causal variants.

Results

Association with component traits of YSS. We detected genome-wide associated SNP, including 9,162 SNP for BP1, 12,173 SNP for BP2, 8,709 SNP for HP1, and 14,096 SNP for HP2. Of the 14,567 SNP that were significant for YSS index, 8,856 SNP showed significant association with BP1, 11,688 SNP for BP2, 8,339 SNP for HP1, and 13,375 SNP for HP2. In addition, there were 8,268 SNP on BTA6, 3 SNP on BTA17, and 266 SNP on BTA23 were associated with both BP1 and HP1. There were 182 SNP on BTA4, 11,655 SNP on BTA6, and 2 SNP on BTA14 associated with both BP2 and HP2. There were 8,400 SNP on BTA6 were associated with both BP1 and BP2; and 8,243 SNP on BTA6 associated with both HP1 and HP2.

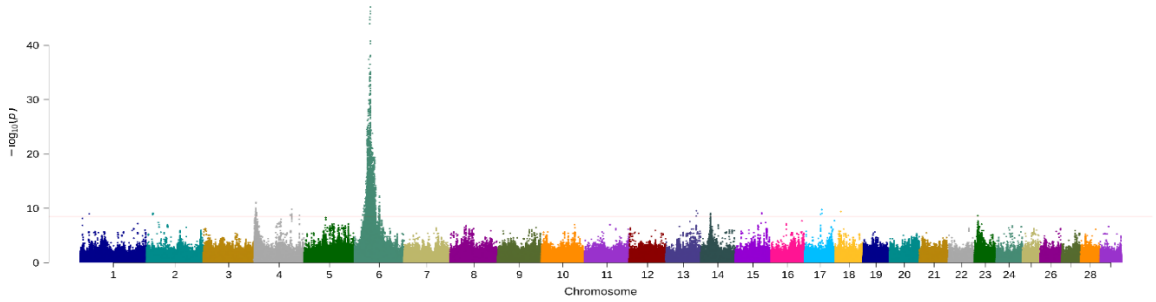

Figure 1. Manhattan plot for association of SNP with young stock survivor in Nordic Red cattle. Red horizontal line indicates genome-wide significance level [$-\log_{10}(P) = 8.5$].

Association with YSS. We found a major QTL for YSS at chromosome 6 (Figure 1). The lead SNP of this QTL is 6:38065928 (rs133470751) with $-\log_{10}$(p value) equal to 47.02. The annotation of

this variant is intergenic variant. The lead SNP is located at the gene desert (Figure 2). There are several genes in a far distance which are worth to investigate.

Checking variants annotation of the flanking area of lead SNP. We checked the flanking 1 Mb region for high-impact (given by variants effect predictor) of the lead SNP. Two genes with several functionally annotated SNPs showed potential to be causal genes and/or causal variants. *FAM184B* have a stop gained variants at 6:37201641. *LCORL* have three frameshift variants at 6:37401110, 6:37401770, 6:37402602 and 6:37403627 and one splice donor variant at 6:37412364. The segregation of these variants in non-Nordic Red WGS animals and Nordic Red WGS animals were listed in Table 1. *LCORL* belong to GO term "regulation of transcription by RNA polymerase II". Both genes do not show any clear functional meaning to YSS from GO, KEGG or mammalian phenotype database.

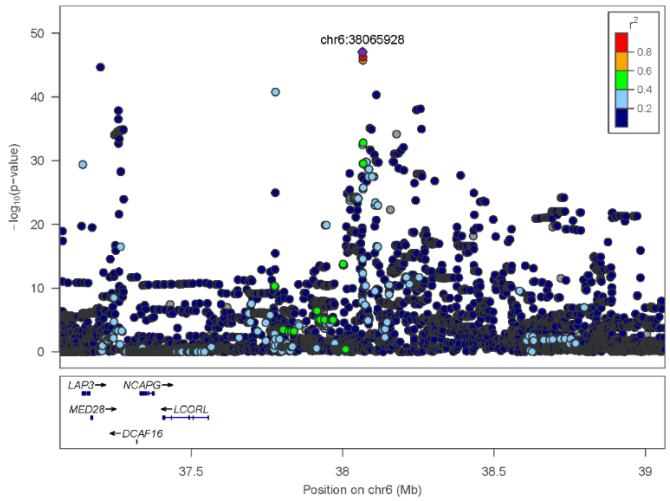

Figure 2. LocusZoom plot (Pruim et al. 2010) showing the QTL region on chromosome 6 in Nordic Dairy cattle; colour code is to show the linkage disequilibrium with the lead SNP.

Table 1. The segregation of the high-impact variants located within two candidate causal genes in 1000 bulls genome sequence without and within the Nordic Red cattle (RDC) sequences.

variants	gene	All except RDC			RDC		
		HOM-A1	HET	HOM-A2	HOM-A1	HET	HOM-A2
6:37201641	*FAM184B*	0	0	2913	1	9	153
6:37401110	*LCORL*	2	28	2882	0	0	164
6:37401770	*LCORL*	643	253	1996	4	24	136
6:37402602	*LCORL*	0	2	2916	0	32	129
6:37403627	*LCORL*	2	12	2892	0	0	162
6:37412364	*LCORL*	1	39	2866	0	2	160

Unveiling the gene desert region around lead SNP. The Repeatmasker profile was downloaded from NCBI, there is a LINE element very close to the lead SNP, which is L2c. The copy in the reference genome is not a full length element, which means it is not active. However, we currently

don't have the right dataset, long read sequences, to further investigate the repeat content of this region in our RDC animals. The sequence reads obtained with next generation sequencing platforms are short and simply do not span long repetitive sequences. We checked for the presence of regulatory elements in the flanking 1 Mb region of the lead SNP, and the result showed the region from 38 Mb to 39 Mb contains several regulatory elements even though the region is devoid of genes.

We fine-mapped a major QTL affecting young-stock survival in Nordic Red dairy cattle. The lead SNP is located in a gene desert, indicating a regulatory variant may be the causal factor for this QTL. The high-impact variants in neighbouring genes were also studied. These variants are included in a custom made SNP array used in routine genotyping.

Acknowledgements

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 1000 Bull Genomes Project is acknowledged for sharing WGS data.

References

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, et al. (2000) Nat Genet 25(1):25-29. https://doi.org/10.1038/75556

Browning, B.L., Zhou Y., and Browning S.R. (2018) AJHG 103(3):338-348. https://doi.org/10.1016/j.ajhg.2018.07.015

Carlen E., Pedersen J., Pösö J., Eriksson J-Å., Nielsen U.S., et al. (2016) Interbull Bulletine No. 50. Puerto Varas, Chile. https://journal.interbull.org/index.php/ib/article/view/1408/1471

Chen, N. (2004) Curr Proto Bioinformatics 5(1):4.10.11-14.10.14. https://doi.org/10.1002/0471250953.bi0410s05

De Bustos, A., Cuadrado, A. and Jouve, N. (2016) Sci Rep 6, 36665. https://doi.org/10.1038/srep36665

Delaneau, O., Marchini J., and Zagury J. F. (2011) Nat Methods 9(2):179-181. https://doi.org/10.1038/nmeth.1785

Fuchsberger, C., Abecasis G.R., and Hinds D.A.. (2015) Bioinformatics 31(5):782-784. https://doi.org/10.1093/bioinformatics/btu704

Kang, H.M., Sul J.H., Service S.K., Zaitlen N.A., Kong S.Y., et al. (2010) Nat Genet 42(4):348-354. https://doi.org/10.1038/ng.548

Kern, C., Wang Y., Xu X., Pan Z., Halstead M., et al. (2021) Nat Commun 12(1):1821. https://doi.org/10.1038/s41467-021-22100-8

McLaren, W., Gil L., Hunt S.E., Riat H.S., Ritchie G.R., A. et al. (2016) Genome Biol 17(1):122. https://doi.org/10.1186/s13059-016-0974-4

Mesbah-Uddin, M., Guldbrandtsen B., Iso-Touru T., Vilkki J., De Koning D.J., et al. (2018) DNA Res 25(1):49-59. https://doi.org/10.1093/dnares/dsx037

Pruim, R.J., Welch R.P., Sanna S., Teslovich T.M., Chines P.S., et al. (2010) Bioinformatics 26(18):2336-2337. https://doi.org/10.1093/bioinformatics/btq419

Wu, X., Guldbrandtsen B., Nielsen U.S., Lund M.S., and Sahana G. (2017) JDS 100(8):6356-6370. https://doi.org/10.3168/jds.2017-12688

Yang, J., Lee S.H., Goddard M.E., and Visscher P.M. (2011) AJHG 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011