Exploring genetics underlying respiratory disorders in dairy calves using producer-recorded data

N. Vukasinovic[1], D. Gonzalez-Peña[1], J. J. Brooker[1], P. Desai[1] & S.K. DeNise[1]

[1] VMRD Genetics , Zoetis, Kalamazoo, MI, 49007 USA
natascha.vukasinovic@zoetis.com (Corresponding Author)

Summary

Genetic analysis of respiratory disorders in Holstein calves have been developed using producer-recorded data and the single-step GBLUP (ssGBLUP) methodology. Phenotypic data on respiratory events up to one year of age have been obtained from herd management software backup files provided directly from about 300 dairies upon obtaining owners' permission. The following traits were included in the analysis: respiratory diseases from 0-3 (RESP1), 4-14 (RESP2), 15-50 (RESP3), 51-365 (RESP4) and 0-365 days of age (RESPA). Each trait was defined as a binary event, having a value of one if an animal has been recorded with a disorder, and zero otherwise. The number of phenotypic records ranged from about 224,000 for RESP1 to approximately 874,000 for RESPA, with the prevalence ranging from 4.6% (RESP1) to 20.2% (RESPA). All traits were analyzed using a univariate threshold animal model, including fixed effect of year of birth x calving season x region and random effects of herd x year of birth and animal. Heritabilities, estimated on the liability scale, ranged from 0.039 (RESP4) to 0.097 (RESP1). A total of 45,425 SNPs were used in genomic analyses. Over 292K genotyped animals were included in the evaluation. Animals genotyped with low density chips were imputed to the required number of SNPs using Fimpute software. All analyses were performed using the ssGBLUP implementing the 'algorithm for proven and young animals' (APY) designed to accommodate very large number of genotypes. Predicted transmitting abilities (PTA) were expressed in percentage points as deviations from the average estimated probability of a disorder in the base population. Reliabilities of breeding values were obtained by approximation in which the genotype contribution was approximated using the diagonal values of the genomic relationship matrix. Estimated genomic PTAs and reliabilities were approximately normally distributed for all traits, with reliabilities for genotyped animals without own phenotypes reaching average values between 38% (RESP4) and 50% (RESP1). Genome-wide association analyses were conducted with a subset of genotyped animals using single-step GWAS (ssGWAS) approach with a 10 SNP sliding window. Regions explaining most genetic variance were found on chromosomes 1, 2, 9, 14, 27, and X. Some of those regions contained genes that may be involved in physiological pathways affecting respiratory health of dairy calves.

Keywords: calf health, respiratory diseases, genomic selection, GWAS

Introduction

Costs of raising a calf from birth to first calving have been estimated at $1,200 to over $2,000 (Rossini, 2004). In a study by Sischo et al. (1990) calf disease costs represented 4% of the total cost a cow incurred during her lifetime. Diarrhea and pneumonia were responsible for 86% of calf disease costs. Occurrence of a calfhood disease increases age at first calving,

reduces survival, and increases culling due to mastitis and other diseases (Rossini, 2004).

In addition to improving management practices, genetically improving calf disease resistance is important for the success of dairy operations. Genetic evaluation for calf respiratory traits is considered less effective because of low heritabilities. Most heritability estimates for calf respiratory diseases were low: depending on the data, trait definition, and methodology, the heritability estimates ranged from 0.04 in Holstein calves in Ontario (McCorquodale et al., 2013), to 0.05 in Norwegian Red calves (Heringstad et al, 2008), to 0.09 in Holstein calves in New York (Henderson et al, 2011) and up to 0.21 when estimated using genomic relationship matrix (Neibergs et al., 2014). Another obstacle in the implementation of genetic evaluation for calf health, especially in the United States, is the lack of routinely recorded data. Producers usually record calf health events and treatment using herd management software, but many factors influencing calf health and survival, such as birth weight, amount of colostrum fed, and blood protein level are not routinely measured and recorded. However, previous studies have shown that producer-recorded data can be successfully used in genetic and genomic evaluation of dairy wellness traits (e.g., Zwald et al., 2004; Parker-Gaddis et al., 2014; Vukasinovic et al., 2017). The inclusion of genomic data substantially improves reliabilities for these traits (Parker-Gaddis et al., 2014; Vukasinovic et al., 2017).

The objectives of this study were to explore genetic background of respiratory disorders in dairy calves and develop a genetic and genomic evaluation for those traits in Holstein dairy calves based on producer-recorded data. Further, we attempted to identify genomic regions associated with dairy calf respiratory disorders and uncover candidate genes located in those regions that may be included in the physiological pathways regulating respiratory health.

Material and methods

Phenotypic data was obtained directly from about 300 dairies in form of the farm management software backup files upon obtaining owners' permission. Pedigree information and health events up to one year of age were extracted using proprietary scripts. Respiratory disorders were defined as events coded 'RESP' or 'PNEU'. Traits were defined based on the distribution of the incidence of respiratory disorders by age, shown in Figure 1.

The distribution of respiratory disorders shows four distinct peaks: the first, and the highest one, occurs very shortly after birth; the second during the first week of life; the third around 35 days of age, and the fourth and the largest peak occurs between about 60 and 100 days of age. After that, the incidence of respiratory diseases is very low up to the time when animals start calving. Based on the distribution of the events by the age of animals, traits RESP1, RESP2, RESP3, and RESP4 were defined as shown in Table 1: Trait RESPA was defined to cover the entire period from birth to one year of age.

The information on respiratory events was merged with the information on healthy herdmates. All traits were treated as binary events, having value of "1" if the animal was recorded as having a disorder or "0" otherwise. Animals sold in the first week of life, males, non-Holstein calves, and incorrect records were excluded. Animals that did not reach the 'opportunity period', which was defined as the upper age limit for each trait, were also removed. Herd x year groups with less than 10 records or with the incidence of disorder smaller than 0.5% or larger than 95% were excluded.

Animals were genotyped with various versions of Zoetis proprietary chips. Genotypes obtained using chips with < 40,000 markers were imputed to 45,425 markers, using the program FImpute (Sargolzaei et al, 2011). As of April 2017, 292,998 genotyped animals were

available for analysis.

Each trait was analyzed separately, using the following univariate threshold animal model:

, [1]

where λ represents a vector of the animals' unobserved liabilities to the given disorder; β is the vector of fixed effects of year of birth x calving season x region combination; h is the random herd-year effect, where , with the variance ; a is the random animal effect, with where is the additive genetic variance and H is the pedigree relationship matrix augmented using genotypes X, , and , are incidence matrices corresponding to the fixed effects in and the random effects of herd x year and animal, respectively, and I is the identity matrix.

Variance components were estimated using the same model, but without genomic information. The variance components were estimated using ASReml v. 4.1 (Gilmour et al., 2015), using generalized linear models with binomial distribution and a LOGIT link function.

Genetic evaluation was performed using the programs from the BLUPF90 family (Misztal et al., 2014). A univariate threshold model based on single step genomic BLUP methodology (ssGBLUP) was applied to all traits. In ssGBLUP, the inverse of the traditional pedigree relationship matrix, , is replaced by the inverse of H matrix that combines pedigree and genomic relationships (Legarra et al., 2009; Aguilar et al., 2010):

[2]

where is an inverse of the pedigree relationship matrix, is an inverse of the genomic relationship matrix and is an inverse of the pedigree- relationship matrix for genotyped animals only. To accommodate the large number of genotypes, the 'algorithm for proven and young animals' (APY) was applied. The APY algorithm involves creation of using genomic recursion based on a subset of animals ('proven' or 'core' animals). Only a relationship matrix for the core animals needs to be inverted; elements of for all other animals ('young' or 'non-core') are calculated linearly by recursion, thus reducing computational requirements (Fragomeni et al., 2015; Masuda et al., 2016). The program *cblup90iod2* version 3.21 (Misztal et al., 2002) was used to obtain genomic breeding values. The core consisted of 25,000 randomly selected animals. The reliabilities of estimated breeding values were obtained with the program *accf90GS* version 2.42, which approximates reliabilities using contribution from phenotypes, pedigree, and genotypes, represented by the value of the diagonal of the G matrix (D. Lourenco, personal communication, 2016):

For each trait, the solutions from the *cblup90iod* program (raw EBVs) were transformed into probabilities of exceeding the threshold value. Threshold values for all traits were calculated from the current data. For each animal solution, we calculated probability that a standard normal variable with the mean equal to this solution and the variance of one exceeds the threshold. The probabilities were multiplied by 100 (to represent percentages), divided by 2 (to obtain PTAs), and expressed as the deviation of the average PTA of all animals born in 2010 with a phenotypic record for that trait.

Genome-wide association analyses were conducted using single-step GWAS (ssGWAS) approach implemented in the program postGSf90 version 1.29 (Aguilar et al., 2014), based on extracting SNP effects from solutions of genomic evaluations and calculation of variance explained by chromosomal segments. To enable computation, the analysis was performed using a subset of 113,015 genotyped animals that had reliability for each of the analyzed traits

of over 40%. The analyses were run with a sliding window of 10-SNPs. The regions explaining the largest proportion of the trait variance were queried using UCSC Genome Browser (Kent et al., 2002) to identify putative candidate genes and QTL located in those regions.

Results and Discussion

Table 2 shows the total number of records used in genetic evaluation, incidence (percentage of affected animals in the data), and estimated heritabilities for each trait. The numbers of records varied depending on the trait. The variation is caused by differences in herds recording various traits and the imposing of the 'opportunity period' for each trait.

The incidence of respiratory disorders is the lowest for RESP1 and RESP2 (4.6%) and increases as the animals become older and the time interval increases (20.2% for the entire period of 0 to 365 days of age). The estimated heritabilities for respiratory disorders ranged from 0.039 for RESP4 to 0.097 for RESP1. Higher heritability values were observed for disorders occurring at younger age, indicating that the genetic component plays a more important role in the disease susceptibility of young animals, whereas later incidences of respiratory disorders reflect higher influence of environment. These heritability estimates are comparable with values published in the literature (e.g., Henderson et al., 2011 and McCorquodale et al., 2013).

Table 3 shows means and standard deviations for genomic PTA (gPTAs) and reliabilities obtained for calf respiratory disorders for genotyped animals without phenotype or progeny. The gPTAs for all traits were approximately normally distributed. The mean reliabilities of gPTAs for RESP traits range between 38% (RESP4) and 50% (RESP1) depending on the amount of information available and the heritability of the trait.

Figure 2 shows Manhattan plots picturing the results of the ssGWAS analysis. Visual examination found two pronounced peaks – on chromosomes 9 and 14 – explaining the most genetic variance in RESP1. RESP2 showed smaller peaks, the largest ones located on chromosomes 1, 27, and X. RESP3 showed similar picture as RESP1 with peaks on chromosomes 9 and 14, RESP4 had the largest peaks on chromosomes 1, 2, and X. RESPA contained almost all peaks found for respiratory disorders within distinct age periods. Table 4 shows the chromosomal regions identified as having most impact on the calf health disorders based on the proportion of the explained genetic variance. Each of the 10-SNP sliding windows in those regions explained between 0.22% and 0.49% genetic variance in one or more traits. Table 4 also shows the putative genes located in these regions. While not all of the searches were successful, some of the regions did harbour genes that could be involved in animals' respiratory health. TRPC1 and PLS1 gene located on BTA1 and MICU3 on BTA27 influencing RESP2 are important for calcium ion binding. TRPC1 has been found to be involved in host defence against Gram-negative bacteria in mice (Zhou et al., 2015). In an independent study of bovine respiratory diseases in feedlot cattle, ZFX gene was found to have differential expression in healthy animals and animals showing severe lung lesions (Desai et al., unpublished results). Multiple genes located in the regions found to influence RESP traits are involved in growth and cell differentiation (e.g., IGF2, ZFX, etc.) indicating that respiratory disorders may be related to calves' body size.

Conclusions

The results of this study indicate that genetic evaluation for resistance to calf respiratory disorders using producer-recorded data is feasible. The large amount of phenotypes and genotypes contributes to fairly high reliabilities of gPTA, even for animals without own records or progeny. Based on the results of the genome-wide association analysis, calf respiratory disorders are polygenic traits influenced by many genomic regions. Several regions harbouring genes potentially associated with various physiological aspects of respiratory disorders were identified. Fine mapping of these regions would be necessary to uncover genetic variants that could be used to improve accuracy of genomic selection for calf respiratory disorders.

Table 1: Calf respiratory disorders definitions based on farm-recorded health events

Acronym	**Description**
RESP1	RESP + PNEU 0-3 days of age
RESP2	RESP + PNEU 4-14 days of age
RESP3	RESP + PNEU 15-50 days of age
RESP4	RESP + PNEU 51-365 days of age
RESPA	RESP + PNEU 0-365 days of age

Table 2: Total number of records in genetic evaluation and incidence for calf respiratory disorders

Trait	Number of records	Incidence (%)	Heritability
RESP1 (0-3d)	224,112	4.6	0.097
RESP2 (4-14d)	431,570	4.6	0.058
RESP3 (15-50d)	633,123	9.7	0.049
RESP4 (51-365d)	828,695	15.3	0.039
RESPA (0-365d)	874,170	20.2	0.042

Table 3: Means and standard deviations of gPTAs and reliabilities for calf respiratory disorders for genotyped animals without phenotypic records or progeny.

		gPTA		Reliability (%)	
Trait	N	Mean	SD	Mean	SD
RESP1	220,571	-0.08	2.51	50.1	4.5
RESP2	202,946	-0.53	1.65	41.7	5.4
RESP3	198,900	-0.32	2.23	42.3	5.5
RESP4	212,895	0.99	2.50	38.1	5.6
RESPA	212,201	1.12	3.01	40.1	5.4

Table 4: Genomic regions explaining the largest proportion of variance in calf respiratory disorders and the putative genes located in these regions.

BTA	Start (bp)	End (bp)	Impacted traits	Genes
1	26,710,358	27,057,982	RESP4, RESPA	
1	76,792,695	76,994,474	RESPA	
1	127,288,890	127,426,647	RESP2	*TRPC1, PLS1*
2	79,295,953	79,846,105	RESP4, RESPA	*GYPC, GLS*
9	43,616,136	43,960,964	RESP1	*QRSL1, RTN4IP1*
9	94,511,747	94,991,477	RESP3	
9	97,584,669	97,790,864	RESPA	*IGF2R, AIRN, SLC22A1*
9	101,044,571	101,286,417	RESP3	
14	11,490,338	11,816,360	RESP3	*ASAP1, FAM49B*
14	53,082,514	53,415,847	RESP1	
27	2,021,945	2,342,720	RESP2	
27	4,456,590	4,476,626	RESP2	*MCPH1, ANGPT2*
27	19,087,624	19,195,734	RESP2	*MICU3*
X	48,575,295	52,673,047	RESP4, RESPA	*BEX2, SRPX2, TSPAN5, TNMD*
X	109,250,657	109,739,240	RESP2, RESP4, RESPA	
X	126,011,140	127,636,818	RESP2, RESP4, RESPA	*ZFX, EIF2S3, KLHL15, CXHXorf58, APOO, SAT1, ACOT9, PRDX4, PTCHD1*

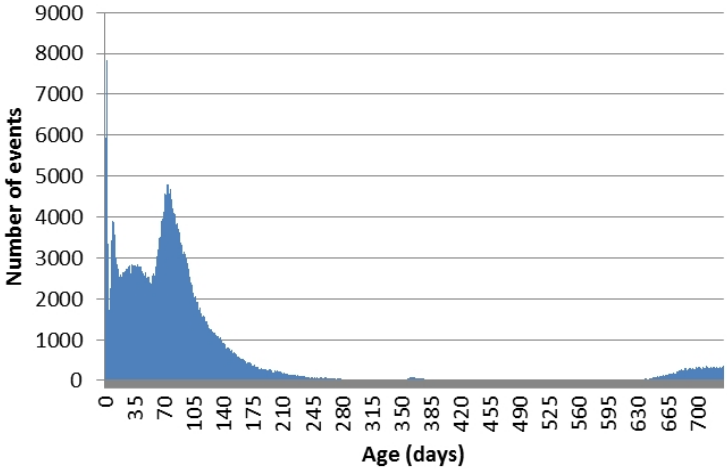
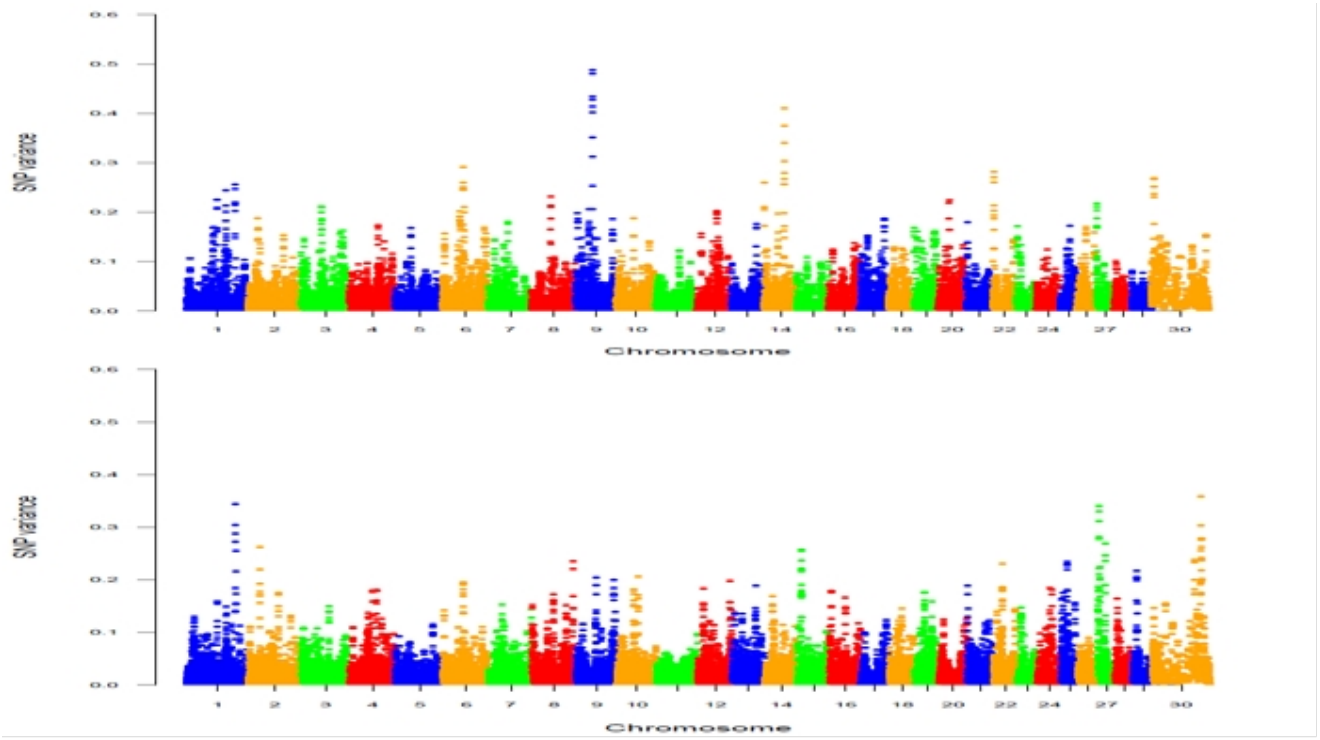
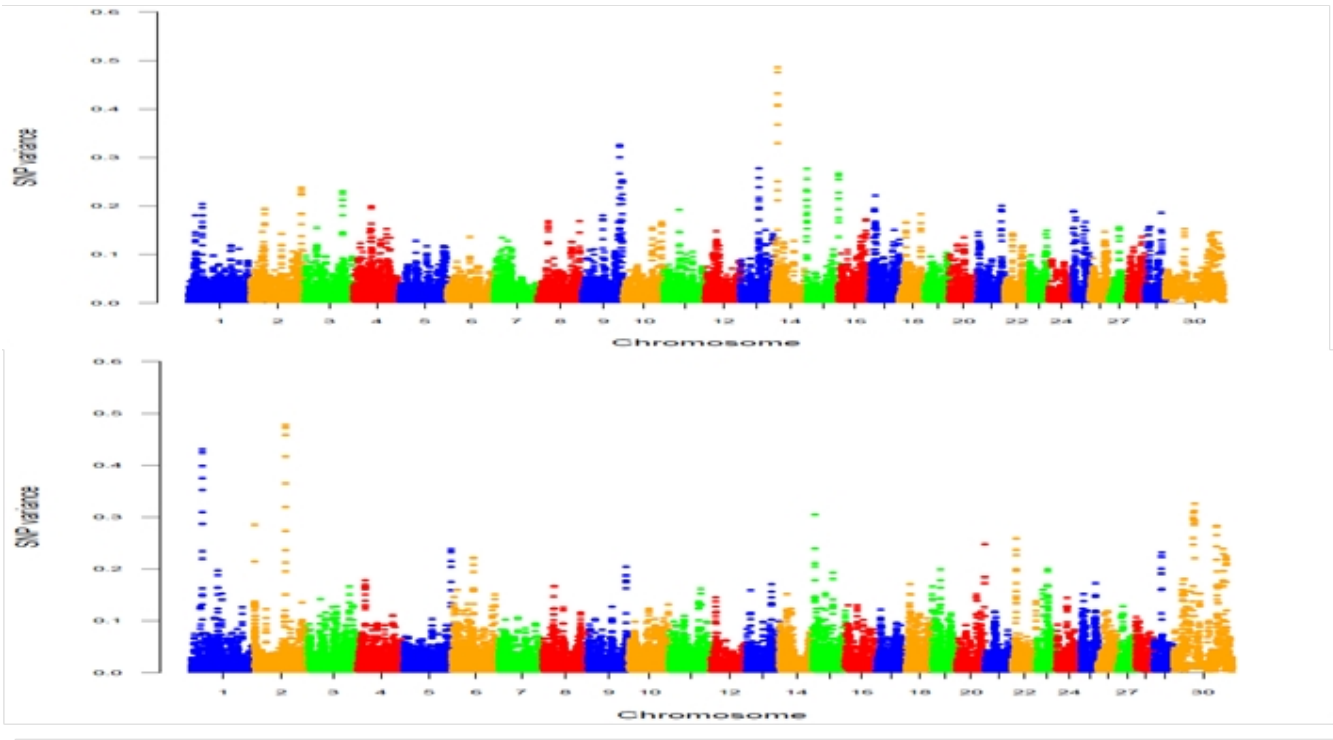


Figure 1: Distribution of calf respiratory disorders (RESP and PNEU) by age.


SNP variance
Chromosome


SNP variance
Chromosome

SNP variance
Chromosome

SNP variance
Chromosome

List of References

Aguilar, I, I, Misztal, S. Tsuruta, A. Legarra & H. Wang, 2014. PREGSF90 – POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada.

Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta & T.J. Lawlor, 2010. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743:752.

Fragomeni, B.O., D. A. L. Lourenco, S. Tsuruta, Y. Masuda, I. Aguilar, A. Legarra, T. J. Lawlor & I. Misztal, 2015. Hot topic: Use of genomic recursions in single-step genomic best linear unbiased predictor (BLUP) with a large number of genotypes. J. Dairy Sci. 98:4090–4094.

Gilmour, A. R., B.J. Gogel, B.R. Cullis, S.J. Welham & R. Thompson, 2015. ASReml User Guide Release 4.1 Structural Specification, VSN International Ltd, Hemel Hempstead, HP1 1ES, UK, www.vsni.co.uk.

Henderson, L., F. Miglior , A. Sewalem, J. Wormuth, D. Kelton, A. Robinson & K. E. Leslie, 2011. Short communication: Genetic parameters for measures of calf health in a population of Holstein calves in New York State. J. Dairy Sci. 94:6181–6187.

Heringstad, B., Y.M. Chang, D. Gianola & O. Østerås, 2008. Short Communication: Genetic analysis of respiratory disease in Norwegian Red calves. J. Dairy Sci. 91: 367–370.

Kent W.J., C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler & D. Haussler, 2002. The human genome browser at UCSC. Genome Res. 12(6):996-1006.

Legarra, A., I. Aguilar & I. Misztal, 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656-4663.

Masuda, Y., I. Misztal, S. Tsuruta, A. Legarra, I. Aguilar, D.A.L. Lourenco, B.O. Fragomeni, & T.J. Lawlor. 2016. Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals. J. Dairy Sci. 99:1968–1974.

McCorquodale, C.E., A. Sewalem, F. Miglior, D. Kelton, A. Robinson, A. Koeck & K.E. Leslie, 2013. Short communication: Analysis of health and survival in a population of Ontario Holstein heifer calves. J. Dairy Sci. 96:1880–1885.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet & D. H. Lee. 2002. BLUPF90 and related programs (BGF90). In Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France. CD-ROM Commun. 28:07.

Misztal, I., S. Tsuruta, D. Lourenco, I. Aguilar, A. Leggara & Z. Vitezica, 2014. Manual for BLUPF90 family of programs. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all4.pdf (accessed 07/27/2016).

Neibergs, H.L., C.M. Seabury, A.J Wojtowicz, Z. Wang, E. Scraggs, J.N Kiser, M. Neupane, J.E. Womack, A. Van Eenennaam, G.R. Hagevoort, T.W. Lehenbauer, S. Aly, J. Davis, J.F. Taylor & The Bovine Respiratory Disease Complex Coordinated Agricultural Project Research Team, 2014. Susceptibility loci revealed for bovine respiratory disease complex in pre-weaned Holstein calves. BMC Genomics 201415:1164.

Parker-Gaddis, K.L., J.B. Cole, J.S. Clay & C. Maltecca, 2014. Genomic selection for producer-recorded health event data in US dairy cattle. J. Dairy Sci. 97:3190-3199.

Rossini, K., 2004. Effects of calfood respiratory and digestive disease on calfhood morbidity and first lactation production and survival rates. M.S. Thesis in Dairy Science. Virginia

Tech, Blacksburg, VA. -

Sargolzaei, M., J.P. Chesnais & F.S. Schenkel, 2011. FImpute: An efficient imputation algorithm for dairy cattle populations. J. Anim. Sci. 89(E Suppl. 1)/J. Dairy Sci. 94(E Suppl. 1): 421 (abstr. 333).

Sischo. W.M., .D.W. Hird, I.A. Gardner, W.W. Utterback K.H. Christiansen, T.E. Carpenter & R. Heron, 1990. Economics of disease occurrence and prevention on California dairy farms: A report and evaluation of data collected for the national animal health monitoring system, 1986–87. Preventive Veterinary Medicine 8:141-156.

Vukasinovic, N., N. Bacciu, C.A. Przybyla, P. Boddhireddy & S.K. DeNise, 2017. Development of genetic and genomic evaluation for wellness traits in US Holstein cows. J. Dairy Sci. 100: 428–438.

Zhou X., Y. Ye, Y. Sun, X. Li, W. Wang, B. Privratsky, S. Tan, Z. Zhou, C. Huang, Y.Q. Wei, L. Birnbaumer, B.B. Singh & M. Wu, 2015. Transient Receptor Potential Channel 1 Deficiency Impairs Host Defense and Proinflammatory Responses to Bacterial Infection by Regulating Protein Kinase Cα Signaling. Mol Cell Biol. 35(16): 2729-2739.

Zwald, N.R., K.A. Weigel, Y.M. Chang, R.D. Welper & J.S. Clay, 2004. Genetic selection for health traits using producer-recorded data. I. Incidence rates, heritability estimates, and sire breeding values. J. Dairy Sci. 87:4287-4294.