



Marker selection and genomic prediction of economically important traits using imputed high-density genotypes for 5 breeds of dairy cattle

A. Al-Khudhair,¹ P. M. VanRaden,^{1*} D. J. Null, and B. Li¹

Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

ABSTRACT

Marker sets used in US dairy genomic predictions were previously expanded by including high-density (HD) or sequence markers with the largest effects for Holstein breed only. Other non-Holstein breeds lacked enough HD genotyped animals to be used as a reference population at that time, and thus were not included in the genomic prediction. Recently, numbers of non-Holstein breeds genotyped using HD panels reached an acceptable level for imputation and marker selection, allowing HD genomic prediction and HD marker selection for Holstein plus 4 other breeds. Genotypes for 351,461 Holsteins, 347,570 Jerseys, 42,346 Brown Swiss, 9,364 Ayrshires (including Red dairy cattle), and 4,599 Guernseys were imputed to the HD marker list that included 643,059 SNP. The separate HD reference populations included Illumina BovineHD (San Diego, CA) genotypes for 4,012 Holsteins, 407 Jerseys, 181 Brown Swiss, 527 Ayrshires, and 147 Guernseys. The 643,059 variants included the HD SNP and all 79,254 (80K) genetic markers and QTL used in routine national genomic evaluations. Before imputation, approximately 91 to 97% of genotypes were unknown for each breed; after imputation, 1.1% of Holstein, 3.2% of Jersey, 6.7% of Brown Swiss, 4.8% of Ayrshire, and 4.2% of Guernsey alleles remained unknown due to lower density haplotypes that had no matching HD haplotype. The higher remaining missing rates in non-Holstein breeds are mainly due to fewer HD genotyped animals in the imputation reference populations. Allele effects for up to 39 traits were estimated separately within each breed using phenotypic reference populations that included up to 6,157 Jersey males and 110,130 Jersey females. Correlations of HD with 80K genomic predictions for young animals averaged 0.986, 0.989, 0.985, 0.992, and

0.978 for Jersey, Ayrshire, Brown Swiss, Guernsey, and Holstein breeds, respectively. Correlations were highest for yield traits (about 0.991) and lowest for foot angle and rear legs–side view (0.981 and 0.982, respectively). Some HD effects were more than twice as large as the largest 80K SNP effect, and HD markers had larger effects than nearby 80K markers for many breed-trait combinations. Previous studies selected and included markers with large effects for Holstein traits; the newly selected HD markers should also improve non-Holstein and crossbred genomic predictions and were added to official US genomic predictions in April 2020.

Key words: marker selection, imputation, minor breeds, high-density genotype, genomic prediction

INTRODUCTION

In the last 2 decades, the fast-paced advancements in genotyping technologies dramatically transformed the era of dairy genomic selection. Genomic selection for dairy cattle is now more accurate because of larger reference populations and increased numbers of markers added to the DNA marker panel used for routine analysis of economically important traits (García-Ruiz et al., 2015; Lund et al., 2016). The accuracy of genomic evaluation can be further improved by (1) increasing the number of genotyped animals in the reference population (Wiggans et al., 2016; Hayes and Daetwyler, 2019), (2) including more accurate phenotypes connected to the genotyped animals directly or by an accurate pedigree record (Berry et al., 2014), and (3) strategically selecting and expanding the number of large-effect genomic markers included in the prediction model (Pausch et al., 2013; Tribout et al., 2020). In the United States, the well-established phenotypic and pedigree records of the Holstein breed and the large number of Holstein animals genotyped with high-density (HD) genomic panels has allowed more SNP to be selected and included in the genomic prediction strategy (VanRaden et al., 2013).

The availability of a large reference population plus whole-genome sequence data from the 1000 Bull Genomes Project (Hayes and Daetwyler, 2019) for hun-

Received July 9, 2020.

Accepted November 22, 2020.

*Corresponding author: Paul.VanRaden@usda.gov

†Current address: Scotland's Rural College, The Roslin Institute Building, Easter Bush, Edinburgh, EH25 9RG, United Kingdom.

Table 1. Number of animals genotyped using different density panels by breed

Genotyping array density	Holstein	Jersey	Brown Swiss	Ayrshire	Guernsey
High (reference population)	4,012	407	181	527	147
Medium (30,000–150,000 SNP)	166,440	60,395	28,061	3,027	1,544
Low (<30,000 SNP)	181,009	286,768	14,104	5,810	2,908
All	351,461	347,570	42,346	9,364	4,599

dreds of Holsteins has allowed further expansion of the SNP set used in routine predictions (Brøndum et al., 2015; VanRaden et al., 2017). Use of either HD chips (>600,000 SNP) or whole-genome sequence also requires the development and validation of computational tools to impute large numbers of genomic variants for animals genotyped with low-density (LD; <30,000 SNP) and medium-density (MD; 30,000–150,000 SNP) chips (García-Ruiz et al., 2015; Lund et al., 2016). Genomic prediction accuracy could be significantly improved by expanding the pool of markers used in the prediction model (VanRaden et al., 2017).

Many non-Holstein breeds such as Ayrshire, Brown Swiss, Guernsey, and Jersey have limited numbers of animals genotyped with HD marker panels. A smaller reference population size would reduce the accuracy of estimating the markers' effects and the overall reliability of the genomic predictions (Cooper et al., 2016; Wiggans et al., 2016; Hayes and Daetwyler, 2019; Tribout et al., 2020). These breeds are still lagging behind in using HD panels for genomic selection because of the cost associated with HD chips compared with LD and MD chips (Berry et al., 2014; García-Ruiz et al., 2015). Recently, more non-Holstein animals have been genotyped with HD marker panels, which provides an opportunity to use them in the reference population as well as imputing HD genotypes from LD and MD marker panels. This should increase the population size available for non-Holstein breeds, facilitate more accurate calling of large-effect markers, and improve the reliability of genome-wide association analysis to select variants for future genomic prediction (Hozé et al., 2013; Ma et al., 2013; Pausch et al., 2013). Predictions can improve by selecting markers from multiple breeds instead of just 1 (Kemper et al., 2015).

The main objectives of this study were to (1) use HD genotypes imputed from LD and MD marker panels to estimate, identify, and select markers with the largest-effect SNP for up to 39 economically important traits of 5 dairy cattle breeds, (2) compare prediction models using 79,254 (80K) markers versus imputed HD markers for the non-Holstein breeds, and (3) allow future DNA chips and routine US predictions to include, for the first time, large-effect markers selected from HD genotypes for Ayrshire, Brown Swiss, Guernsey, and Jersey breeds.

MATERIALS AND METHODS

Study Populations and Imputation

Genomic, phenotypic, and pedigree data were supplied by the Council on Dairy Cattle Breeding (CDCB; Bowie, MD) from the national cooperator database including evaluations and pedigrees of foreign bulls from Interbull (Uppsala, Sweden). Genotypes for 755,340 animals (351,461 Holsteins, 347,570 Jerseys, 42,346 Brown Swiss, 9,364 Ayrshires, and 4,599 Guernseys) were examined and are categorized by chip density (HD, MD, or LD) in Table 1. Millions of Holsteins have been genotyped, but only Holstein bulls, their ancestors, and females with HD genotypes were included, whereas all genotypes of the other breeds were included. All LD and MD genotypes were imputed to HD using Findhap. f90, version 3 (VanRaden et al., 2015).

A new reference assembly of the bovine genome, ARS-UCD1, was applied to imputation, and the new map has improved performances in marker locations, sequence alignment, and genotype imputation compared with the previous UMD 3.1 reference assembly (Null et al., 2019; Rosen et al., 2020). Quality control of the imputed genotypes followed (Wiggans et al., 2010). The imputed genotypes were not pruned for high LD as in previous HD studies for Holsteins (Wiggans et al., 2016; Ye et al., 2019), and all HD markers were kept to improve marker density and accuracy of potential later imputation to sequence. Available SNP were categorized as those from the initial 50K chip, those already included in 80K routine predictions as of April 2019, those from the HD chip, sequence SNP previously selected from Holstein data and added to chips (VanRaden et al., 2017), and QTL included in recent chips by genotyping laboratories (Wiggans et al., 2016). Chip manifests introduced after 2016 were compared with previous SNP lists to identify additional QTL now being genotyped.

Phenotypic Reference Population

The multistep genomic predictions used deregressed national and international PTA as phenotypes. The phenotypic reference population included genotyped bulls and cows with progeny records or their own phe-

Table 2. Numbers of animals and traits for estimation of SNP effects by breed

Animals and traits (number)	Holstein	Jersey	Brown Swiss	Ayrshire	Guernsey
Progeny-tested bulls	43,140	6,157	7,112	766	447
Cows with records	40,835	110,130	3,252	550	1,238
Total reference population	83,975	116,287	10,364	1,316	1,685
Traits	39	26	29	26	26

notypes in the conventional pedigree evaluations, or the conventional multitrait across-country evaluations from Interbull. The remaining young animals without records or progeny were used as a test population to compare predictions. Numbers of animals and traits per breed used for the SNP effect estimation are shown in Table 2. Number of progeny-tested bulls (bulls with daughter milk records) ranged from 447 for Guernseys to 43,140 for Holsteins. Number of cows with milk records of each breed used as a reference ranged from 550 for Guernseys to 40,835 for Holsteins. Reference population size was slightly reduced for some other traits (e.g., when a bull had no progeny or cow had no phenotype for that trait). The phenotypic reference for each trait was used to estimate SNP effects with the 643,059 SNP set and the 80K SNP set used routinely, and then genomic predictions from the 2 sets were compared. The largest SNP effects from the HD marker set were selected for potential inclusion in future chips and in routine evaluations.

Estimation of SNP Effects and Marker Selection

After imputation, the 643,059 SNP were used for estimation of SNP effects for 26 to 39 traits dependent on breed (Table 2). Those markers included the original SNP from the HD chip as well as all the 80K markers. The iteration model to estimate SNP effects is described by VanRaden (2008) and is based on a Bayes A algorithm.

To compare SNP estimates across traits and breeds, absolute effects were divided by the square root of variance of effect sizes within trait and breed to obtain standardized effect sizes (**genSD**). To prevent selecting too many markers for the same QTL, the genome was divided into nonoverlapping 10-Mb windows, and num-

bers of selected SNP were limited within each window. For Holsteins, which have a large reference population, markers were selected if **genSD** was >5 , but with a maximum of the 5 largest SNP selected per window. For Jerseys, **genSD** had to be >7 , and a maximum of only 3 SNP were selected per window. For Brown Swiss, Ayrshires, and Guernseys, **genSD** was required to be >10 , with a maximum of 2 SNP selected per window. The more stringent requirements for the smaller breeds were because smaller reference populations tend to reduce accuracy for imputation and estimation of the SNP effects, and because fewer breeders would benefit from SNP selected for those populations.

RESULTS AND DISCUSSION

Marker Availability Before and After Imputation

Before imputation, percentage of markers that had genotypes available ranged from 3.5 for Jerseys to 10.4 for Ayrshires; after imputation, the percentage of called plus imputed HD markers ranged from 93.3 for Brown Swiss to 98.9 for Holsteins (Table 3). Thus, about 90% of markers available for SNP effect estimation were from imputation rather than from the original called LD, MD, and HD genotypes within each breed. The remaining small fractions of alleles that are not filled by imputation primarily result from genotypes observed at lower density that have no matching haplotype observed at higher density, and the remaining missing alleles within those haplotypes are set equal to allele frequencies during marker effect estimation. This imputation loss affects HD marker selection, but future effects computed in routine prediction should be more accurate after the selected markers are added to chips and genotyped directly.

Table 3. Percentages of medium- and low-density SNP genotypes available for effect estimation before and after imputation¹ by breed

SNP genotype	Holstein	Jersey	Brown Swiss	Ayrshire	Guernsey
Before imputation	6.8	3.5	6.2	10.4	8.6
After imputation	98.9	96.8	93.3	95.2	95.8

¹Genotypes were imputed using Findhap, version 3 (VanRaden et al., 2015).

Table 4. Overview of the selected marker list of large-effect SNP

SNP genotyping category	Available		Selected		Percentage of available markers selected
	Number	%	Number	%	
80,000 SNP list	79,254	12.32	3,127	35.05	3.95
50,000 SNP chip	46,290	7.20	1,298	14.55	2.80
Sequence ¹	750	0.12	86	0.96	11.47
QTL ²	82	0.01	8	0.09	9.76
Original high-density chip	554,014	86.15	5,539	62.08	1.00
All SNP	643,059		8,922		1.39

¹SNP previously selected from run 5 sequence data from 1000 Bull Genomes Project (Hayes and Daetwyler, 2019) and added to chips.

²Known causal mutations previously added to chips.

The percentage available for each breed after imputation appeared to be dependent mainly on the number of HD genotypes in the reference population (Table 1), the percentage of LD chips used, and, to a lesser extent, breed population structure. For example, Jerseys had fewer SNP missing after imputation than did Ayrshires despite slightly more HD genotypes for Ayrshires, perhaps because Ayrshires and other Red dairy cattle may be genetically more diverse than Jerseys (Brøndum et al., 2011). At least 50,000 markers may be needed to achieve accurate imputation to HD (Chud et al., 2015). Other studies (Gao et al., 2013; Hozé et al., 2013; Ma et al., 2013; Pausch et al., 2013) also highlighted the importance of pedigree information for more accurate imputation. Prephasing of the tested population and its reference resulted in faster computation with higher imputation accuracy and less biased genomic prediction. Imputation using Findhap (version 3) with 24 processors took <2 d for each breed. Future research could compare HD imputation and marker selection from all breeds simultaneously versus separately.

Largest SNP Effects from Imputed HD Markers

The total number of large-effect SNP markers selected was 11,045, including 8,922 unique SNP markers because some large-effect SNP overlap among breeds or traits. The high proportion of unique to total markers indicated little overlap across breeds. The selected

marker list consisted of large-effect SNP from imputed HD markers for all of the studied traits in 5 dairy cattle breeds. Among all of the selected large-effect SNP, 35% of the selected SNP were within the official 80K marker set and 62% were from HD chip (Table 4). In proportion to the number of available SNP, the selected SNP included 11.47% of the SNP previously added from large-effect sequence markers and 9.76% of the QTL that are now genotyped on several chips. The number of selected SNP markers among breeds are in Table 5. More than half of selected SNP (55.8%) of the total SNP were from Holsteins, and the rest of the SNP (44.2%) were distributed almost evenly among the other 4 breeds.

The top 5 traits that had the most selected markers were heifer conception rate (3.9%), followed by mastitis (3.7%), net merit (3.7%), metritis (3.6%), and rear udder height (3.3%) as listed in Table 6. All traits had reasonable numbers of SNP selected, ranging from 90 to 426 (Supplemental Table S1, https://www.ars.usda.gov/ARSUserFiles/80420530/Publications/Scientific/Journals/JDS-19260_SupplTableS1.pdf). The traits were expected to be almost evenly represented, but some had more or fewer SNP above the limit. An alternative would be to select SNP in proportion to relative economic value, but that could overlook SNP with a large effect on unique traits of interest but small effects on the main index.

The HD markers had larger effects than nearby 80K markers for many breed-trait combinations. Some were

Table 5. Number of SNP selected per breed

Breed	Number of SNP	Percentage of total
Holstein	6,162	55.8
Jersey	1,423	12.9
Brown Swiss	1,145	10.4
Ayrshire	1,304	11.8
Guernsey	1,011	9.2
Total	11,045	100

Table 6. Number of SNP selected for the top 5 traits

Trait	Number of SNP	Percentage of total
Heifer conception rate	426	3.9
Mastitis	412	3.7
Net merit	409	3.7
Metritis	394	3.6
Rear udder height	366	3.3

Table 7. Correlations (Corr) of high-density (HD) with 80K genomic predictions for young animals (bulls and heifers)

Item	Jersey	Ayrshire	Brown Swiss	Guernsey	Holstein
Corr (HD, 80K)	0.988	0.991	0.970	0.921	0.983

expected simply from the larger number of HD markers, but several had consistently larger effects across traits. For Jerseys, HD SNP on chromosome 11 at 104 Mb had the largest effects for fore udder attachment, front teat placement, rump width, and rump angle. For udder cleft and teat length, HD SNP had larger effects than for the highest 80K SNP. Large new effects were also discovered from HD SNP for daughter pregnancy and cow conception rates. For Brown Swiss, HD markers had the largest effects for fat yield, fat percentage, SCS, cow conception rate, and several type traits. For Ayrshires, HD markers on chromosome X at 118 Mb had the largest effects on milk, fat, and protein yields. Other traits on which HD markers had the largest effects were net merit, productive life, SCS, daughter pregnancy rate, heifer and cow conception rates, and several type traits. For Guernseys, HD markers had larger effects than the previously discovered very large effects on chromosome 19 at 28 Mb for many traits (Cooper et al., 2016).

Gene tests previously reported to have large effects had large effects in the HD analysis and in the routine predictions. The *DGAT1* gene test had about the same effect size as nearby 50K and HD markers, indicating close linkage or some imputation loss because *DGAT1* is not genotyped in most animals (Grisart et al., 2002). Gene tests for the bovine growth hormone receptor GHRJA microsatellite (Blott et al., 2003) had the largest effects on fat and protein percentages in Brown Swiss and Ayrshires. The β -*LG* gene (Ganai et al., 2009) test had the largest effect for fat yield in Brown Swiss, but for protein in Jerseys, it had smaller effects than 3 nearby HD SNP. The *MC1R* gene test for red hair color (Klungland et al., 1995) had large effects for net merit, protein, productive life, final score, rear udder height, and foot angle for Jerseys, and protein for Brown Swiss, perhaps indicating some introgression of Holstein DNA rather than a direct effect of *MC1R*. This report does not focus on Holsteins, but the *ABCG2* (Cohen-Zinder et al., 2005) gene test had the largest effect on net merit; the new cholesterol deficiency mutation test had large effects on SCS and teat length, and new SNP added to track the polled mutation had large effects on SCS and front teat placement.

The selected markers were provided by CDCB to approved genotyping laboratories in September 2019 for

potential inclusion in their future chips. In April 2020, the selected markers from this project were included in routine genomic prediction and SNP effect estimation by CDCB. Additional HD markers selected with the largest effects on residual feed intake in Holsteins in a companion study (Li et al., 2019) were provided to the laboratories and included in routine genomic prediction at these same times. Instead of only adding SNP, thousands of previously used SNP with the smallest effects were also removed from routine predictions to reduce the growth in computation.

Genomic Predictions and Marker Effects (HD vs. 80K)

The genomic predictions using the 80K markers versus HD imputed markers were further compared for the young genotyped animals that did not yet have phenotypes. The prediction correlations in Table 7 were high for Holstein, Jersey, and Ayrshire (0.983, 0.988, and 0.991, respectively), but lower for Brown Swiss (0.970) and Guernsey breeds (0.921). The lower prediction correlations in Brown Swiss and Guernsey breeds could be caused by the smaller HD reference populations available for imputing SNP markers from LD and MD chips. More detailed correlations of the genomic predictions for each breed and trait are in Supplemental Table S2 (https://www.ars.usda.gov/ARSUserFiles/80420530/Publications/Scientific/Journals/JDS-19260_SupplTableS2.pdf).

An advantage of the imputed HD markers over the 80K markers was that imputed HD markers often had larger SNP effects when compared with those previously used from the 80K list. Figure 1 shows an example of SNP effects calculated from the imputed HD and from 80K. The marker effects are for productive life and are calculated using data for Guernseys, which have the smallest breed population size and a major QTL on chromosome 19. The Manhattan plot in Figure 1 panel A shows the BTA19 SNP effects calculated for the HD chip (green) versus for the 80K (red); the HD SNP (green) are very condensed and form a higher peak compared with a lower condensed and shorter peak for the 80K SNP (red). The same result is confirmed when comparing variances explained by these markers. The green peak for the HD imputed markers forms a higher

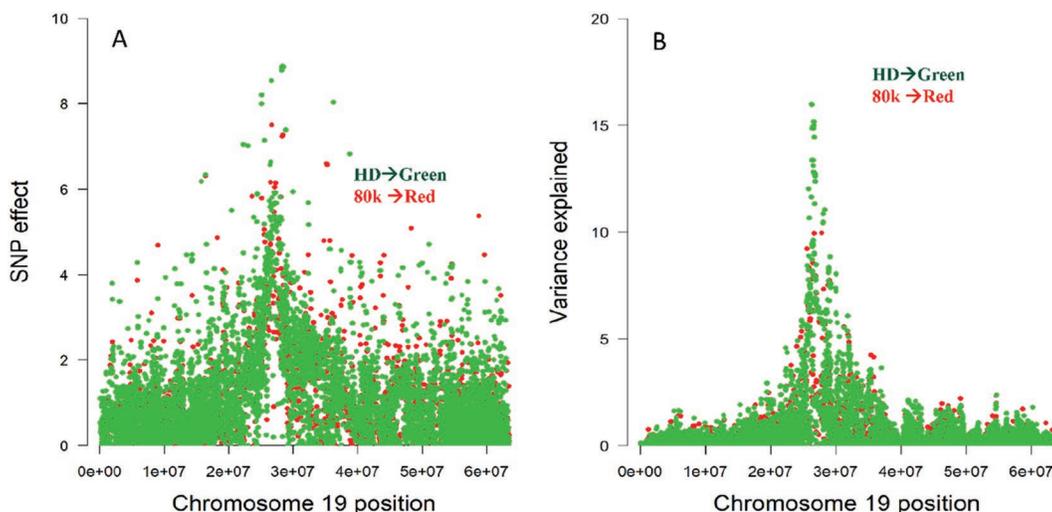


Figure 1. Manhattan plots of (A) marker effects and (B) explained variance across BTA19 for productive life of Guernseys based on high-density (HD; green) and 80K (red) SNP.

and more condensed peak, representing a larger effect and better quality compared with the same markers from the 80K.

Overall, the SNP selection could identify important new SNP to include in future chips and in the genomic prediction, despite the limited HD genotyped reference population. Directly using all HD markers often gives little advantage, even for across-breed prediction (Harris et al., 2011). The HD chip is useful for discovery, but routine predictions cannot afford to include all HD markers or sequence variants available that provide only small benefits, especially for the millions of Holsteins. Genomic prediction reliabilities from imputed data could perhaps have higher accuracy by utilizing a Bayesian mixture model or exponential power model (Su et al., 2012; Gao et al., 2013), or restricting the HD markers based on gene annotation (Erbe et al., 2012).

CONCLUSIONS

Markers with the largest effects for multiple economically important traits were selected from imputed HD genotypes for 5 dairy breeds using methods previously applied to Holsteins. Smaller HD reference populations in those non-Holstein populations limited the imputation accuracy, but many of the imputed HD markers showed larger estimated effects than the nearby 80K markers used previously. High correlations, averaging 0.986 for HD with 80K predictions, did not indicate major benefits from additional markers. Given that both marker sets resulted in very similar breeding val-

ues, using the 80K set is preferable because it has a cost advantage over the HD chip. Additionally, including selected markers and QTL directly on genotyping chips will reduce imputation losses and improve prediction accuracy in the future. As indicated by previous studies, adding large-effect SNP or replacing 80K markers with nearby HD SNP, sequence SNP, or QTL that have larger effects is anticipated to gradually improve reliability for all breeds and crossbreds.

ACKNOWLEDGMENTS

The authors thank the Council on Dairy Cattle Breeding (Bowie, MD) and dairy industry cooperators for supplying data. Al-Khudhair, VanRaden, and Null were supported by appropriated project 8042-31000-002-00-D, “Improving Dairy Animals by Increasing Accuracy of Genomic Prediction, Evaluating New Traits, and Redefining Selection Goals,” of USDA’s Agricultural Research Service (ARS, Washington, DC). Li was supported in part by an appointment to the ARS Research Participation Program, administered by the Oak Ridge Institute for Science and Education (Oak Ridge, TN) through an interagency agreement between the US Department of Energy (Washington, DC) and ARS. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. The USDA is an equal opportunity provider and employer. The authors have not stated any conflicts of interest.

REFERENCES

- Berry, D. P., M. C. McClure, and M. P. Mullen. 2014. Within- and across-breed imputation of high-density genotypes in dairy and beef cattle from medium- and low-density genotypes. *J. Anim. Breed. Genet.* 131:165–172. <https://doi.org/10.1111/jbg.12067>.
- Blott, S., J.-J. Kim, S. Moiso, A. Schmidt-Küntzel, A. Cornet, P. Berzi, N. Cambisano, C. Ford, B. Grisart, D. Johnson, L. Karim, P. Simon, R. Snell, R. Spelman, J. Wong, J. Vilkki, M. Georges, F. Farnir, and W. Coppieters. 2003. Molecular dissection of a quantitative trait locus: A phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* 163:253–266.
- Brøndum, R. F., E. Rius-Vilarrasa, I. Strandén, G. Su, B. Guldbbrandtzen, W. F. Fikse, and M. S. Lund. 2011. Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. *J. Dairy Sci.* 94:4700–4707. <https://doi.org/10.3168/jds.2010-3765>.
- Brøndum, R. F., G. Su, L. Janss, G. Sahana, B. Guldbbrandtzen, 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>.
- Chud, T. C. S., R. V. Ventura, F. S. Schenkel, R. Carvalheiro, M. E. Buzanskas, J. O. Rosa, M. A. Mudadu, M. V. G. B. da Silva, F. B. Mokry, C. R. Marcondes, L. C. A. Regitano, and D. P. Munari. 2015. Strategies for genotype imputation in composite beef cattle. *BMC Genet.* 16:99. <https://doi.org/10.1186/s12863-015-0251-7>.
- Cohen-Zinder, M., E. Seroussi, D. M. Larkin, J. J. Loor, A. Evertsvan der Wind, J.-H. Lee, J. K. Drackley, M. R. Band, A. G. Hernandez, M. Shani, H. A. Lewin, J. I. Weller, and M. Ron. 2005. Identification of a missense mutation in the bovine *ABCG2* gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Res.* 15:936–944. <https://doi.org/10.1101/gr.3806705>.
- Cooper, T. A., S. A. E. Eaglen, G. R. Wiggans, J. Jenko, H. J. Huson, D. R. Morrice, M. Bichard, W. G. L. Luff, and J. A. Woolliams. 2016. Genomic evaluation, breed identification, and population structure of Guernsey cattle in North America, Great Britain, and the Isle of Guernsey. *J. Dairy Sci.* 99:5508–5515. <https://doi.org/10.3168/jds.2015-10445>.
- Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, and M. E. Goddard. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.* 95:4114–4129. <https://doi.org/10.3168/jds.2011-5019>.
- Ganai, N. A., H. Bovenhuis, J. A. M. Van Arendonk, and M. H. P. W. Visker. 2009. Novel polymorphisms in the bovine β -lactoglobulin gene and their effects on β -lactoglobulin protein concentration in milk. *Anim. Genet.* 40:127–133. <https://doi.org/10.1111/j.1365-2052.2008.01806.x>.
- Gao, H., G. Su, L. Janss, Y. Zhang, and M. S. Lund. 2013. Model comparison on genomic predictions using high-density markers for different groups of bulls in the Nordic Holstein population. *J. Dairy Sci.* 96:4678–4687. <https://doi.org/10.3168/jds.2012-6406>.
- García-Ruiz, A., F. J. Ruiz-Lopez, G. R. Wiggans, C. P. Van Tassell, and H. H. Montaldo. 2015. Effect of reference population size and available ancestor genotypes on imputation of Mexican Holstein genotypes. *J. Dairy Sci.* 98:3478–3484. <https://doi.org/10.3168/jds.2014-9132>.
- Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. *Genome Res.* 12:222–231. <https://doi.org/10.1101/gr.224202>.
- Harris, B., F. Creagh, A. Winkelman, and D. Johnson. 2011. Experiences with the Illumina High Density Bovine BeadChip. *Interbull Bull.* 44:3–7.
- Hayes, B. J., and H. D. Daetwyler. 2019. 1000 Bull Genomes Project to map simple and complex genetic traits in cattle: Applications and outcomes. *Annu. Rev. Anim. Biosci.* 7:89–102. <https://doi.org/10.1146/annurev-animal-020518-115024>.
- Hozé, C., M.-N. Fouilloux, E. Venot, F. Guillaume, R. Dassonneville, S. Fritz, V. Ducrocq, F. Phocas, D. Boichard, and P. Croiseau. 2013. High-density marker imputation accuracy in sixteen French cattle breeds. *Genet. Sel. Evol.* 45:33. <https://doi.org/10.1186/1297-9686-45-33>.
- Kemper, K. E., C. M. Reich, P. J. Bowman, C. J. vander Jagt, A. J. Chamberlain, B. A. Mason, B. J. Hayes, and M. E. Goddard. 2015. Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. *Genet. Sel. Evol.* 47:29. <https://doi.org/10.1186/s12711-014-0074-4>.
- Klungland, H., D. I. Vage, L. Gomez-Raya, S. Adalsteinsson, and S. Lien. 1995. The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mamm. Genome* 6:636–639. <https://doi.org/10.1007/BF00352371>.
- Li, B., L. Fang, D. J. Null, J. L. Hutchison, E. E. Connor, P. M. VanRaden, M. J. VandeHaar, R. J. Tempelman, K. A. Weigel, and J. B. Cole. 2019. High-density genome-wide association study for residual feed intake in Holstein dairy cattle. *J. Dairy Sci.* 102:11067–11080. <https://doi.org/10.3168/jds.2019-16645>.
- Lund, M. S., I. van den Berg, P. Ma, R. F. Brøndum, and G. Su. 2016. Review: How to improve genomic predictions in small dairy cattle populations. *Animal* 10:1042–1049. <https://doi.org/10.1017/S1751731115003031>.
- Ma, P., R. F. Brøndum, Q. Zhang, M. S. Lund, and G. Su. 2013. Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle. *J. Dairy Sci.* 96:4666–4677. <https://doi.org/10.3168/jds.2012-6316>.
- Null, D. J., P. M. VanRaden, B. D. Rosen, J. R. O'Connell, and D. M. Bickhart. 2019. Using the ARS-UCD1.2 reference genome in U.S. evaluations. *Interbull Bull.* 55:30–34.
- Pausch, H., B. Aigner, R. Emmerling, C. Edel, K.-U. Götz, and R. Fries. 2013. Imputation of high-density genotypes in the Fleckvieh cattle population. *Genet. Sel. Evol.* 45:3. <https://doi.org/10.1186/1297-9686-45-3>.
- Rosen, B. D., D. M. Bickhart, R. D. Schnabel, S. Koren, C. G. Elsik, E. Tseng, T. N. Rowan, W. Y. Low, A. Zimin, C. Coudrey, R. Hall, W. Li, A. Rhie, J. Ghurye, S. D. McKay, F. Thibaud-Nissen, J. Hoffman, B. M. Murdoch, W. M. Snelling, T. G. McDaniel, J. A. Hammond, J. C. Schwartz, W. Nandolo, D. E. Hagen, C. Dreischer, S. J. Schultheiss, S. G. Schroeder, A. M. Phillipy, J. B. Cole, C. P. Van Tassell, G. Liu, T. P. L. Smith, and J. F. Medrano. 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* 9:giaa021. <https://doi.org/10.1093/gigascience/giaa021>.
- Su, G., R. F. Brøndum, P. Ma, B. Guldbbrandtzen, G. P. Aamand, and M. S. Lund. 2012. Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. *J. Dairy Sci.* 95:4657–4665. <https://doi.org/10.3168/jds.2012-5379>.
- Tribout, T., P. Croiseau, R. Lefebvre, A. Barbat, M. Boussaha, S. Fritz, D. Boichard, C. Hoze, and M.-P. Sanchez. 2020. Confirmed effects of candidate variants for milk production, udder health, and udder morphology in dairy cattle. *Genet. Sel. Evol.* 52:55. <https://doi.org/10.1186/s12711-020-00575-1>.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>.
- VanRaden, P. M., D. J. Null, M. Sargolzaei, G. R. Wiggans, M. E. Tooker, J. B. Cole, T. S. Sonstegard, E. E. Connor, M. Winters, J. B. C. H. M. van Kaam, A. Valentini, B. J. Van Doormaal, M. A.

- Faust, and G. A. Doak. 2013. Genomic imputation and evaluation using high density Holstein genotypes. *J. Dairy Sci.* 96:668–678. <https://doi.org/10.3168/jds.2012-5702>.
- VanRaden, P. M., C. Sun, and J. R. O'Connell. 2015. Fast imputation using medium or low-coverage sequence data. *BMC Genet.* 16:82. <https://doi.org/10.1186/s12863-015-0243-7>.
- VanRaden, P. M., M. E. Tooker, J. R. O'Connell, J. B. Cole, and D. M. Bickhart. 2017. Selecting sequence variants to improve genomic predictions for dairy cattle. *Genet. Sel. Evol.* 49:32. <https://doi.org/10.1186/s12711-017-0307-4>.
- Wiggans, G. R., T. A. Cooper, P. M. VanRaden, C. P. Van Tassell, D. M. Bickhart, and T. S. Sonstegard. 2016. Increasing the number of single nucleotide polymorphisms used in genomic evaluation of dairy cattle. *J. Dairy Sci.* 99:4504–4511. <https://doi.org/10.3168/jds.2015-10456>.
- Wiggans, G. R., P. M. VanRaden, L. R. Bacheller, M. E. Tooker, J. L. Hutchison, T. A. Cooper, and T. S. Sonstegard. 2010. Selection and management of DNA markers for use in genomic evaluation. *J. Dairy Sci.* 93:2287–2292. <https://doi.org/10.3168/jds.2009-2773>.
- Ye, S., N. Gao, R. Zheng, Z. Chen, J. Teng, X. Yuan, H. Zhang, Z. Chen, X. Zhang, J. Li, and Z. Zhang. 2019. Strategies for obtaining and pruning imputed whole-genome sequence data for genomic prediction. *Front. Genet.* 10:673. <https://doi.org/10.3389/fgene.2019.00673>.

ORCIDS

- A. Al-Khudhair  <https://orcid.org/0000-0002-4944-6404>
P. M. VanRaden  <https://orcid.org/0000-0002-9123-7278>
B. Li  <https://orcid.org/0000-0002-8528-1889>