J. Dairy Sci. 104
https://doi.org/10.3168/jds.2020-19879

Genome-wide association study between copy number variants and hoof health traits in Holstein dairy cattle

Adrien M. Butty,[1] Tatiane C. S. Chud,[1] Diercles F. Cardoso,[1] Lucas S. F. Lopes,[1] Filippo Miglior,[1] Flavio S. Schenkel,[1] Angela Cánovas,[1] Irene M. Häfliger,[2] Cord Drögemüller,[2] Paul Stothard,[3] Francesca Malchiodi,[1,4] and Christine F. Baes[1,2]*

[1]Department of Animal Biosciences, Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, Ontario N1G 2W1, Canada
[2]Vetsuisse Faculty, Institute of Genetics, University of Bern, Bern 3012, Switzerland
[3]Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton T6G 2R3, Canada
[4]The Semex Alliance, Guelph, Ontario N1H 6J2, Canada

ABSTRACT

Genome-wide association studies based on SNP have been completed for multiple traits in dairy cattle; however, copy number variants (CNV) could add genomic information that has yet to be harnessed. The objectives of this study were to identify CNV in genotyped Holstein animals and assess their association with hoof health traits using deregressed estimated breeding values as pseudophenotypes. A total of 23,256 CNV comprising 1,645 genomic regions were identified in 5,845 animals. Fourteen genomic regions harboring structural variations, including 9 deletions and 5 duplications, were associated with at least 1 of the studied hoof health traits. This group of traits included digital dermatitis, interdigital dermatitis, heel horn erosion, sole ulcer, white line lesion, sole hemorrhage, and interdigital hyperplasia; no regions were associated with toe ulcer. Twenty candidate genes overlapped with the regions associated with these traits including *SCART1*, *NRXN2*, *KIF26A*, *GPHN*, and *OR7A17*. In this study, an effect on infectious hoof lesions could be attributed to the *PRAME* (Preferentially Expressed Antigen in Melanoma) gene. Almost all genes detected in association with noninfectious hoof lesions could be linked to known metabolic disorders. The knowledge obtained considering information of associated CNV to the traits of interest in this study could improve the accuracy of estimated breeding values. This may further increase the genetic gain for these traits in the Canadian Holstein population, thus reducing the involuntary animal losses due to lameness.

Received November 7, 2020.
Accepted January 31, 2021.
*Corresponding author: cbaes@uoguelph.ca

Key words: genotype, dairy cattle, structural variant, functional analysis

INTRODUCTION

Since the implementation of genomic selection in dairy cattle, millions of animals have been genotyped and evaluated. Single nucleotide polymorphisms have been the main type of genetic variant used in dairy cattle genomic research. Genomic studies have greatly improved our understanding of the genetic architecture of many economically important traits and diseases and have contributed to maximizing genetic gain through selection (Goddard et al., 2016). Several GWAS have been performed to elucidate the genetic background of dairy cattle traits. However, only a few GWAS have been conducted with different types of variants, such as copy number variants (**CNV**). The CNV are inheritable chromosomal structural variations in the form of deletions or insertions greater than 50 bp (Sudmant et al., 2015), which cover a greater percentage of the genome than SNP (Fadista et al., 2010; Stothard et al., 2011). In addition, the lack of linkage disequilibrium between any SNP and 25% of the detected CNV led to the conclusion that CNV carry information that cannot be detected solely by SNP (Xu et al., 2014; Hay et al., 2018). In other words, SNP can be used to tag three-quarters of the CNV information, but one-quarter remains untagged. Therefore, CNV are proposed as an additional information source to explain the genetic variance of complex traits not accounted for by SNP alone (Hay et al., 2018).

Identification and genotyping of CNV is challenging, which may explain the limited number of CNV association studies in livestock (Bickhart and Liu, 2014). As a result, imputation methods for CNV are not well established (Butty et al., 2019). Most commonly used methods to detect CNV include comparative genome

hybridization, obtaining CNV information from SNP arrays, and whole-genome sequence (**WGS**) approaches (Alkan et al., 2011; Yang et al., 2018; Butty et al., 2020). The latter might be more precise in detecting CNV boundaries than array methods (Alkan et al., 2011). Array-based approaches would suit scenarios in which the downstream CNV analyses include phenotype association studies, due to the high number of available samples (Spencer et al., 2009; Yang et al., 2018). Each CNV detection method handles the control of false discovery rates differently; therefore, common results between different identification methods and types of information may represent CNV with higher confidence (Zhan et al., 2011; Rafter et al., 2020). Another factor affecting the accurate identification of structural variants is the quality of the reference assembly used to map either WGS or SNP array information (Winchester et al., 2009; Baes et al., 2014; Pirooznia et al., 2015). In this sense, the recently released bovine reference genome, ARS-UCD1.2 (Rosen et al., 2018), might enable more precise CNV identification in cattle and benefit further association tests.

Copy number variants have been associated with economically important traits in dairy cattle, including reproduction, health, and conformation traits (Glick et al., 2011; Xu et al., 2014; Ben Sassi et al., 2016; Durán Aguilar et al., 2017; Prinsen et al., 2017; Zhou et al., 2018; Liu et al., 2019). However, associations between CNV and traits related to hoof health are yet to be assessed. Hoof disorders are the third major reason for premature culling in dairy cattle worldwide, after mastitis and reproductive failures (Heringstad et al., 2018). Canadian data showed that approximately 40% of Holstein cows presented to hoof trimmers had at least 1 foot disorder (Chapinal et al., 2013; Malchiodi et al., 2017). Hoof health can be genetically improved; since 2018, genetic evaluations in Canada include infectious and noninfectious lesion traits, recorded according to the claw health atlas developed by the International Committee for Animal Recording (Egger-Danner et al., 2015). Infectious traits mostly refer to skin injuries, such as digital dermatitis (**DD**), interdigital dermatitis (**ID**), and heel horn erosion (**HHE**), whereas noninfectious traits refer to claw horn lesions, including sole ulcer (**SU**), toe ulcer (**TU**), white line lesion (**WL**), sole hemorrhage (**SH**), and interdigital hyperplasia (**IH**). As these traits have low heritability estimated in Canadian Holsteins (Chapinal et al., 2013; Malchiodi et al., 2017), analysis of association involving in silico–identified CNV and a large number of phenotypes provides an initial basis for a better comprehension of genetic mechanisms behind hoof health traits.

This study aimed to identify CNV with high confidence in a large sample of genotyped Holstein animals and to assess the association between the identified CNV and hoof health traits, followed by functional annotation of the associated CNV regions.

MATERIALS AND METHODS

Animal Genotypes and CNV Identification

Genotypic information of 10,682 Holstein animals genotyped with different platforms (Table 1), along with Log R ratio (**LRR**) and B allele frequency (**BAF**) information for every SNP, were used for CNV identification. The samples comprised 70 animals genotyped with the BovineHD Beadchip (**HD**; Illumina Inc.), 587 with the Genome Profiler Bovine 150K (Neogen Corp.), 807 with Genome Profiler Bovine HD (Neogen Corp.), 9,035 with BovineSNP50 (**50K**; Illumina Inc.), and 183 with Genome Profiler Bovine 50K (Neogen Corp.). The SNP positions were updated from the bovine reference genome assembly UMD3.1 (Zimin et al., 2009) to ARS-UCD1.2 (Rosen et al., 2018) using the information made available on the National Animal Genome Research Program (NAGRP) data repository (https://www.animalgenome.org/repository/cattle/UMC_bovine_coordinates/). Nonautosomal SNP and SNP with a GenCall score below 0.15 were removed on a per-sample basis. After these edits, the average number of markers was 680,557; 136,968; 76,009; 46,683; and 46,909 for the HD, Genome Profiler Bovine150K, Genome Profiler HD, 50K, and Genome Profiler Bovine 50K panels, respectively (Table 1).

The CNV identification was performed per panel using the PennCNV software (version 1.0.3; Wang et al., 2007), which integrates LRR and BAF on a per-sample basis into a hidden Markov model to determine the number of copies and genotypes of each CNV. The LRR values were corrected to the guanine-cytosine content at 500Kb upstream and downstream of each SNP based on a regression model (Diskin et al., 2008) to reduce waviness due to the correlation between LRR and guanine-cytosine contents of the genomic regions. After CNV calling, a sample-based quality control was performed to filter out possible false-positive CNV using ParseCNV software package (release 20, Glessner et al., 2013). Samples that were filtered out had a low genotype call rate (<0.97), a high intensity noise (LRR SD >0.3), extreme intensity waviness (absolute waviness factor >0.05, after LRR correction), BAF drift <0.01, more than 9 CNV identified per animal, or shared more than 50% of their genotypes with another

Table 1. Number of markers before and after quality control (QC) and number of samples for each genotype array used in the study

	Number of markers		Number of samples	
Panel[1]	Before QC	After QC[2]	Genotyped (cows; bulls)	Association analyses[3]
HD	777,962	680,557 (43,730)	70 (12; 58)	5
GGP-150K	138,892	136,968 (4,692)	587 (497; 90)	19
GGP-HD	76,883	76,009 (1,282)	807 (653; 154)	35
50K	54,001	46,683 (1,322)	9,035 (4,007; 5,028)	1,827
GGP-50K	49,463	46,909 (307)	183 (177; 9)	3

[1]Genotyping platforms: HD = Illumina BovineHD Beadchip (Illumina Inc.); GGP-150K = Neogen Corp. Genome Profiler Bovine 150K; GGP-HD = Neogen Corp. Genome Profiler Bovine HD, 50k = Illumina BovineSNP50 Beadchip; GGP-50k = Neogen Corp. Genome Profiler Bovine 50K.

[2]Average number of SNP per array after QC. Standard deviation in parentheses.

[3]Number of genotyped bulls presenting great genomic EBV reliability, then kept for association analyses.

animal. Nine CNV per animal was chosen as the upper limit as it represented the mean number of CNV found per animal plus 3 standard deviations. In addition, the minimum number of SNP covered by a CNV was set to 10 for samples genotyped with HD panel, and 3 for all other panels. Finally, 5,845 samples and 23,256 CNV (out of 56,561 detected with PennCNV) remained for further analyses.

Phenotypes

Genomic estimated breeding values (**GEBV**) and heritability estimates for 8 hoof health traits including DD, ID, HHE, SU, TU, WL, SH, and IH were retrieved from the April 2019 routine genetic evaluation performed by the Canadian Dairy Network (Table 2). Phenotypes used for GEBV estimation comprised 345,436 observations of each trait, recorded by 54 trimmers on 206,417 cows from 1,312 herds. The following animal model was fitted to estimate the GEBV for each trait:

$$Y = HD + P + T + S + a + pe + e, \qquad [1]$$

where Y was 0 or 1 in the absence or presence of each lesion, and HD, P, T, and S were the fixed effects of herd by date of trimming, parity, trimmer, and stage of lactation at trimming, respectively. The random effects were the animal additive effect a, the permanent environmental effect pe, and the residual effect e.

The GEBV were deregressed following the method presented in VanRaden et al. (2009). The deregressed GEBV (**dGEBV**) were used as the pseudophenotype for the association analyses. The dGEBV were computed for 1,889 bulls that all had phenotyped daughters, whose GEBV had a reliability above 0.25, and for which CNV could be detected and were thus used for association analyses. The average and range values of the dGEBV are presented in Table 2.

Association Analyses

The software program ParseCNV was used to identify associations between the CNV identified and dGEBV of 1,889 Holstein bulls. ParseCNV converts the CNV calls into probe-based genotypes. In other words, it separates the markers depending on their CNV genotype (deletions or duplications), correcting at the same time for family structure based on the parents of each sample. These probe statistics, independent for deletions or duplications, were then used for an association

Table 2. Heritability estimates (and SD) as published by the Canadian Dairy Network used for deregression of the hoof health EBV and descriptive statistics of the deregressed EBV (dEBV)

		dEBV		
Trait	Heritability (SD)	Mean	Min.	Max.
Digital dermatitis	0.08 (0.004)	0.27	−0.02	0.86
Interdigital dermatitis	0.05 (0.003)	0.16	−0.04	0.62
Heel horn erosion	0.08 (0.005)	0.26	−0.67	1.14
Sole ulcer	0.05 (0.003)	0.41	−1.00	1.00
Toe ulcer	0.04 (0.003)	−0.01	−0.62	1.00
White line lesion	0.04 (0.003)	−0.06	−0.62	0.75
Sole hemorrhage	0.03 (0.003)	0.61	−0.33	1.25
Interdigital hyperplasia	0.07 (0.004)	0.03	−0.73	0.45

analysis as implemented in Plink (version 1.07; Purcell et al., 2007). Correction for population structure was also carried out at this stage using the *covar* option and multidimensional scaling between the samples.

The model used for association testing was:

$$y = Xb + e, \quad [2]$$

where y was a vector containing the dEBV, X was the design matrix of the fixed effect of 1 CNV genotype at a time, b was the CNV effect, and e was the vector of random residual effects. The output of the association tests was used to merge neighboring SNP (less than 1Mb apart) reaching a similar significance level to CNV regions (**CNVR**). This method to create CNVR was shown to be flexible, and thus appropriate to define the breakpoints of the significantly associated regions (Glessner et al., 2013). Significance of each CNVR were computed with a Wald test based on the regression coefficients and the standard errors of each single CNVR. To account for multiple testing, a stringent criterion was required ($P < 0.0005$) to consider a CNVR significantly associated with the studied trait, as suggested by the ParseCNV developers (Glessner et al., 2013).

To reinforce the control of false-positive results, only significantly ($P < 0.0005$) associated regions that had overlap with CNVR previously identified with WGS information of 80 Holstein bulls (Butty et al., 2020) were kept to functional annotation.

Description of Associated Regions

Peptide sequences of the associated regions were retrieved from the Ensembl Gene database (release 99, Cunningham et al., 2019) with the Ensembl Biomart tool (Kinsella et al., 2011). The OmicsBox (version 1.1.0, Götz et al., 2008) was used to annotate the significantly associated regions. The gene ontology (**GO**) analyses were performed by taking the 3 GO categories (biological processes, molecular function, and cellular component) into account and using OmicsBox (Götz et al., 2008). Coding sequences were annotated with blastx and the OmicsBox mapping and GO annotation routines as in Conesa et al., (2005). Query sequences were compared against all the sequences found in the database of the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov, last accessed May 31, 2019). A significance level of at least 0.001 (e-value) and similarity of at least 70% were needed to consider a reported match for further analysis. The GO significance levels were computed following Fisher's exact test for multiple testing in OmicsBox. As described by Cánovas et al. (2013) and Li et al. (2016), the OmicsBox suite was used to examine associations between the sequences and biological pathways from the Kyoto Encyclopedia of Genes and Genomes (**KEGG**; Ogata et al., 1999). Information about the genes annotated in the significantly associated sequences was retrieved from GeneCards (Safran et al., 2010).

RESULTS

The CNV Identification

On average, 4 CNV (min: 1, max: 9) were identified per sample on the 5,845 samples remaining after quality control on a per sample basis. Of the 23,256 CNV included in the association analysis, 13,724 were deletions and 9,532 were duplications. The length of the CNV was not parametrically distributed ($P < 0.05$; Shapiro-Wilk normality test) and ranged from 76bp to 4.17Mb with an average length of 168.52Kb. The distribution of the length of the CNV showed no statistical differences ($P > 0.05$, Wilcoxon rank-sum test with continuity correction) between the cattle chromosomes. The CNV were found on all autosomes with a maximum of 2,775 CNV on BTA12 and a minimum of 106 CNV on BTA24. All CNV had a frequency below 2% in our sample set. Merging CNV with at least 1bp overlap to nonoverlapping CNVR reduced the number of variants to 1,645. Accounting for redundancy of variants over the genome, 9.43% of the total bovine genome was found to be within a CNVR (Figure 1).

The CNVR Associated With Hoof Health Traits

Association analyses between hoof health traits and the discovered CNV of 1,889 bulls led to the identification of 23 CNVR significantly associated with at least 1 of the evaluated traits ($P < 0.0005$; Supplemental Table S1, https://data.mendeley.com/datasets/xwmnrgn97f/1; Butty, 2021). Fourteen of these regions overlapped with CNVR previously identified with the WGS information of 80 Holstein bulls (Butty et al., 2020), corresponding to the most likely true-positive results. The 14 regions represented 9 deletions and 5 duplications, distributed on 13 chromosomes (Table 3, Figure 2), and had an average length of 104Kb (ranging from 9.8–343.3Kb). The number of samples on which the associated CNVR were detected ranged between 1 and 807. Detection of CNV can be biased by the quality of the reference genome assembly. Although ARS-UCD1.2 is of high quality, some gaps are still present that could not be sequenced; therefore, CNV detected in those gaps cannot be controlled for their veracity. None of the 14 CNVR associated with hoof health traits were found

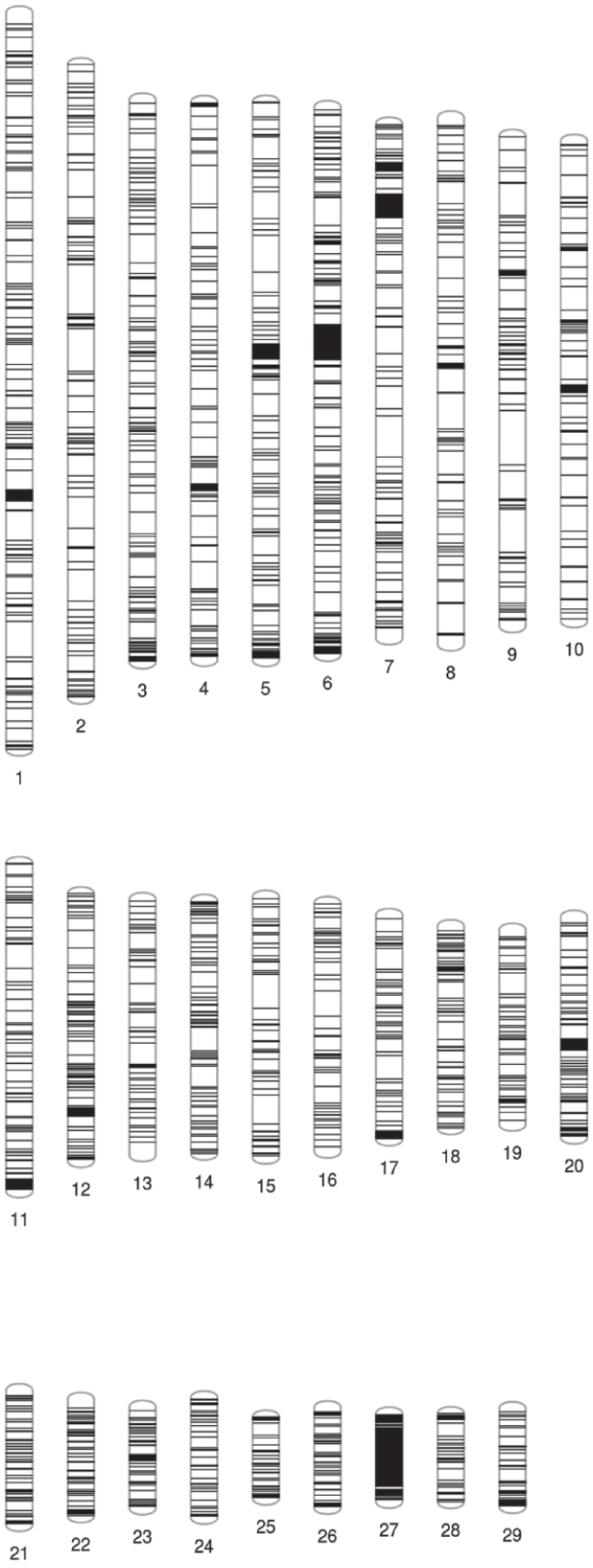

Figure 1. Distribution of the copy number variant regions identified on 5,845 samples over the bovine autosomes (black stripes).

in any gap of the reference assembly ARS-UCD1.2. All of the traits showed association with at least 1 CNVR, except TU. Only 2 CNVR were associated with more than 1 trait: CNVR3 on BTA9 associated with ID and IH, and CNVR14 on BTA29 associated with DD and SU.

As the CNV and the resulting CNVR were based on genotype array information, the associated regions' breakpoints had to be at a SNP position. Use of the WGS information of 80 Holstein bulls (Butty et al., 2020) to filter the associated CNVR showed that the true breakpoint of the associated CNVR is probably a few bases next to the breakpoints given by the array information. Figure 3 depicts the read depth over CNVR7 on BTA16 in 3 sequenced samples; the 2 top samples represent a deletion that starts before and ends after the region as described with genotype array (red bar). Moreover, the red-colored reads observed in the sequences of the CNV carriers mark reads that were split at the time of alignment, a further hint on the presence of a CNV in this region. The bottom sample in Figure 3 has no deletion in CNVR7. Definition of the region breakpoint could therefore be more precise with additional sequencing of a selection of animal carriers of deletions or duplications at each of the significantly associated CNVR.

The 14 regions significantly associated with hoof health traits comprised 54 Ensembl peptide sequences. Performing analyses using the OmicsBox mapping and annotation routines, 43 sequences were found to have basic local alignment search tool (BLAST) hits, and genes could be annotated for 11 associated CNVR (Table 3). We found CNVR9 on BTA20 contained the highest number of genes (6), whereas only 1 gene was found in the regions of CNVR1, CNVR2, CNVR3, CNVR4, CNVR13, and CNVR14.

Associated GO terms in the 3 main GO categories (biological processes, molecular functions, and cellular component) were identified. At the most informative level of the biological processes, 11% of the GO terms were related to biological regulation, 14% to metabolic, and 25% to cellular processes. The remaining 50% were distributed over multiple categories that never reached more than 4% of the terms. Regarding the molecular function terms, 51% were related to binding, 27% were related to catalytic activity, 16% were related to transporter activity, and 5% were related to receptor activity. Of the cellular component terms, 46% related to cell parts, 45% related to membrane parts, and 9% related to protein-containing complex. Enzyme codes were retrieved for 13 sequences and associated with 5 KEGG pathways. Among them, the folate biosynthesis pathway was associated with CNVR4, whereas CNVR10 was associated with purine metabolism; alanine, aspartame,

Table 3. Copy number variant regions (CNVR) significantly ($P < 0.0005$) associated with hoof health traits, their type (duplication, DUP, or deletion, DEL), and their gene content

CNVR	BTA	Start	End	Type	Trait[1]	Gene
CNVR1	7	10,422,889	10,432,630	DEL	WL	*OR7A17*
CNVR2	8	23,776,015	23,878,364	DEL	IH	*MLLT3*
CNVR3	9	44,794,304	44,864,222	DUP	ID, IH	*POPDC3*
CNVR4	10	78,557,712	78,830,390	DUP	SH	*GPHN*
CNVR5	12	86,121,984	86,338,161	DEL	SU	*ATP11A*, *TUBGCP3*, *SPACA7*
CNVR6	15	79,760,818	79,808,157	DEL	HHE	*OR5M11*, *OR5M3*
CNVR7	16	54,477,653	54,495,676	DEL	HHE	*PRAME8*
CNVR8	18	31,109,599	31,125,563	DUP	HHE	
CNVR9	20	70,834,509	71,177,834	DEL	SH	*LPCAT1*, *CLPTM1L*, *NDUSF6*, *TERT*, *SLC6A18*, *MRPL36*
CNVR10	21	68,617,018	68,743,664	DEL	DD	*ASPG*, *KIF26A*
CNVR11	23	25,984,486	26,166,446	DEL	WL	
CNVR12	26	25,491,013	25,509,679	DUP	HHE	*SORCS3*, *SCART1*
CNVR13	29	42,865,742	42,882,539	DUP	HHE	*NRXN2*
CNVR14	29	49,648,648	49,670,956	DEL	DD, SU	*SYT8*

[1]WL = white line lesion; IH = interdigital hyperplasia; ID = interdigital dermatitis; SH = sole hemorrhage; SU = sole ulcer; HHE = heel horn erosion; DD = digital dermatitis.

and glutamate metabolism; cyanoamino metabolism; and thiamine metabolism pathways.

DISCUSSION

In this study, 23,256 CNV were identified relying on the genotype array data of 5,845 Holstein individuals aligned to the bovine reference genome ARS-UCD1.2. Association analysis between the identified CNV and deregressed GEBV of 8 hoof health traits were performed for 1,889 Canadian bulls. The CNVR significantly associated with hoof health traits were analyzed for their gene content and putative functions related to the traits. The large number of samples included in this study for CNV identification, the use of updated SNP position to ARS-UCD1.2, and the discovery of associations between CNV and hoof health traits make this study novel on multiple levels. Moreover, conservative approaches were applied: (1) use of strict quality thresholds for CNV identification; (2) deregression of highly reliable GEBV only; and (3) removal of associated CNVR that were not overlapping with CNV identified on a set of partly similar samples but relying on WGS information. Therefore, the presented associated CNVR and genes are highly reliable candidates for their effects on hoof health traits.

Identified CNV

Although a high number of CNV were identified in this study, the average number of 4 CNV per sample can be compared with results presented in other studies relying on genotype array data (e.g., Butty et al., 2020). The density of the genotype array used is known to affect the number and length of the variants identified. The average length of the identified CNV (168.52Kb) was equivalent to the average distance between markers of the 50K panel after quality control (174.53Kb; 3 SNP were needed to consider a CNV valid). Eighty-five percent of the samples on which CNV identification relied were genotyped with the 50K marker panel. This showed that the higher number of samples genotyped with this panel truly influenced the final CNV set.

In line with the expectations, more deletions than duplications were detected in the Holstein population; current CNV detection methods relying on genotype array information often behave this way (Boussaha et al., 2015; Sasaki et al., 2016; Letaief et al., 2017; Mielczarek et al., 2017; Prinsen et al., 2017). In addition, the CNV were not distributed equally over the bovine autosomes due to CNV formation mechanisms such as nonallelic homologous recombination, fork stalling and template switching, nonhomologous end-joining, and mobile element insertion. These would take place more often in some genomic regions than others, in a similar way that recombination events occur more often in hotspots of the genome (Fadista et al., 2010; Bickhart and Liu, 2014). The genome coverage of the CNVR observed in this study (9.43%) is higher than reported by previous studies that show coverage values below 8% in the cattle genome (Fadista et al., 2010; Hou et al., 2011; Stothard et al., 2011; Boussaha et al., 2015; Letaief et al., 2017). This can be explained by the fact that variants in this study were identified on a higher number of samples in comparison with these studies. Also, our samples were mostly genotyped with medium-density marker panels. Due to the lower number of possible breakpoints than with higher density genotype array information, the CNV identified were longer and covered a greater part of the genome.

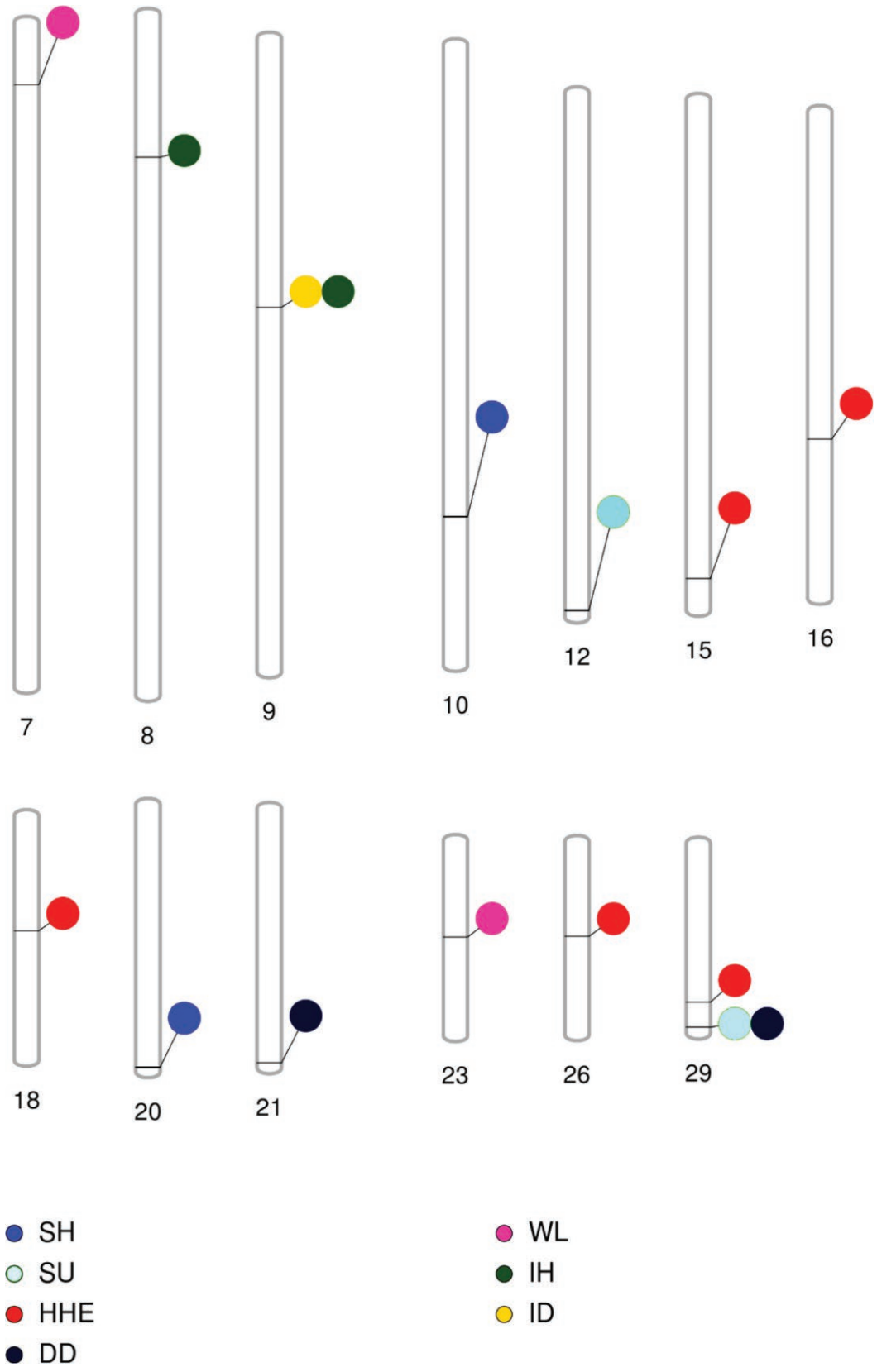

Figure 2. Copy number variant regions associated with hoof health traits: sole hemorrhage (SH), sole ulcer (SU), heel horn erosion (HHE), digital dermatitis (DD), white line lesion (WL), interdigital hyperplasia (IH), and interdigital dermatitis (ID).

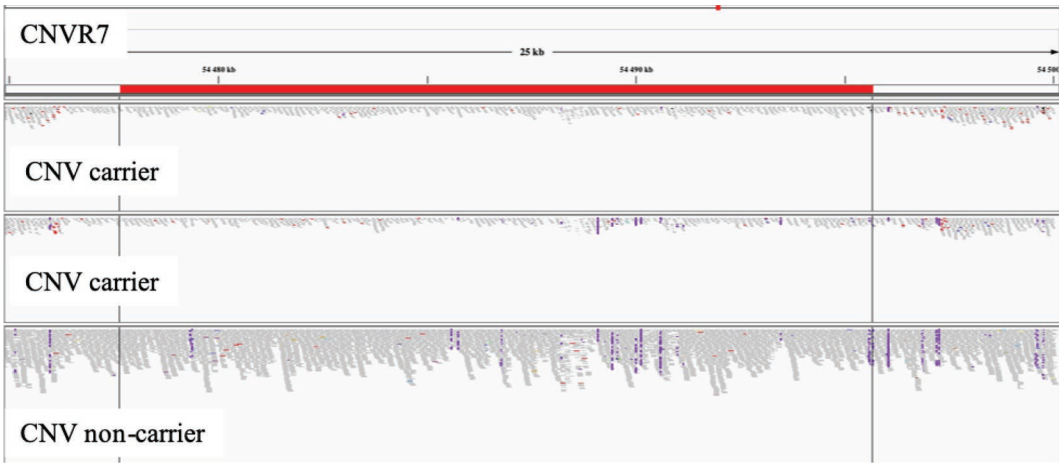

Figure 3. Read depth around the copy number variant region no. 7 (CNVR7; 16:54,474,882–54,500,285) for 3 sequenced samples. The red bar shows the breakpoints defined for CNVR7 with genotype array information. Deletions can be observed on the 2 top samples. Red-colored reads in the copy number variant (CNV) carrier sequences represent reads split at the time of alignment. No CNV is observed in the bottom sample.

Associations Between CNVR and Hoof Health Traits

Several CNVR were found associated with hoof health traits. Two regions were significantly associated with 2 traits, whereas all other associated CNVR were linked to a single trait. This discovery rate can be explained by the more stringent conditions adopted to consider a CNVR significantly associated in this analysis. First, the CNV had to cover at least 3 SNP. Second, they had to pass the PennCNV defaults and the ParseCNV-adjusted filter values. Third, only CNVR that overlapped with CNVR identified on WGS information were kept. Further analyses with less stringent conditions at the time of CNV identification or at the time of CNVR association would likely result in a higher number of associated regions. However, the risk for false positives would also be higher.

No region associated with hoof health traits in the current study coincided with genomic regions described in a GWAS on the same trait in the same Canadian Holstein cattle population but using SNP (Malchiodi et al., 2018). The lack of concordance between studies might be due the stringent criteria adopted in this study to identify CNVR and consider them significantly associated with the hoof health traits. However, such inconsistency might also reflect the potential of CNV to provide additional information not captured by SNP (Xu et al., 2014; Hay et al., 2018).

Gene Content and Putative Function of the Associated CNVR

Immune-related genes are expected to be related to infectious traits, whereas noninfectious traits are often related to metabolic or mechanic processes (Heringstad et al., 2018). Therefore, it was expected that different genomic regions would be associated with these 2 groups of traits. Interestingly, both CNVR that were associated with 2 hoof health traits included 1 trait of each category, indicating that common mechanisms affect infectious and noninfectious traits. However, the roles of the genes mapped within these 2 CNVR in resistance to hoof lesions are not fully understood. The *POPDC3* gene, located at CNVR3 that was associated with ID and IH, encodes a membrane protein associated with muscle dystrophy and serum level of creatine kinase (Vissing et al., 2019). The CNVR14 that was associated with DD and SU comprises the *SYT8* gene, a critical regulator of exocytosis and insulin secretion (Xu et al., 2011). The deletion of this gene might be associated with propensity to negative energy balance and a declining metabolism, which are presumed effects leading to increased susceptibility for hoof diseases (Collard et al., 2000; Heringstad et al., 2018).

Immune-related genes were identified in the CNVR associated with infectious lesions DD and HHE, including *SCART1*, *NRXN2,* and *KIF26A*. The *SCART1* gene encodes a protein only expressed in a specific type of delta gamma T-cell, acting in recognition of important pathogens (Baldwin et al., 2014). Thus, an effect of the number of *SCART1* copies on HHE (an infectious trait) can exist. The *NRXN2* was mapped to CNVR13, which is also associated with HHE, and it affects leukocyte adhesion deficiency type 3 (Safran et al., 2010). With a higher number of copies, this gene could lead to an increased ability of the leukocytes to act in the case of the presence of a pathogen in the organism. *KIF26A* is the gene in CNVR10 that can be related to the other

infectious trait DD. Kinesin proteins are part of the microtubules used to form vacuoles in the cells and affect their stability (Jancsik et al., 1996). The less solidified vacuoles could be more prone to fail their purpose of isolating pathogens in the cell, leading to a susceptibility to infections (Mostowy and Shenoy, 2015).

Metabolic diseases often lead to poor hoof quality, and thus higher incidence of lesions, as nutrients are not supplied to the dermal-epidermal junction between the live and the horn tissues of the hoof, which slowly degenerate and lead to a lack of support inside the hoof. This can be followed by the appearance of ulcers, hemorrhages, and white line diseases (Lischer and Ossent, 2007). The CNVR detected on BTA10 and BTA20 were significantly associated with SH. The *GPHN* gene on BTA10 was associated with the folate biosynthesis KEGG pathway. Changes in the folate metabolism lead to an increase of metabolites in the blood that may affect hoof quality (Lischer and Ossent, 2002). The genomic region of CNVR9 on BTA20 was previously identified in Holstein and associated with SCS (Durán Aguilar et al., 2017), suggesting influence of its gene content (*LPCAT1*, *NDUFS6*, *CLPTM1L*, *TERT*, *SLC6A18, MRPL36*) on the resistance to mastitis and metabolic disorders, 2 causes of economic losses in dairy industry. Interestingly, both CNVR1 and CNVR9 were also associated with ketosis traits (data not shown), a common metabolic disease of dairy cattle (Duffield, 2000). The only gene mapped in CNVR1 associated with WL was the olfactory receptor (**OR**) *OR7A17*. The other 2 OR genes were also detected in the CNVR5, which was associated with an infectious lesion, HHE. The OR represent a superfamily of genes crucial to the development of the sensory system in mammals that underwent high evolutionary selection pressure (Bickhart and Liu, 2014). Several studies report the expression of OR genes across nonolfactory issues, and unclear additional functions unrelated to olfaction, including the control of metabolic health and energy homeostasis, recognition of pathogens, and activation of immune cell responses (Li et al., 2013; Ferrer et al., 2016; Riera et al., 2017).

CONCLUSIONS

This study describes CNV associated with hoof health traits using pseudophenotypes and provides information on functional annotation of the associated CNVR. Fourteen CNVR were found to be significantly associated with infectious and noninfectious hoof lesions using strict quality control parameters and can therefore be considered high-confidence associations. Genes were mapped to the associated CNVR that had previously described functions related to the recorded hoof health traits in Canada. This study is a good foundation for the analysis of association between hoof health traits and in silico–identified CNV. Nevertheless, additional data will be needed to strengthen the analysis. Inclusion of the associated CNV identified in this study into national genetic evaluations could lead to greater genetic improvement rates in the Holstein dairy cattle population, thus reducing the involuntary animal losses due to lameness on farms.

ACKNOWLEDGMENTS

The Semex Alliance (Guelph, Canada), Genex (Shawano, WI), Alta Genetics (Balzac, Canada), Select Sires (Plain City, OH), Holstein Canada (Brantford, Canada), and Neogen Corporation (Lansing, MI) are acknowledged for providing the array genotype signal intensities. Phenotypic information was provided by the Canadian Dairy Network (Guelph, Canada). Computations were done at the server facilities provided by the Centre for Genetic Improvement of Livestock, Department of Animal Biosciences at the University of Guelph (Guelph, Canada). We gratefully acknowledge funding by the Efficient Dairy Genome Project, funded by Genome Canada (Ottawa, Canada), Genome Alberta (Calgary, Canada), Ontario Genomics (Toronto, Canada), Alberta Ministry of Agriculture (Edmonton, Canada), Ontario Ministry of Research and Innovation (Toronto, Canada), Ontario Ministry of Agriculture, Food and Rural Affairs (Guelph, Canada), Canadian Dairy Network (Guelph, Canada), GrowSafe Systems (Airdrie, Canada), Alberta Milk (Edmonton, Canada), Victoria Agriculture (Melbourne, Australia), Scotland's Rural College (Edinburgh, United Kingdom), USDA Agricultural Research Service (Beltsville, MD), Qualitas AG (Zug, Switzerland), Aarhus University (Aarhus, Denmark). AB is especially grateful to the Qualitas AG team and the Swiss Association for Animal Science (Zug, Switzerland) for their support. CB acknowledges financial support from Natural Sciences and Engineering Research Council (NSERC, Ottawa, Canada). The authors have not stated any conflicts of interest.

REFERENCES

Alkan, C., B. P. Coe, and E. E. Eichler. 2011. Genome structural variation discovery and genotyping. Nat. Rev. Genet. 12:363–376. https://doi.org/10.1038/nrg2958.

Baes, C. F., M. Dolezal, J. Koltes, B. Bapst, E. Fritz-Waters, S. Jansen, C. Flury, H. Signer-Hasler, C. Stricker, R. Fernando, R. Fries, J. Moll, D. Garrick, J. Reecy, and B. Gredler. 2014. Evaluation of variant identification methods for whole genome sequencing data in dairy cattle. BMC Genomics 15:948. https://doi.org/10.1186/1471-2164-15-948.

Baldwin, C. L., H. Hsu, C. Chen, M. Palmer, J. McGill, W. R. Waters, and J. C. Telfer. 2014. The role of bovine γδ T cells and their WC1 co-receptor in response to bacterial pathogens and promoting vac-

cine efficacy: A model for cattle and humans. Vet. Immunol. Immunopathol. 159:144–155. https://doi.org/10.1016/j.vetimm.2014.02.011.

Ben Sassi, N., Ó. González-Recio, R. de Paz-Del Río, S. T. Rodríguez-Ramilo, and A. I. Fernández. 2016. Associated effects of copy number variants on economically important traits in Spanish Holstein dairy cattle. J. Dairy Sci. 99:6371–6380. https://doi.org/10.3168/jds.2015-10487.

Bickhart, D. M., and G. E. Liu. 2014. The challenges and importance of structural variation detection in livestock. Front. Genet. 5. https://doi.org/10.3389/fgene.2014.00037.

Boussaha, M., D. Esquerré, J. Barbieri, A. Djari, A. Pinton, R. Letaief, G. Salin, F. Escudié, A. Roulet, S. Fritz, F. Samson, C. Grohs, M. Bernard, C. Klopp, D. Boichard, and D. Rocha. 2015. Genome-wide study of structural variants in bovine Holstein, Montbéliarde and Normande dairy breeds. PLoS One 10:e0135931. https://doi.org/10.1371/journal.pone.0135931.

Butty, A. 2021. 20200923_SuppTable1_Buttyetal_JDS. Mendeley Data V1. https://doi.org/10.17632/xwmnrgn97f.1.

Butty, A. M., T. Chud, F. Miglior, F. Schenkel, A. Kommadath, K. Krivushin, J. Grant, I. Häfliger, C. Drögemüller, A. Canovas, P. Stothard, and C. F. Baes. 2020. High confidence copy number variants identified in Holstein dairy cattle from whole genome sequence and genotype array data. Sci. Rep. 10:8044. https://doi.org/10.1038/s41598-020-64680-3.

Butty, A. M., M. Sargolzaei, F. Miglior, P. Stothard, F. Schenkel, B. Gredler-Grandl, and C. F. Baes. 2019. Optimizing selection of the reference population for genotype imputation from array to sequence variants. Front. Genet. 10:510. https://doi.org/10.3389/fgene.2019.00510.

Cánovas, A., G. Rincón, A. Islas-Trejo, R. Jimenez-Flores, A. Laubscher, and J. F. Medrano. 2013. RNA sequencing to study gene expression and single nucleotide polymorphism variation associated with citrate content in cow milk. J. Dairy Sci. 96:2637–2648. https://doi.org/10.3168/jds.2012-6213.

Chapinal, N., A. Koeck, A. Sewalem, D. F. Kelton, S. Mason, G. Cramer, and F. Miglior. 2013. Genetic parameters for hoof lesions and their relationship with feet and leg traits in Canadian Holstein cows. J. Dairy Sci. 96:2596–2604. https://doi.org/10.3168/jds.2012-6071.

Collard, B. L., P. J. Boettcher, J. C. M. Dekkers, D. Petitclerc, and L. R. Schaeffer. 2000. Relationships between energy balance and health traits of dairy cattle in early lactation. J. Dairy Sci. 83:2683–2690. https://doi.org/10.3168/jds.S0022-0302(00)75162-9.

Conesa, A., S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles. 2005. Blast2GO: A universal annotation and visualization tool in functional genomics research. Bioinformatics 21:3674–3676. https://doi.org/10.1093/bioinformatics/bti610.

Cunningham, F., P. Achuthan, W. Akanni, J. Allen, M. R. Amode, I. M. Armean, R. Bennett, J. Bhai, K. Billis, S. Boddu, C. Cummins, C. Davidson, K. J. Dodiya, A. Gall, C. G. Girón, L. Gil, T. Grego, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, M. Kay, M. R. Laird, I. Lavidas, Z. Liu, J. E. Loveland, J. C. Marugán, T. Maurel, A. C. McMahon, B. Moore, J. Morales, J. M. Mudge, M. Nuhn, D. Ogeh, A. Parker, A. Parton, M. Patricio, A. I. Abdul Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, H. Sparrow, E. Stapleton, M. Szuba, K. Taylor, G. Threadgold, A. Thormann, A. Vullo, B. Walts, A. Winterbottom, A. Zadissa, M. Chakiachvili, A. Frankish, S. E. Hunt, M. Kostadima, N. Langridge, F. J. Martin, M. Muffato, E. Perry, M. Ruffier, D. M. Staines, S. J. Trevanion, B. L. Aken, A. D. Yates, D. R. Zerbino, and P. Flicek. 2019. Ensembl 2019. Nucleic Acids Res. 47(D1):D745–D751. https://doi.org/10.1093/nar/gky1113.

Diskin, S. J., M. Li, C. Hou, S. Yang, J. Glessner, H. Hakonarson, M. Bucan, J. M. Maris, and K. Wang. 2008. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res. 36:e126. https://doi.org/10.1093/nar/gkn556.

Duffield, T. 2000. Subclinical ketosis in lactating dairy cattle. Vet. Clin. North Am. Food Anim. Pract. 16:231–253. https://doi.org/10.1016/S0749-0720(15)30103-1.

Durán Aguilar, M., S. I. Román Ponce, F. J. Ruiz López, E. González Padilla, C. G. Vásquez Peláez, A. Bagnato, and M. G. Strillacci. 2017. Genome-wide association study for milk somatic cell score in Holstein cattle using copy number variation as markers. J. Anim. Breed. Genet. 134:49–59. https://doi.org/10.1111/jbg.12238.

Egger-Danner, C., P. Nielsen, A. Fiedler, K. Müller, T. Fjeldaas, D. Döpfer, V. Daniel, C. Bergsten, G. Cramer, A. M. Christen, K. F. Stock, G. Thomas, M. Holzhauer, A. Steiner, J. Clarke, N. Capion, N. Charfeddine, J. E. Pryce, E. Oakes, J. Burgstaller, B. Heringstad, C. Ødegård, and J. Kofler. 2015. ICAR claw health atlas. ICAR Tech. Ser. 18:45.

Fadista, J., B. Thomsen, L. E. Holm, and C. Bendixen. 2010. Copy number variation in the bovine genome. BMC Genomics 11:284. https://doi.org/10.1186/1471-2164-11-284.

Ferrer, I., P. Garcia-Esparcia, M. Carmona, E. Carro, E. Aronica, G. G. Kovacs, A. Grison, and S. Gustincich. 2016. Olfactory receptors in non-chemosensory organs: The nervous system in health and disease. Front. Aging Neurosci. 8. https://doi.org/10.3389/fnagi.2016.00163.

Glessner, J. T., J. Li, and H. Hakonarson. 2013. ParseCNV integrative copy number variation association software with quality tracking. Nucleic Acids Res. 41:e64. https://doi.org/10.1093/nar/gks1346.

Glick, G., A. Shirak, E. Seroussi, Y. Zeron, E. Ezra, J. I. Weller, and M. Ron. 2011. Fine mapping of a QTL for fertility on BTA7 and its association with a CNV in the Israeli Holsteins. G3 (Bethesda) 1:65–74. https://doi.org/10.1534/g3.111.000299.

Goddard, M. E., K. E. Kemper, I. M. MacLeod, A. J. Chamberlain, and B. J. Hayes. 2016. Genetics of complex traits: Prediction of phenotype, identification of causal polymorphisms and genetic architecture. Proc. Biol. Sci. 283:20160569. https://doi.org/10.1098/rspb.2016.0569.

Götz, S., J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, M. Talón, J. Dopazo, and A. Conesa. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36:3420–3435. https://doi.org/10.1093/nar/gkn176.

Hay, E. H. A., Y. T. Utsunomiya, L. Xu, Y. Zhou, H. H. R. Neves, R. Carvalheiro, D. M. Bickhart, L. Ma, J. F. Garcia, and G. E. Liu. 2018. Genomic predictions combining SNP markers and copy number variations in Nellore cattle. BMC Genomics 19:441. https://doi.org/10.1186/s12864-018-4787-6.

Heringstad, B., C. Egger-Danner, N. Charfeddine, J. E. Pryce, K. F. Stock, J. Kofler, A. M. Sogstad, M. Holzhauer, A. Fiedler, K. Müller, P. Nielsen, G. Thomas, N. Gengler, G. de Jong, C. Ødegård, F. Malchiodi, F. Miglior, M. Alsaaod, and J. B. Cole. 2018. Invited review: Genetics and claw health: Opportunities to enhance claw health by genetic selection. J. Dairy Sci. 101:4801–4821. https://doi.org/10.3168/jds.2017-13531.

Hou, Y., G. E. Liu, D. M. Bickhart, M. F. Cardone, K. Wang, E-S. Kim, L. K. Matukumalli, M. Ventura, J. Song, P. M. VanRaden, T. S. Sonstegard, and C. P. Van Tassell. 2011. Genomic characteristics of cattle copy number variations. BMC Genomics 12:127. https://doi.org/10.1186/1471-2164-12-127.

Jancsik, V., D. Filliol, and A. Rendon. 1996. Tau proteins bind to kinesin and modulate its activation by microtubules. Neurobiology (Bp.) 4:417–429.

Kinsella, R. J., A. Kähäri, S. Haider, J. Zamora, G. Proctor, G. Spudich, J. Almeida-King, D. Staines, P. Derwent, A. Kerhornou, P. Kersey, and P. Flicek. 2011. Ensembl BioMarts: A hub for data retrieval across taxonomic space. Database (Oxford) 2011:bar030. https://doi.org/10.1093/database/bar030.

Letaief, R., E. Rebours, C. Grohs, C. Meersseman, S. Fritz, L. Trouilh, D. Esquerré, J. Barbieri, C. Klopp, R. Philippe, V. Blanquet, D. Boichard, D. Rocha, and M. Boussaha. 2017. Identification of copy number variation in French dairy and beef breeds using next-generation sequencing. Genet. Sel. Evol. 49:77. https://doi.org/10.1186/s12711-017-0352-z.

Li, J. J., H. L. Tay, M. Plank, A.-T. Essilfie, P. M. Hansbro, P. S. Foster, and M. Yang. 2013. Activation of olfactory receptors on mouse pulmonary macrophages promotes monocyte chemotactic protein-1 production. PLoS One 8:e80148. https://doi.org/10.1371/journal.pone.0080148.

Li, M., S. Riddle, H. Zhang, A. D'Alessandro, A. Flockton, N. J. Serkova, K. C. Hansen, R. Moldovan, B. A. McKeon, M. Frid, S. Kumar, H. Li, H. Liu, A. Caánovas, J. F. Medrano, M. G. Thomas, D. Iloska, L. Plecitá-Hlavatá, P. Ježek, S. Pullamsetti, M. A. Fini, K. C. El Kasmi, Q. H. Zhang, and K. R. Stenmark. 2016. Metabolic reprogramming regulates the proliferative and inflammatory phenotype of adventitial fibroblasts in pulmonary hypertension through the transcriptional corepressor c-terminal binding protein-1. Circulation 134:1105–1121. https://doi.org/10.1161/CIRCULATIONAHA.116.023171.

Lischer, C. J., and P. Ossent. 2002. Pathogenesis of sole lesions attributed to laminitis in cattle. Pages 82–89 in Proceedings of the 12th international symposium on lameness in ruminants. International Conference on Lameness in Ruminants.

Liu, M., L. Fang, S. Liu, M. G. Pan, E. Seroussi, J. B. Cole, L. Ma, H. Chen, and G. E. Liu. 2019. Array CGH-based detection of CNV regions and their potential association with reproduction and other economic traits in Holsteins. BMC Genomics 20:181. https://doi.org/10.1186/s12864-019-5552-1.

Malchiodi, F., L. F. Brito, F. S. Schenkel, A. M. Christen, D. F. Kelton, and F. Miglior. 2018. Genome-wide association study and functional analysis of infectious and horn type hoof lesions in Canadian Holstein cattle. Page 694 in Proceedings of the World Congress on Genetics Applied to Livestock Production. World Congress on Genetics Applied to Livestock Production.

Malchiodi, F., A. Koeck, S. Mason, A. M. Christen, D. F. Kelton, F. S. Schenkel, and F. Miglior. 2017. Genetic parameters for hoof health traits estimated with linear and threshold models using alternative cohorts. J. Dairy Sci. 100:2828–2836. https://doi.org/10.3168/jds.2016-11558.

Mielczarek, M., M. Frąszczak, R. Giannico, G. Minozzi, J. L. Williams, K. Wojdak-Maksymiec, and J. Szyda. 2017. Analysis of copy number variations in Holstein-Friesian cow genomes based on whole-genome sequence data. J. Dairy Sci. 100:5515–5525. https://doi.org/10.3168/jds.2016-11987.

Mostowy, S., and A. R. Shenoy. 2015. The cytoskeleton in cell-autonomous immunity: Structural determinants of host defence. Nat. Rev. Immunol. 15:559–573. https://doi.org/10.1038/nri3877.

Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 27:29–34. https://doi.org/10.1093/nar/27.1.29.

Pirooznia, M., F. Goes, and P. P. Zandi. 2015. Whole-genome CNV analysis: Advances in computational approaches. Front. Genet. 6. https://doi.org/10.3389/fgene.2015.00138.

Prinsen, R. T. M. M., A. Rossoni, B. Gredler, A. Bieber, A. Bagnato, and M. G. Strillacci. 2017. A genome wide association study between CNV and quantitative traits in Brown Swiss cattle. Livest. Sci. 202:7–12. https://doi.org/10.1016/j.livsci.2017.05.011.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575. https://doi.org/10.1086/519795.

Rafter, P., I. C. Gormley, A. C. Parnell, J. F. Kearney, and D. P. Berry. 2020. Concordance rate between copy number variants detected using either high- or medium-density single nucleotide polymorphism genotype panels and the potential of imputing copy number variants from flanking high density single nucleotide polymorphism haplotypes in cattle. BMC Genomics 21:205. https://doi.org/10.1186/s12864-020-6627-8.

Riera, C. E., E. Tsaousidou, J. Halloran, P. Follett, O. Hahn, M. M. A. Pereira, L. E. Ruud, J. Alber, K. Tharp, C. M. Anderson, H. Brönneke, B. Hampel, C. D. M. Filho, A. Stahl, J. C. Brüning, and A. Dillin. 2017. The sense of smell impacts metabolic health and obesity. Cell Metab. 26:198–211.e5. https://doi.org/10.1016/j.cmet.2017.06.015.

Rosen, B. D., D. M. Bickhart, R. D. Schnabel, S. Koren, C. G. Elsik, A. Zimin, C. Dreischer, S. Schultheiss, R. Hall, S. G. Schroeder, C. P. Van Tassell, T. P. L. Smith, and J. F. Medrano. 2018. Modernizing the bovine reference genome assembly. Page 802 in Proceedings of the World Congress on Genetics Applied to Livestock Production. World Congress on Genetics Applied to Livestock Production.

Safran, M., I. Dalah, J. Alexander, N. Rosen, T. Iny Stein, M. Shmoish, N. Nativ, I. Bahir, T. Doniger, H. Krug, A. Sirota-Madi, T. Olender, Y. Golan, G. Stelzer, A. Harel, and D. Lancet. 2010. GeneCards version 3: The human gene integrator. Database (Oxford) 2010:baq020. https://doi.org/10.1093/database/baq020.

Sasaki, S., T. Watanabe, S. Nishimura, and Y. Sugimoto. 2016. Genome-wide identification of copy number variation using high-density single-nucleotide polymorphism array in Japanese Black cattle. BMC Genet. 17:26. https://doi.org/10.1186/s12863-016-0335-z.

Spencer, C. C. A., Z. Su, P. Donnelly, and J. Marchini. 2009. Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. PLoS Genet. 5:e1000477. https://doi.org/10.1371/journal.pgen.1000477.

Stothard, P., J. W. Choi, U. Basu, J. M. Sumner-Thomson, Y. Meng, X. Liao, and S. S. Moore. 2011. Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. BMC Genomics 12:559. https://doi.org/10.1186/1471-2164-12-559.

Sudmant, P. H., T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov, J. Huddleston, Y. Zhang, K. Ye, G. Jun, M. Hsi-Yang Fritz, M. K. Konkel, A. Malhotra, A. M. Stütz, X. Shi, F. Paolo Casale, J. Chen, F. Hormozdiari, G. Dayama, K. Chen, M. Malig, M. J. P. Chaisson, K. Walter, S. Meiers, S. Kashin, E. Garrison, A. Auton, H. Y. K. Lam, X. Jasmine Mu, C. Alkan, D. Antaki, T. Bae, E. Cerveira, P. Chines, Z. Chong, L. Clarke, E. Dal, L. Ding, S. Emery, X. Fan, M. Gujral, F. Kahveci, J. M. Kidd, Y. Kong, E.-W. Lameijer, S. McCarthy, P. Flicek, R. A. Gibbs, G. Marth, C. E. Mason, A. Menelaou, D. M. Muzny, B. J. Nelson, A. Noor, N. F. Parrish, M. Pendleton, A. Quitadamo, B. Raeder, E. E. Schadt, M. Romanovitch, A. Schlattl, R. Sebra, A. A. Shabalin, A. Untergasser, J. A. Walker, M. Wang, F. Yu, C. Zhang, J. Zhang, X. Zheng-Bradley, W. Zhou, T. Zichner, J. Sebat, M. A. Batzer, S. A. McCarroll, R. E. Mills, M. B. Gerstein, A. Bashir, O. Stegle, S. E. Devine, C. Lee, E. E. Eichler, and J. O. Korbel. 2015. An integrated map of structural variation in 2,504 human genomes. Nature 526:75–81. https://doi.org/10.1038/nature15394.

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24. https://doi.org/10.3168/jds.2008-1514.

Vissing, J., K. Johnson, A. Töpf, S. Nafissi, J. Díaz-Manera, V. M. French, R. F. Schindler, P. Sarathchandra, N. Løkken, S. Rinné, M. Freund, N. Decher, T. Müller, M. Duno, T. Krag, T. Brand, and V. Straub. 2019. POPDC3 gene variants associate with a new form of limb girdle muscular dystrophy. Ann. Neurol. 86:832–843. https://doi.org/10.1002/ana.25620.

Wang, K., M. Li, D. Hadley, R. Liu, J. Glessner, S. F. A. Grant, H. Hakonarson, and M. Bucan. 2007. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 17:1665–1674. https://doi.org/10.1101/gr.6861907.

Winchester, L., C. Yau, and J. Ragoussis. 2009. Comparing CNV detection methods for SNP arrays. Brief. Funct. Genomic Proteomics 8:353–366. https://doi.org/10.1093/bfgp/elp017.

Xu, L., J. B. Cole, D. M. Bickhart, Y. Hou, J. Song, P. M. VanRaden, T. S. Sonstegard, C. P. Van Tassell, and G. E. Liu. 2014. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. BMC Genomics 15:683. https://doi.org/10.1186/1471-2164-15-683.

Xu, Z., G. Wei, I. Chepelev, K. Zhao, and G. Felsenfeld. 2011. Mapping of *INS* promoter interactions reveals its role in long-range regulation of *SYT8* transcription. Nat. Struct. Mol. Biol. 18:372–378. https://doi.org/10.1038/nsmb.1993.

Yang, T. L., R. H. Hao, Y. Guo, C. J. Papasian, and H. W. Deng. 2018. Copy Number Variation. Elsevier Inc.

Zhan, B., J. Fadista, B. Thomsen, J. Hedegaard, F. Panitz, and C. Bendixen. 2011. Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. BMC Genomics 12:557. https://doi.org/10.1186/1471-2164-12-557.

Zhou, Y., E. E. Connor, G. R. Wiggans, Y. Lu, R. J. Tempelman, S. G. Schroeder, H. Chen, and G. E. Liu. 2018. Genome-wide copy number variant analysis reveals variants associated with 10 diverse production traits in Holstein cattle. BMC Genomics 19:314. https://doi.org/10.1186/s12864-018-4699-5.

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.

ORCIDS

Adrien M. Butty https://orcid.org/0000-0003-2320-8405
Tatiane C. S. Chud https://orcid.org/0000-0001-7559-1165
Diercles F. Cardoso https://orcid.org/0000-0002-4618-6221
Filippo Miglior https://orcid.org/0000-0003-2345-8842
Flavio S. Schenkel https://orcid.org/0000-0001-8700-0633
Angela Cánovas https://orcid.org/0000-0002-0036-0757
Irene M. Häfliger https://orcid.org/0000-0002-5648-963X
Cord Drögemüller https://orcid.org/0000-0001-9773-522X
Paul Stothard https://orcid.org/0000-0003-4263-969X
Francesca Malchiodi https://orcid.org/0000-0002-1151-0845
Christine F. Baes https://orcid.org/0000-0001-6614-8890