J. Dairy Sci. TBC
https://doi.org/10.3168/jds.2023-24295

Genome-wide association analyses reveals copy number variant regions associated with reproduction and disease traits in Canadian Holstein cattle

Hinayah R. Oliveira,[1,2]* Tatiane C. S. Chud,[2] Gerson A. Oliveira Jr.,[2] Isis C. Hermisdorff,[2] Saranya G. Narayana,[2,5] Christina M. Rochus,[2] Adrien M. Butty,[3] Francesca Malchiodi,[2,4] Paul Stothard,[6] Filippo Miglior,[2,5] Christine F. Baes,[2,7] and Flavio S. Schenkel[2]*

[1]Department of Animal Sciences, Purdue University, West Lafayette, Indiana, USA;
[2]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada;
[3]Qualitas AG, Zug, Switzerland;
[4]Semex, Guelph, Ontario, Canada;
[5]Lactanet, Guelph, Ontario, Canada;
[6]University of Alberta, Edmonton, Alberta, Canada;
[7]Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland.

Abstract:

This study aimed to evaluate the impact of copy number variants (**CNVs**) on 13 reproduction and 12 disease traits in Holstein cattle. Intensity signal files containing Log R ratio and B allele frequency information from 13,730 Holstein animals genotyped with a 95K SNP panel, and 8,467 Holstein animals genotyped with a 50K SNP panel were used to identify the CNVs. Subsequently, the identified CNVs were validated using whole genome sequence data from 126 animals, resulting in 870 high-confidence CNV regions (**CNVRs**) on 12,131 animals. Out of these, 54 CNVRs had frequencies higher than or equal to 1% in the population and were used in the genome-wide association analysis (one CNVR at a time, including the **G** matrix). Results revealed that 4 CNVRs were significantly ($p\text{-value} < 3.7 \times 10^{-5}$) associated with at least one of the traits analyzed in this study. Specifically, 2 CNVRs were associated with 3 reproduction traits (i.e., calf survival, first service to conception, and non-return rate), and 2 CNVRs were associated with 2 disease traits (i.e., metritis and retained placenta). These CNVRs harbored genes implicated in immune response, cellular signaling, and neuronal development, supporting their potential involvement in these traits. Further investigations to unravel the mechanistic and functional implications of these CNVRs on the mentioned traits are warranted.

Keywords: dairy cattle, deletion, duplication, GWAS, structural variation

INTRODUCTION

Although the initial focus of most dairy cattle breeding programs has been to improve milk production, official genetic and genomic evaluations performed in several countries have already included other economically important traits (e.g., Fleming et al. 2018; Miglior et al., 2017). For instance, in Canada, reproduction traits have been evaluated since 2004, when genetic evaluations for 16 female reproduction traits were launched (Jamrozik et al., 2005). Moreover, various metabolic and reproductive diseases have been routinely evaluated since 2016 and 2020, respectively (Jamrozik et al., 2016a,b; Oliveira Jr. et al., 2021). Despite the high economic impact of reproduction and diseases in the dairy industry, achieving genetic progress for these traits is slower compared with production traits, because they usually have lower heritabilities (Oliveira Jr. et al., 2021). Therefore, understanding the genetic basis of variation in the expression of these traits across individuals can help to accelerate genetic progress.

Single nucleotide polymorphisms (**SNPs**) have been the main type of genetic variant used to identify quantitative trait loci (**QTLs**) and candidate genes associated with traits of interest in several livestock species, due to their wide distribution across the genome and relative low cost of genotyping (e.g., Weller et al., 2017; VanRaden 2020). However, genetic variation is not only identified by SNP markers. Recent studies involving human subjects have shown that structural variants (**SV**), mainly copy number variants (**CNVs**), also have an important impact on phenotypes of various traits, including reproduction and diseases (e.g., Sudmant et al., 2015; Beyter et al., 2021). Briefly, CNVs are variations in the number of DNA segments, usually deletions or duplications, which are more than 50 bp in length (Mills et al., 2011; Sudmant et al.,

Received October 10, 2023.
Accepted April 1, 2024.
*Correspondent authors: *hinayah@purdue.edu* and *schenkel@uoguelph.ca*

The list of standard abbreviations for JDS is available at adsa.org/jds-abbreviations-24. Nonstandard abbreviations are available in the Notes.

2015). As CNVs involve larger genomic regions compared with SNPs, they can likely affect gene expression through a variety of mechanisms (Levy et al., 2007). For instance, human studies have also shown that CNVs can affect gene structure and function by changing coding sequences and regulatory elements (Stranger et al. 2007; Zhang et al., 2009). In this context, Stranger et al. (2007) estimated that CNVs can account for up to ~18% of the genetic variation in gene expression of various human related-traits. Specifically in cattle, CNVs have already been found to be associated with feed intake, mastitis, and hoof health traits (Butty et al., 2021; Lee at al., 2021, 2023). The low linkage disequilibrium between SNPs and CNVs suggests that CNVs contain additional information not previously captured in traditional genome-wide association studies (**GWAS**) based solely on SNPs (e.g., Hay et al., 2018; Xu et al., 2014).

The main limitation of using CNVs in GWAS is the lack of consensus regarding CNV detection methods. In general, CNVs can be detected using genotypes obtained from either SNP panels or whole-genome sequence (**WGS**) data (Butty et al., 2020). Using WGS data enhances the precision of identifying CNV regions (**CNVRs**) compared with methods using SNP panels (Jiang et al., 2013), offering increased accuracy and the ability to detect shorter CNVs; nevertheless, the cost of WGS data for use in association studies can still be prohibitive. As an alternative, WGS data can be used to validate CNVRs identified using SNP arrays, to ascertain high-confidence CNVRs in a large number of animals (Butty et al., 2020; Rafter et al., 2020; Zhan et al., 2011). Consequently, this study aimed to: 1) identify CNVs using genotypes obtained from 2 different SNP arrays (i.e., 50K and 95K); 2) validate the CNVRs identified using WGS data; and 3) assess the association between the CNVs identified within the high-confidence CNVRs and several reproduction and disease traits in Canadian Holstein cattle.

MATERIALS AND METHODS

No Animal Care Committee approval was necessary for the purposes of this study, as all information required was obtained from existing databases.

SNP panels and quality control

Final reports containing the intensity signal files from 13,730 Holstein cows, genotyped using the 95K SNP panel (Illumina®, San Diego, CA), and 8,467 Holstein animals (4,726 bulls and 3,741 cows) genotyped using the 50K SNP panel (Illumina®, San Diego, CA), were available for this study. All SNP marker positions were updated to the ARS-UCD1.2 bovine reference genome assembly (Rosen et al., 2020). Non-autosomal SNPs, SNPs with unknow genome position, and SNPs with a GenCall score below 0.15 were removed on a per-sample basis during the genomic quality control. After the quality control, a total of 88,593 (95K) and 46,689 (50K) SNPs remained for further analyses.

CNV Identification. The CNV identification was performed independently for each SNP panel, using the PennCNV software (version 1.0.3; Wang et al., 2007) and the intensity signal files containing the log R ratio (**LRR**) and B allele frequency (**BAF**) information. To reduce the waviness due to the high correlation between LRR and the content of guanine-cytosine (**GC**) in the genomic regions, the genomic waves were adjusted using the *-gcmodel* option available in the PennCNV software. The cattle *gcmodel file* was generated by calculating the GC content of each marker. Thereafter, the LRR values of each SNP were adjusted for the genomic waves along the genomic regions, taking into account the expected GC content ± 500Kb around each SNP (Diskin et al., 2008).

After the CNV calling, a sample-based quality control was performed to remove possible false positive CNVs. Samples with high intensity noise (LRR SD >0.3), extreme intensity waviness (absolute waviness factor >0.05, after LRR correction), BAF drift <0.01, and more than 9 CNVs identified per animal were removed. The threshold criterion of 9 CNV per animal was chosen because it represents the mean number of CNVs found per animal (i.e., 4.73 between the 2 SNP panels) plus 3 standard deviations (SD = 1.46), which is similar to what was previously recommended by Butty et al. (2021). Moreover, only CNVs containing at least 3 SNPs were retained for further analysis.

The copy number of the CNV calls obtained from PennCNV are coded as 0 (deletion of 2 copies), 1 (deletion of one copy), 2 (normal state), 3 (single-copy duplication), and 4 (double-copy duplication). However, due to the lower frequencies of the codes 0 and 4 compared with the others, we have combined the deletion of one and 2 copies in the same group, as well as the duplication of one and 2 copies in another group. Thereafter, the states 0–1, 2, and 3–4 were re-coded using a biallelic format for the association analysis (please see details under the *"Association Analyses"* section). Finally, a total of 10,101 animals from the 95K SNP array genotype data and 5,353 animals from the 50K SNP array genotype data remained after quality control. This included 46,732 CNVs (26,398 deletions and 20,334 duplications) and 26,396 CNVs (15,630 deletions and 10,766 duplications) identified from the 95K and 50K SNP panels, respectively. Details on the CNV calls in each chromosome, for the 2 SNP panels, is provided in the Supplementary Table S1.

CNVRs Creation and Validation. The CNVRs were formed by collating overlapping and/or contiguous CNVs from both SNP panels (within and across animals) as originally described by Redon et al. (2006). The CNVRs that had more than one state for the same animal (i.e., both deletion and duplication in the same CNVR, for the same animal) were removed from the analyzed sample, as the probability to have 2 different events in the same region is low (Lin et al., 2013). Moreover, to remove the CNVs with very low frequency in the population (possible false positives), only continuous CNVRs and/or CNVRs that minimally overlapped (i.e., 1% overlap) between at least 2 animals were considered to be validated using the WGS data. Consequently, a total of 12,336 animals and 2,490 CNVRs remained in the data set to be validated.

The CNVRs originated from SNP genotypes were validated using CNVRs identified based on whole genome sequence data (**$CNVR_{WGS}$**), following the approach described by Butty et al. (2020). For instance, only CNVRs that reciprocally overlapped at least 50% with the $CNVR_{WGS}$ were kept for the association analysis (i.e., high-confidence CNVRs). The $CNVR_{WGS}$ were identified by collating overlapping and/or contiguous CNVs (Redon et al., 2006). The CNVs of the sequence data were identified using the CNVnator software (Abyzov, et al., 2011) and 2 independent data sets: one containing 23 Holsteins cows, and other containing 103 Holstein animals (88 bulls and 15 cows). Details about CNV detection using the sequence data are available in a previous study (Butty et al., 2020). In both data sets, paired-end reads were aligned to the ARS-UCD1.2 bovine genome assembly, following the protocol of the 1,000 Bull Genomes Project (http://www.1000bullgenomes.com/). Furthermore, only CNVs from regions with more than 50% of the reads mapped with a quality greater than zero were kept.

At the end, non-validated CNVRs were excluded from the data set, and in cases where none of the CNVRs of an animal were validated after aligning the CNVRs detected from the SNP genotypes with the $CNVR_{WGS}$, the animal was removed from the data set. These final quality control criteria led to a total of 870 high-confidence CNVRs, identified on 12,131 animals, which were available to be used in the association analyses. The CNVRs were classified as deletions when the animals showed a region with loss of a chromosomal segment, duplication for repeated chromosomal regions, and mixed, when it was identified deletions and duplications in the same genomic region.

Breeding Values and Deregressed Breeding Values. Estimated breeding values (and their theoretical reliabilities) for several reproduction and disease traits from the official genetic evaluation performed by Lactanet (Guelph, Ontario, Canada) in August 2021 were available for the association analyses. A total of 25 traits were analyzed in this study, which were categorized as reproduction: 1) age at first service (**AFS**), 2) calf survival [measured in heifers (**CSh**), cows (**CSc**), and bulls mated with either heifers (**CSsh**) or cows (**CSsc**)], 3) calving ease [measured in heifers (**CEh**) or cows (**CEc**)], 4) calving to first service (**CTFS**), 5) days open (**DO**), 6) first service to conception [measured in heifers (**FSTCh**) or cows (**FSTCc**)], 7) non return rate [measured in heifers (**NRRh**) or cows (**NRRc**)]; and disease: 8) cystic ovaries [measured in heifers (**COh**) or cows (**COc**)], 9) displaced abomasum [measured in heifers (**DAh**) or cows (**DAc**)], 10) clinical ketosis [measured in heifers (**CKh**) or cows (**CKc**)], 11) subclinical ketosis [measured in heifers (**SCKh**) or cows (**SCKc**)], 12) metritis [measured in heifers (**METh**) or cows (**METc**)], and 13) retained placenta [measured in heifers (**RPLh**) or cows (**RPLc**)] traits. A summary of the 25 traits evaluated in this study and their abbreviations are shown in Table 1.

Traits were defined as currently done by Lactanet Canada (Guelph, ON, Canada) in the official genetic and genomic evaluations for Holstein cattle. In this context, reproduction and disease traits were evaluated by Lactanet using multiple-trait animal models, and the traits measured in cows (i.e., CSc, CSsc, CEc, FSTCc, NRRc, COc, DAc, CKc, SCKc, METc, RPLc) were evaluated considering measurements from different lactations as repeated records. Additional details regarding the statistical models used by Lactanet can be found in the literature for reproduction (Jamrozik et al., 2005; Oliveira Jr. et al., 2021), and disease traits (Jamrozik et al., 2016; Oliveira Jr. et al., 2021). For all traits, EBVs were standardized to a mean of 100 and standard deviation of 5, as usually performed by Lactanet to facilitate the comparison of EBVs among traits.

The EBV of animals predicted during the August 2021 official genetic evaluation run by Lactanet were used to derive deregressed breeding values (**dEBVs**) for animals with CNVRs identified. The dEBV were calculated following the method presented in VanRaden et al. (2009), and subsequently used as pseudo-phenotypes in the association analyses. Only animals with theoretical reliability above 0.10 for the analyzed trait and effective record contributions higher than zero (i.e., animals whose EBVs were different from their parent average) were included in the association analyses. The heritability estimates used in the deregression procedure ranged from 0.003 (CSsc) to 0.15 (SCKh). The heritability used in the deregression process for each trait is shown in the Supplementary Table S2.

Association Analyses

The association analyses between the CNVs contained in the high-confidence CNVRs and the dEBVs were

performed using the blupf90+ software (Misztal et al., 2022). They were performed for one CNVR at a time, and the CNVs located within each CNVR were coded as −1 (deletion of either one or 2 copies), 0 (normal state), and 1 (duplication of either one or 2 copies). Only high-confidence CNVRs identified in at least 1% of animals in the population (i.e., 54 CNVRs) were included in the association analyses. Animals that did not have a CNV identified within the CNVR used in the association analysis (for example, due to differences in the SNP panel used), were excluded from the analysis. The final number of animals included in the association analysis for each trait, along with descriptive statistics of their dEBV, is shown in the Supplementary Table S2. The number of SNPs within each CNVR for each SNP panel is showed in the Supplementary Table S3.

The general model used for the association analysis of each trait is described as:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}b + \mathbf{Z}\mathbf{g} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of pseudo-phenotypes (i.e., dEBV), $\mathbf{1}$ is a vector of ones, μ is the average of the dEBVs (close to 100 for all traits; Supplementary Table S2), $\mathbf{X}$ is the incidence matrix for the fixed regression on the recoded CNV of one CNVR at a time, b is the linear regression coefficient (the CNV effect) in the analyzed CNVR, g is the vector of additive genomic random effects, $\mathbf{Z}$ is the incidence matrix for the vector $\boldsymbol{g}$, and $\mathbf{e}$ is the vector of residual effects. The model assumptions were:

$$\mathrm{E}[\mathbf{y}] = \mathbf{1}\mu + \mathbf{X}\mathrm{b} \text{ and } \quad \mathrm{Var}\begin{bmatrix} g \\ e \end{bmatrix} = \begin{bmatrix} \mathbf{G}\sigma_g^2 & 0 \\ 0 & \mathbf{R}\sigma_e^2 \end{bmatrix},$$

where $\mathbf{G}$ is the is genomic relationship matrix created as in the first method presented in VanRaden (2008), including all SNP markers that were not included in the CNVR under investigation (to avoid double-counting) and blended with the traditional pedigree-based additive relationship matrix $\mathbf{A}$ (10%) to make $\mathbf{G}$ invertible to account for the polygenic effects. The σ_g^2 is the additive genetic variance, $\mathbf{R}$ is a matrix of weights to account for the differences in dEBV reliabilities (i.e., 1/dEBV reliability −1), and σ_e^2 is the residual variance. The variance components were previously estimated by fitting the above-described model excluding the CNVR as a fixed effect in the model and using only the $\mathbf{A}$ matrix. The estimation of variance components was also performed using the blupf90+ software (Misztal et al., 2022). All other terms in the model were previously described.

The effect size and standard error (SE) estimated for each CNVR was used to calculate the t statistic (t = effect size/SE). Thereafter, p-values were obtained using a t-distribution with "n– 2" degrees of freedom, where n

Table 1. Summary of the traits evaluated in this study and their abbreviations

Category	Trait	Abbreviation	[1]Animal group (Final abbreviation)
Reproduction	**Age at first service**	AFS	Heifers (AFS)
	Calf survival	CS	Heifers (CSh)
			Cows (CSc)
			Bulls mated with heifers (CSsh)
			Bulls mated with cows (CSsc)
	Calving ease	CE	Heifers (CEh)
			Cows (CEc)
	Calving to first service	CTFS	Heifers (CTFS)
	Days open	DO	Cows (DO)
	First service to conception	FSTC	Heifers (FSTCh)
			Cows (FSTCc)
	Non return rate	NRR	Heifers (NRRh)
			Cows (NRRc)
Disease	**Cystic ovaries**	CO	Heifers (COh)
			Cows (COc)
	Displaced abomasum	DA	Heifers (DAh)
			Cows (DAc)
	Clinical ketosis	CK	Heifers (CKh)
			Cows (CKc)
	Subclinical ketosis	SCK	Heifers (SCKh)
			Cows (SCKc)
	Metritis	MET	Heifers (METh)
			Cows (METc)
	Retained placenta	RPL	Heifers (RPLh)
			Cows (RPLc)

[1]Animal group in which the phenotypic records were measured, as currently used in the official genetic evaluations performed by Lactanet (Guelph, Ontario, Canada).

is the number of animals used in the association analysis for each trait (Supplementary Table S2). To correct for multiple testing, a Bonferroni correction at a significance level of $\alpha = 0.05$ was applied by dividing α by the total number of tests (i.e., $0.05/(54x25) = 3.7 \times 10^{-5}$).

Functional Analyses. The CNVRs that were significant after the Bonferroni correction were further investigated. Positional candidate genes located within the associated regions were retrieved from the Ensembl Gene Database (Howe et al., 2021), using the Ensembl Biomart tool and the latest bovine reference genome assembly available (i.e., ARS-UCD1.2). Information about the quantitative trait loci (QTL) annotated in the significant CNVR was retrieved from the Animal QTL Database (Hu et al., 2013), and the complete list of gene functions were obtained from the National Center for Biotechnology Information database (NCBI, 2018). Biological pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG, Ogata et al., 1999) and the GO analyses (considering the main categories of biological processes, molecular function, and cellular component) were used to identify the most likely candidate genes. The NetworkAnalyst web server (Zhou, et al., 2019) was used for construction of gene networks, and the PhenoGram option available in the Ritchie Lab Visualization tool (http://visualization.ritchielab.org/phenograms/plot), from Perelman School of Medicine at the University of Pennsylvania, was used to visualize the position of the CNVRs in the chromosomes. Genome size was obtained from the NCBI (https://www.ncbi.nlm.nih.gov/assembly/GCF_002263795.1/#/st), using the ARS-UCD1.2 reference genome.

RESULTS AND DISCUSSION

CNV Identification

After quality control, the average number of CNVs per animal detected using both the 95K and 50K SNP panels was comparable (4.63 and 4.93, respectively; before CNV validation). The length of CNVs ranged from 66 bp to 6.19 Mb for the 95K SNP panel, with an average length of 112.93 Kb. For the 50K SNP panel, the length ranged from 76 bp to 4.53 Mb, with an average length of 154.54 Kb. The possible implications of these differences in our results, and recommendations for future studies are discussed in the *"Limitations of this Study and Future Directions"* sub-topic.

Interestingly, the number of CNVs was consistently higher on BTAs 8 and 12, regardless of the SNP panel used (see Supplementary Table S1). These findings, specially related to BTA12, corroborate with those in Lee et al. (2020). The length of the high-confidence CNVRs ranged from 5.11 Kb to 3.09 Mb, with an average length of 160.39 Kb, where approximately 5.61% of the entire bovine genome was encompassed by CNVRs. This finding aligns with a previous study performed by Mielczarek et al. (2017), which analyzed 29 Polish Holstein-Friesian cows and reported that CNVs covered an average of 5.89% of their genome. It is important to highlight that although there is a common trend in genome coverage by CNVs across these studies, variability in the precise extent to which CNVs impact the animal genome exists, a variation that can be observed among different individuals and breeds (Bickhart and Liu, 2014; Peripolli et al., 2017). Moreover, it is noteworthy that deletions tend to be more frequent than duplications in the overall genome (Fan et al., 2007), a pattern that was also observed in our study (i.e., 42,028 deletions versus 31,100 duplications).

CNVRs Creation and Validation

The analysis of high-confidence CNVRs reinforced the evidence that the majority of CNVRs exhibit low frequency within the Holstein population. Notably, a considerable proportion, approximately 37%, were identified in only one individual, highlighting their rarity within the population. This observation aligns with the understanding that CNVRs, akin to SNPs, undergo a process of evolutionary selection, where de novo mutations arise over thousands of generations, with only a fraction becoming fixed while the majority disappears. In addition some of the CNVRs identified are likely attributed to recent genetic events, including recent de novo mutations (Itsara et al., 2010; Belyeu et al., 2021). In this context, hypothesizing that CNVRs have emerged recently, it is possible that these CNVRs have not had sufficient time to become prevalent within the Holstein population.

While the majority of the identified CNVRs exhibit low frequencies in the population (most CNVRs appeared in less than 1% of the animals), it is worth noting that a subset of CNVRs, accounting for 6.2% of the total (i.e., 54 out of 870 CNVRs), displayed frequencies higher than or equal to 1%. These findings corroborate those reported by Sassi et al. (2016), who identified 90 CNVRs out of 823 segregating in more than 1% of the Spanish Holstein population. Moreover, these higher-frequency CNVRs, although still relatively scarce, offer potential for genetic diversity and could serve as valuable resources for breeding programs. However, the scarcity of high-frequency CNVRs further reinforces the prevailing notion that most of these genomic variants are infrequent in the Holstein population. For detailed information regarding the number and location of all high-confidence CNVRs identified in each autosome and their corresponding genome coverage, please refer to the Supplementary material (Table S4 and Figure S1)

Figure 1 provides a visual representation of the specific locations of high-confidence CNVRs with fre-

quency higher than or equal to 1% in the population. It is interesting to observe that previous studies have also consistently reported that BTA12 had the highest proportional CNV length compared with all bovine autosomes (Zhou et al., 2016; Durán Aguilar et al., 2017, Lee et al., 2020). Specifically in this study, the BTA12 not only exhibited the highest proportional CNV length, covering 12.04 Mb, but also contained the most frequently occurring CNVR in the population (i.e., this CNVR was present in 46.74% of the animals). Similarly, Lee et al. (2020) reported BTA12 as the autosome with the densest coverage of CNVRs in the bovine genome. Previous research has linked a region on this autosome to a significant QTL impacting reproduction and milk production in Nordic Red Cattle (Kadri et al., 2014). Please refer to the Supplementary material (Table S5) for the distribution and frequency of the identified CNVRs in this study.

Association Analyses

In our study, we examined only high-confidence CNVRs with frequency higher than or equal to 1% in the population (i.e., 54 CNVRs). Out of these, 8 CNVRs were tested solely based on the CNVRs identified in the 100K SNP panel: CNVRs located on BTA2 (123,701,258 to 123,887,590), BTA3 (37,275,225 to 37,702,081), BTA4 (115,319,018 to 115,367,660), BTA6 (51,856,578 to 52,320,441), BTA9 (1,657,206 to 1,813,012), BTA9 (91,813,809 to 91,996,013), BTA16 (79,509,462 to 79,953,581), and BTA18 (13,195,401 to 13,466,019), as they were not identified using the 50K SNP panel (Supplementary Table S3). A total of 4 unique CNVRs showed statistical significance for at least one of the analyzed traits. Specifically, 2 CNVRs (CNVR2 and CNVR3) were associated with 3 reproduction (CSh, FSTCc, NRRh), and 2 CNVRs (CNVR1 and CNVR4) were associated with 2 disease (METc, and RPLc) traits (Figure 2). No CNVRs were found to be significantly associated with AFS, CEc, CEh, CSc, CSsh, CSsc, CTFS, DO, FSTCh, and NRRc (reproduction traits), and COh, COc, DAh, DAc, CKh, CKc, METh, RPLh, SCKh, and SCKc (disease traits). Detailed information about the significant CNVRs for each trait, along with their frequency in the population, can be found in Table 2. To the best of our knowledge, this is the first time CNV-based GWAS has been performed for the specific traits analyzed in this study.

Significant CNVRs were detected on 2 autosomes (i.e., BTA7 and BTA21) for reproduction traits, and 2 autosomes (BTA5 and BTA23) for disease traits. From these, 2 CNVRs (i.e., CNVR1 and CNVR3) showed significant associations with more than one trait, suggesting the presence of pleiotropic effects. On the other hand, CNVR2 and CNVR4 had significance for a single trait (i.e., CSh and RPLc, respectively). Notably, from the 4 CNVRs found significant, 2 were found to be associated with RPLc in cows.

The shortest significant CNVRs identified covered 23 Kb (CNVR1 located on BTA5), for which we found significant association with RPLc and METc. This CNV is near the *CD163L1* gene (located within CNVR1), a gene found to be associated with inflammatory response (Moeller et al., 2012; Schieffer et al., 2021). The relationship between RPL and MET with the inflammatory response is complex, as these disorders can be influenced by multiple factors (Miyoshi et al., 2002; Dervishi et al., 2016). Metritis, an inflammation of the uterine lining commonly observed post-calving, shares overlapping pathophysiological mechanisms with RPL. Remarkably, MET and RPL show a high genetic correlation of 0.55 (Weller et al., 2019), illustrating their strong relationship. For instance, during the process of calving, there is an activation of immune cells and release of inflammatory mediators to facilitate the expulsion of the placenta. The prolonged inflammation can impair the normal physiological mechanisms involved in the separation and expulsion of the placenta, leading to its retention (Lye, 1996). The inflammatory environment in the uterus predisposes it to bacterial contamination, delaying uterine involution and increasing the risk of metritis (Chastant and Saint-Dizier, 2019; Amin and Hussein, 2022). Conversely, MET can also exacerbate the inflammation in the uterus, contributing to RPL in subsequent calvings (Amin and Hussein, 2022).

Notably, a specific CNVR located on BTA7 (CNVR2) was significantly associated with CSh, and it was observed in 36.28% of the studied animals. Within this region, there were 92 genes with uncertain functions and one olfactory receptor (*OR7A17*). Olfactory receptors exhibit a remarkable degree of diversity (Spehr and Munger, 2009) and serve essential functions in social interaction (Kotajima-Murakami et al., 2022) and reproduction (Spehr et al., 2006). Given the significance of maternal recognition for calf survival, it is plausible to hypothesize that olfactory receptor genes may contribute to this process. However, the precise role of *OR7A17* in reproductive disorders remains mainly unknown.

The CNVR located on BTA21 (i.e., CNVR3) exhibited significance for FSTCc and NRRh. This CNVR4, detected in 2.52% of the animals in the analyzed Holstein population, spans a genomic region (BTA21: 41,221,141 bp - 41,648,365 bp) that encompasses the *G2E3* gene. Previous studies have linked *G2E3* to the interval between calving and first insemination (Liu et al., 2017), suggesting its crucial involvement in reproductive processes. Notably, *G2E3* plays a pivotal role in regulating the cell cycle (Brooks et al., 2007), which holds particular importance during the estrous cycle in cows. Its impact

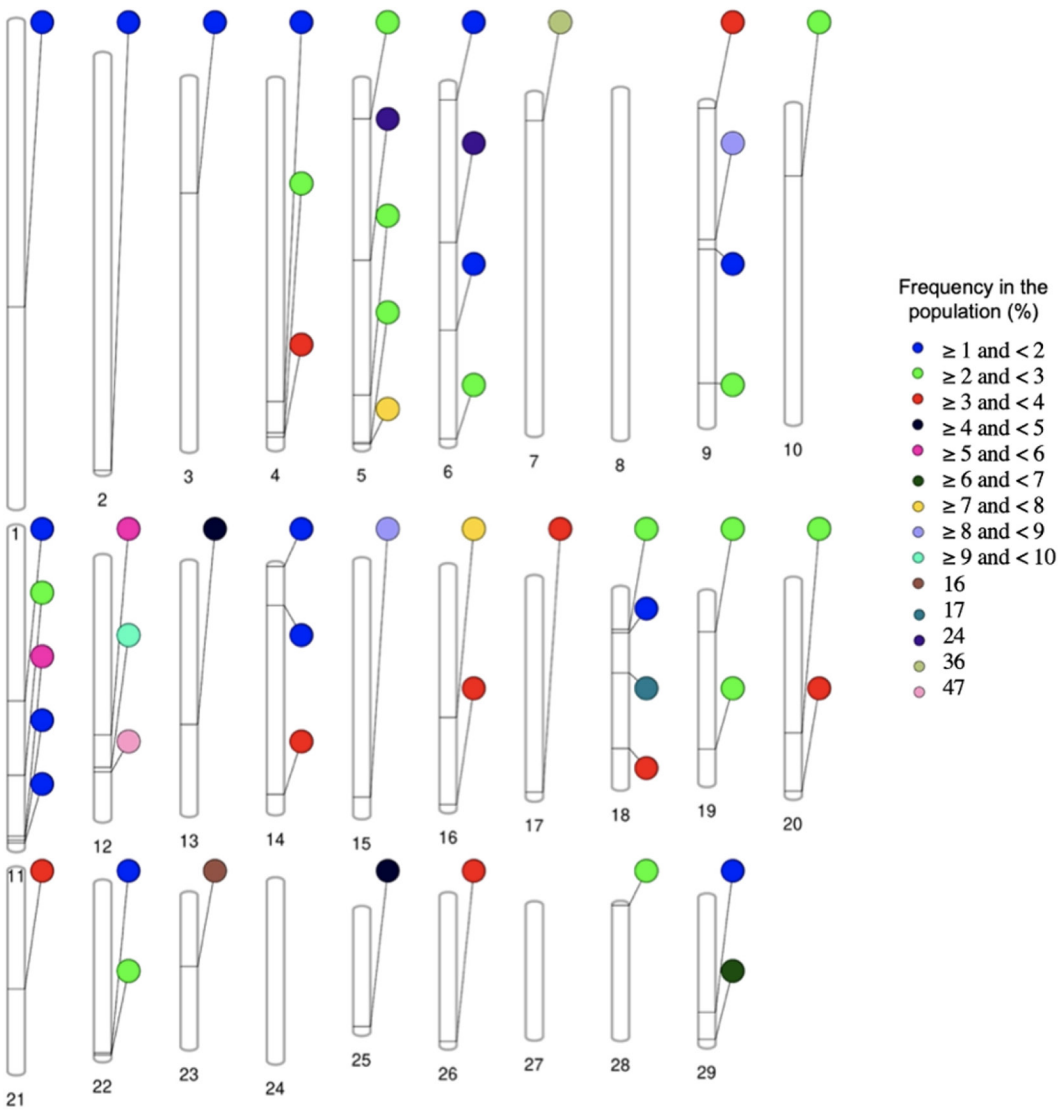

Figure 1. Distribution across the genome of the high-confidence copy number variant regions (CNVR) with frequency higher than or equal to 1% in the population. Colors represent the frequency.

on follicular development, ovulation, and corpus luteum function directly influences reproductive traits such as NRR and FSTC.

Our study also revealed that the CNVR located on BTA23 (CNVR4), present in 16.32% of the evaluated animals, was significantly associated with RPLc. This genomic region contains 44 genes, with 21 of them having unknown functions. Among these genes, we observed the presence of the *MCM3* gene, which may play a key role on RPL disorder in cows. This gene is known for its role in DNA replication and cell cycle regulation (Madine, et al., 1995). Therefore, we hypothesize that alterations in cellular processes or genetic factors related to DNA replication and cell cycle regulation can indirectly contribute to the development of reproductive disorders, including RPL. A protein–protein interaction (PPI) network (Zhou et al., 2019) was created and identified *MCM3* as a key gene for progesterone-mediated oocyte maturation (Figure 3). Imbalances or deficiencies in progesterone affect the normal maturation of oocytes and subsequently impact reproduction (López-Gatius and Garcia-Ispierto, 2022). Moreover, the decline in progesterone levels before uterine contraction during parturition is a pivotal process (Janszen et al., 1990; Mesiano, 2022), that may have implications for the proper release or detachment of the placenta after calving, thereby heightening the risk of RPL.

Table 2. High-confidence copy number variant regions (CNVR) significantly (*p-value* < $3.7x10^{-5}$) associated with the reproduction and disease traits analyzed in this study

CNVR	BTA	Start (bp)	End (bp)	Type[1]	Freq[2] (%)	Traits[3]
CNVR1	5	103,010,571	103,046,878	DUP	2.23	RPLc, METc
CNVR2	7	8,325,833	10,619,615	DEL	36.28	CSh
CNVR3	21	41,221,141	41,648,365	DUP	2.52	FSTCc, NRRh
CNVR4	23	24,697,169	26,863,188	DEL	16.32	RPLc

[1]Type: deletion (DEL) or duplication (DUP). [2]Freq (%): frequency of the CNVR in the population. [3]Traits: retained placenta measured in cows (RPLc), metritis measured in cows (METc), calf survival measured in heifers (CSh), first service to conception measured in cows (FSTCc), and non-return rate measured in heifers (NRRh).

Limitations of this Study and Future Directions

The reliance on SNP panels used for CNV detection may lead to underrepresentation or incomplete characterization of CNVRs across the genome, particularly in regions poorly covered by SNP arrays. This limitation underscores the importance of incorporating technologies like WGS to achieve more comprehensive CNV detection. Indeed, if all animals had WGS data to create CNVRs, we would expect a higher resolution and accuracy in CNV identification, potentially uncovering additional CNVs not captured by the SNP arrays. Moreover, the use of different SNP panels may introduce variability in CNV calling and subsequent association analyses.

In this study, we performed GWAS using CNV data obtained from 2 different SNP panels: one containing 46,689 SNPs (50K) and another one containing 88,593 SNPs (95K). After the quality control of the CNV calls, the average number of CNVs per animal detected using both SNP panels was comparable, indicating robustness across different SNP densities. However, the possible discrepancy in SNP coverage between the 2 panels raised concerns regarding potential biases in CNV calling and subsequent analysis. To address this, we used WGS data to validate CNVRs identified based on the SNP genotypes. This comprehensive validation approach enhances the robustness and reliability of our CNVR data, minimizing potential biases introduced by differences in SNP panel density and strengthening the validity of our GWAS analysis. However, it is essential to acknowledge the possibility of false positives and false negatives in detections (Lepamets et al., 2022). Future studies should consider testing different minimum numbers of SNPs in a CNV for various SNP panel densities, instead of the fixed number of 3 SNPs used in this study. Additionally, incor-

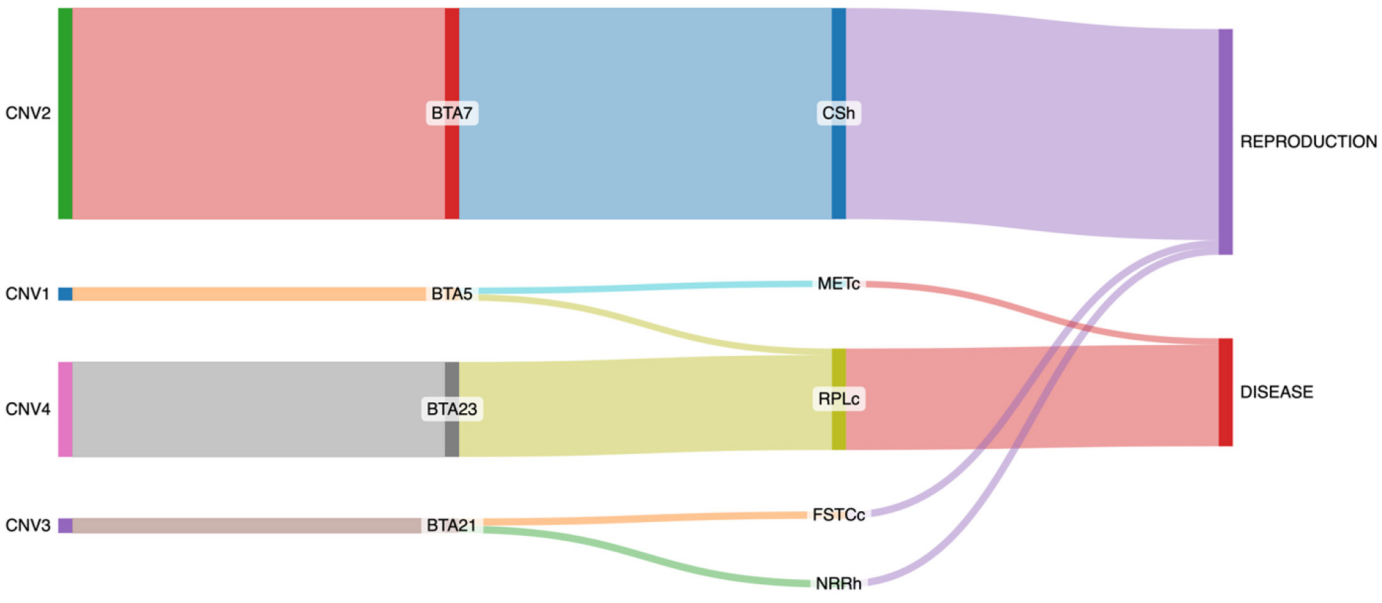

Figure 2. Sankey diagram illustrating the relationship between high-confidence copy number variant regions (CNVRs) with frequency higher than or equal to 1% in the population and the analyzed traits. The diagram displays the proportions of significantly associated CNVRs across chromosomes, organized by trait categories (i.e., reproduction and disease). Reproduction traits: calf survival measured in heifers (CSh); first service to conception measured in cows (FSTCc); and non-return rate measured in heifers (NRRh). Disease traits: retained placenta measured in cows (RPLc); and metritis measured in cows (METc).

porating multi-omics information could help improving power for CNV associations (Lepamets et al., 2022).

It is also paramount to acknowledge the wide range of methodological approaches and statistical models used for GWAS of CNVs. Notably, many studies in the literature have overlooked the incorporation of polygenic effects and potential adjustments for population stratification, which can introduce biases and hinder the identification of true associations. Recognizing this gap, it becomes imperative for future studies to integrate these aspects into their analytical frameworks to ensure robust and reliable findings. For instance, in a previous version of this study, we initially fit a statistical model incorporating the **A** matrix and the first few principal components of the **G** matrix in the association analysis, to avoid the possible double-counting of information generated by the fact that CNVs were defined using the SNP information. However, during the review process of this paper, we enhanced our methodology by integrating the **G** matrix derived from SNP information alongside CNVs in the model, but excluding the SNPs within the CNVR under evaluation. We believe this refinement helped us to avoid possible false positives and mitigate potential confounding effects. Regardless, incorporating polygenic effects and accounting for population stratification remain paramount in GWAS, and should also be considered in studies involving CNVs.

CONCLUSIONS

Our investigation on CNVs in Holstein cattle showed that, while high-confidence CNVRs were relatively uncommon in the population, a small subset exhibited higher frequencies (54 out of 870 CNVRs had frequencies higher than or equal to 1% in the population). Association analyses revealed that a total of 4 CNVRs were significantly associated with reproduction (CSh, FSTCc, NRRh) and/or disease (METc and RPLc) traits. These CNVRs harbored genes implicated in immune response, cellular signaling, and neuronal development, supporting their potential involvement in these traits. Further investigations to unravel the mechanistic and functional implications of these CNVRs on the mentioned traits are warranted.

ACKNOWLEDGMENTS

This research was financially supported by Agriculture and Agri-Food Canada, and by additional contributions from Dairy Farmers of Canada, the Canadian Dairy Network and the Canadian Dairy Commission under the Agri-Science Clusters Initiative. We also gratefully acknowledge the funding from the Resilient Dairy Genome Project supported by the organizations listed at http://www.resilientdairy.ca/funders-and-partners/, as administered by Genome Canada, Genome Alberta, Ontario Genomics, Genome Quebec, and Genome British Columbia. As per the research agreements, aside from providing financial support, the funders have no role in the design and conduct of the studies, data collection and analysis or interpretation of the data. Researchers maintain independence in conducting their studies, own their data, and report the outcomes regardless of the results. The decision to publish the findings rests solely with the researchers. The authors have not stated any conflicts of interest.

We also would like to acknowledge the two anonymous reviewers and the section editor for their valuable suggestions and comments.

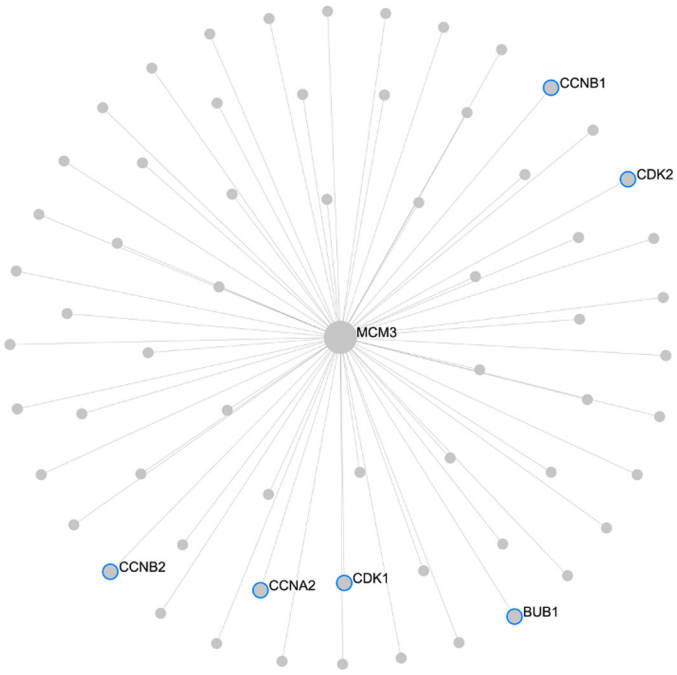

Figure 3. Protein-protein interaction network identified using the Network Analyst database. The network highlights the key gene, *MCM3*, and identifies hub genes, including *BUB1, CDK1, CCNA2, CCNB2, CDK1*, and *CCNB1*, involved in progesterone-mediated oocyte maturation.

REFERENCES

Abyzov, A., A. E. Urban, M. Snyder, and M. Gerstein. 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 21:974–984. https://doi.org/10.1101/gr.114876.110.

Amin, Y. A., and H. A. Hussein. 2022. Latest update on predictive indicators, risk factors and 'Omic' technologies research of retained placenta in dairy cattle–A review. Reprod. Domest. Anim. 57:687–700. https://doi.org/10.1111/rda.14115.

Belyeu, J. R., H. Brand, H. Wang, X. Zhao, B. S. Pedersen, J. Feusier, M. Gupta, T. J. Nicholas, J. Brown, L. Baird, B. Devlin, S. J. Sanders, L. B. Jorde, M. E. Talkowski, and A. R. Quinlan. 2021. De

novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families. Am. J. Hum. Genet. 108:597–607. https://doi.org/10.1016/j.ajhg.2021.02.012.

Beyter, D., H. Ingimundardottir, A. Oddsson, H. P. Eggertsson, E. Bjornsson, H. Jonsson, B. A. Atlason, S. Kristmundsdottir, S. Mehringer, M. T. Hardarson, S. A. Gudjonsson, D. N. Magnusdottir, A. Jonasdottir, A. Jonasdottir, R. P. Kristjansson, S. T. Sverrisson, G. Holley, G. Palsson, O. A. Stefansson, G. Eyjolfsson, I. Olafsson, O. Sigurdardottir, B. Torfason, G. Masson, A. Helgason, U. Thorsteinsdottir, H. Holm, D. F. Gudbjartsson, P. Sulem, O. T. Magnusson, B. V. Halldorsson, and K. Stefansson. 2021. Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. Nat. Genet. 53:779–786. https://doi.org/10.1038/s41588-021-00865-4.

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

Brooks, W. S., S. Banerjee, and D. F. Crawford. 2007. G2E3 is a nucleocytoplasmic shuttling protein with DNA damage responsive localization. Exp. Cell Res. 313:665–676. https://doi.org/10.1016/j.yexcr.2006.11.020.

Butty, A. M., T. C. Chud, F. Miglior, F. S. Schenkel, A. Kommadath, K. Krivushin, J. R. Grant, I. M. Häfliger, C. Drögemüller, A. Cánovas, 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., T. C. S. Chud, D. F. Cardoso, L. S. F. Lopes, F. Miglior, F. S. Schenkel, A. Cánovas, I. M. Häfliger, C. Drögemüller, P. Stothard, F. Malchiodi, and C. F. Baes. 2021. Genome-wide association study between copy number variants and hoof health traits in Holstein dairy cattle. J. Dairy Sci. 104:8050–8061. https://doi.org/10.3168/jds.2020-19879.

Chastant, S., and M. Saint-Dizier. 2019. Inflammation: friend or foe of bovine reproduction? Anim. Reprod. 16:539–547. https://doi.org/10.21451/1984-3143-AR2019-0057.

Dervishi, E., G. Zhang, D. Hailemariam, S. M. Dunn, and B. N. Ametaj. 2016. Occurrence of retained placenta is preceded by an inflammatory state and alterations of energy metabolism in transition dairy cows. J. Anim. Sci. Biotechnol. 7:26. https://doi.org/10.1186/s40104-016-0085-9.

Durán Aguilar, M., S. i. Román Ponce, F. Ruiz López, E. González Padilla, C. Vásquez Peláez, A. Bagnato, and M. 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.

Fan, Y., W. Wang, G. Ma, L. Liang, Q. Shi, and S. Tao. 2007. Patterns of Insertion and Deletion in Mamm. Genomes. Curr. Genomics 8:370–378. https://doi.org/10.2174/138920207783406479.

Fleming, A., E. A. Abdalla, C. Maltecca, and C. F. Baes. 2018. Invited review: Reproductive and genomic technologies to optimize breeding strategies for genetic progress in dairy cattle. Arch. Tierzucht 61:43–57. https://doi.org/10.5194/aab-61-43-2018.

Gao, Y., J. Jiang, S. Yang, Y. Hou, G. E. Liu, S. Zhang, Q. Zhang, and D. Sun. 2017. CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. BMC Genomics 18:265. https://doi.org/10.1186/s12864-017-3636-3.

Hay, E. H. A., Y. T. Utsunomiya, L. Xu, Y. Zhou, H. H. 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 19:1–8. https://doi.org/10.1186/s12864-018-4787-6.

Hellwege, J., J. Keaton, A. Giri, X. Gao, D.R. Velez Edwards, and T.L. Edwards. 2017. Population Stratification in Genetic Association Studies. Curr Protoc Hum Genet 95:1.22.1–1.22.23. doi:https://doi.org/10.1002/cphg.48.

Hu, Z. L., C. A. Park, X. L. Wu, and J. M. Reecy. 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Res. 41(D1):D871–D879. https://doi.org/10.1093/nar/gks1150.

Itsara, A., H. Wu, J. D. Smith, D. A. Nickerson, I. Romieu, S. J. London, and E. E. Eichler. 2010. De novo rates and selection of large copy number variation. Genome Res. 20:1469–1481. https://doi.org/10.1101/gr.107680.110.

Jamrozik, J., J. Fatehi, G. Kistemaker, and L. R. Schaeffer. 2005. Estimates of genetic parameters for Canadian Holstein female reproduction traits. J. Dairy Sci. 88:2199–2208. https://doi.org/10.3168/jds.S0022-0302(05)72895-2.

Jamrozik, J., G. Kistemaker, B. Van Doormaal, A. Fleming, A. Koeck, and F. Miglior. 2016b. Genetic evaluation for resistance to metabolic diseases in Canadian dairy breeds. Interbull Bull. •••:50.

Jamrozik, J., A. Koeck, G. Kistemaker, and F. Miglior. 2016a. Multiple-trait estimates of genetic parameters for metabolic disease traits, fertility disorders, and their predictors in Canadian Holsteins. J. Dairy Sci. 99:1990–1998. https://doi.org/10.3168/jds.2015-10505.

Janszen, B. P., M. M. Bevers, S. J. Dieleman, G. C. van der Weyden, and M. A. Taverne. 1990. Synchronised calvings after withdrawal of norgestomet implants from cows treated near term with prostaglandin. Vet. Rec. 127:405–407.

Jiang, L., J. Jiang, J. Yang, X. Liu, J. Wang, H. Wang, X. Ding, J. Liu, and Q. Zhang. 2013. Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. BMC Genomics 14:1–10. https://doi.org/10.1186/1471-2164-14-131.

Kadri, N. K., G. Sahana, C. Charlier, T. Iso-Touru, B. Guldbrandtsen, L. Karim, U. S. Nielsen, F. Panitz, G. P. Aamand, N. Schulman, M. Georges, J. Vilkki, M. S. Lund, and T. Druet. 2014. A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic Red Cattle: Additional Evidence for the Common Occurrence of Balancing Selection in Livestock. PLoS Genet. 10:e1004049. https://doi.org/10.1371/journal.pgen.1004049.

Kotajima-Murakami, H., H. Hagihara, A. Sato, Y. Hagino, M. Tanaka, Y. Katoh, Y. Nishito, Y. Takamatsu, S. Uchino, T. Miyakawa, and K. Ikeda. 2022. Exposure to GABAA Receptor Antagonist Picrotoxin in Pregnant Mice Causes Autism-Like Behaviors and Aberrant Gene Expression in Offspring. Front. Psychiatry 13:13. https://doi.org/10.3389/fpsyt.2022.821354.

Lee, Y. L., M. Bosse, E. Mullaart, M. A. Groenen, R. F. Veerkamp, and A. C. Bouwman. 2020. Functional and population genetic features of copy number variations in two dairy cattle populations. BMC Genomics 21:1–15. https://doi.org/10.1186/s12864-020-6496-1.

Lee, Y. L., M. Bosse, H. Takeda, G. C. M. Moreira, L. Karim, T. Druet, C. Oget-Ebrad, W. Coppieters, R. F. Veerkamp, M. A. Groenen, M. Georges, A. C. Bouwman, and C. Charlier. 2023. High-resolution structural variants catalogue in a large-scale whole genome sequenced bovine family cohort data. BMC Genomics 24:1–17. https://doi.org/10.1186/s12864-023-09259-8.

Lee, Y. L., H. Takeda, G. Costa Monteiro Moreira, L. Karim, E. Mullaart, W. Coppieters, R. Appeltant, R. F. Veerkamp, M. A. Groenen, M. Georges, M. Bosse, T. Druet, A. C. Bouwman, and C. Charlier. 2021. A 12 kb multi-allelic copy number variation encompassing a GC gene enhancer is associated with mastitis resistance in dairy cattle. PLoS Genet. 17:e1009331. https://doi.org/10.1371/journal.pgen.1009331.

Lepamets, M., C. Auwerx, M. Nõukas, A. Claringbould, E. Porcu, M. Kals, T. Jürgenson, A. P. Morris, U. Võsa, M. Bochud, and S. Stringhini. 2022. Omics-informed CNV calls reduce false-positive rates and improve power for CNV-trait associations. Human Genetics and Genomics Advances, 3(4).

Levy, S., G. Sutton, P. C. Ng, L. Feuk, A. L. Halpern, B. P. Walenz, N. Axelrod, J. Huang, E. F. Kirkness, G. Denisov, Y. Lin, J. R. MacDonald, A. W. C. Pang, M. Shago, T. B. Stockwell, A. Tsiamouri, V. Bafna, V. Bansal, S. A. Kravitz, D. A. Busam, K. Y. Beeson, T. C. McIntosh, K. A. Remington, J. F. Abril, J. Gill, J. Borman, Y.-H. Rogers, M. E. Frazier, S. W. Scherer, R. L. Strausberg, and J. C. Venter. 2007. The diploid genome sequence of an individual human. PLoS Biol. 5:254. https://doi.org/10.1371/journal.pbio.0050254.

Lin, C. F., A. C. Naj, and L. S. Wang. 2013. Analyzing copy number variation using SNP array data: protocols for calling CNV and association tests. Curr. Protoc. Hum. Genet. 79:1–27. https://doi.org/10.1002/0471142905.hg0127s79.

Liu, A., Y. Wang, G. Sahana, Q. Zhang, L. Liu, M. S. Lund, and G. Su. 2017. Genome-wide Association Studies for Female Fertility Traits in Chinese and Nordic Holsteins. Sci. Rep. 7:8487. https://doi.org/10.1038/s41598-017-09170-9.

López-Gatius, F., and I. Garcia-Ispierto. 2022. Clinical Overview of Luteal Deficiency in Dairy Cattle. Animals (Basel) 12:1871. https://doi.org/10.3390/ani12151871.

Lye, S. J. 1996. Initiation of parturition. Anim. Reprod. Sci. 42:495–503. https://doi.org/10.1016/0378-4320(96)01529-1.

Madine, M. A., C.-Y. Khoo, A. D. Mills, and R. A. Laskey. 1995. MCM3 complex required for cell cycle regulation of DNA replication in vertebrate cells. Nature 375:421–424. https://doi.org/10.1038/375421a0.

Mesiano, S. 2022. Progesterone withdrawal and parturition. J. Steroid Biochem. Mol. Biol. 224:106177. https://doi.org/10.1016/j.jsbmb.2022.106177.

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.

Miglior, F., A. Fleming, F. Malchiodi, L. F. Brito, P. Martin, and C. F. Baes. 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. J. Dairy Sci. 100:10251–10271. https://doi.org/10.3168/jds.2017-12968.

Mills, R. E., K. Walter, C. Stewart, R. E. Handsaker, K. Chen, C. Alkan, A. Abyzov, S. C. Yoon, K. Ye, R. K. Cheetham, A. Chinwalla, D. F. Conrad, Y. Fu, F. Grubert, I. Hajirasouliha, F. Hormozdiari, L. M. Iakoucheva, Z. Iqbal, S. Kang, J. M. Kidd, M. K. Konkel, J. Korn, E. Khurana, D. Kural, H. Y. K. Lam, J. Leng, R. Li, Y. Li, C.-Y. Lin, R. Luo, X. J. Mu, J. Nemesh, H. E. Peckham, T. Rausch, A. Scally, X. Shi, M. P. Stromberg, A. M. Stütz, A. E. Urban, J. A. Walker, J. Wu, Y. Zhang, Z. D. Zhang, M. A. Batzer, L. Ding, G. T. Marth, G. McVean, J. Sebat, M. Snyder, J. Wang, K. Ye, E. E. Eichler, M. B. Gerstein, M. E. Hurles, C. Lee, S. A. McCarroll, and J. O. Korbel. 2011. Mapping copy number variation by population-scale genome sequencing. Nature 470:59–65. https://doi.org/10.1038/nature09708.

Misztal, I., S. Tsuruta, D. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica. Manual for BLUPF90 family of programs. University of Georgia, 2022.

Miyoshi, M., Y. Sawamukai, and T. Iwanaga. 2002. Reduced Phagocytotic Activity of Macrophages in the Bovine Retained Placenta. Reprod. Domest. Anim. 37:53–56. https://doi.org/10.1046/j.1439-0531.2002.00332.x.

Moeller, J. B., M. J. Nielsen, M. P. Reichhardt, A. Schlosser, G. L. Sorensen, O. Nielsen, I. Tornøe, J. Grønlund, M. E. Nielsen, J. S. Jørgensen, O. N. Jensen, J. Mollenhauer, S. K. Moestrup, and U. Holmskov. 2012. CD163–L1 Is an Endocytic Macrophage Protein Strongly Regulated by Mediators in the Inflammatory Response. J. Immunol. 188:2399–2409. https://doi.org/10.4049/jimmunol.1103150.

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.

Oliveira, G. O. Jr., F. S. Schenkel, L. Alcantara, K. Houlahan, C. Lynch, and C. F. Baes. 2021. Estimated genetic parameters for all genetically evaluated traits in Canadian Holsteins. J. Dairy Sci. 104:9002–9015. https://doi.org/10.3168/jds.2021-20227.

Peripolli, E., D. P. Munari, M. V. G. B. Silva, A. L. F. Lima, R. Irgang, and F. Baldi. 2017. Runs of homozygosity: current knowledge and applications in livestock. Anim. Genet. 48:255–271. https://doi.org/10.1111/age.12526.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

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 21:1–10. https://doi.org/10.1186/s12864-020-6627-8.

Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, W. Chen, and E. K. Cho. 2006. Global variation in copy number in the human genome. nature, 444:444–454. doi: https://doi.org/10.1038/nature05329.

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. Couldrey, R. Hall, W. Li, A. Rhie, J. Ghurye, S. D. McKay, F. Thibaud-Nissen, J. Hoffman, B. M. Murdoch, W. M. Snelling, T. G. McDaneld, J. A. Hammond, J. C. Schwartz, W. Nandolo, D. E. Hagen, C. Dreischer, S. J. Schultheiss, S. G. Schroeder, A. M. Phillippy, 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.

Schieffer, K. M., S. M. Emrich, G. S. Yochum, and W. A. Koltun. 2021. CD163L1+CXCL10+ Macrophages are Enriched Within Colonic Lamina Propria of Diverticulitis Patients. J. Surg. Res. 267:527–535. https://doi.org/10.1016/j.jss.2021.06.016.

Spehr, M., and S. D. Munger. 2009. Olfactory receptors: G protein-coupled receptors and beyond. J. Neurochem. 109:1570–1583. https://doi.org/10.1111/j.1471-4159.2009.06085.x.

Spehr, M., K. Schwane, J. A. Riffell, R. K. Zimmer, and H. Hatt. 2006. Odorant receptors and olfactory-like signaling mechanisms in mammalian sperm. Mol. Cell. Endocrinol. 250:128–136. https://doi.org/10.1016/j.mce.2005.12.035.

Stranger, B. E., M. S. Forrest, M. Dunning, C. E. Ingle, C. Beazley, N. Thorne, R. Redon, C. P. Bird, A. Grassi, C. Lee, and C. Tyler-Smith. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315:848–853. https://doi.org/10.1126/science.1136678.

Sudmant, P. H., S. Mallick, B. J. Nelson, F. Hormozdiari, N. Krumm, J. Huddleston, B. P. Coe, C. Baker, S. Nordenfelt, M. Bamshad, L. B. Jorde, O. L. Posukh, H. Sahakyan, W. S. Watkins, L. Yepiskoposyan, M. S. Abdullah, C. M. Bravi, C. Capelli, T. Hervig, J. T. S. Wee, C. Tyler-Smith, G. van Driem, I. G. Romero, A. R. Jha, S. Karachanak-Yankova, D. Toncheva, D. Comas, B. Henn, T. Kivisild, A. Ruiz-Linares, A. Sajantila, E. Metspalu, J. Parik, R. Villems, E. B. Starikovskaya, G. Ayodo, C. M. Beall, A. Di Rienzo, M. F. Hammer, R. Khusainova, E. Khusnutdinova, W. Klitz, C. Winkler, D. Labuda, M. Metspalu, S. A. Tishkoff, S. Dryomov, R. Sukernik, N. Patterson, D. Reich, and E. E. Eichler. 2015. Global diversity, population stratification, and selection of human copy-number variation. Science 349:6253. https://doi.org/10.1126/science.aab3761.

Therneau, T. M. 2015. coxme: Mixed Effects Cox Models. *R package version* 2.3–18. Website: https://CRAN.R-project.org/package=coxme

VanRaden, P. M. 2020. Symposium review: How to implement genomic selection. J. Dairy Sci. 103:5291–5301. https://doi.org/10.3168/jds.2019-17684.

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.

Wang, K., M. Li, D. Hadley, R. Liu, J. Glessner, S. F. 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.

Weller, J. I., E. Ezra, and M. Ron. 2017. Invited review: A perspective on the future of genomic selection in dairy cattle. J. Dairy Sci. 100:8633–8644. https://doi.org/10.3168/jds.2017-12879.

Weller, J. I., E. Ezra, and M. van Straten. 2019. Genetic and environmental analysis of diseases with major economic impact in Israeli Holsteins. J. Dairy Sci. 102:10030–10038. https://doi.org/10.3168/jds.2018-16026.

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 15:1–10. https://doi.org/10.1186/1471-2164-15-683.

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 12:1–20. https://doi.org/10.1186/1471-2164-12-557.

Zhang, F., W. Gu, M. E. Hurles, and J. R. Lupski. 2009. Copy number variation in human health, disease, and evolution. Annu. Rev. Genomics Hum. Genet. 10:451–481. https://doi.org/10.1146/annurev.genom.9.081307.164217.

Zhou, G., O. Soufan, J. Ewald, R. E. W. Hancock, N. Basu, and J. Xia. 2019. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta- analysis. Nucleic Acids Res. 47(W1):W234–W241. https://doi.org/10.1093/nar/gkz240.

Zhou, Y., Y. T. Utsunomiya, L. Xu, E. H. Hay, D. M. Bickhart, T. S. Sonstegard, C. P. Van Tassell, J. F. Garcia, and G. E. Liu. 2016. Comparative analyses across cattle genders and breeds reveal the pitfalls caused by false positive and lineage-differential copy number variations. Sci. Rep. 6:29219. https://doi.org/10.1038/srep29219.