Tran, Q., Macleod, I., Nguyen, T., Wang, J., Tahir, M., Thakur, P., Reich, C., Chamberlain, A. (2026) Structural variants are associated with changes in gene expression in blood of Holstein cows, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 23898: https://doi.org/10.31274/wcgalp.23898 ©2026 Tran, et al. CC BY-NC-ND

Structural variants are associated with changes in gene expression in blood of Holstein cows

Q.H. Tran[1,2], I.M. Macleod[1,2], T.V. Nguyen[1], J. Wang[1], M.S. Tahir[1], P. Thakur[1], C.M. Reich[1], and A.J. Chamberlain[1,2]

[1]Agriculture Victoria, Centre for AgriBioscience, Bundoora, Australia; [2]School of Applied Systems Biology, La Trobe University, Bundoora, Australia; Corresponding author: quanhong.tran@agriculture.vic.gov.au

Introduction

Structural variants (SVs) are an important source of genetic variation. These variants are large changes in DNA (>50 bp), including deletions, duplications, insertions, inversions, or other rearrangements. Because SVs often affect large genome segments, they can influence gene dosage, disrupt coding sequences, alter regulatory elements, which in turn can have a large impact on phenotypes. In cattle, studies have reported that SVs cover a 5 to 6 times larger fraction of the genome than SNPs and small indels (Grant *et al.*, 2024), Several structural variants in cattle have been linked to embryonic or neonatal mortality, as well as growth, milk production and coat colour (reviewed by Chen *et al*., 2024). However, the application of SVs in genome-wide analyses remains limited.

Traditional short-read sequencing and genotyping approaches have limited sensitivity for detecting SVs, which are often large, complex, and sometimes multiallelic. Often, short reads do not span the entire SV, leaving some of them difficult to detect or inaccurately genotyped. Recently, long-read sequencing technologies, which typically produce reads in excess of 10 kb, have begun to change this landscape. These methods improve the accuracy of SV genotyping and provide more precise breakpoint information. Projects such as The Bovine Long Read Consortium (BovLRC) are applying these technological advances to build a reference population of SV in cattle to enable their integration into genome-wide association studies (GWAS) and genomic prediction models (Chamberlain *et al*., 2025)**.**

While SV detection and genotyping in cattle have improved, the functional roles of most SVs are not yet characterized. Computational annotation tools, such as the Ensembl Variant Effect Predictor (VEP) (McLaren *et al*., 2016)**,** can predict potential functional consequences of SVs that occur in or across annotated genes. However, these are in silico predictions and cannot determine the specific effects of SVs in noncoding regions.

Most GWAS for important traits in cattle such as milk yield, fertility, growth, and health have generally not included SV, however it is plausible that SVs underlie at least some of the reported QTL. In this study, we used both whole-genome long-read sequence and transcriptome data (RNAseq) from the same 28 Holstein cattle to examine whether their SV genotypes were associated with changes in expression of nearby genes. Although our sample size is small and statistical power for association testing is limited, we identified some SVs with potential functional impact on gene expression, and these candidate SV–gene pairs may guide further validation in larger cohorts when more data becomes available.

Materials & Methods

DNA and RNA sequencing, SV genotyping, and gene expression analysis. 86 Holstein and 74 Jersey animals were used to generate a comprehensive catalogue of genome wide SVs. DNA extraction, sequencing, basecalling, quality control, alignment and SV detection are described in Chamberlain *et al (*2025). SVs were filtered to retain only segregating, biallelic deletions, insertions, and duplications in the 28 Holstein females that also had transcriptomic data. SVs

longer than 100 kb or present in fewer than 3 animals were removed. These SVs were annotated using VEP v115.0.
RNA was extracted from whole blood sampled in early lactation (<14 days in milk) from the 28 animals (with approval from the Department of Energy, Environment and Climate Action Agricultural Research and Extension Animal Ethics Committee application 2022-04 approved 11 May 2022) using the RNeasy Protect Animal Blood Kit (Qiagen), libraries were prepared with the Nextflex Rapid Directional RNA-Seq Kit (Revvity), and sequenced as 150 bp paired-end reads on DNBSEQ-G400 (MGI Tech) DNA sequencer. Reads were trimmed with QuadTrim (Robinson & Ross, 2019), aligned to the ARS-UCD2.0 bovine genome using STAR in dual pass mode (Dobin *et al*., 2013), uniquely mapped reads were quantified with FeatureCounts (Liao *et al*., 2014). Genes with counts <3 CPM in at least three samples or with zero counts in more than half of the samples were removed and the remaining normalised for library size.

Association test for effect of SV genotype on gene expression. Individual SV-gene associations were tested using a simple linear regression model: $\mathbf{y} = \mu + \mathbf{x}b + \mathbf{e}$, where $\boldsymbol{y}$ is the vector of normalised gene expression values across animals, μ is the overall mean, $\mathbf{x}$ is the vector of SV genotypes coded as 0 (no SV), 1 (heterozygous), or 2 (homozygous for SV), b is the fixed effect of the SV genotype, and $\mathbf{e}$ is the vector of residuals. Significance of the SV effect was assessed using a Wald test (Wald, 1943), and p-values were adjusted using the Benjamini–Hochberg FDR method (Benjamini & Hochberg, 1995). SV–gene expression association tests were restricted to genes overlapping an SV or within ±1 Mb of the SV boundary.

Results & Discussion

In total 34,628 segregating structural variants (SVs), including 16,660 deletions, 17,851 insertions, and 117 duplications were tested for their association with the expression of 13,623 genes. The association tests between SVs and overlapping or nearby genes resulted in 450,125 tests. Among these, 1,398 SV-gene pairs were significant after multiple testing correction (adjusted p-value < 0.05). These pairs involved 662 genes and 1,156 SVs, including 6 duplications, 561 deletions, and 589 insertions.
Among these significant associations, 86 SV overlapped part or all of a gene, including 15 pairs where the SVs overlapped exons and were annotated as high impact by VEP and 71 pairs where SVs were within introns. SVs located upstream and downstream of genes were involved in 667 and 645 significant pairs, respectively. Exonic SV-gene pairs showed the highest proportion of associations, with 3.4% of tested exon-overlapping pairs being significant, compared with 0.9% for intronic pairs and approximately 0.3% for upstream or downstream pairs (Fig 1A).
For the 1,312 significant SV–gene pairs in which the SV was located outside the gene, 641 were associated with deletions, 665 with insertions, and only six with duplications. Among these pairs, 726 (~55%) showed a negative effect of the SV genotype on expression (expression decreases with increasing SV copies), while the remaining pairs showed a positive effect. This pattern was generally consistent across SV positions (upstream, downstream) and SV types (deletion or insertion). Examination of SV–gene distances showed an enrichment of significant SV–gene pairs at shorter distances. Compared with all tested pairs, SV–gene pairs with significant associations ($p < 0.05$ or $p < 0.01$) were more frequent within the first ~0.2 Mb of the gene, indicating that SVs closer to genes are more likely to be associated with expression changes (Fig. 1B). This was similar to observations from SNP-based eQTLs, where SNPs within a few hundred kb of a gene tend to have stronger effects on expression (Liu *et al*., 2022). Among these SV–gene pairs, 229 genes had more than one SV associated with expression

levels, and 163 SVs were associated with the expression of multiple genes. Six SVs were associated with clusters of at least five genes, highlighting potential hotspot regions with regulatory effects.

Among the 15 significant exon-overlapping SV–gene pairs, one duplication encompassed the entire *CA5A* gene (extending ~13 kb upstream and ~29 kb downstream) and animals with the duplication (heterozygous form, no homozygotes detected) showed a positive effect of the SV on gene expression (Fig. 1C). Eleven deletions affecting exons resulted in six pairs with a negative effect and five pairs with a positive effect, while three insertions in exons were associated with two pairs showing a negative effect and one pair showing a positive effect. For the 71 intronic SV–gene pairs, 37 deletions corresponded to 20 pairs with a negative effect and 17 with a positive effect, and 34 insertions corresponded to 14 pairs with a negative effect and 20 with a positive effect. Notably, the frequency of 1/1 genotypes was lower for exon-overlapping SVs compared with other SVs (43% of exonic SVs lacked a 1/1 genotype vs. 38% overall), thus many tests for exon-overlapping SVs included only 0/0 and 0/1 individuals.

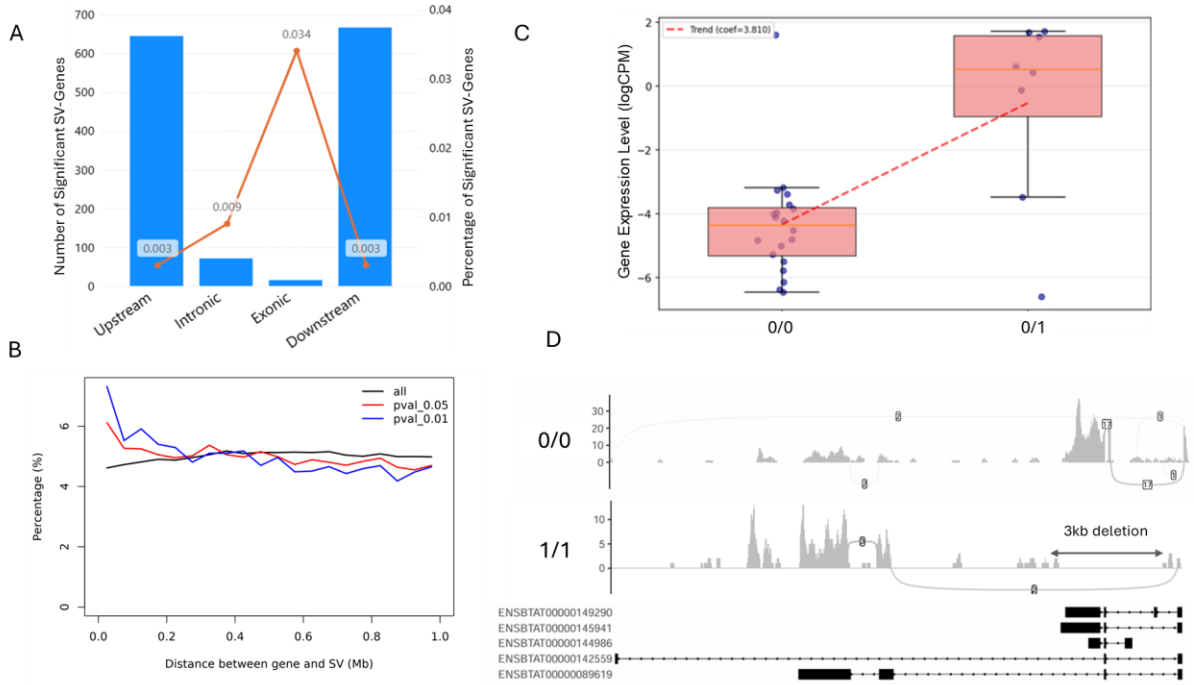

Figure 1. (A) Number of SVs (blue bars, left y-axis) and proportion of significant SV–gene associations (orange line, right y-axis) by SV genomic location (Upstream, Intronic, Exonic, Downstream). **(B)** Histogram showing the distribution of distances between SVs and genes. Lines represent all SV-gene pairs, pairs with $p < 0.05$, and pairs with $p < 0.01$. **(C)** Boxplot showing normalized expression levels of *CA5A* gene by SV genotype (0/0 vs 0/1). **(D)** RNA-seq alignments for ENSBTAG00000068759 showing transcript changes from an exon-deleting SV in 0/0 and 1/1 genotypes. The y-axis shows read coverage (coverage read counts). The 3 kb deletion is marked by the arrow.

We note that some SVs and their associated genes may be promising causal variants for cattle traits, particularly those overlapping genic regions, and some concur with previously reported quantitative trait loci (QTL) in cattle. For example, *CA5A* gene, has been implicated in a genomic region associated with postpartum blood calcium concentration in dairy cattle, a trait influencing periparturient hypocalcemia risk and related metabolic outcomes (Novo *et al.*, 2023). Other genes in significant SV–gene pairs include *GSTO1* and *GRAMD2*B which are

located in or near QTL regions associated with traits such as milk fatty acid composition, and stillbirth in cattle (Bai *et al*., 2025; Palombo *et al*., 2018).

Detecting an association with gene expression depended strongly on the genomic position of the SV relative to the gene. SVs overlapping exons showed the highest rate of association with gene expression. Indeed, for some exon deletions, visualization of RNA-seq alignments showed clear and expected transcript structure changes in animals with homozygous deletions (Fig. 1D). In addition, the absence of homozygous alternate genotypes for many exonic SVs likely reduces the power to detect their true effects on gene expression. When only heterozygous individuals are present, statistical tests may underestimate the functional impact of these SVs, suggesting that their true contribution to gene regulation could be greater than observed in this cohort. The lower allele frequency of the exonic SVs, combined with few or no homozygous alternate genotypes (also observed in our larger dataset of 160 animals) suggests they may be under negative selection pressure, potentially due to detrimental effects on fitness or survival. SVs in noncoding regions showed low rates of association with gene expression compared to exonic variants that directly impact coding regions. Some of these variants may be located in regulatory regions and could therefore influence gene expression. Across all significant SV–gene pairs, negative effects on gene expression were slightly more common than positive effects, consistent with loss-of-function mechanisms. However, both directions were observed across SV types and genomic contexts, highlighting the diverse regulatory consequences of structural variation. This analysis only tested associations between SVs and genes within ±1 Mb of the SV, so more some regulatory effects may have been missed, as regulatory elements can act at greater distances or in *trans*. Small variants (< 50 bp) were not tested in this study.

Conclusions

Despite the limited sample size, direct association testing between SV genotypes and gene expression provided some evidence that structural variants contribute to expression variability. Results indicate that structural variants overlapping exons are more likely to affect gene expression than intergenic SV. These findings motivate future studies using larger cohorts with SV, SNV genotypes and targeted validation studies to improve statistical power and to confirm causal relationships between specific SVs and genes in cattle.

Acknowledgements

The authors acknowledge funding from the DairyBio project: a joint venture between Agriculture Victoria (Melbourne, Australia), Dairy Australia (Melbourne, Australia) and the Gardiner Foundation (Melbourne, Australia).

References

Chamberlain A., Nguyen T., Wang J., *et al*. (2025) Proc Assoc Advmt Anim Breed Genet 25:274-277.
Bai Z., Chen W., Cai Z., *et al*. (2025) Genomics 117:111134.
Benjamini Y., and Hochberg Y. (1995) J R Stat Soc B 57:289-300.
Chen Y., Khan M.Z., Wang X., *et al*. (2024.) Front Vet Sci.
Dobin A., Davis C.A., Schlesinger F., *et al*. (2013) Bioinformatics 29:15-21.
Grant J.R., Herman E.K., Barlow L.D., *et al*. (2024) BMC Genomics 25:903.
Liu S., Gao Y., Canela-Xandri O., *et al*. (2022) Nat Genet 54:1438–1447.
McLaren W., Gil L., Hunt S.E., *et al*. (2016) Genome Biol 17:122.
Novo L.C., Poindexter M.B., Rezende F.M., *et al*. (2023) Sci Rep 13:21900.
Palombo V., Milanesi M., Sgorlon S., *et al*. (2018) J Dairy Sci 101:11004-11019.
Robinson A., and Ross E. (2019) bioRxiv.
Wald A. (1943) Trans Am Math Soc 54:426-482.