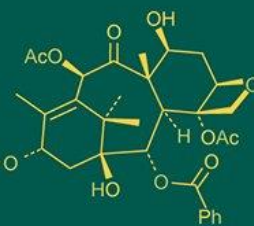
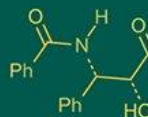


International Journal of Advanced Biochemistry Research



ISSN Print: 2617-4693

ISSN Online: 2617-4707

NAAS Rating (2026): 5.29

IJABR 2026; 10(5): 613-616

www.biochemjournal.com

Received: 01-02-2026

Accepted: 06-03-2026

Saalom King J

Ph.D. Scholar, Department of
Animal Breeding, Genetics and
Biostatistics, Veterinary
College and Research Institute
(TANUVAS), Namakkal,
Tamil Nadu, India

Jeyakumar M

Assistant Professor,
Department of Animal
Genetics and Breeding,
Veterinary College and
Research Institute
(TANUVAS), Namakkal,
Tamil Nadu, India

Jannathul Firthous I

PG Scholar, Department of
Livestock Production and
Management, College of
Veterinary and Animal
Sciences, (KVASU),
Mannuthy, Kerala, India

Corresponding Author:**Saalom King J**

Ph.D. Scholar, Department of
Animal Breeding, Genetics and
Biostatistics, Veterinary
College and Research Institute
(TANUVAS), Namakkal,
Tamil Nadu, India

High-impact variants on the X chromosome in Italian Holstein cattle: Implications for male cattle health

Saalom King J, Jeyakumar M and Jannathul Firthous I

DOI: <https://www.doi.org/10.33545/26174693.2026.v10.i5h.8462>

Abstract

Whole exome sequencing (WES) has emerged as a powerful approach for identifying functional genetic variants associated with economically important traits and health conditions in livestock. The present study aimed to identify and characterize high-impact variants located on the X chromosome in Italian Holstein cattle using publicly available whole exome sequencing data (ERR2015687). Since male cattle possess a single X chromosome, deleterious mutations on this chromosome may exert pronounced phenotypic effects due to the absence of a compensatory allele. Raw paired-end sequencing reads were retrieved from the NCBI-European Nucleotide Archive and subjected to quality assessment using FastQC, followed by quality trimming and adapter removal using Trimmomatic. High-quality reads were aligned to the *Bos taurus* reference genome (ARS-UCD2.0) using BWA-MEM and alignment files were processed using SAMtools. Variant calling was performed using BCftools and functional annotation of variants was carried out using SnpEff. A total of 1,391,294 variants were identified, among which 23 high-impact variants were located on the X chromosome. Detailed exploratory analysis identified six frameshift INDELs and four high-impact SNPs affecting functionally important genes including *GPR119*, *IGSF1*, *CCNB3*, *POLA1*, *OFD1* and *HUWE1*. Frameshift mutations identified in *GPR119* may affect glucose metabolism and insulin regulation, whereas variants in *IGSF1* may influence reproductive and neuroendocrine development. Mutations in *CCNB3* could impair cell cycle progression and spermatogenesis, while *POLA1* variants may disrupt DNA replication and genomic stability. Additionally, splice-site variants in *OFD1* and *HUWE1* may interfere with transcript processing and cellular regulatory pathways. The findings highlight the potential biological significance of X-linked high-impact variants in cattle and emphasize the importance of investigating sex-linked genomic alterations in livestock breeding and health management programmes.

Keywords: Whole exome sequencing, X chromosome, high-impact variants, Italian Holstein cattle, frameshift INDELs, functional genomics

Introduction

The Holstein Friesian (HF) is one of the most widely distributed and economically important dairy cattle breeds worldwide. The breed originated in the Netherlands through the crossing of black Batavian and white Friesian cattle, resulting in animals with superior milk-producing ability even under limited feeding conditions (Elischer, 2014) [2]. Among HF populations, the Italian Holstein has gained particular importance due to its exceptional milk yield and major contribution to the dairy industry, especially in the production of high-value dairy products such as Parmigiano-Reggiano and Mozzarella cheese. Consequently, genomic studies in Holstein cattle are highly relevant for improving productivity, fertility, disease resistance and overall breeding efficiency.

Fertility and production traits in dairy cattle are strongly influenced by genetic factors. Earlier studies demonstrated that variation in fertility traits possesses a substantial genetic component, making genomic evaluation an important tool for livestock improvement programmes (Jansen, 1985) [7]. With the advancement of next-generation sequencing technologies, identification of functional genomic variants associated with economically important traits has become increasingly feasible. Whole exome sequencing (WES) is a targeted sequencing approach that focuses specifically on protein-coding regions of the genome, where the majority of functional mutations are located (Ng *et al.*, 2009; Mamanova *et al.*, 2010) [9, 8].

Compared with whole genome sequencing, WES provides a cost-effective strategy for detecting potentially deleterious variants, including single nucleotide polymorphisms (SNPs) and insertion-deletion mutations (INDELs), particularly in genes associated with metabolism, reproduction, development and disease susceptibility.

In livestock genomics, exploratory analysis of single-animal exome datasets can provide preliminary insights into candidate functional variants and facilitate screening of potentially important mutations in breeding animals. Such analyses are especially important for investigating X-linked variants, since males possess only one X chromosome and therefore may exhibit stronger phenotypic effects from deleterious mutations due to the absence of a compensatory allele. Therefore, the present study aimed to perform an exploratory whole exome analysis of an Italian Holstein bull to identify high-impact variants located on the X chromosome and to investigate their potential biological significance in relation to metabolism, reproduction, development and cellular regulation.

Materials and Methods

Data Acquisition

Publicly available whole exome sequencing (WES) data of Italian Holstein cattle were obtained from the European Nucleotide Archive (ENA) under project accession PRJEB21516. One paired-end sequencing dataset (ERR2015687) generated using the Illumina HiSeq 2500 platform was selected for exploratory analysis. The dataset consisted of 8,273,140 paired-end reads comprising approximately 1.7 Gb of sequence data.

Quality Control and Read Preprocessing

The quality of raw sequencing reads was assessed using FastQC to evaluate sequence quality scores, GC content, duplication levels and adapter contamination. Low-quality bases and adapter sequences were removed using Trimmomatic. Reads containing ambiguous bases and short reads were filtered to improve downstream alignment accuracy. Following preprocessing, high-quality reads were retained for subsequent analysis. Quality-filtered reads exhibited improved Q20 and Q30 scores, indicating high sequencing accuracy.

Read Alignment

Filtered paired-end reads were aligned to the *Bos taurus* reference genome (ARS-UCD2.0/bosTau9.fa) using the Burrows-Wheeler Aligner MEM (BWA-MEM) algorithm. The generated Sequence Alignment Map (SAM) file was converted into Binary Alignment Map (BAM) format and subsequently sorted using SAM tools for downstream analysis.

BAM File Processing and Coverage Analysis

The sorted BAM file was indexed using SAM tools to facilitate efficient data access during variant analysis. Sequencing depth and genome coverage statistics were assessed to evaluate the quality and reliability of exome coverage across the reference genome.

Variant Calling

Variant calling was performed using BCF tools to identify single nucleotide polymorphisms (SNPs) and insertion-deletion mutations (INDELs) from the aligned sequencing data. The identified variants were stored in Variant Call Format (VCF) for downstream functional annotation and

analysis.

Variant Annotation

Functional annotation of identified variants was performed using SnpEff implemented through the Galaxy platform. The X chromosome (NC_037357.1) was extracted from the annotated dataset for further analysis. High-impact X-linked variants were filtered using the vcfr package in R based on the following criteria: mapping quality (MQ) > 30, read depth (DP) > 10, quality score (QUAL) > 20 and high-impact functional annotation.

Results and Discussion

Sequencing Quality and Read Filtering

Quality assessment of the raw paired-end sequencing data indicated high overall sequencing quality prior to filtering. In Read1, a total of 8,273,140 reads were obtained, with Q20 and Q30 values of 99.16% and 97.21%, respectively. Similarly, Read2 contained 8,273,140 reads with Q20 and Q30 values of 97.47% and 95.43%, respectively. These results demonstrated that the sequencing dataset was of high quality, with the majority of bases exceeding acceptable quality thresholds. Following preprocessing and filtering, 8,130,049 reads were retained in both forward and reverse reads. Quality metrics improved after trimming, with Q20 values increasing to 99.60% in Read1 and 99.54% in Read2, while Q30 values improved to 97.73% and 97.52%, respectively. The improvement in quality parameters indicated successful removal of low-quality regions, poly-G sequences and adapter contamination.

A total of 16,260,098 reads passed filtering criteria. Approximately 16,092 reads were removed due to low quality, 4,052 reads were discarded because of excessive ambiguous bases (N) and 266,038 reads were excluded for being shorter than 36 bp. Adapter trimming was observed in 137,682 reads, resulting in the removal of 1,444,084 adapter-associated bases. These findings confirmed efficient preprocessing and generation of high-confidence reads suitable for downstream analysis. The duplication rate of the sequencing dataset was estimated at 6.53%, indicating minimal redundancy among reads. Additionally, the insert size peak was observed at approximately 160 bp, which was consistent with expected fragment size distributions for exome sequencing libraries and suitable for accurate alignment and variant detection.

Variant analysis identified a total of 1,391,294 variants across the exome dataset. Among these, 32,245 variants corresponded to the X chromosome (NC_037357.1). Filtering based on mapping quality (MQ > 30), read depth (DP > 10), quality score (QUAL > 20) and predicted functional consequence identified 10 high-impact variants on the X chromosome. The identified high-impact variants included six INDELs and four SNPs. Among the INDELs, five were classified as frameshift variants and one as a splice variant. The SNPs included two splice donor variants and two splice acceptor variants. Only characterised variants were considered for in biologically interpretations including: *GPR119*, *IGSF1*, *CCNB3*, *POLA1*, *OFD1* and *HUWE1* mentioned in Table 1. Since male cattle possess only one X chromosome, deleterious mutations affecting X-linked genes may exert stronger phenotypic effects due to the absence of a second compensatory allele. Therefore, identification of high-impact X-linked variants is important for understanding reproductive performance, growth, metabolism and developmental traits in cattle.

Table 1: High-impact variants identified on the X chromosome of the Italian Holstein bull.

Position (NC_037357.1)	REF	ALT	Variant Category	Gene
14348583	CGAGAGA	CGAGA	Frameshift INDEL	<i>GPR119</i>
15290343	TGGGGGGG	TGGGGGGGG	Frameshift INDEL	<i>IGSF1</i>
87970147	GAA	GA	Frameshift INDEL	<i>CCNB3</i>
1.19E + 08	ATTTTTTT	ATTTTT	Frameshift INDEL	<i>POLA1</i>
1.3E + 08	C	T	Splice-associated SNP	<i>OFD1</i>
90970458	T	C	Splice-associated SNP	<i>HUWE1</i>

Fertility is one of the most economically important traits in dairy cattle and previous studies have shown that reproductive efficiency can be improved through better management practices and hormonal interventions (Darwash *et al.*, 2001) [4]. However, a substantial proportion of variation in fertility traits is also influenced by genetic factors (Jansen, 1985) [7]. In this context, identification of high-impact genomic variants associated with reproductive and physiological functions is important for understanding the molecular basis of fertility and productivity in cattle. In the present study, exploratory analysis identified high-impact variants on the X chromosome affecting genes including *POLA1*, *GPR119*, *IGSF1*, *CCNB3*, *OFD1* and *HUWE1*. The biological significance of these variants may be particularly important in male cattle because bulls possess only one X chromosome, increasing the likelihood of phenotypic expression of deleterious mutations.

Frameshift variants identified in *POLA1* may have important implications for cellular proliferation and genomic stability. The *POLA1* gene encodes the catalytic subunit of DNA polymerase alpha, which is essential for DNA replication and cell division. Previous studies have shown that mutations in *POLA1* can alter gene splicing and reduce transcript and protein levels, resulting in immune dysfunction and developmental abnormalities (Starokadomskyy *et al.*, 2016) [11]. In affected males, *POLA1*-associated disorders have been linked to growth retardation, recurrent infections and systemic developmental defects (Pezzani *et al.*, 2013) [10]. Therefore, disruption of *POLA1* in cattle may potentially influence growth, immunity and reproductive performance. The identified variant in *GPR119* may also possess functional importance due to the gene's role in metabolic regulation. *GPR119* belongs to the G-protein-coupled receptor superfamily and is involved in glucose-dependent insulin secretion as well as incretin hormone release (Yang *et al.*, 2018) [13]. Alterations in this gene could impair glucose homeostasis, feed efficiency and metabolic adaptation in cattle, potentially affecting growth and production traits.

Similarly, variants identified in *IGSF1* may influence growth and reproductive physiology. Previous studies in mice have highlighted *IGSF1* as a candidate gene associated with skeletal muscle development and muscle mass regulation (Verbrugge *et al.*, 2018) [12]. Since the gene is also involved in endocrine signalling pathways, functional disruption may affect developmental processes and reproductive efficiency in cattle. A frameshift mutation identified in *CCNB3* may have reproductive significance due to its role in cell cycle progression and meiotic regulation. Earlier reports demonstrated that *CCNB3* is indispensable for female fertility in mice and is associated with successful meiotic progression (Chotiner *et al.*, 2022) [1]. Although the present study focused on male cattle, disruption of this gene may still affect germ cell

development and spermatogenesis, potentially influencing bull fertility.

In addition to the identified frameshift INDELs, SNP variants detected in *OFD1* and *HUWE1* may also possess important biological significance in cattle development and reproduction. The *OFD1* gene, also known as *CXORF5*, has been associated with several developmental syndromes in humans, including the X-linked dominant Oral-Facial-Digital type 1 syndrome (Ferrante *et al.*, 2001) [5]. Previous studies demonstrated that *OFD1* exhibits stage-specific expression during embryonic development, particularly in genital ridges, nervous tissues and craniofacial structures. The gene is also highly expressed in epithelial tissues lining the oral and nasal cavities (Ferrante *et al.*, 2001) [5]. Since *OFD1* plays a major role in ciliary function and organ development, mutations affecting this gene may potentially contribute to developmental abnormalities, reproductive defects and impaired organogenesis in cattle.

Similarly, variants identified in *HUWE1* may influence important cellular regulatory mechanisms. The *HUWE1* gene encodes an E3 ubiquitin protein ligase involved in protein degradation, apoptosis, DNA damage response and cell cycle regulation. Previous reports in humans demonstrated that increased copy number variations in *HUWE1* are associated with cognitive impairment and neurological abnormalities in males (Froyen *et al.*, 2012) [6]. Since *HUWE1* participates in multiple cellular homeostasis pathways, functional alterations in this gene may affect cellular proliferation, reproductive physiology and developmental stability in cattle.

Conclusion

The present exploratory whole exome analysis identified high-impact variants on the X chromosome of an Italian Holstein bull, including frameshift INDELs and splice-associated SNPs. The genes associated with the variants are associated with metabolism, reproductive function, DNA replication, cellular regulation and developmental processes. The findings provide preliminary insights into candidate functional variants that may influence fertility, growth and overall health in cattle, highlighting the importance of whole exome sequencing for screening breeding animals and identifying potentially important genomic variants.

References

- Chotiner JY, Leu NA, Xu Y, Wang PJ. Recurrent pregnancy loss in mice lacking the X-linked *Ccnb3* gene. *Biol Reprod.* 2022;106:382-384.
- Elischer M. History of dairy cow breeds: Holstein. Michigan State University Extension. Howell (MI): Michigan State University Extension; 2014. p. 1-12.
- Jansen J. Genetic aspects of fertility in dairy cattle based on analysis of artificial insemination data: A review with emphasis on areas for further research. *Livest Prod Sci.* 1985;12:1-12.

4. Darwash AO, Lamming GE, Royal MD. A protocol for initiating oestrus and ovulation early post partum in dairy cows. *Anim Sci.* 2001;72:539-546.
5. Ferrante MI, Giorgio G, Feather SA, Bulfone A, Wright V, *et al.* Identification of the gene for oral-facial-digital type I syndrome. *Am J Hum Genet.* 2001;68:569-576.
6. Froyen G, Belet S, Martinez F, Santos-Reboucas CB, Declercq M, Verbeeck J, *et al.* Copy-number gains of HUWE1 due to replication-and recombination-based rearrangements. *Am J Hum Genet.* 2012;91:252-264.
7. Jansen J. Genetic aspects of fertility in dairy cattle based on analysis of artificial insemination data: A review with emphasis on areas for further research. *Livest Prod Sci.* 1985;12:1-12.
8. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, *et al.* Target-enrichment strategies for next-generation sequencing. *Nat Methods.* 2010;7:111-118.
9. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature.* 2009;461:272-276.
10. Pezzani L, Brena M, Callea M, Colombi M, Tadini G. X-linked reticulate pigmentary disorder with systemic manifestations: A new family and review of the literature. *Am J Med Genet A.* 2013;161A:1414-1420.
11. Starokadomskyy P, Gemelli T, Rios JJ, Xing C, Wang RC, Li H, *et al.* DNA polymerase-alpha regulates the activation of type I interferons through cytosolic RNA:DNA synthesis. *Nat Immunol.* 2016;17:495-504.
12. Verbrugge SA, Schönfelder M, Becker L, Yaghoob Nezhad F, Hrabě de Angelis M, Wackerhage H. Genes whose gain or loss-of-function increases skeletal muscle mass in mice: A systematic literature review. *Front Physiol.* 2018;9:553.
13. Yang JW, Kim HS, Choi YW, Kim YM, Kang KW. Therapeutic application of GPR119 ligands in metabolic disorders. *Diabetes Obes Metab.* 2018;20:257-269.