Acta Scientific Veterinary Sciences (ISSN: 2582-3183)

Volume 5 Issue 7 July 2023

Review Article

Solutions to Post-GWAS Regulatory Variants in Bovine

Nidhi Sukhija[1], Kanaka K K[1], Jayakumar Sivalingam[2]*, Komal Jaglan[3], Pallavi Rathi[1], Rangasai Chandra Goli[1] and Chethan Raj[4]

[1]ICAR-National Dairy Research Institute, Karnal, Haryana, India

[2]ICAR-Directorate of Poultry Research, Hyderabad, India

[3]Lala Lajpat Rai University of Veterinary and Animal Sciences, Hisar, Haryana, India

[4]ICAR-Indian Veterinary Research Institute, Izatnagar, UP, India

***Corresponding Author:** Jayakumar Sivalingam, ICAR-Directorate of Poultry Research, Hyderabad, India.

DOI: 10.31080/ASVS.2023.05.0681

Received: May 23, 2023

Published: June 03, 2023

© All rights are reserved by **Jayakumar Sivalingam., *et al.***

Abstract

Advancements in Next Generation Sequencing have led to an increased exploration of the genome and transcriptome to uncover genetic variants associated with phenotypic traits. Genome-wide association studies (GWAS) and genomic predictions play a crucial role in identifying significant genetic variants that contribute to complex traits. However, most of these variants are found in non-coding regions of the genome, making their functional annotation and interpretation challenging. This review highlights the importance of characterizing and prioritizing non-coding variants and their effects on regulatory elements in livestock genomics. Regulatory elements such as promoters, enhancers, silencers, and non-coding RNAs coordinate gene expression and are critical for understanding the underlying mechanisms of traits. The abstract also discusses various tools and methods for annotating and predicting the effects of regulatory variants, as well as validation platforms for studying their functional impact. Comprehensive functional annotation of non-coding variants is essential for gaining insights into the genetic architecture of complex traits and improving genetic selection strategies in livestock breeding programs.

Keywords: GWAS; Regulatory Variants, SNPs; Tools; RNA

Introduction

Exploration of genome and transcriptome are increasing day by day to accelerate genetic gain with the advent of Next generation Sequencing (NGS), there is a hike in studies where genetic variants are tested for associations with phenotypic traits. In the era of genomics and phenomics, carrying out genome-wide association studies (GWAS) and genomic predictions, is the key to account for significant genetic variants leading to variations in complex traits, thereby modifying the phenotypes. GWAS detects mutations that explain variance enough that surpass threshold p-values [1]. Quantitative/Complex/Multifactorial/Polygenic traits are affected by a number of variants having small effects, which may be coding or non-coding, leading to the phenotypic variations [2]. For instance using GWAS approach seven genes were identified in Indian Buffalo viz., NCBP1, FOXN3, TPK1, XYLT2, CPXM2, HERC1, and OPCML associated with mastitis [34] and AQP1, TRNAE-CUC, NRIP1, CPNE4 and VOPP1 have role in different fertility traits [35]. This is because non-coding areas of the genome are where the majority of GWAS-identified SNPs are found. The GWAS-identified SNPs must be viewed as merely representative of all SNPs in the same haplotype block, and it is equally possible that additional SNPs in high linkage disequilibrium (LD) with the array-identified SNPs are causal for the disease [3]. Most of the time we do not hit the gene directly, instead we hit the region surrounding it. Due to lack of fine-scale mapping, unfortunately the nearby gene is assumed as causal for the trait. This is where linkage disequilibrium comes into play. A total of 88% of disease-associated variants lie in non-coding regions [4]. The regulatory variants should be annotated further for prioritization of variants in genomic studies [5]. Searches pertaining to coding regions may fail to yield causal variants, if this is the case [6]. Causal variants underlying quantitative traits often have regulatory effects on the expression of target genes and that these expression effects might be modest and cell-type specific [7]. Lack of comprehensive functional annotations across a wide range of tissues and cell types severely hinders the biological interpretations in livestock. Thus, it is hi-time to characterize, annotate and prioritize non-coding variants as well as their effects [8].

Key regulatory elements

Regulatory elements coordinate the precise expression of genes as per- cell type, developmental stage and stimuli. Genomic elements that regulate gene expression, generally, located within noncoding regions are known as regulatory elements. Promoters are located in the 5' region of genes that activate transcription via RNA Polymerase II (RNAPII). Enhancers are bound by activators and concerned with upregulation while silencers are bound by repressors and concerned with downregulation. Insulators prevent interaction between promoters and enhancers. Promoters, enhancers, silencers and repressors are the key cis-regulatory elements (CREs). The Non-coding RNAs (ncRNAs) are subject to post-transcriptional modification [9]. In *Bos taurus* cattle, 28.3 million SNPs and InDels have been detected, which can be imputed further into larger datasets for GWAS and genomic predictions [10]. Moreover, the current map of bovine regulatory variants is also limited [11].

Promoter region of Holstein Friesian cattle, a total of 16 alleles (R1A/B to R16A/B). Amongst these, allele R5A/B at position -204 (G>C) from the transcription start site holds importance as it lies within the binding site of milk protein binding factor (MPBF), and might affect the activity of the gene product [12].

Methodology

Map and characterize the circuitry of non-coding elements including *cis*-regulatory regions (promoters, enhancers, insulators and silencers) and ncRNAs. These elements can be identified by a combination of functional genomics approaches and sequence conservation [13]. Then identify disease-relevant tissues, annotate variants and regulators (Table 1). Combine the genetic and epigenetic variation in the study. Then, uncover and manipulate trait mechanism and circuitry. High-throughput perturbations and therapeutic delivery should be done for validation, as described in Table 2.

Name	Uses	Data sources	Limitations	References
RegulomeDB	Score-based prioritization	ENCODE, Roadmap Epigenomics	Difficult to interpret	[14]
HaploReg	Variants in LD, within or next to regulatory elements	ENCODE, GTEx Roadmap Epigenomics	Not updated periodically	[15]
FunciSNP	Identification and prioritization of putative regulatory SNPs	ENCODE, Roadmap Epigenomics	A minimum knowledge of R is needed	[16]
ENlight	Annotation of GWAS variants and analyzing their putative effects by plot visualization.	GWAS, ENCODE, GTEx	Not updated periodically	[17]

Table 1: Regulatory variant annotation tools.

Effect prediction tools

Prediction algorithms to calculate the probability of this variant to affect regulatory motifs and hence, affect the traits. GWAVA (Genome-Wide Annotation of Variants): is a tool that facilitates noncoding variant prioritisation through the incorporation of various genomic and epigenomic annotations. Compared to a conventional variant predictor, combined annotation-dependent depletion (CADD) performs better on regulatory variants [18]. CADD (combined annotation-dependent depletion), a process for integrating a variety of different annotations into a single, objective measurement (C score) for each variant. A support vector machine called CADD has been trained to distinguish between 14.7 million high-frequency alleles derived from humans and 14.7 million simulated variants. C scores rank known pathogenic variants within individual genomes highly and correlate with allelic diversity, annotations of functionality, pathogenicity, disease severity, experimentally measured regulatory effects, and complex trait associations. Through a variety of functional categories, effect sizes, and genetic architectures, CADD can prioritize functional, harmful, and pathogenic variants [19]. DANN, (The Deleterious Annotation of genetic variants using Neural Networks tool) forecasts the effects of non-coding variants, it makes use of a Deep Neural Network (DNN) algorithm that captures linear relationships among various annotations, including evolutionary features. This tool was created to enhance the CADD SVM algorithm results; by utilizing DNN, it can capture more relationships between annotated objects. DANN has been shown to outperform CADD results using the same annotations and training data sets [20]. LINSIGHT predicts the potential effects of regulatory variants and ranks them; it combines probabilistic and linear models with functional and evolutionary conservation data. LINSIGHT identifies harmful regulatory variants linked to inherited diseases by analyzing data for various genomic features from sources like ENCODE and FANTOM5. This method is used to determine the selective pressure on regulatory regions, assess the fitness effects of regulatory variants, and forecast their effects [21]. FATHMM-MKL: (Functional Analysis through Hidden Markov Models, http://fathmm.biocompute.org.uk/): It is based on a machine learning algorithm that employs annotations from ENCODE to predict the potential effects of regulatory variants using a multiple kernel (MK) learning technique. In order to classify input variants and ultimately predict their potential effects, it weights all the annotations according to their relevance during training and generates matrices that will be used by an MK algorithm. The pathogenic variants from the HGMD and benign variants from the 1000 Genomes Project are both included in the gold-standard data set. The p-values for the predictions made by FATHMM-MKL are provided for use in other integrative studies. The FATHMM-XF method, which trains a supervised machine learning approach with additional genetic and epigenetic features from ENCODE and the Roadmap Epigenomics Project and assigns a confidence score to all predictions, has recently improved the prediction system of FATHMM-MKL. Recent research has shown that FATHMM-XF performs better than other predictors, such as CADD and DANN [22].

Challenges

Genomic predictions are not practical due to computing limitations. For GWAS, very stringent significance thresholds are required to avoid false positives. There is also a need to annotate the variants into classes and prioritize them for testing with a higher a priori probability of containing trait associated variants (TAV). However, a large number of variants with significant associations are found in the non-protein coding regions of the genome [8]. Category-based Bonferroni adjustment based on the enrichment was implemented in Nordic Holstein cattle was carried out where upstream and downstream classes were most enriched, for more dairy traits. Intergenic and intragenic variants constituted ~67% and 32% of the total number of variants, respectively [23]. Using functional annotations to prioritize variants within the QTL interval has become a popular strategy. It was recently demonstrated that the use of a variant annotation tool and its evolutionary conservation score [24]. Due to many reasons, such as LD, inaccuracy of imputation, random sampling errors, etc., the lead single nucleotide polymorphism (SNP) may not be the causative one [23].

Technique	Description
Chromosome conformation capture (3C)	Analyse chromatin structure by quantifyng interactions between two selected loci
Chromosome conformation capture-on-chip (4C)	Between a specific locus and other loci
Chromosome conformation capture carbon copy (5C)	All possible interactions within different genomic regions
Hi-C	genome-wide chromatin structure using high-throughput sequencing techniques
Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET)	Combination of ChIP-based methods with 3C and sequencing
Luciferase reporter assay	activity of genomic functional element
DNA fluorescence in situ hybridization (FISH)	locating specific DNA sequences within chromosomes
CRISPR/Cas9	target mutations to specific regulatory elements in experimental models

Table 2: Validation platforms for regulatory variants.

Applications

Genetic diversity plays a massive role in combating abiotic stress [38]. Diversity is indicated in polymorphism the causal regulatory polymorphisms, rSNPs may be used in Marker-trait association followed by Marker-assisted Selection. They can help to select young male calves especially for traits with low heritability (h^2). These variants can be of transgenic use and improve the accuracy of several prediction models, thus enabling functional dissection of traits. The variants mined may help understand novel target regulatory functions and navigate choice of novel therapeutics and personalized medicine. Comparative epigenomics in conjunction with large-scale GWAS for more reliable results. The findings may be extrapolated for changes associated with immune and reproduction in cattle to further advance human research. The findings may be extrapolated for changes associated with immune and reproduction in livestock to further advance human research like immunotherapy which is generally recognised as a viable treatment option for food allergies [39].

The polymorphism analysis of bovine BLG promoter region by Lum., *et al*. (1997) [25] revealed 10 polymorphic sites. They confirmed the functional importance of transversion (G to C) within a consensus binding site for activator protein-2 (AP-2) at position –430 bp from the transcription initiation site [25]. In the Braunvieh cattle, two coat colour variations are noted *viz.,* colour-sided and belted [26]. Artificial insemination was extensively done in the 1960s. Besides *KIT* and *MITF* genes, an intronic regulatory single nucleotide variant was found in bovine MITF in Holstein and Simmental cattle, related to coat colour genetics [27,28]. Hauswirth., *et al*. (2012) [29]; Korberg., *et al*. (2014) [30] and Negro., *et al*. (2017) [31] also reported such variants for white spots on the head and the body in dogs and horses. Brown Swiss cattle with white spots on the abdomen and/or on the head reported more frequently. Genotyping of 172 Brown Swiss cattle revealed two significantly associated completely linked single nucleotide variants (rs722765315 and rs719139527). Both variants are located in the 5′-regulatory regions of the bovine *MITF* gene. Comparative sequence analysis (DNaseI hypersensitive site and a H3K27ac cluster) showed that the variant rs722765315, located 139 kb upstream of the transcription start site of the bovine melanocyte-specific *MITF* transcript [32]. In depth studies of quantitative trait loci (QTL) at chicken chromosome 1 associated with growth traits and contributed 14.4% of the genetic variance for growth. Many candidate genes reside in the associated region, including *Retinoblastoma 1* (*RB1*), *Forkhead box O1* (*FOXO1*). The SIRT6 promoter variants significantly affect transcriptional levels and subsequently significantly influence bovine intramuscular fat content (c.-1100 A > G) [33]. In the chicken genome Kanaka et al. (2021)[38] has identified polymorphism in SERPINB14 gene promoter regions which were associated with egg quality and age at sexual maturity. Identifying key regulatory variants using GWAS approach and functional genomics can also be used to provide conclusive evidence where there will be conflict between different schools of thoughts, for example impact of Beta Casomorphin-7 in A1 milk and its effects on human well being [39].

Conclusion

Researchers have to put great efforts into the annotation of regulatory elements (e.g., promoters and enhancers) across multiple tissues and cell types in cattle, parallel to ENCODE projects (in human, mouse, and Drosophila) and Roadmap Epigenomics Project. By integrating such functional annotations with GWAS from large cohorts (e.g., 1000 bulls project), investigators can gain novel biological insights into regulatory genetic architecture underlying complex traits and diseases. The generally conserved sequences across species can help to explore the biological basis of complex outcomes and adaptive evolution in the target species (e.g., cattle and swine) by borrowing functional annotations from well-studied species such as humans and mice.

Bibliography

1. Weng L., *et al.* "SNP-based pathway enrichment analysis for genome-wide association studies". *BMC Bioinformatics* 12.1 (2011): 1-9.
2. Glazier A M., *et al.* "Finding genes that underlie complex traits". *Science* 298.5602 (2002): 2345-2349.
3. Tak Y G and Farnham P J. "Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome". *Epigenetics and Chromatin* 8.1 (2015): 1-18.
4. Edwards S L., *et al.* "Beyond GWASs: illuminating the dark road from association to function". *The American Journal of Human Genetics* 93.5 (2013): 779-797.
5. Schaub M A., *et al.* "Linking disease associations with regulatory information in the human genome". *Genome Research* 22.9 (2012): 1748-1759.
6. Mackay TF., *et al.* "The genetics of quantitative traits: challenges and prospects". *Nature Reviews Genetics* 10.8 (2009): 565-577.
7. Gallagher MD and Chen-Plotkin AS. "The post-GWAS era: from association to function". *The American Journal of Human Genetics* 102.5 (2018): 717-730.
8. Koufariotis L., *et al.* "Regulatory and coding genome regions are enriched for trait associated variants in dairy and beef cattle". *BMC Genomics* 15.1 (2014): 1-16.
9. Riethoven J J M. "Regulatory regions in DNA: promoters, enhancers, silencers, and insulators". *Computational Biology of Transcription Factor Binding* (2010): 33-42.
10. Daetwyler H D., *et al.* "Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking". *Genetics* 193.2 (2013): 347-365.
11. Bickhart D M and Liu GE. "Identification of candidate transcription factor binding sites in the cattle genome". *Genomics, Proteomics and Bioinformatics* 11.3 (2013): 195-198.
12. Wagner VA., *et al.* "DNA variants within the 5'-flanking region of milk-protein-encoding genes II. The β-lactoglobulin-encoding gene. TAG. Theoretical and applied genetics". *Theoretische und angewandte Genetik* 89.1 (1994): 121–126.
13. Khurana E., *et al.* "Role of non-coding sequence variants in cancer". *Nature Reviews* Genetics 17.2 (2016): 93-108.
14. Boyle A P., *et al.* "Annotation of functional variation in personal genomes using RegulomeDB". *Genome Research* 22.9 (2012): 1790-1797.

15. Ward LD and Kellis M. “HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants”. *Nucleic Acids Research* 40.D1 (2012): D930-D934.

16. Coetzee S G., *et al.* “FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs”. *Nucleic Acids Research* 40.18 (2012): e139-e139.

17. Guo Y., *et al.* “Enlight: web-based integration of GWAS results with biological annotations”. *Bioinformatics* 31.2 (2015): 275-276.

18. Ritchie G R and Flicek P. “Computational approaches to interpreting genomic sequence variation”. *Genome Medicine* 6 (2014): 1-11.

19. Kircher M., *et al.* “A general framework for estimating the relative pathogenicity of human genetic variants”. *Nature Genetics* 46.3 (2014): 310-315.

20. Quang D., *et al.* “DANN: a deep learning approach for annotating the pathogenicity of genetic variants”. *Bioinformatics* 31.5 (2014): 761-763.

21. Huang YF., *et al.* “Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data”. *Nature Genetics* 49.4 (2017): 618-624.

22. Rogers M F., *et al.* “FATHMM-XF: accurate prediction of pathogenic point mutations via extended features”. *Bioinformatics* 34.3 (2018): 511-513.

23. Cai Z., *et al.* “Dissecting closely linked association signals in combination with the mammalian phenotype database can identify candidate genes in dairy cattle”. *BMC Genetics* 20 (2019): 1-12.

24. Nishizaki S S and Boyle AP. “Mining the unknown: assigning function to noncoding single nucleotide polymorphisms”. *Trends in Genetics* 33.1 (2017): 34-45.

25. Lum L S., *et al.* “Polymorphisms of bovine β-lactoglobulin promoter and differences in the binding affinity of activator protein-2 transcription factor”. *Journal of Dairy Science* 80.7 (1997): 1389-1397.

26. Drögemüller C., *et al.* “Genetic mapping of the belt pattern in Brown Swiss cattle to BTA3”. *Animal Genetics* 40.2 (2009): 225-229.

27. Fontanesi L., *et al.* “Haplotype variability in the bovine MITF gene and association with piebaldism in Holstein and Simmental cattle breeds”. *Animal Genetics* 43.3 (2012): 250-256.

28. Jansen S., *et al.* “Assessment of the genomic variation in a cattle population by re-sequencing of key animals at low to medium coverage”. *BMC Genomics* 14 (2013): 1-9.

29. Hauswirth R., *et al.* “Mutations in MITF and PAX3 cause “splashed white” and other white spotting phenotypes in horses”. *PLoS Genetics* 8.4 (2012): e1002653.

30. Baranowska Körberg I., *et al.* “A simple repeat polymorphism in the MITF-M promoter is a key regulator of white spotting in dogs”. *PLoS One* 9.8 (2014): e104363.

31. Negro S., *et al.* “Association analysis of KIT, MITF, and PAX3 variants with white markings in Spanish horses”. *Animal Genetics* 48.3 (2017): 349-352.

32. Hofstetter S., *et al.* “A non-coding regulatory variant in the 5′-region of the MITF gene is associated with white-spotted coat in Brown Swiss cattle”. *Animal genetics* 50.1 (2019): 27-32.

33. Gui L S., *et al.* “Detection of polymorphisms in the promoter of bovine SIRT1 gene and their effects on intramuscular fat content in Chinese indigenous cattle”. *Gene* 700 (2019): 47-51.

34. Jaglan, K., et al. Genomic clues of association between clinical mastitis and SNPs identified by ddRAD sequencing in Murrah buffaloes. Animal Biotechnology (2023):1-9.

35. GEORGE, L., et al. Genetic improvement of economic traits in Murrah buffalo using significant SNPs from genome wide association study (2023).

36. Kumar, H., et. al. A review on epigenetics: Manifestations, modifications, methods & challenges. Journal of Entomology and Zoology studies, 8(4) (2020): 01-06.

37. Murugesan, K. D., et. al. Profiling and integrated analysis of whole-transcriptome changes in uterine caruncles of pregnant and non-pregnant buffaloes. Genomics, 113(4) (2021): 2338-2349.

38. Kanaka, K. K., et. al. On the concepts and measures of diversity in the genomics era. Current Plant Biology, 33 (2023): 100278.

39. Nidhi Sukhija., et al. “Epitope tailored vaccine construct for goat milk allergy". Acta Scientific Veterinary Sciences 5.3 (2023): 23-30.

40. Kanaka, K. K., et. al. Cloning, characterisation and expression of the SERPINB14 gene, and association of promoter polymorphisms with egg quality traits in layer chicken. British Poultry Science, 62(6) (2021): 783-794.

41. Kanaka KK.,et al. “Network analysis of beta casomorphin-7 revealed genes and pathways associated with human diseases". Acta Scientific Veterinary Sciences 5.3 (2023): 31-46