SUMMARY & CONCLUSION

Milk production is a quantitative trait as it involves several genes. Individual quantitative trait loci (QTL) or selection of linked genetic markers may be used to increase the rate of genetic gain through the use of DNA-based genetic markers or Quantitative trait markers (QTM). The development of DNA-level genetic markers or Quantitative trait Markers (QTM) would revolutionize and make detection of QTLs routine in livestock breeding by commercial farmers. This study aims at determining the polymorphisms in the lactation pathway genes in cattle and its association with milk performance traits. This is achieved by in-silico identification of the lactation pathway genes in cattle. Further, these genes were analyzed for their expression levels by qPCR in the five major cattle breeds in the country. Subsequently, the functional domains were identified.

Milk samples were collected from the major milk producing regions of the country. A total of 104 samples were collected which belonged to the five major cattle breeds in the country namely Sahiwal (24), Tharparkar (15), Jersy (20), Holstein Friesian (25) and Hariana (20). From each animal, 100 ml of milk was collected. Somatic cells were isolated and RNA was extracted. RNA was then converted to cDNA. An extensive review of literature was carried out to determine the significant genes that could be used as markers of milk production in cattle. Gene expression was carried out using primers and probes for the shortlisted genes.

The cattle breeds Sahiwal, Tharparkar, Jersey, Holstein Friesian and Hariyana, form the major and most significant milk producing breeds in India, especially in the states
of Delhi, Haryana and Punjab. The genes CEL, MFGE8, TGF-β and Lactoferrin do not directly affect the milk productivity, but are indispensible for their significant roles. CEL showed a ~67 fold increase in expression in Holeisten Fries. The protein helps in an efficient absorption in neonates. Significance of the gene is also evident from the fact that it is expressed in abundance in human milk. Milk fat globule–EGF factor 8 (MFGE8), a milk mucin associated protein provides protection against enteric pathogens. Its expression was ~3 fold in Jersey as compared to Hariana, and was the highest MFGE-8 gene expression observed among all the breeds. It forms an important component of Milk fat globule. This was followed by ~1.2 fold in Holeisten Fries. TGF-β (Transforming Growth Factor-β) is involved in development of an immature GI tract as in newborns, and is also involved in development of the immune system. A ~1.5 fold increase in TGF-β gene expression was observed in Holeisten Fries.

Lactoferrin, another immune defensive protein, is found in cow milk and human breast milk. Lactoferrin levels are seven times higher in the colostrum as compared to the regular breast milk. Lactoferrin can provide protection against several bacterial and viral infections. Holeisten Fries exhibited a ~2 fold higher lactoferrin gene expression, the highest among all the breeds.

In terms of the total lactation yield, Holstein Friesan is reported to exhibit the highest yield of 4295 kg, which is followed by 2727 kg by Jersey, 2334 kg by Tarparkar, 2266 kg by Sahiwal and the least is given by Hariana, 1500 kg.
CSN (Casein) and LALBA (Lactalbumin) refer to milk producing genes exclusive to the mammary gland, and also form the two highest expressing genes of human mammary epithelium during lactation. Significance of CSN is evident from mice experiments which report abnormal lactation and milk composition in knockout experiments. Human milk CSN is low, but mice studies reveal a high CSN. CSN2 (Casein beta), and CSN-3 (Casein-kappa) are majorly involved in formation of casein, that forms 80% of milk. Casein is important as its increased expression increase the nutritional value as well as processing properties. Thus both form important markers to decide on milk quality. CSN-2 expression has been related to milk calcium levels, whereas CSN-3 is linked with properties such as heat stability and cheese making properties. Sahiwal exhibited a ~167 fold increased expression of CSN2, followed by ~30 fold in Tharparkar, and ~17 fold in Holeisten Fries. This probably is the cause for a higher SNF% in Sahiwal (9.2%). CSN3 is expressed about ~20 fold in Jersey, and ~17 fold in Holeisten Fries. This may be involved in increasing the SNF% in Jersey (15%). LALBA is involved in milk protein and lactose synthesis, and is highest (8 fold) in Sahiwal, causing an increased total lactation yield, as well as SNF%.

FDFT1 (farnesyl-diphosphate farnesyltransferase1), a membrane linked enzyme involved in biosynthesis of cholesterol, and Lipoprotein lipase (LPL is involved in fat synthesis, and regulation of triglycerides. These are expressed ~99 fold and ~1.8 fold in H. Fries, exhibiting the highest expression among the breeds.

XDH (xanthine dehydrogenase) is involved in milk fat synthesis and lipid droplet formation. Our study reports a ~3 fold higher expression in Jersey, followed by 1.7
fold expression in Tharparkar, and \(~0.13\) fold expression. The highest fat\% is reported from Jersey (5.3\%), Tharparkar (5.3\%) and Sahiwal (5.1\%).

Prolactin (PRL) is involved in synthesis of milk proteins, lactose, lipids and all major milk components. It plays significant role in development, differentiation and regulation of mammary gland. Holeisten Fries, exhibits highest expression of this gene, and is known for its high milk yield, and an average Fat\% though the SNF\% is low.

Holeisten Fries and Jersy, are observed to cluster closely, along with Hariana, and the breeds Sahiwal and Tharpakar cluster together. This clustering is on the basis of the expression levels of the 12 marker genes identified as significant.

Studying functional domains/markers is of relevance in understanding the domains which are indispensible for the gene’s milk producing capacity or other milk quality associated attributes. Prolactin (PRL) is a significant contributor to major milk components including milk proteins, lactose and lipids. It is also involved in development, differentiation and regulation of the mammary gland. Previous studies have proposed several genes as potential candidates for dairy traits, and prolactin gene (PRL) seems to be the most promising since it plays a crucial role in mammary gland development, initiation and maintenance of lactation, and expression of milk protein genes. Prolactin stimulates mammary development and promotes the formation and action of the corpus luteum during the female reproductive cycle in mammals. Allelic variations in the structural or regulatory sequences of PRL and variations in genes upstream and downstream to PRL in lactation pathway are of interest because of their possible direct or indirect effect on milk production.
Holeisten Fries, the cattle breed known for its high milk yield, average fat percentage, and low SNF percentage, raised in the Northern part of India exhibited highest expression of the prolactin gene, and is known for its high milk yield.

Structure is primarily responsible for the functional features displayed by a protein. The conservation of tertiary structure clearly implies the functional significance of a motif. The functional motif remains conserved during the process of evolution, and experience a selection pressure, which preserves the functional motif. This study involves identification of functional domains in the prolactin gene in four mammals namely cow (Bos taurus), chimpanzee (Pan troglodytes), sheep (Ovis Aries) and human (Homo sapiens).

Prolactin amino acid sequences from chimpanzee, sheep, cow and human were obtained from genbank (NCBI). We created protein structures using the Swift Modeler software [6]. We created 10 models for each sequence and selected the one with lowest DOPE score. The selected models were analyzed using VMD software version1.9 (Visual Molecular Dynamics). Multiseq (Multiple Sequence Alignment) was used for comparing structures and sequences. Multiseq uses STAMP for structural alignment and Clustal W for sequence based alignment. The aligned models were colored on the basis of their Qres and entropy values for each residue. Phylogenetic analysis was carried out with sequence based (Clustal W) and structure based alignments (STAMP). Amino acid composition was analyzed using the ExPASy portal in Chimpanzee, Sheep, Human and Cow. Subsequently, we also calculated the molecular weight and theoretical PI for all the four proteins. Functional
motif was determined by the PROSITE tool in ExPASY. The functional motif is further confirmed by analysis of the 3D structures, using structural similarity determinants such as RMSD and Qres, using VMD software. Structure is primarily responsible for the functional features displayed by a protein. Therefore, structural similarity can be observed in proteins which have a common function. This is because proteins with common functional properties will have common interaction with molecular factors, receptor, etc. The conservation of tertiary structure clearly implies the functional significance of a motif. The functional motif remains conserved during the process of evolution, and experience a selection pressure, which preserves the functional motif. Prolactin amino acid sequences of chimpanzees, sheep, cow and human were retrieved from genbank (NCBI). These were used to construct protein models using Swift Modeller software. The sequences and structures were aligned using Visual Molecular Dynamics (VMD) software. The protein models were analysed. Bioinformatics tools form a practical and inexpensive solution for discovering functional motifs in a protein, unlike the labour intensive and costly experimental approaches. Most approaches that are involved in prediction and confirmation of functional motifs involve examining conserved structural motifs, and these are involved in the protein function.

In conclusion, Holeisten Fries forms the breed with the highest milk production yield, and contains an average fat percentage of 3%. But, in terms of SNF %, the Jersey is the leading breed with 15%. The markers analyzed are of significance in milk yield and milk quality especially CSN, LALBA, XDH, PRL, MFGE8, LTF, LPL and FDFT1. These give a good idea about the milk quality and quantity, and should be considered while deciding on the quality of any breed. Additionally, CEL, MFGE8,
TGF-β and Lactoferrin genes do not directly affect the milk productivity, but are indispensible for their significant roles, as discussed above. The structural and sequence based analysis confirmed the presence of two functional motifs in the prolactin gene viz. ‘CHTSSLPTPEDKEQAQQTHHEVLMLSLILGLLRSW’ and ‘CLRRDSSKIDTYLKLNNC’. CSN2 showed the presence of single functional motif viz. CLVALALA. CEL showed the presence of two functional motifs viz. EDCLYLNIWVP and FGGPDNITLFGESAG. FDFT1 showed the presence of two functional motifs viz. YCHYVAGLVGIGLSRL and MGLFLQKTNINDQREGREFWP. LALBA showed the presence of one functional motif CNISCDKFLDDDTDDIMC. The proteins in cow, sheep and chimpanzees exhibit a similar protein structure which is evident from the RMSD and Qres values, but interestingly, the functional motifs remain majorly conserved.