# A COMPLEX INTERPLAY OF GENES AND BIOLOGICAL PROCESSES UNDERLYING MILK PRODUCTION: REVELATIONS THROUGH GENOME-WIDE ASSOCIATIONS IN SAHIWAL CATTLE

Aneet Kour* and S.M. Deb

ICAR-National Dairy Research Institute, Karnal, Haryana-132001, India

*Present Address: ICAR-National Research Centre on Yak, Dirang-790101, Arunachal Pradesh

## ABSTRACT

The genetic uniqueness with respect to the performance and sustainability in *Bos indicus* breeds has been a subject of curiosity for researchers worldwide. Sahiwal is one of these breeds native to Indian-subcontinent which has traversed far and wide around the world due to its superior features. However, the genomic regions and biological processes underlying its production performance have not been studied extensively. We performed a genome-wide association study (GWAS) in the *Bos indicus* breed Sahiwal to elucidate the genetic architecture of complex milk production traits. Estimated Breeding Values (EBVs) for all eight traits i.e. First lactation 305 days or less milk yield (FL305MY), First lactation average fat percentage (FLAFP), First lactation average protein percentage (FLAPP), First lactation average solid non-fat percentage (FLASNFP), First lactation length (FLL), First lactation average fat yield (FLAFY), First lactation average protein yield (FLAPY), First lactation average solid non-fat yield (FLASNFY). were used as pseudo-phenotype to determine genome-wide associations. In all, 22 SNPs were significantly influencing the traits with the maximum identified for FLL while 3 SNPs were found to be pleiotropic for different traits. On further mapping of these SNPs to the genomic locations, candidate genes affecting the traits were revealed which were associated with diverse range of functions. Further, candidate SNP enrichment was performed to highlight the enriched gene ontology terms and these results also presented a somewhat similar picture with respect to traits being influenced by biological processes underlying a wide spectrum of functions. The findings of this study embolden the point that a complex interplay of biological mechanisms is influencing milk production and further research on this aspect may improvise selection for milk production traits in Sahiwal cattle.

**Keywords:** GWAS, Sahiwal, Milk production, Biological processes

***Corresponding Author**:aneetk25@gmail.com

## INTRODUCTION

Genome-wide Association study (popularly known as GWAS), initially conceptualized and implemented for identifying human disease variants, has energized the field of animal genomics as an improvised approach to Quantitative Trait Loci (QTL) mapping by complementing, if not completely obliterating the traditional candidate gene analysis (Wilkening et al., 2009). Since the major economic traits in livestock being quantitative in nature are governed by polygenes and exhibit low to medium heritability, the importance of additive SNP effects for selecting superior animals could not be understated. Therefore, a GWAS is considered to be crucial preliminary investigation for undertaking genomic prediction and ultimately, genomic selection for complex traits in livestock (Zhang et al., 2014). With a view to unravel the underlying genetic architecture, GWA studies in cattle have encompassed almost all the performance, functional, conformation and even some experimental traits generated by in-depth phenotyping.

Sahiwal is a predominant breed of Indian subcontinent constituting around 12% and 21% of the descript cattle population respectively, in India and Pakistan, its two major breeding tracts (Breed survey 2013; Khan et al. 2008). Owing to the optimum performance along with superior adaptability in tropical conditions, the breed has traversed continents as far as Africa and Australia (Illatsia et al. 2011; Colditz and Brown 1976). The breed is highly preferred by Indian farmers due to its economic sustainability with respect to milk production traits (Joshi et al. 2001), which play a major role in deciding market price dynamics of animal as well as milk in the country. In India, milk pricing is based on fat and solids-non-fat (SNF) content which necessitates the continued improvement in the breed for milk production traits while maintaining optimum fertility and health. However, there is no reported literature on the underlying genetic architecture of milk production traits in Sahiwal cattle of India till date. This prompted us to

attempt a Genome-wide Association Study (GWAS) related to economically important milk production traits in Sahiwal cattle.

## METHODOLOGY

### Sample collection and data recording

Blood samples were collected from 242 adult female Sahiwal animals in two different experimental herds *viz.* National Dairy Research Institute (NDRI), Karnal (213) and Lala Lajpat Rai University of Veterinary and Animal Sciences, Hisar (29) (IAEC Approval number: 44-IAEC-19-15). Standard Phenol-chloroform procedure (Sambrook and Russel, 2006) was used for DNA isolation. Genotyping was performed for 54K SNPs using INDUSCHIP designed by NDDB, Anand Gujarat, India (https://www.nddb.coop/ services/ animal breeding/ geneticimprovement/ genomic). The quality check was done in PLINK1.9 (Chang et al., 2015) and its criteria included sample and SNP call rate <95%, H-WE<$1X10^{-6}$ and minor allele frequency<1%.

The recorded phenotypic traits comprised of: First lactation 305 days or less milk yield (FL305MY), First lactation average fat percentage (FLAFP), First lactation average protein percentage (FLAPP), First lactation average solid non-fat percentage (FLASNFP), First lactation length (FLL), First lactation average fat yield (FLAFY), First lactation average protein yield (FLAPY), First lactation average solid non-fat yield (FLASNFY). The phenotypic as well as pedigree data was edited by removing outliers – FL305MY less than 500kg and FLL less than 100 days were set as exclusion criteria. Similarly, animals having abnormal values for other traits were excluded. Before proceeding for further analysis, normalization of data was done to make it statistically appropriate.

### Genetic parameters estimation

MCMCglmm package in R 3.6.2 (Hadfield, 2010) was used to obtain posterior variance estimates for the milk production traits. Subsequently, Estimated Breeding Values (EBVs) for all the traits were calculated by using animal model in WOMBAT (Meyer, 2007). Relevant non-genetic factors like age group, season and period of first calving were included in the animal model. Herd was considered as an influencing factor for FL305MY as it was the only milk phenotype available for animals from Hisar herd. BLUP Animal model for different traits is shown below:

FL305MY

$Y_{ijklmn} = \mu + season_i + period_j + age_k + herd_l + animal_m + e_{ijklmn}$

Where $Y_{ijklmn}$ is the vector of phenotype for mth individual in ith season, jth period, kth age group and lth herd, $\mu$ = overall mean, $season_i$ = ith season of calving, $period_j$ = jth period of calving, $age_k$ = kth age group at first calving, $herd_l$ = lth herd in which animal was raised, $animal_m$= random effect of mth individual, $e_{ijklmn}$ = random residual in the ith season, jth period and kth age group of mth individual from lth herd

All Milk production and composition traits (except FL305MY)

$Y_{ijklm} = \mu + season_i + period_j + age_k + animal_l + e_{ijklm}$

Where $Y_{ijklm}$ is the vector of phenotype for lth individual in ith season, jth period and kth age group, $\mu$ = overall mean, $season_i$ = ith season of calving, $period_j$ = jth period of calving, $age_k$ = kth age group at first calving, $animal_l$= random effect of lth individual, $e_{ijklm}$ = random residual in ith season, jth period and kth age group for lth individual

### GWAS and downstream studies

Genome-wide associations were performed using a single marker model approach in GAPIT package in R (Lipka et al., 2010). Two thresholds *viz.* $p \leq 10^{-5}$ and $10^{-5} < p \leq 10^{-4}$ were used to depict highly significant and significant associations respectively. Significant SNP associations were plotted using manhattanly package in R (https://github.com/sahirbhatnagar/manhattanly/). Further functional annotation was performed using candidate SNP enrichment analysis in SNP2GO package (Szkiba et al., 2014) with FDR set at 0.05.

## RESULTS

### Phenotypic analysis

The posterior mean heritability of FLASNFP was the highest and was the only one falling in mid-heritability range amongst all the traits. All other traits had posterior heritability estimates falling in the lower range. As far as HPD interval of heritability estimates were concerned, all the traits were falling in the low to medium range except FL305MY, FLASNFP and FLAFY for which the interval range was between low to high. The posterior variance and heritability estimates for the traits have been presented in Table 1. Considering the posterior variance and heritability estimates, EBVs were estimated for all the traits.

**Table 1** Posterior means variance and heritability estimates for milk production and composition traits

| TRAIT | POSTERIOR MEANS | | | |
|---|---|---|---|---|
| | ADDITIVE VARIANCE | RESIDUAL VARIANCE | HERITABILITY | HPD INTERVAL OF HERITABILITY |
| **MILK PRODUCTION AND COMPOSITION TRAITS** | | | | |
| FL305MY | 76627 | 344636 | 0.179±0.10 | 0.025-0.403 |
| FLAFP | 0.018 | 0.092 | 0.164±0.09 | 0.023-0.357 |
| FLAPP | 0.002 | 0.011 | 0.152±0.08 | 0.021-0.313 |
| FLASNFP | 0.001 | 0.006 | 0.214±0.10 | 0.040-0.430 |
| FLL | 620.1 | 3498 | 0.149±0.08 | 0.024-0.326 |
| FLAFY | 23.57 | 107 | 0.177±0.12 | 0.019-0.427 |
| FLAPY | 6.412 | 39.1 | 0.139±0.08 | 0.021-0.320 |
| FLASNFY | 42.86 | 252.7 | 0.143±0.09 | 0.018-0.322 |

**Genotypic data analysis**

43,483 SNPs passed the quality check filters and were queried for association with milk production traits. In all, 22 SNPs (1 SNP at $p \leq 10^{-5}$ and 21 SNPs at $10^{-5} < p \leq 10^{-4}$) were found significantly associated with various traits. The number of significant SNPs were highest for the trait FLL (36.36%) whereas none of the SNPs were found to be significant for FLAPP and FLAPY traits.Three SNPs *viz.* rs110333764, rs134503291 and rs109895198) were pleiotropic for different groups of traits- FL305MY, FLL & FLASNFY, FLAFP & FLASNFP and FLL & FLASNFY respectively. The complete list of significant SNPs along with their distribution along chromosomes and nearby putative genes (within ± 60 kb) has been presented in Table 2 and Manhattan plots for significant SNPs have been shown in Figure 1.

**Table 2** Significant SNPs found to be having association ($p \leq 10^{-4}$) with various milk production traits

| TRAIT | SNP Name | BTA | POSITION (bp) | p-VALUE | MAF | GENE ID | DISTANCE FROM SNP (bp) | LOCATIONa |
|---|---|---|---|---|---|---|---|---|
| FL305MY | rs110976676 | 18 | 9864773 | 4.74E-05 | 0.286 | CDH13 | Within | |
| | rs110333764 | 4 | 27292592 | 7.66E-05 | 0.144 | HDAC9 | Within | |
| | rs41647618 | 26 | 3578544 | 8.49E-05 | 0.443 | - | | |
| | rs110638352 | 16 | 79180948 | 9.12E-05 | 0.395 | NEK7 | 44554 | Upstream |
| FLAFP | rs109670479 | 11 | 15465224 | 6.73E-05 | 0.462 | LTBP1 | - | Downstream |
| | rs135507832 | 11 | 14089816 | 7.82E-05 | 0.447 | - | 28079 | |
| | rs134503291 | 17 | 44558725 | 8.79E-05 | 0.483 | GUCY1B1 | - | |
| | | | | | | GUCY1A1 | Within | |
| | rs134266584 | 19 | 18743422 | 1.00E-04 | 0.433 | | 41783 | Downstream |
| **FLAPP** | | | | | | | | |
| FLASNFP | rs134503291 | 17 | 44558725 | 2.53E-05 | 0.483 | GUCY1B1 | Within | |
| | | | | | | GUCY1A1 | 41783 | Downstream |
| | rs110671163 | 5 | 73755321 | 7.03E-05 | 0.441 | ISX | 25126 | Upstream |
| | rs41665148 | 14 | 3660751 | 1.00E-04 | 0.021 | GPR20 | 20034 | Upstream |
| | | | | | | SLC45A4 | 51085 | Downstream |
| FLL | rs135181323 | 22 | 27935111 | 1.03E-05 | 0.288 | - | | |
| | rs135923322 | 13 | 60209596 | 1.74E-05 | 0.445 | MC3R | 23493 | Upstream |
| | rs42198780 | 25 | 30986679 | 3.50E-05 | 0.403 | SIRPD | Within | |
| | rs110333764 | 4 | 27292592 | 5.80E-05 | 0.144 | NSFL1C | 35292 | Downstream |
| | rs110917829 | 5 | 100800335 | 6.16E-05 | 0.293 | - | | |
| | rs109895198 | 4 | 22867471 | 8.06E-05 | 0.437 | HDAC9 | Within | |
| | rs133597786 | 13 | 61156374 | 8.83E-05 | 0.354 | CD69 | 52663 | Upstream |
| | | | | | | CLECL1 | 21175 | Upstream |
| | | | | | | DGKB | Within | |
| | | | | | | CSNK2A1 | 49876 | Upstream |

| TRAIT | SNP Name | BTA | POSITION (bp) | p-VALUE | MAF | GENE ID | DISTANCE FROM SNP (bp) | LOCATIONa |
|---|---|---|---|---|---|---|---|---|
| | | | | | | TBC1D20 | 13927 | Upstream |
| | | | | | | RBCK1 | Within | |
| | | | | | | TRIB3 | 19843 | Downstream |
| | | | | | | NRSN2 | 55287 | Downstream |
| | rs137402690 | 6 | 79260207 | 9.07E-05 | 0.489 | ADGRL3 | Within | |
| FLAFY | rs41949203 | 20 | 41135365 | 2.18E-05 | 0.351 | SUB1 | Within | |
| | rs41597320 | 19 | 24876192 | 2.43E-05 | 0.366 | SPATA22 | 51897 | Upstream |
| | | | | | | ASPA | 29084 | Upstream |
| | | | | | | TRPV3 | Within | |
| | | | | | | TRPV1 | 14558 | Downstream |
| | | | | | | SHPK | 57557 | Downstream |
| | rs41949210 | 20 | 41124297 | 1.00E-04 | 0.218 | SUB1 | Within | |
| FLAPY | | | | | | | | |
| FLASNFY | rs121919184 | 28 | 43882171 | 3.50E-05 | 0.155 | C28H10orf7 | 48256 | Downstream |
| | rs110333764 | 4 | 27292592 | 4.25E-05 | 0.144 | 1 | Within | |
| | rs109298840 | 13 | 75273574 | 6.81E-05 | 0.222 | HDAC9 | 42464 | Upstream |
| | | | | | | SPINT4 | Within | |
| | | | | | | WFDC3 | 10957 | Downstream |
| | | | | | | DNTTIP1 | 33182 | Downstream |
| | | | | | | UBE2C | 42848 | Downstream |
| | | | | | | TNNC2 | 53864 | Downstream |
| | rs109895198 | 4 | 22867471 | 7.90E-05 | 0.437 | ACOT8 | Within | |
| | | | | | | DGKB | | |

**First Lactation 305 Days or less Milk Yield (FL305MY)**

Four SNPs were found to be significantly associated with FL305MY at $10^{-5}<p\leq10^{-4}$ whereas no association was found to be significant at $p\leq10^{-5}$. SNPs rs110976676 and rs110333764 were mapped to locations within the protein-coding genes- CDH13 and HDAC9 genes respectively whereas NEK7 gene was 44,554 bp upstream from the SNP rs110638352.

**First Lactation Average Fat percentage (FLAFP)**

4 SNPs were having significant association with FLAFP (10-5<p≤10-4) and were mapped to 3 protein coding genes –LTBP1, GUCY1B1 and GUCY1A1. SNP rs134503291 was present within GUCY1B1 (Guanylate cyclase soluble subunit beta-1) gene and was mapped nearly 40kb upstream of GUCY1A1 (Guanylate cyclase soluble subunit alpha-1) gene.

**First Lactation Average Protein percentage (FLAPP)**

None of the SNPs could cross the significance threshold for FLAPP.

**First Lactation Average SNF Percentage (FLASNFP)**

3 SNPs were significant for FLASNFP including one pleiotropic SNP rs134503291 which was also found significant for FLAFP and mapped to GUCY1B1 and GUCY1A1 genes. SNP rs110671163 was mapped closer to ISX gene whereas genes GPR20 and SLC45A4 were present in the vicinity of SNP rs41665148.

**First Lactation Length (FLL)**

8 SNPs significant for FLL were mapped to 13 protein-coding genes including MC3R, SIRPD, NSFL1C, HDAC9, CD69, CLECL1, DGKB, CSNK2A1, TBC1D20, RBCK1, TRIB3, NRSN2 and ADGRL3. SNP rs135181323 found to be highly significant ($p\leq10^{-5}$) could not be mapped to any nearby gene within ±60 kb region.

**First Lactation Average Fat Yield (FLAFY)**

SNP rs41949203 on chromosome 20 was mapped to SUB1 (SUB1 regulator of transcription, also known as positive cofactor 4) gene which protects DNA against oxidative damage and coordinates cellular response to DNA strand breaks (Yu et al., 2016). The other SNP rs41597320 on chromosome 19 was mapped to four other genes *viz.* SPATA22, ASPA, TRPV3, TRPV1, SHPK besides SUB1.

**First Lactation Average Protein yield (FLAPY)**

No significant SNP was identified for the trait FLAPY.

**First Lactation Average SNF yield (FLASNFY)**

Four SNPs were found significant for FLASNFY out of which SNP rs110333764 was also pleiotropic for FL305MY and FLL while rs109895198 was pleiotropic with respect to FLL as well. These pleiotropic SNPs were mapped to HDAC9 and DGKB genes, respectively while the other two SNPs - rs121919184 and rs109298840 were mapped to C28H10orf71 and SPINT4, WFDC3, DNTTIP1, UBE2C, TNNC2 and ACOT8 genes, respectively.

**Candidate SNP Enrichment**

Candidate SNP Enrichment analysis highlighted that GO terms for biological processes were enriched only for the traits FL305MY, FLL and FLAFY wherein two GO terms were enriched each for FL305MY and FLL whereas a maximum of thirteen GO terms were enriched for FLAFY. Interestingly, enrichment analysis revealed a complex interplay of biological pathways and mechanisms underlying milk production. All the enriched GO terms related to various traits have been elucidated in Table 3.

**Table 3:** Significant Gene Ontology (GO) terms related to biological processes enriched for various milk production traits in Sahiwal cattle

| TRAIT | GO TERM | FDR | GO TERM DESCRIPTION |
|---|---|---|---|
| FL305MY | GO:0030859 | 0.012 | Polarised epithelial cell differentiation |
| | GO:0001738 | 0.012 | Morphogenesis of polarized epithelium |
| FLL | GO:0042220 | 0.049 | Response to Cocaine |
| | GO:0042445 | 0.049 | Hormone metabolic process |
| FLAFY | GO:0032330 | 0.004 | Regulation of chondrocyte differentiation |
| | GO:0035729 | 0.001 | Cellular Response to Hepatocyte Growth Factor stimulus |
| | GO:0048012 | 0.0001 | Hepatocyte growth factor receptor signaling pathway |
| | GO:2000177 | 0.007 | Regulation of neural precursor cell proliferation |
| | GO:0006516 | 0.001 | Glycoprotein catabolic process |
| | GO:0032331 | 4.143e-05 | Negative regulation of chondrocyte differentiation |
| | GO:0061037 | 5.371e-05 | Negative regulation of cartilage development |
| | GO:0061154 | 4.143e-05 | Endothelial tube morphogenesis |
| | GO:0006029 | 0.015 | Proteoglycan metabolic process |
| | GO:0098693 | 0.047 | Regulation of synaptic vesicle cycle |
| | GO:0090288 | 0.001 | Negative regulation of cellular response to growth factor stimulus |
| | GO:0061035 | 0.009 | Regulation of cartilage development |
| | GO:0003158 | 0.007 | Endothelium development |

**Manhattan Plot for FL305MY** **Manhattan Plot for FLAFP**

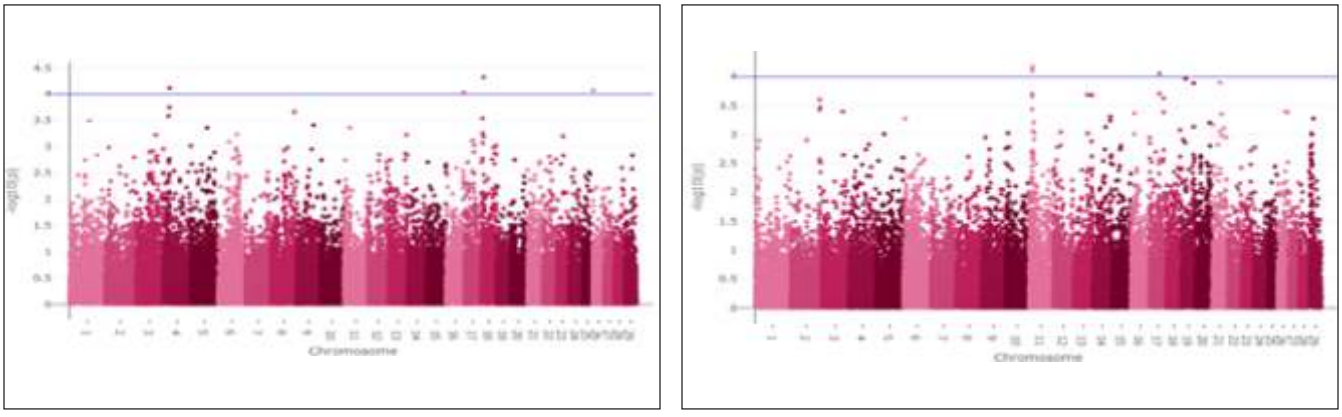
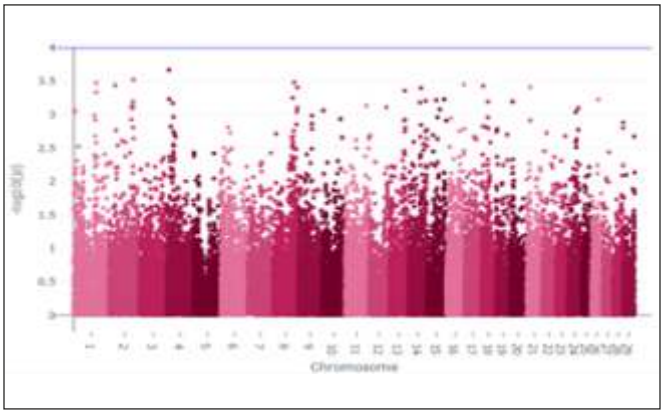
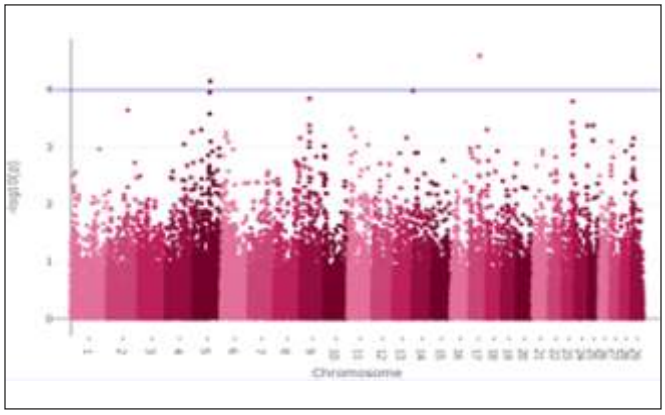


**Figure 1: Manhattan plots corresponding to various milk production traits in Sahiwal cattle**

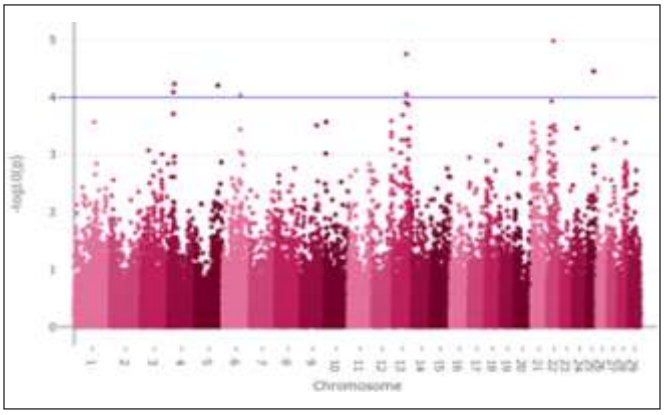
**Manhattan Plot for FLAPP**



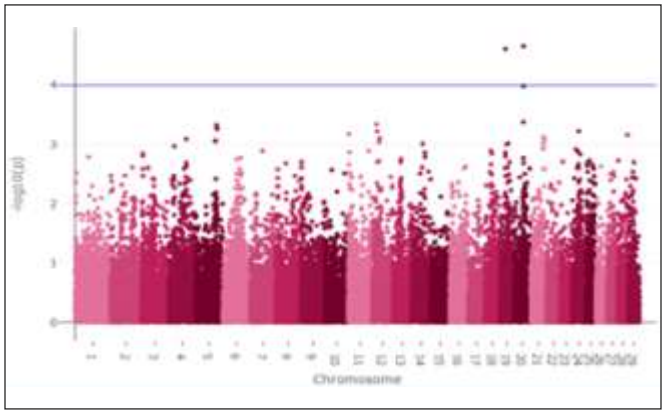
**Manhattan Plot for FLASNFP**



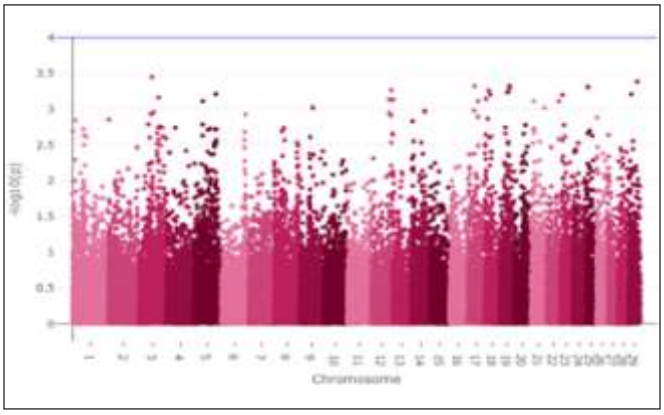
**Manhattan Plot for FLL**



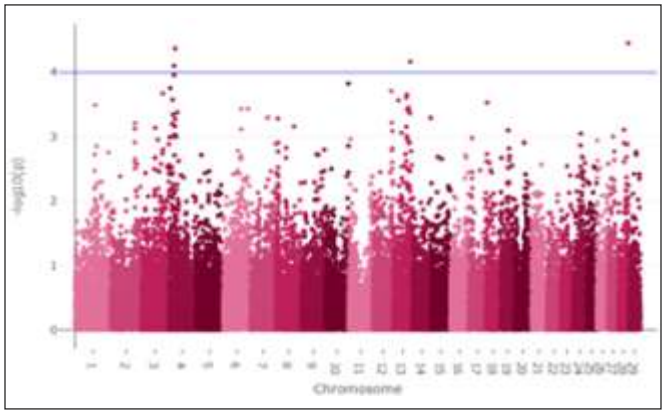
**Manhattan Plot for FLAFY**



**Manhattan Plot for FLAPY**



**Manhattan Plot for FLASNFY**



## DISCUSSION

Generally, GWAS is reported to be highly powered for detecting variants with high effect i.e. traits with higher heritability and is underpowered for traits with moderate effect (~0.03-0.3 phenotypic SD) (Caballero et al., 2015). In our study, all the traits were falling in the low to moderate range, which is again in consonance with the previously reported findings (Gebreyohannes et al., 2013; Verma et al., 2017). Hence, very few SNPs were identified in our study with none found significant for FLAPP and FLAPY. The lower heritability of these traits may be contributing to smaller effect sizes for the variants which remain 'hidden' and subsequently, require a larger sample size to be unveiled.

Mapping of SNPs to their genomic locations revealed certain protein-coding genes and the ones with huge importance were the pleiotropic genes like GUCY1B1 and GUCY1A1 for FLAFP and FLASNFP and HDAC9

gene for FL305MY and FLASNFY. These results are in agreement with Nayeri et al. (2016) who also found significant overlaps within production traits. The reason may be that since almost all the milk production traits are interlinked, these genes may be performing a generalized or a specific role in the common pathways responsible for phenotypic expression of these traits. GUCY1B1 and GUCY1A1 genes possess soluble guanylate cyclase activity which has been found to protect against weight gain in obesity and proven helpful for regulation of energy expenditure and fighting obesity in humans (Hoffman et al., 2015). HDAC9 (Histone Deacetylase 9) gene is responsible for transcriptional modulation and cell cycle progression, thus possessing a control over all biological processes, including milk production. It has also been implicated as a critical factor regulating adipogenic differentiation with its overexpression resulting in ectopic lipid deposition in the liver (Chatterjee et al., 2018).

The picture portrayed by the putative candidate genes was very diverse with respect to their functions. Genes like *GPR20* and *SLC45A4* are important candidate genes for milk production traits whereas *LTBP1* and *TRIB3* are involved in regulating fatty acid biosynthesis in milk (Rogers et al. 1996; Cui et al. 2014). GPR20 (G-protein coupled receptor 20) gene involved in regulating intracellular cAMP levels (Hase et al., 2008) has also been found to be associated with all five major milk production traits including milk yield, milk fat percentage, milk fat yield, milk protein percentage and milk protein yield. SLC45A4 (Solute Carrier Family 45 Member 4) gene is involved in transport of sugars like glucose, fructose and sucrose ( Bartölke et al., 2014) and has also been evidenced to be regulating all the five major milk production traits. LTBP1 (Latent-Transforming Growth Factor Beta-binding Protein 1) gene mapped to a region surrounding the SNP rs109670479, codes for one of the most predominant growth factors in milk and is involved in the synthesis of unsaturated fatty acids in milk (Rogers et al., 1996; Lung et al., 2019). In beef cattle, SNP having significant associations with residual intake gain and residual average daily gain has been mapped to a location closer to LTBP1 gene (Serao et al., 2013). TRIB3 (Tribbles homolog 3) gene is a protein kinase negatively regulating fatty acid biosynthesis which has been proposed as a promising candidate gene affecting milk protein and fat percentage in cattle and intramuscular fatty acid composition in pigs (Cui et al., 2014; Muñoz et al., 2013). Genes like *CSNK2A1, ADGRL3* and *CDH13* are known to influence milk production and its quality. CSNK2A1 (Casein Kinase 2 subunit alpha) gene is involved in casein protein phosphorylation while ADGRL3 (Adhesion G protein coupled receptor L3) gene is believed to influence milk yield and milk protein yield. CDH13 (Cadherin 13) gene is involved in cell-cell adhesion mediated by calcium as well as cadherins, thus facilitating the clustering of alveolar epithelial cells for milk production (Gorewit, 1988). It is also known to have been involved in regulating dry matter intake and cholesterol content in milk (Do et al., 2018).

On the other hand, genes like *MC3R, TNNC2* and *ACOT8* are responsible for body weight and growth of animals ((Demidowich et al., 2017; Skorczyk et al., 2011; Rosiane et al., 2019; Salleh et al., 2018) whereas *CD69* and *CLECL1* have major immunoregulatory roles (Zieglar et al., 1994; Wallace et al., 2012). Also, genes like *TBC1D20, SPATA22* and *UBE2C* are putative candidate genes regulating reproduction and hence, fertility traits in cows and bulls (Nicolini et al., 2018; Salle et al., 2012; León et al., 2019). So, a close underlying link between various economic traits is perceptible and change in phenotypic expression of one trait may clearly be having physiological ramifications on all the traits. Interestingly, the two reported QTNs related to milk production traits in cattle in DGAT1 and ABCG2 genes (Grisart et al., 2002; Cohen-Zinder et al., 2005) were not present within or in the vicinity of any of the significant SNPs. This further stresses on the fact that every population has its own unique SNPs, genomic regions and QTNs related to various traits, as validated in few more studies (Yodklaew et al., 2017).

Most of the GO terms had direct or indirect association with the studied traits while some of the biological processes were found to be acting in conjunction to perform a particular action. GO terms for related biological processes like polarized epithelial cell differentiation (GO:0030859) and morphogenesis of polarized epithelium (GO:0001738) for FL305MY, hepatocyte growth factor regulation (GO:0035729) and endothelium development and morphogenesis (GO:0061154, GO:0003158) for FLAFY, chondrocyte differentiation and cartilage development (GO:0032330, GO:0032331, GO:0061037, GO:0061035) for FLAFY embody this fact.

The polarized epithelium has crucial underpinnings in milk synthesis and secretion into the lumen of mammary gland during lactation (Anderson et al., 2007). Massive alveolar cell differentiation occurs from mid-pregnancy as the developing mammary alveoli cleave resulting in

polarisation of the alveolar cells. Subsequently, they form a sphere-like single layer of epithelial cells enveloping the circular lumen and connected by a small duct to the entire ductal network (Fata et al., 2004). At the time of parturition, secretory activation occurs in which preparation for active milk secretion starts - alveolar tight junctions are closed and milk and colostrum proteins move into the lumen (Oakes et al., 2006).

As far as the trait AFY is concerned, Hepatocyte growth factor in milk promotes cell proliferation and angiogenesis and is also known to facilitate morphogenesis of endothelial cells (Kobata et al., 2008), which further act as blood-milk barrier by effecting an exchange of solutes and macromolecules, thus leading to optimal milk production (Ryman et al., 2015).

Chondrocytes can differentiate into bone forming cells osteoblasts and during lactation, bone resorption occurs to divert calcium for milk production (Salari and Abdollahi, 2014). Additionally, this may exert an indirect effect on milk production by directly regulating growth and body weight of the animal.

## CONCLUSIONS

To summarise, certain novel insights pertaining to Sahiwal cattle were highlighted in this study which though, suffered from the menace of small data size, nonetheless strengthened the already known fact that milk production is a highly complex process with genes and biological processes encompassing hormonal, immunomodulatory, growth and reproductive pathways, all of which seem to be interconnected and work in synchrony to yield the desired result *viz*. the complex phenotype. Moreover, the relevance of this study is due to the reason that this type of work has never been attempted before in Sahiwal cattle. So, it opens the realm of genetic introspection into this indispensible *Bos indicus* germplasm, though further validation of the results with additional data may bring greater clarity.

## ACKNOWLEDGEMENTS

The financial support rendered by National Dairy Development Board (NDDB), Anand, Gujarat, in genotyping of samples is deeply acknowledged. The first author is grateful to Director, National Dairy Research Institute (NDRI) for providing necessary facilities. The first author is also thankful to members of the Advisory committee for their relentless support and guidance and Indian Council of Medical Research (ICMR) for providing Junior Research Fellowship (JRF).

## FUNDING

No financial help was received for conducting this study

## CONFLICT OF INTERESTS

All the authors declare that they have no conflict of interest.

## REFERENCES

Anderson SM, Rudolph MC, McManaman JL, Neville MC. Key stages in mammary gland development. Secretory activation in the mammary gland: it's not just about milk protein synthesis**!** Breast Cancer Res. 2007;9(1):204. https://doi.org/10.1186/bcr1653

Bartolke R, Heinisch JJ, Wieczorek H, Vitavska O. Proton-associated sucrose transport of mammalian solute carrier family 45: an analysis in Saccharomyces cerevisiae". *Biochem J*. 464 (2): 193–201. https://doi.org/10.1042/BJ20140572

Breed survey. Ministry of Agriculture and Farmers Welfare, Department of Animal Husbandry, Dairying and Fisheries, Government of India. 2013

Caballero A, Tenesa A, Keightley PD. The Nature of Genetic Variation for Complex Traits Revealed by GWAS and Regional Heritability Mapping Analyses. *Genetics*. 2015;201(4):1601-13.https://doi.org/10.1534/genetics.115.177220

Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7. https://doi.org/10.1186/s13742-015-0047-8

Chatterjee TK, Basford JE, Yiew KH, Stepp DW, Hui DY, Weintraub NL. Role of histone deacetylase 9 in regulating adipogenic differentiation and high fat diet-induced metabolic disease. Adipocyte.2018;3(4):333-8. https://doi.org/10.4161/adip.28814

Cohen-Zinder M, Seroussi E, Larkin DM, et al. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the

QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Res.* 2005;15(7):936-44. https://doi.org/10.4161/adip.28814

Colditz P.J., Brown G.W. Performance of Australian Friesian Sahiwal dairy cattle. In: *Proceedings of Australian Society of Animal Production,* 1976; 11:29P.

Cui X, Hou Y, Yang S, Xie Y, Zhang S, Zhang Y, et al. Transcriptional profiling of mammary gland in Holstein cows with extremely different milk protein and fat percentage using RNA sequencing. BMC Genom . 2014;15:226.

Demidowich AP, YunJun J, Yanovski JA. Polymorphisms and mutations in the melanocortin-3 receptor and their relation to human obesity. Biochim Biophys Acta. 2017;2468-76.

Do DN, Schenkel FS, Miglior F, Zhao X, Awemu EMI.Genome wide association study identifies novel potential candidate genes for bovine milk cholesterol content. Sci Rep. 2018;8:13239.

Fata, J.E., Werb, Z., Bissell, M.J. Regulation of mammary gland branching morphogenesis by the extracellular matrix and its remodeling enzymes. Breast Cancer Res. 2004; 6:1-11.

Gebreyohannes G, Koonawootrittriron S, Elzo MA, Suwanasopee T. Variance components and genetic parameters for milk production and lactation pattern in an ethiopian multibreed dairy cattle population. Asian-Australas J Anim Sci. 2013;26(9):1237-46.

Gorewit RC. Lactation Biology and Methods of Increasing Efficiency. In: National Research Council (US) Committee on Technological Options to Improve the Nutritional Attributes of Animal Products. Designing Foods: Animal Product Options in the Marketplace. Washington (DC): National Academies Press (US); 1988.

Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, et al. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 2002;12(2):222-31.

Hadfield JD. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. J Stat Softw. 2010;33(2):1-22.

Hase M, Yokomizo T, Shimizu T, Nakamura M. Characterization of an Orphan G Protein-coupled Receptor, GPR20, That Constitutively Activates $G_i$ Proteins. J Biol Chem. 2008;283(19):12747-55.

Hoffman LS, Etzrodt J, Willkomm L, Sanyal A, Scheja L, Fischer AWC, et al. Stimulation of Soluble Guanylyl Cyclase Protects Against Obesity by Recruiting Brown Adipose Tissue. Nat Commun. 2015;6:7235.

https://github.com/sahirbhatnagar/manhattanly/last accessed on 10-04-2020

https://www.nddb.coop/services/animalbreeding/geneticimprovement/genomic last accessed 09-04-2020

Ilatsia E.D., Roessler R., Kahi A.K., Piepho H.P., Valle Z. Production objectives and breeding goals of Sahiwal cattle keepers in Kenya and implications for a breeding programme. Trop. Anim. health and prod. 2011; 44:519-530.

Joshi B.K., Singh A., Gandhi R.S. Performance evaluation, conservation and improvement of Sahiwal cattle in India. Anim. Genet. Res. 2001; 31:43-54.

Khan M.S., Rehman Z.U., Khan M.A., Ahmad S. Genetic resources and diversity in Pakistani cattle. Pak. Vet. J. 2008; 28:95-102.

Kobata R, Tsukahara H, Ohshima Y, Ohta N, Tokuriki S, Tamura S, et al. High levels of growth factors in human breast milk. Early Hum Dev. 2008, 84(1):67–9.

León CD, Manrique C, Martínez R, Rocha JF. Genomic association study for adaptability traits in four Colombian cattle breeds. Genet Mol Res. 2019; 18 (3): gmr18373.

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: genome association and prediction integrated tool. *Bioinformatics.*

2012;28(18):2397–2399.

Lung LHS, Petrini J, Diaz JR, Salvian M, Rovadoscki GA, Pilonetto F, et al. Genome-wide association study for milk production traits in a Brazilian Holstein population. J Dairy Sci. 2019;102(6):5305-14.

Meyer K. WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). J Zhejiang Univ Sci B. 2007;8(11):815–21.

Muñoz M, Rodríguez MC, Alves E, Folch JM, Escriche NI, Silió L. Genome-wide analysis of porcine backfat and intramuscular fat fatty acid composition using high-density genotyping and expression data. BMC Genom. 2013;14:845.

Nayeri S, Sargolzaei M, Abo-Ismail MK, et al. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genet. 2016;17(1):75.

Nicolini P, Amorin R, Han Y, Penagaricano F. Whole-genome scan reveals significant non-additive effects for sire conception rate in Holstein cattle. BMC Genet. 2018;19:14.

Oakes, S.R., Hilton, H.N., Ormandy, C.J. Key stages in mammary gland development - The alveolar switch: coordinating the proliferative cues and cell fate decisions that drive the formation of lobuloalveoli from ductal epithelium. Breast Cancer Res. 2006; 8:207.

Rogers ML, Goddard C, Regester GO, Ballard FJ, Belford DA. Transforming Growth Factor Beta in Bovine Milk: Concentration, Stability and Molecular Mass Forms. J Endocrinol. 1996;151(1):77-86.

Rosiane PS, Berton MP, Grigoletto L, Carvalho FE, Silva RMO, Peripolli E, et al. Genomic regions and enrichment analyses associated with carcass composition indicator traits in Nellore cattle. J Anim Breed Genet. 2019;136(2):118–33.

Ryman VE, Packiriswamy N, Sordillo LM. Role of endothelial cells in bovine mammary gland health and disease. Anim Health Res Rev. 2015;16(2):135-49.

Salari P, Abdollahi M. The influence of pregnancy and lactation on maternal bone health: a systematic review. J Family Reprod Health. 2014; 8(4), 135–48.

Salle SL, Palmer K, O'Brien M, Schimenti JC, Eppig J, Handel MA. *Spata22*, a Novel Vertebrate-Specific Gene, Is Required for Meiotic Progress in Mouse Germ Cells. 2012; 86(2): 45Biol Reprod..

Salleh MS, Mazzoni G, Nielson MO, Løvendahl P, Kadarmideen H. Identification of Expression QTLs Targeting Candidate Genes for Residual Feed Intake in Dairy Cattle Using Systems Genomics. J Genet Genome Res. 2018; 5(1): 035.

Sambrook J., Russel D.W. Rapid isolation of yeast DNA. *CSH protocols*, 2006; 2006(1), pdb.prot4039. doi:10.1101/ pdb.prot4039

Serao NVL, Pena DG, Beever JE, Faulkner DB, Southey BR, Zas SLR. Single nucleotide polymorphisms and haplotypes associated with feed efficiency in beef cattle. BMC Genet. 2013;14:94.

Skorczyk A, Flisikowski K, Szydlowski M, Cieslak J, Fries R, Switonski M. Association of MC3R Gene Polymorphisms With Body Weight in the Red Fox and Comparative Gene Organization in Four Canids. Anim Genet. 2011;42(1):104-7.

Szkiba D, Kapun M, von Haeseler A, Gallach M. SNP2GO: Functional analysis of genome-wide association studies. *Genetics*. 2014;197(1):285-9.

Verma RK, Gupta AK, Kumar M, Ratwan P. Estimates of genetic parameters for milk and milk constituent's yield traits in Sahiwal cattle. Indian J Dairy Sci. 2017;70(6):751-4.

Wallace C, Rotival M, Cooper JD, Rice CM, Yang JHM, McNeill M, et al. Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. *Hum Mol Genet. 2012;21(12):2815-24.*

Wilkening S, Chen B, Bermejo JL, Canzian F. Is there still a need for candidate gene approaches in the era of genome-wide association studies?

Genomics. 2009;93(5):415-9.

Yodklaew P, Koonawootrittriron S, Elzo MA, Suwanasopee T, Laodim T. Genome-wide association study for lactation characteristics, milk yield and age at first calving in a Thai multibreed dairy cattle population. Agric Nat Resour. 2017;51(3):223-30.

Zhang Z, Ober U, Erbe M, Zhang H, Gao N, He J, et al. Improving the Accuracy of Whole Genome Prediction for Complex Traits Using the Results of Genome Wide Association Studies. PLoS One. 2014; 9(3): e93017.

Zieglar SF, Levin SD, Johnson L, Copeland NG, Gilbert DJ, Jenkins NA, et al. The mouse CD69 gene. Structure, expression, and mapping to the NK gene complex. J Immunol. 1994; 152 (3) 1228-36.