Maternal genes and pathways affecting birth weight and weaning weight traits in sheep: A GWAS and pathway enrichment analysis

Mehdi Esmaeilifard (mehdi.esmaeilifard@gmail.com)
Sari Agricultural Sciences and Natural Resources University https://orcid.org/0000-0001-7509-5030
Seyed Hasan Hafezian
Sari Agricultural Sciences and Natural Resources University
Mohsen Gholizadeh
Sari Agricultural Sciences and Natural Resources University
Rostam Abdollahi-Arpanahi
University of Florida

Research article

Keywords: Ewe productivity, Maternal genes, Maternal pathways, GWAS, Gene set enrichment

Posted Date: February 11th, 2020

DOI: https://doi.org/10.21203/rs.2.23075/v1

Version of Record: A version of this preprint was published at Frontiers in Genetics on July 28th, 2021.
See the published version at https://doi.org/10.3389/fgene.2021.710613.

Maternal genes and pathways affecting birth weight and weaning weight traits in sheep: A GWAS and pathway enrichment analysis

Mehdi Esmaeilifard[1], Seyed Hasan Hafezian[2], Mohsen Gholizadeh[2], Rostam Abdollahi-Arpanahi[3]

[1] Ph.D. Graduate, Sari Agricultural Sciences and Natural Resources University (SANRU), Sari, Iran

[1, 2] Associate Professor, Sari Agricultural Sciences and Natural Resources University (SANRU), Sari, Iran

[3] Research Assistant, Department of Animal Sciences, University of Florida, Gainesville, FL, United States

Abstract

Background: Ewe productivity is considered as the most important economic trait in sheep meat production. Due to very limited reports, the objective of this study was the application of alternative GWAS approaches followed by gene set enrichment analysis (GSEA) on the maternal genome to unravel the genomic architecture underlying ewe productivity in Iranian Baluchi sheep. Six maternal composite traits including progeny birth weight (PBW), litter mean weight at birth (LMWB), total litter weight at birth (TLWB), progeny weaning weight (PWW), litter mean weight at weaning (LMWW) and total litter weight at weaning (TLWW) were studied.

Results: Genes such as *RDX*, *FDX1*, *ARHGAP20*, *ZC3H12C*, *THBS1*, and *EPG5* on OAR6, OAR7, OAR15, and OAR23 were identified for composite traits at birth. The genes are involved in pregnancy, including autophagy in the placenta, progesterone production by the placenta, maternal immune response and placenta formation. Some maternal pathways, related to calcium ion transport, signal transduction, neurogenesis, and immune response were also identified for birth weight traits. Moreover, many genes including *NR2C1*, *VEZT*, *HSD17B4*, *RSU1*, *CUBN*, *VIM*, *PRLR*, and *FTH1* were located on OAR2, OAR3, OAR5, OAR7, OAR13, OAR16, and OAR25 identified as maternal genes affecting weaning weight traits. Most of the identified genes

were involved in mammary glands development and milk components production. Also, many GO terms related to protein processing and transport, ion transport and homeostasis, proteins and lipid phosphorylation, and phospholipid translocation were identified in association with weaning weight traits.

Conclusions: The results of the present study revealed that calcium ion homeostasis and transport and the maternal immune system could have an important role in progeny's birth weight. Also, the results showed that genes and pathways affecting mammary glands development during pregnancy and milk components production have the most impact on lambs weaning weight. These findings contribute to a better understanding of the genetic architecture of the studied traits and providing opportunities for improving ewe productivity via marker-assisted selection.

Keywords: Ewe productivity, Maternal genes, Maternal pathways, GWAS, Gene set enrichment

Background

In sheep breeding, ewe productivity is the most important trait affecting profitability and genetic progress in this complex trait can lead to more efficient lamb production [1]. In some countries such as Iran, where meat is the main sheep product, the productivity of the ewe flock has the greatest influence on profitability per ewe [2]. An increase in meat output in the sheep production system could be achieved by increasing the number and weight of lambs weaned per ewe within a specific year [3]. Ewe productivity, which was defined as total lamb weight weaned (TLWW) per ewe, is a most common composite trait that is affected by many cooperative components linked to reproduction and growth, including age at puberty, ovulation, pregnancy, parturition, lactation, mothering ability and lamb survival and growth [4] and often is regarded as an overall measure of lamb production capacity by ewes [5]. Composite traits are a combination of growth and

reproductive traits. Therefore, genetic improvement of ewe productivity is a key target in sheep breeding programs [3]. Common composite reproductive traits in sheep are total litter weight at birth (TLWB) and total litter weight at weaning (TLWW) and have the excellence of being better coordinate to the market, where producers are paid on a per kilogram and not on a per head [6].

Although estimates of genetic parameters have been reported for composite traits of different Iranian sheep breeds [6, 7], reports on genes and pathways affecting these traits are limited. To our knowledge, there is only one published report of genes and genomic regions associated with composite traits in sheep [6]. A whole-genome scan carried out by the authors and five genes neighboring the top SNPs (on OAR2, OAR3, OAR15, and OAR16), including *TEX12*, *BCO2*, *WDR70*, *INHBE*, and *INHBC* reported as possible candidate genes affecting composite traits of the Lori–Bakhtiari sheep.

Due to the strict threshold used in Genome-wide association studies (GWAS) to finding significant SNPs, several poorly associated SNPs are always ignored. An alternative strategy is to add gene set analysis as a complement approach after GWAS [8]. In this approach, a set of genes with some common functional features (e.g. being a member of a specific pathway) which are identified by significant (with comfortable threshold) SNPs of GWAS, is tested for over-representation in a specific pathway [9]. Recently a growing interest in gene set enrichment analysis has been observed in dairy cattle [8, 10, 11]. However, there are no published reports of gene set analysis of sheep GWAS dataset.

Baluchi sheep breed is the largest sheep breed in Iran and well adapted to a wide range of arid subtropical environments from north-east to south-east of the country [12]. Due to very limited reports on maternal genes and pathways affecting composite traits, the objective of this study was

68 to conduct GWAS and gene set enrichment analysis (GSEA) to unravel the genomic architecture
69 underlying ewe productivity in Iranian Baluchi sheep.

70 Results

71 GWA analysis of the maternal composite traits at birth

72 For these set of traits, we looked for maternal genes and pathways that influence progeny's birth
73 weight during pregnancy. The results of GWAS analysis are shown in a Circular Manhattan plot
74 in Figure 1. Nine significant SNPs located on chromosomes OAR6, OAR7, OAR15 and OAR23
75 were identified for ewe reproductive traits at birth (Table 1). Three common significant SNPs,
76 rs428350449, rs422482383 and rs423274340 on OAR15 (19.7-20.3 Mb) harbors four candidate
77 genes, *RDX*, *FDX1*, *ZC3H12C*, and *ARHGAP20*. SNPs rs422482383 and rs423274340 were
78 significantly associated with the three traits in both GWAS approaches which rs422482383
79 identified within the ovine known gene *ARHGAP20*. Another SNP (rs427207318) located on
80 OAR15 at 69.5 Mb had a significant association with LMWB trait which was not contained any
81 gene. In addition, we found three marginally significant SNPs (rs408063438, rs399067974, and
82 rs400684038) located on OAR7 (30.2-35.1 Mb) associated with PBW trait.

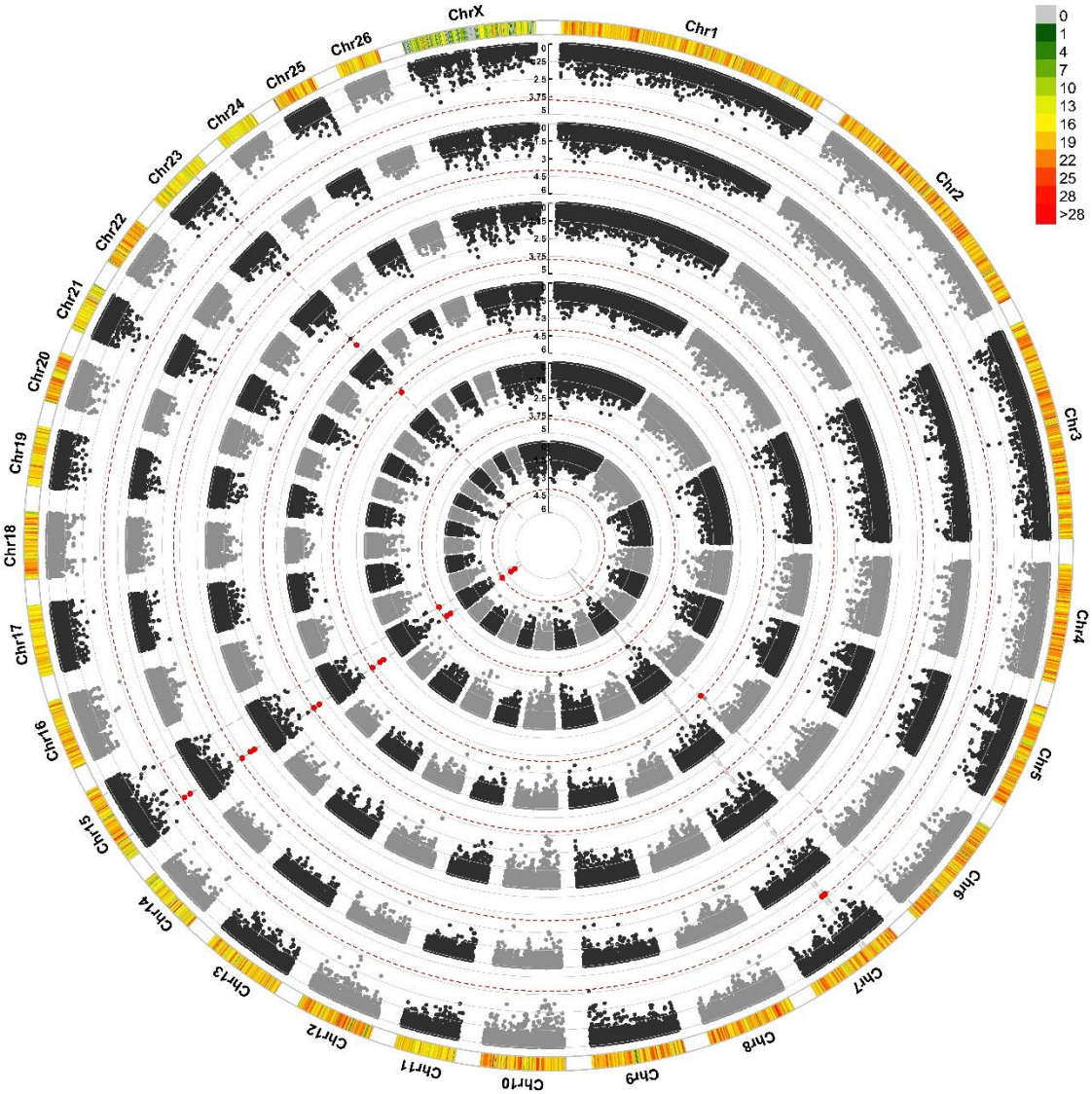


Fig. 1 Circular Manhattan plot for associations of SNPs with ewe composite traits at birth by two GWAS approaches. The 6 circles from outside to inside represent Progeny Birth Weight (PBW): pGWAS and eGWAS; Total Litter Weight at Birth (TLWB): pGWAS and eGWAS; Litter Mean Weight at Birth (LMWB): pGWAS and eGWAS. X-axis: SNPs positions on chromosomes, Y-axis: -Log10 P-value. The dashed lines indicate the threshold for statistical significance ($P<10^{-4}$). The outermost circle show SNPs density in 1 Mb window for each chromosome. pGWAS: GWAS using phenotypes as response variable; eGWAS: GWAS using EBVs as response variable.

Our BLAST search identified 12 genes in this region and SNP rs400684038 was located within the *TTBK2* gene. Also, two SNPs (rs430043751 and rs426428997) on OAR23 and OAR6 were

significantly related to TLWB trait. SNP rs430043751 on OAR23 identified in both GWAS approaches harbors six genes. This SNP was very close to the threshold line for PBW and LMWB traits in both GWAS approaches. SNP rs426428997 on OAR6 did not contain any gene in the searched region.

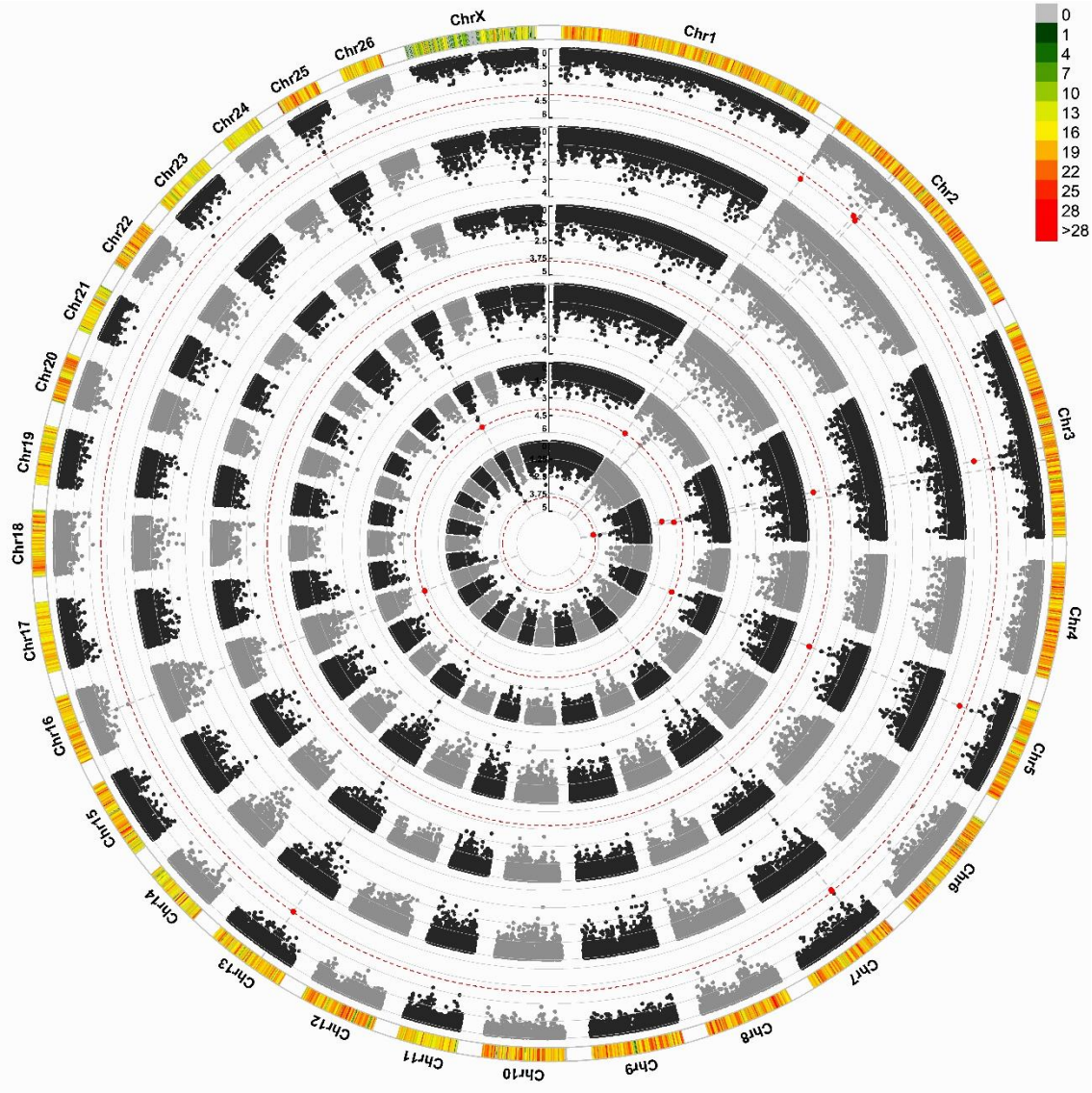
Table 1 SNPs significantly associated with ewe composite traits at birth in Baluchi sheep.

Chr	SNP	Position	Genes in 300kb interval	Analysis method	Adjusted P-value	Trait(s)
15	rs422482383	20125491	RDX, FDX1, ARHGAP20 (Within)	pGWAS & eGWAS	2.06×10^{-5} 3.34×10^{-6}	PBW, TLWB & LMWB
15	rs423274340	20304472	ARHGAP20	pGWAS & eGWAS	6.43×10^{-5} 8.20×10^{-6}	PBW, TLWB & LMWB
15	rs428350449	19740174	ZC3H12C, RDX, FDX1	eGWAS	6.34×10^{-5}	PBW & LMWB
15	rs427207318	69550331	Without gene	pGWAS	8.43×10^{-5}	LMWB
7	rs408063438	30207038	Without gene	pGWAS	9.37×10^{-5}	PBW
7	rs399067974	32062261	ZCRB1, THBS1, FSIP1	pGWAS	9.37×10^{-5}	PBW
7	rs400684038	35113511	ZNF106, SNAP23, LRRC57, HAUS2, STARD9, CDAN1, TTBK2 (Within), UBR1, TMEM62	pGWAS	9.37×10^{-5}	PBW
23	rs430043751	46167308	EPG5, PSTPIP2, ATP5F1A, HAUS1, RNF165, LOXHD1	pGWAS & eGWAS	2.52×10^{-5} 5.21×10^{-5}	TLWB
6	rs426428997	109147722	Without gene	eGWAS	7.20×10^{-5}	TLWB

Chr: Chromosome number; pGWAS: GWAS using phenotypes as response variable; eGWAS: GWAS using EBVs as response variable; PBW: Progeny Birth Weight; TLWB: Total Litter Weight at Birth; LMWB: Litter Mean Weight at Birth. P-values are presented just for the first trait in the Trait(s) column. P-values are adjusted based on the Genomic Control value.

GWA analysis of the maternal composite traits at weaning

For maternal composite traits at weaning, we looked for maternal genes and pathways that influence progeny's weaning weight. The circular Manhattan plot for associations of SNP markers with the traits for both GWAS approaches is shown in Figure 2.



104

105 **Fig. 2** Circular Manhattan plot for associations of SNPs with ewe composite traits at weaning by two GWAS
106 approaches. The 4 circles, from outside to inside, represent Progeny Weaning Weight (PWW): pGWAS and eGWAS;
107 Total Litter Weight at Weaning (TLWW): pGWAS and eGWAS; Litter Mean Weight at Weaning (LMWW): pGWAS
108 and eGWAS. X-axis: SNPs positions on chromosomes, Y-axis: -Log10 P-value. The dashed lines indicate the
109 threshold for statistical significance ($P<10^{-4}$). The outermost circle show SNPs density in 1 Mb window for each
110 chromosome. pGWAS: GWAS using phenotypes as response variable; eGWAS: GWAS using EBVs as response
111 variable.

A total of 10 significant SNPs located on OAR2, OAR3, OAR5, OAR7, OAR13, OAR16 and OAR25 were related to maternal composite traits at weaning (Table 2).

Table 2 SNPs significantly associated with ewe composite traits at weaning in Baluchi sheep.

Chr	SNP	Position	Genes in 300kb interval	Analysis method	Adjusted P-value	Trait(s)
3	rs428404187	131255497	NDUFA12, NR2C1, FGD6, VEZT (Within), MIR331, METAP2	pGWAS & eGWAS	4.72×10^{-6} 8.60×10^{-5}	PWW, TLWW & LMWW
5	rs398620273	32383306	HSD17B4 (+461434), DMXL1, DTWD2 (Within)	pGWAS	2.72×10^{-5}	PWW, TLWW & LMWW
2	rs412011189	1426911	EIPR1 (TSSC1)	pGWAS	3.39×10^{-5}	PWW & LMWW
2	rs411656768	81968762	NFIB	pGWAS	4.74×10^{-5}	PWW
7	rs430218107	23778602	EDDM3B, ANG1, RNASE1, RNASE6, RNASE4, ANG2, UQCRFS1, RNASE12, RNASE11, RNASE9, RNASE10, PIP4P1, APEX1, OSGEP, KLHL33, TEP1, PARP2, RPPH1, SNORA79B, CCNB1IP1, TTC5, OR11H4, OR11H7, OR11H6	pGWAS	7.42×10^{-5}	PWW
2	rs403459195	77075145	RPLP0 (+435045)	pGWAS	7.88×10^{-5}	PWW
13	rs401393221	30320719	PTER, C1ql3, RSU1 (within), CUBN, VIM (-426809)	pGWAS	9.50×10^{-5}	PWW
3	rs404069303	143726957	SNORA62	pGWAS	1.90×10^{-5}	LMWW
16	rs409558668	39225407	PRLR, AGXT2, DNAJC21, BRIX1, RAD1, TTC23L (Within), RAI14	pGWAS	7.95×10^{-5}	LMWW
25	rs405045517	16553906	CDK1, FTH1, RHOBTB1 (within)	pGWAS	9.89×10^{-5}	LMWW

Chr: Chromosome number; pGWAS: GWAS using phenotype; eGWAS: GWAS using EBVs; PWW: Progeny Weaning Weight; TLWW: Total Litter Weight at Weaning; LMWW: Litter Mean Weight at Weaning. P-values are presented just for the first trait in the Trait(s) column. In a few spatial cases, genes with a distance more than 300kb from significant SNP are also shown. P-values are adjusted based on the Genomic Control value.

The results of two GWAS approaches showed similar profiles with two common significant SNPs on OAR3 (rs428404187) and OAR5 (rs398620273). The most significant SNP was rs428404187 ($p=4.72\times10^{-6}$), which is located at 131.2 Mb on OAR3 and was significant for the three composite traits at weaning in pGWAS approach. This SNP had a significant association with LMWW in both GWAS approaches and is located within the *VEZT* gene. Other SNPs were significant just in pGWAS approach. For PWW and TLWW traits, there were no significant SNPs using eGWAS and for LMWW only one SNP located on OAR3 identified by this approach. Another common significant SNP (rs398620273) on OAR5 was identified for the three composite traits at weaning and was located within the *DTWD2* gene. SNP rs412011189 on OAR2 was associated with PWW and LMWW and was close to *EIPR1* (*TSSC1*) gene. In addition, we found four SNPs on OAR2 (rs411656768 and rs403459195), OAR7 (rs430218107), and OAR13 (rs401393221) that were significantly associated with PWW. A SNP on OAR7 was located on 23.7 Mb and harbors 24 genes in 300Kb interval, which seven of them were RNase genes. A SNP on OAR13 was located in the *RSU1* gene, as well as, *VIM* gene is located downstream of this SNP at a distance of 4.2kb. Three significant SNPs related to LMWW were identified on OAR3 (rs404069303), OAR16 (rs409558668), and OAR25 (rs405045517). A SNP located on OAR16 harbors seven genes. This SNP located in ovine Known gene *TTC23L*. Additionally, the *Prolactin receptor* (*PRLR*) gene located close to this SNP. Another significant SNP was rs405045517 located on OAR25 and harbor three genes (*CDK1*, *FTH1*, and *RHOBTB1*) in the searched region. This SNP located within the *RHOBTB1* gene. Also, the *FTH1* gene is located close to this SNP.

Gene-Set enrichment Analysis

The results of GWAS were complemented with gene set enrichment analysis using the GO database. In GSA we looked for maternal pathways affecting lambs' weights at birth and weaning.

In total, 23,462 of 45,342 SNPs tested in the GWAS, were located within or 15 kb upstream or downstream of 15815 annotated genes in the Oar.v3.1 ovine genome assembly. On average, 1310 out of 15815 genes (ranging between 1,291 for LMWW to 1,351 for TLWW) contained at least one significant SNP (P-value ≤ 0.05) and defined as significantly associated genes with maternal composite traits. GO terms with a nominal P-value ≤ 0.01 were reported as significant terms. GWAS results using direct phenotypes (pGWAS) were used for the analysis and each trait was analyzed separately.

Gene-set enrichment analysis of the maternal composite traits at birth

Table 3 shows a set of GO terms that were significantly ($P \leq 0.01$) enriched with genes associated with maternal composite traits at birth. Several GO terms related to neural system showed an overrepresentation of significant genes, including *postsynaptic density* (GO:0014069), *Schaffer collateral-CA1 synapse* (GO:0098685), *glutamatergic synapse* (GO:0098978), *synapse* (GO:0045202), *neuron projection development* (GO:0031175), *neurogenesis* (GO:0022008), and many others terms that were not included in Table5 (see Additional file1: Table S1-Table S3). We identified *calcium ion transport* (GO:0000045) associated with the composite traits at birth. Also, *calcium channel inhibitor activity* (GO:0019855) showed an overrepresentation of significant genes associated with TLWB. Many GO terms related to immune system also showed significant enrichment of genes associated with composite traits at birth, including *cellular response to chemokine* (GO:1990869), *positive regulation of T-helper 1 type immune response* (GO:0002827), *positive regulation of interleukin-12 production* (GO:0032735), and *positive regulation of T cell activation* (GO:0050870). Several significant GO terms were related to the signaling process.

Table 3 Most related Gene Ontology (GO) terms significantly ($p \leq 0.01$) enriched using genes associated with maternal composite traits at birth.

Trait	Cat. No.	Term	No. of genes in the term	No. of sig. genes	P_value	Ont.
PBW	GO:0014069	postsynaptic density	82	21	2.36E-06	CC
	GO:0098685	Schaffer collateral - CA1 synapse	42	12	1.10E-04	CC
	GO:0098978	glutamatergic synapse	165	28	2.21E-04	CC
	GO:0006816	calcium ion transport	61	13	1.30E-03	BP
	GO:0045202	synapse	141	23	1.35E-03	CC
	GO:0008237	metallopeptidase activity	77	15	1.51E-03	MF
	GO:0031175	neuron projection development	59	12	2.98E-03	BP
	GO:0060395	SMAD protein signal transduction	25	7	3.43E-03	BP
	GO:0045746	negative regulation of Notch signaling pathway	14	5	4.19E-03	BP
	GO:1990869	cellular response to chemokine	10	4	6.66E-03	BP
	GO:0007155	cell adhesion	169	24	6.81E-03	BP
	GO:0022008	neurogenesis	29	7	8.31E-03	BP
	GO:0002827	positive regulation of T-helper 1 type immune response	6	3	9.49E-03	BP
TLWB	GO:0098685	Schaffer collateral - CA1 synapse	45	15	7.05E-07	CC
	GO:0098978	glutamatergic synapse	165	31	1.48E-05	CC
	GO:0014069	postsynaptic density	82	18	1.17E-04	CC
	GO:0045746	negative regulation of Notch signaling pathway	14	6	5.60E-04	BP
	GO:0006816	calcium ion transport	61	13	1.34E-03	BP
	GO:0045202	Synapse	141	23	1.42E-03	CC
	GO:0032735	positive regulation of interleukin-12 production	10	4	6.75E-03	BP
	GO:0004222	metalloendopeptidase activity	73	13	6.95E-03	MF
	GO:0007155	cell adhesion	169	24	7.13E-03	BP
	GO:0002827	positive regulation of T-helper 1 type immune response	6	3	9.59E-03	BP
	GO:0019855	calcium channel inhibitor activity	6	3	9.59E-03	MF
	GO:1901222	regulation of NIK/NF-kappaB signaling	6	3	9.59E-03	BP
LMWB	GO:0060395	SMAD protein signal transduction	25	10	1.64E-05	BP
	GO:0014069	postsynaptic density	82	18	1.18E-04	CC
	GO:0045202	synapse	141	25	2.52E-04	CC
	GO:0008237	metallopeptidase activity	77	16	5.35E-04	MF
	GO:0098685	Schaffer collateral - CA1 synapse	42	10	1.99E-03	CC
	GO:0098978	glutamatergic synapse	165	25	2.61E-03	CC
	GO:0006816	calcium ion transport	16	12	4.12E-03	BP
	GO:0045746	negative regulation of Notch signaling pathway	14	5	4.27E-03	BP
	GO:0001764	neuron migration	64	12	6.17E-03	BP
	GO:1990869	cellular response to chemokine	10	4	6.77E-03	BP
	GO:0007155	cell adhesion	169	24	7.19E-03	BP
	GO:0022008	neurogenesis	29	7	8.52E-03	BP
	GO:1901222	regulation of NIK/NF-kappaB signaling	6	3	9.61E-03	BP
	GO:0050870	positive regulation of T cell activation	11	4	9.94E-03	BP

Cat. No.: Category number; Ont.: Ontology; PBW: Progeny Birth Weight; TLWB: Total Litter Weight at Birth; LMWB: Litter Mean Weight at Birth. Complete associated GO terms with these traits are provided in Additional file1: Table S1-Table S3.

Especially, *SMAD protein signal transduction* (GO:0060395), *negative regulation of Notch signaling pathway* (GO:0045746), and *regulation of NIK/NF-kappaB signaling* (GO:1901222) showed an overrepresentation of significant genes. In addition, we identified *cell adhesion* (GO:0007155) and *metallopeptidase activity* (GO:0008237) GO terms as significant processes that were associated with the composite traits at birth. Several other GO terms were also significant for composite traits at birth. The full list is provided in the Additional file1: Table S1-Table S3.

Gene-Set enrichment Analysis of the maternal composite traits at weaning

Table 4 shows a set of GO terms that were significantly ($p \leq 0.01$) enriched for significant genes associated with weaning traits. *Filopodium* (GO:0030175) was significantly associated with the composite traits at weaning. Moreover, many GO terms related to protein metabolism were identified, including *protein catabolic process* (GO:0030163), *positive regulation of intracellular protein transport* (GO:0090316), protein processing (GO:0016485), and *protein localization to plasma membrane* (GO:0072659). Several GO terms related to GTPase activity were significant. Among these, *GTPase activator activity* (GO:0005096) showed an overrepresentation of significant genes associated with the composite traits at weaning. Many significant GO terms were related to ion transport and homeostasis and also channel activity, including *cellular calcium ion homeostasis* (GO:0006874), *ion channel activity* (GO:0005216), *ion transmembrane transport* (GO:0034220), and *ion transport* (GO:0006811). Also, many GO terms related to lipids, cholesterol, and fatty asides metabolism showed an overrepresentation of genes associated with the traits at weaning, including *phospholipid translocation* (GO:0045332), *lipid phosphorylation* (GO:0046834), *cholesterol homeostasis* (GO:0042632), and *fatty acid beta-oxidation* (GO:0006635).

Table 4 Most related Gene Ontology (GO) terms significantly ($p \leq 0.01$) enriched using genes associated with ewe composite traits at weaning.

Trait	Cat. No.	Term	No. of genes in the term	No. of sig. genes	P_value	Ont.
PWW	GO:0030175	filopodium	25	11	1.59E-06	CC
	GO:0030163	protein catabolic process	19	8	6.51E-05	BP
	GO:0005096	GTPase activator activity	121	21	7.63E-04	MF
	GO:0005911	cell-cell junction	107	19	1.00E-03	CC
	GO:0098978	glutamatergic synapse	165	25	1.94E-03	CC
	GO:0016310	phosphorylation	241	33	2.35E-03	BP
	GO:0090316	positive regulation of intracellular protein transport	13	5	2.67E-03	BP
	GO:0016485	protein processing	31	8	2.81E-03	BP
	GO:0045332	phospholipid translocation	9	4	4.01E-03	BP
	GO:0016757	transferase activity, transferring glycosyl groups	112	18	4.18E-03	MF
	GO:0010628	positive regulation of gene expression	207	28	5.70E-03	BP
	GO:0008283	cell proliferation	82	14	6.33E-03	BP
	GO:0042755	eating behavior	6	3	9.03E-03	BP
	GO:0098609	cell-cell adhesion	61	11	9.78E-03	BP
TLWW	GO:0005096	GTPase activator activity	121	24	7.69E-05	MF
	GO:0046834	lipid phosphorylation	9	5	4.25E-04	BP
	GO:0043547	positive regulation of GTPase activity	163	26	1.40E-03	BP
	GO:0043565	sequence-specific DNA binding	291	40	1.83E-03	MF
	GO:0090316	positive regulation of intracellular protein transport	13	5	3.24E-03	BP
	GO:0016485	protein processing	31	8	3.69E-03	BP
	GO:0006874	cellular calcium ion homeostasis	38	9	3.96E-03	BP
	GO:0030175	filopodium	25	7	3.98E-03	CC
	GO:0005216	ion channel activity	141	22	4.13E-03	MF
	GO:0045332	phospholipid translocation	9	4	4.71E-03	BP
	GO:0035556	intracellular signal transduction	215	30	5.13E-03	BP
	GO:0034220	ion transmembrane transport	110	18	5.43E-03	BP
	GO:0006811	ion transport	295	34	7.84E-03	BP
	GO:0042632	cholesterol homeostasis	35	8	8.11E-03	BP
	GO:0007275	multicellular organism development	107	17	9.09E-03	BP
LMWW	GO:0030175	filopodium	25	9	9.60E-05	CC
	GO:0005096	GTPase activator activity	121	23	1.07E-04	MF
	GO:0042127	regulation of cell proliferation	91	18	3.49E-04	BP
	GO:0016485	protein processing	31	9	6.06E-04	BP
	GO:0016310	phosphorylation	241	34	1.19E-03	BP
	GO:0072659	protein localization to plasma membrane	74	14	2.38E-03	BP
	GO:0030163	protein catabolic process	19	6	3.14E-03	BP
	GO:0098978	glutamatergic synapse	165	24	3.97E-03	CC
	GO:0035556	intracellular signal transduction	215	29	5.06E-03	BP
	GO:0010628	positive regulation of gene expression	207	28	5.59E-03	BP
	GO:0016301	kinase activity	238	31	6.43E-03	MF
	GO:0006635	fatty acid beta-oxidation	18	5	1.25E-02	BP

Cat. No.: Category number; Ont.: Ontology; PWW: Progeny Weaning Weight; TLWW: Total Litter Weight at Weaning; LMWW: Litter Mean Weight. Complete associated GO terms with these traits are provided in Additional file1: Table S4-Table S6.

194 In addition, terms related to cell proliferation (e.g. GO:0008283 and GO:0042127), gene
195 expression (e.g. GO:0010628), cell adhesion (e.g. GO:0098609), cell junction (e.g. GO:0005911),
196 Protein kinase activity (GO:0016301), and phosphorylation (GO:0016310) were also enriched
197 with significant genes. Several other GO terms were also significant for the traits at weaning. The
198 full list is provided in the Additional file1: Table S4-Table S6.

199 **Discussion**

200 **GWAS and GSA of maternal composite traits at birth**

201 To attain more consistent findings, two different GWAS approaches were used. Both GWAS
202 approaches identified similar regions that may be explain some part of the genetic variation of
203 studied traits. Four genes, namely *RDX*, *FDX1*, *ZC3H12C*, and *ARHGAP20* were identified as
204 maternal genes affecting composite traits at birth on OAR15 (19.7-20.3 Mb). *RDX* (Radixin) is
205 part of the ERM (*EZR-RDX-MSN*) cytoskeleton linker protein family. The expression of ERM
206 proteins in the blastocyst and the uterus has been reported and linked to the implantation potential
207 in mice [13]. Recently *EZR* and *MSN* proteins have been detected in the yolk sac [14] and
208 blastocoel fluid of pregnant horses [15] as proteins involved in embryo-maternal interaction.
209 *Ferredoxin* (*FDX1*) is an electron transport intermediate which is functional in mitochondrial
210 cytochromes P450 and is found mainly in the steroidogenic tissues, including testis, adrenal,
211 ovaries, and placenta [16]. *ARHGAP20* gene was identified as the candidate gene in both GWAS
212 approaches. High expression of ARHGAP20 in the brain has been reported which proposes a role
213 for this gene product in neurogenesis [17]. *Zc3h12c* is an endogenous inhibitor of TNFα-induced
214 inflammatory signaling in human umbilical vein endothelial cells. It seems the *Zc3h12c* gene plays
215 a role in immune regulation in pregnancy [18]. Abdoli et al. (2019) identified a SNP on OAR15

located on 22.02 Mb as a significantly associated SNP with the TLWB trait, which is close to the region identified in this study and reinforces this possibility that this region on OAR15 likely has an effect on fetal development during pregnancy.

Our GWA analysis identified a region on OAR7 at 30.2-35.1 Mb that contains three significantly associated SNPs with PBW. This region harbors 12 genes, such as *THBS1* and *TTBK2*. One of these significant SNPs, rs400684038, located within *TTBK2 gene*. It has been reported that expression of *THBS1* by placental cells is crucial for formation of the placental structure [19]. *TTBK2* gene encodes a serine-threonine kinase that phosphorylates tau and tubulin proteins and is a critical regulator of the initiating the assembly of primary cilia in the embryo [20]. Both GWAS approaches identified a SNP, rs430043751, on OAR23 significantly associated with TLWW. This SNP harbors six genes, including *EPG5*, *PSTPIP2*, *ATP5F1A*, *HAUS1*, *RNF165a,* and *LOXHD1*. *EPG5* gene encodes a protein with a crucial role in the autophagy pathway and early differentiation events in human embryonic stem cells [21].

Our gene set analysis identified several significantly associated GO terms with maternal composite trait at birth (Table 3). Interestingly many GO terms related to the neural system showed an overrepresentation of significant genes. Numerous studies have reported that the neural alterations in pregnant women's brains are extensive [22, 23]. Noticeably, two pathways (GO:0000045 and GO:0019855) related to calcium ion metabolism enriched by significant genes. It has been shown that placental calcium transfer increases over pregnancy to match fetal needs and ensure appropriate fetal skeletal mineralization [24]. However, recently inconsistent evidence of effects of maternal calcium on birth weight has been reported [25]. Many GO terms related to the immune system, were among significant functional categories. The maternal immune system plays an essential role in a successful pregnancy. A vital balance is needed to protect the mother and fetus

against possible viral or bacterial infections while avoiding fetal rejection [26]. GO terms related to signaling pathways also showed an overrepresentation of significant genes. *SMAD protein signal transduction* (GO:0060395) was one of these pathways. SMAD proteins transduce signals from TGF-β superfamily ligands and as a result, regulate target gene expression. TGF-β superfamily signaling is vital for female reproduction (Figure 3).

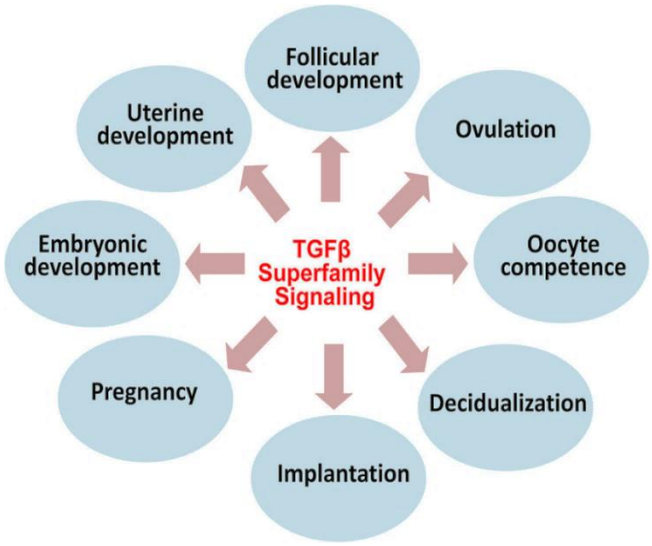


Fig. 3 Main functions of TGF-β family signaling in female reproduction [27]. SMAD proteins transduce signals from TGF-β superfamily ligands.

It has been reported that SMAD proteins have a role in maintaining the structural and functional integrity of oviduct and uterus and are essential for the establishment and maintenance of pregnancy [28]. Another signaling pathway was Notch signaling that exerts effects throughout the pregnancy and playing an important role in placental angiogenesis and normal function and also in trophoblast function [29]. *NIK/NF-kappaB signaling* (GO:1901222) works as a transcription factor involved in inflammatory and immune responses [30]. The effects of NF-κB and its signaling pathway in the human myometrium during pregnancy and parturition are well reviewed [31].

GWAS and GSA of maternal composite traits at weaning

Genes including *NDUFA12*, *NR2C1*, *FGD6*, *VEZT*, *MIR331*, and *METAP2* were identified on OAR3 around significantly associated SNP (rs428404187) with the composite traits at weaning. This SNP is located in the *VEZT* gene which has a main role in the cell adhesion process. Cell adhesion process has a widespread effect on mammary glands development and mainly occurs in late pregnancy and partially in the onset of lactation (reviewed in Shamir and Ewald, 2015). Noticeably, we identified the *cell-cell adhesion* (GO:0098609) pathway as a significant GO term associated with the PWW trait in our gene set analysis. Another gene in this region, *NR2C1*, is a nuclear steroid hormone receptor. This gene acts as a transcription factor and plays an important role in mammary glands differentiation and development in late pregnancy and during lactation [33].

A SNP, rs398620273, on OAR5 was significantly related to composite traits at weaning and is located within the *DTWD2* gene. It is suggested that this gene may be involved in RNA processing [34]. *HSD17B4* is another gene that was identified around this SNP and plays an important role in feed intake and food conservation [35]. In dairy cows, it has been reported that feed intake is the major factor limiting milk production in high-yielding dairy cows in early lactation [36]. We identified *RPLP0* gene on OAR2.Dominant expression of RPLP0 in mammary vasculature tissue has been reported during lactation [37]. The *SNORA62* gene was identified on OAR3 as a candidate gene affecting LMWW trait. Recently, a GWAS of milk fatty acid composition identified the *SNORA62* gene as a candidate gene affecting milk fatty acid content [38].

We identified 24 genes on OAR7 associated with PWW which seven of these genes belong to the pancreatic ribonuclease A family (*RNases*). It seems that this region plays an important role in

RNA processing. It has been reported that *RNASE5* is a functional gene in milk production, specifically milk protein percent [39]. We identified Significant SNP, rs401393221, within the RSU1 gene on OAR13. Using a meta-analysis of microarray data in combination with supervised machine learning models, the *RSU1* gene has been identified as DEG during the lactation process in both approaches [40]. It worth to be noted that the *RSU1* gene is a member of the *milk proteins* KEGG pathway. In addition, we identified *CUBN* and *VIM* genes around this SNP on OAR13. It has been shown that variation in vitamin B-12 content of bovine milk is associated with the *CUBN* gene [41]. *VIM* is a cytoskeletal type III intermediate and has a critical role in mammary gland development [42]. A four-fold increase of *VIM* protein in lactating tissue compared to resting tissue has been reported [43].

A SNP, rs409558668 on OAR16 was identified as significantly ($p=7.95\times10^{-5}$) associated SNP with LMWW trait. This SNP located in the *TTC23L* gene and also harbors *PRLR* (Prolactin receptor) gene. *TTC23L* gene is highly expressed during lactation [44] and also identified as a candidate gene that can affect mastitis in Holstein cows [45]. Expression of the *PRLR* gene in lactating mammary glands has been reported and it is showed that polymorphism in exon 3 and 7 of the *PRLR* gene is correlated with milk production in Holstein cows [46]. In addition, we identified the *FTH1* gene on OAR16. The *FTH1* gene encodes for the heavy subunit of ferritin. The presence of ferritin in cow's and buffalo's milk has been reported [40, 47].

Through gene set analysis for composite traits at weaning, several maternal functional categories were identified. Many GO terms related to protein metabolism and transport and fatty asides metabolism were identified. Recently, in a transcriptome analysis study on buffalo milk, *protein metabolism* (GO:0019538) pathway identified as a significant term [47]. Changes in the amount of fatty acid synthesis during late-pregnancy and lactation have been long and newly reported for

a variety of species including the rat, rabbit, pig, and cow [48, 49]. Terms related to calcium and other ions metabolism and transport were also significant. Several different comparative transcriptome analyses have reported the role of calcium metabolism-related pathways in the lactation process [47, 50]. *Phosphorylation* (GO:0016310) term was identified as a significant pathway associated with PWW and LMWW traits. Caseins comprising 80% of the proteins in cow's and sheep's milk and phosphorylation by Casein Kinase enzyme is a crucial step for milk production in the lactating mammary gland [51]. Noticeably, the *kinase activity* (GO:0016301) pathway was another significant term for LMWW that catalysis the transfer of a phosphate group to a substrate molecule. The c*ell-cell adhesion* pathway was significantly associated with PWW. The effects of cell adhesion molecules on the lactogenesis process have been well-reviewed [52]. Noticeably, the *VEZT* gene, one of our identified genes in GWA analysis, is a member of this pathway. *GTPase activator activity* (GO:0005096) GO term was identified as a significant pathway related to the composite traits at weaning. GTPases are known to be involved in numerous secretory processes and play an important role in translation and translocation of proteins, secretion of the milk fat globule, and probably other milk components [47]. Many GO terms associated with cell proliferation and differentiation were also detected as significant. Most mammary growth takes place through pregnancy. Mammary gland cell proliferation and differentiation have a great impact on milk yield and lactation persistency [53]. As expected, must identified genes and GO terms have a role in milk production or mammary gland development which means feeding lambs by milk have the most impact on weight gain rather than other maternal effects.

Conclusions

In this study, we complemented the GWAS with gene set enrichment analysis (GSEA) to find genes and pathways affecting maternal composite traits at birth and weaning in sheep. Some maternal genes *RDX*, *FDX1*, *ARHGAP20*, *ZC3H12C*, *THBS1*, and *EPG5* were associated with composite traits at birth. These genes play roles in pregnancy e.g. autophagy, immune response, angiogenesis, and placental formation. Gene set analysis identified calcium ion transport GO term as a significant pathway affecting the composite traits at birth. In addition, we identified many genes (e.g., *NR2C1*, *VEZT*, *HSD17B4*, *RSU1*, *CUBN*, *VIM*, *PRLR*, and *FTH1*) as maternal genes affecting composite traits at weaning. Our gene set analysis for these traits identified several significantly associated GO terms such as protein processing and transport, phospholipid translocation, ion transport, and cell-cell adhesion. Most of the identified genes and pathways are involved in mammary glands differentiation and development and also in milk components production. The results provide a good insight into how maternal genes and pathways influence progeny weight at birth and weaning and elucidate the mechanisms underlying these complex traits in sheep.

Methods

Phenotypic and Genotypic Data

The data set consisted of 3916 and 3635 birth weight and weaning weight (90 days of age) records, respectively. Progeny birth weight (PBW), litter mean weight at birth (LMWB) and total litter weight at birth (TLWB) were used as maternal composite traits at birth. Also, progeny weaning weight (PWW), litter mean weight at weaning (LMWW) and total litter weight at weaning (TLWW) were considered as maternal composite traits at weaning. LMWB and LMWW are arithmetic mean of TLWB and TLWW traits and were calculated for each lambing per ewes. The

pedigree file included 4727 animals with 178 sires, 1509 dams and 818 founders. Data were collected from 2004 to 2012 (9-year span) at Abbas Abad Baluchi sheep Breeding Station, located in Sarakhs city, Khorasan Razavi province, Iran. Descriptive statistics of studied traits are presented in Table 5.

Table 5 Descriptive statistics of studied traits.

Trait	N	Ave	Min	Max	CV	Total N (EBV analysis)
PBW	436	4.26	2.30	6.80	0.17	3916
TLWB	317	5.90	2.80	13.00	0.30	3063
LMWB	317	4.40	2.70	6.80	0.17	3063
PWW	398	20.31	9.10	34.60	0.20	3635
TLWW	294	27.52	9.90	57.80	0.30	2869
LMWW	294	20.97	9.90	34.60	0.19	2869

PBW, Progeny Birth Weight; TLWB, Total Litter Weight at Birth; LMWB, Litter Mean Weight at Birth; PWW, Progeny Weaning Weight; TLWW, Total Litter Weight at Weaning; LMWW, Litter Mean Weight at Weaning. N, number of records; Ave, Average; Var, Variance; SD, Standard deviation; CV, Coefficient of variation; Total N, Total number of observations used for EBVs calculation.

SNP genotype data of 54,241 genomic position were provided for 91 ewes by animal genetic group of Sari Agriculture Science and Natural Resource University (SANRU), Iran [54]. GenABEL package [55] was used for quality control in R software (R Core Team, 2019). Those SNP markers with unknown genomic location, or were monomorphic, or had minor allele frequency and genotype call rate less than 0.01 and 93%, respectively, and SNPs that departed from Hardy–Weinberg equilibrium (for a P-value cut-off of 1×10^{-6}) were removed from the dataset. As well as, ewes with a genotyping call rate less than 95%, removed from the dataset. After quality control, 84 ewes and 45342 SNPs were retained for the analysis.

Genome-wide association study

Long-term monitoring of some traits, commonly include repeated measures of individuals and allows estimation of year and age effects over time. In this study, we performed the GWAS in two different ways, (i) Phenotypes were used as the response variable, (ii) EBVs were used as the response variable.

Genome-wide association mapping using phenotypes (pGWAS)

For the GWAS using phenotypes, repeatability model was extended as below,

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{X_{SNP}\beta_{SNP}} + \boldsymbol{Zu} + \boldsymbol{Wpe} + \boldsymbol{e}$$

where $\boldsymbol{y}$ is a vector of ewe's progenies weights; $\boldsymbol{b}$ is a vector of fixed effects; $\boldsymbol{u}$ is the vector of random direct additive genetic effects, $\boldsymbol{pe}$ is the vector of permanent environmental effects, and $\boldsymbol{e}$ is the vector of random residual effects. The matrices **X**, **Z**, and **W** are the design matrices relating individuals' phenotypic records to their fixed and random effects, respectively. $\mathbf{X}_{\text{SNP}}$ is the incidence matrix for the SNP markers and $\boldsymbol{\beta}_{\text{SNP}}$ is the regression coefficient. In this case, the random effects have multivariate Gaussian (co)variance,

$$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{pe} \\ \boldsymbol{e} \end{pmatrix} | \sigma_u^2, \sigma_{pe}^2, \sigma_e^2 \Big) \sim N \left[0, \begin{pmatrix} \boldsymbol{G}\sigma_u^2 & 0 & 0 \\ 0 & \boldsymbol{I}_n\sigma_{pe}^2 & 0 \\ 0 & 0 & \boldsymbol{I}_N\sigma_e^2 \end{pmatrix} \right]$$

Where **G** is the genomic relationship matrix, **I** is the identity matrix, n is the number of genotyped individuals with reproductive records (n = 84, after QC) and N is the total number of observations for genotyped individuals (N = 294-436, depends on the trait). We can write the extended repeatability model as below,

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{X_{SNP}\beta_{SNP}} + \boldsymbol{e}$$

This model is the same as the above model if,

$$\boldsymbol{e} \sim N(0,\ \boldsymbol{V}) \text{ where } \mathbf{V} = \boldsymbol{ZGZ'}\sigma_u^2 + \boldsymbol{WW'}\sigma_{pe}^2 + \boldsymbol{I}_N\sigma_e^2.$$

In this approach, the P-value for the SNP effects in the original model can be calculated from the ratio of the β_{SNP} and its standard error (Wald test). An alternative approach is to use the following

score test statistic that is more computationally efficient, asymptotically standard normal and approximates the Wald test,

$$\boldsymbol{Z} = \frac{\boldsymbol{X}'_{SNP}\boldsymbol{V}_{\circ}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})}{\sqrt{\boldsymbol{X}'_{\boldsymbol{SNP}}\boldsymbol{V}_{\circ}^{-1}\boldsymbol{X}_{SNP}}}$$

but here $\boldsymbol{V}_{\circ}$ computed same as $\boldsymbol{V}$ using a model where the SNP effects $\boldsymbol{X}_{SNP}\boldsymbol{\beta}_{SNP}$ has been excluded and $\widehat{\boldsymbol{\beta}}$ is computed from the model $\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{X}_{\boldsymbol{SNP}}\boldsymbol{\beta}_{\boldsymbol{SNP}} + \boldsymbol{e}$, assuming $\boldsymbol{e} \sim N(0,\ \boldsymbol{V}_{\circ}\,\sigma_e^2)$. The analyses were performed using the R package RepeatABEL [56].

Genome-wide association mapping using estimated breeding values (eGWAS)

In this approach, at first, we ran a pedigree-BLUP analysis using the classical repeatability animal model in BLUPF90 software [57] and breeding values of animals were estimated for all traits. Lambs' sex, lamb birth year, and dam's age at lambing included in the model as fixed effects. Animal direct additive genetic and ewe permanent environmental effects were used as random effects. Variance components were estimated using the Restricted Maximum Likelihood (REML) approach, implemented in the AIREMLF90 software [57]. At the next step, EBVs were considered as the response variable and SNPs genotypes fitted in a GLM model as shown below,

$$\boldsymbol{EBVs} = \boldsymbol{X}_{\boldsymbol{SNP}}\boldsymbol{\beta}_{\boldsymbol{SNP}} + \boldsymbol{e}$$

where $\boldsymbol{X}_{SNP}$ is the design matrix relating EBVs to SNP genotypes and $\boldsymbol{\beta}_{\text{SNP}}$ is the regression coefficient. GenABEL [55] package in the R environment was used for this analysis. Due to the use of the genomic and pedigree-based relationship matrix in GWA analysis, p-values were almost non-inflated ($1.01 \leq \lambda \leq 1.07$) for all traits, however, partial inflation was corrected using genomic

control (GC) method and all p-values presented without any inflation (λ=1). CMplot (https://github.com/YinLiLin/R-CMplot) R package was used to drawing Manhattan plots.

Gene annotation

Some well-known databases including BioMart-Ensembl (www.ensembl.org/biomart), UCSC Genome Browser (http://genome.ucsc.edu) and National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov) were used along with the Ovis aries reference genome assembly (Oar_v3.1) to identify candidate genes within a window of 300 kb up and downstream of the significant SNPs. The distance of 300 kb was selected according to the results of linkage disequilibrium analysis.

Gene-set enrichment and pathway-based analysis

Gene set enrichment analysis can be performed in three steps: (i) the assignment of SNPs to the known genes, (ii) the assignment of genes to functional categories, (iii) the association analysis between each functional category and studied phenotypes.

For each trait, a threshold of P-values < 0.05 was applied to determine significant SNPs (based on the results of the pGWAS) for enrichment analysis. The Bioconductor R package biomaRt [58, 59] and the Oar_v3.1 ovine reference genome assembly were used to flagging genes by significant SNPs. SNPs were assigned to genes if they were within the genomic region or 15 kb up and downstream of an annotated gene. Genes harboring at least one significant SNP were considered as significantly associated genes.

Gene Ontology (GO) database [60] was used to define the functional sets of genes. The GO database classifies genes into three functional categories (biological process, molecular function,

and cellular component) based on their common properties. Finally, the significant association of a particular GO term with maternal composite traits was calculated using Fisher's exact test based on the hypergeometric distribution. The P-value of observing *g* significant genes in the term was computed using the following formula,

$$P - value = 1 - \sum_{i=0}^{g-1} \frac{\binom{s}{i}\binom{m-s}{k-i}}{\binom{M}{k}}$$

where *s* is the total number of significant genes associated with a given maternal composite trait at birth or weaning, *m* is the total number of analyzed genes, and *k* is the total number of genes in the term under consideration [10]. The GO enrichment analysis was performed using the R package goseq [61]. GO terms with more than 5 and less than 500 genes were tested. Functional categories with a nominal P-value less than or equal to 0.01 ($p \leq 0.01$) considered as significant categories.

Abbreviations

Cat. No: Category number

Chr: Chromosome

eGWAS: GWAS using EBVs as response variable.

GO: Gene Ontology

GSEA: Gene set enrichment analysis

GWAS: Genome-wide association study

LMWB: litter mean weight at birth

LMWW: litter mean weight at weaning

Ont: Ontology

PBW: progeny birth weight

pGWAS: GWAS using phenotypes as response variable

PWW: progeny weaning weight

TLWB: total litter weight at birth

TLWW: total litter weight at weaning

Declarations

Ethics approval and consent to participate

Consent for publication

Not applicable.

Availability of data and materials

Competing interests

The authors declare that they have no competing interests.

Funding

Authors' contributions

ME, RAA designed the research. ME performed data analysis, interpreted the data and wrote the manuscript. MG, collected the samples and Genotyped. RAA, MG, SHH edited the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

Not applicable

Authors' information (optional)

Affiliations

Mehdi Esmaeilifard, Seyed Hasan Hafezian, Mohsen Gholizadeh, Rostam Abdollahi-Arpanahi

Ph.D. Graduate, Department of animal science and fisheries, Sari Agricultural Sciences and Natural Resources University (SANRU), Sari, Iran.

Associate Professor, Department of animal science and fisheries, Sari Agricultural Sciences and Natural Resources University (SANRU), Sari, Iran.

Associate Professor, Department of animal science and fisheries, Sari Agricultural Sciences and Natural Resources University (SANRU), Sari, Iran.

Researcher assistant, Department of Animal Sciences, University of Florida, Gainesville, FL, United States.

Corresponding author

Correspondence to Mehdi Esmaeilifard.

Supplementary information

Additional file1: List of significantly associated GO terms with the studied traits.

References

1. Hanford KJ, Van Vleck LD, Snowder GD. Estimates of genetic parameters and genetic change for reproduction, weight, and wool characteristics of Targhee sheep. J Anim Sci. 2003;81(3):630–40.

2. Wang CT, Dickerson GE. Simulation of life-cycle efficiency of lamb and wool production for genetic levels of component traits and alternative management options. J Anim Sci. 1991 Nov;69(11):4324–37.

3. Duguma G, Schoeman SJ, Cloete SWP, Jordaan GF. Genetic and environmental parameters for ewe productivity in Merinos. S Afr J Anim Sci. 2002;32(3):154–9.

4. Snowder GD, Fogarty NM. Composite trait selection to improve reproduction and ewe productivity: a review. Anim Prod Sci. 2009;49(1):9–16.

5. Bromley CM, Van Vleck LD, Snowder GD. Genetic correlations for litter weight weaned with growth, prolificacy, and wool traits in Columbia, Polypay, Rambouillet, and Targhee sheep. J Anim Sci. 2001;79(2):339–46.

6. Abdoli R, Mirhoseini S, Ghavi Hossein-Zadeh N, Zamani P, Ferdosi M, Gondro C. Genome-wide association study of four composite reproductive traits in Iranian fat-tailed sheep. Reprod Fertil Dev [Internet]. 2019 [cited 2019 Aug 31];31:1127–33. Available from: http://www.publish.csiro.au/RD/RD18282

7. Abbasi MA, Abdollahi-Arpanahi R, Maghsoudi A, Torshizi RV, Nejati-Javaremi A. Evaluation of models for estimation of genetic parameters and maternal effects for early growth traits of Iranian Baluchi sheep. Small Rumin Res [Internet]. 2012 May 1 [cited 2019 Aug 21];104(1–3):62–9. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0921448811004056

8. Dadousis C, Pegolo S, Rosa GJM, Gianola D, Bittante G, Cecchinato A. Pathway-based genome-wide

association analysis of milk coagulation properties, curd firmness, cheese yield, and curd nutrient recovery in dairy cattle. J Dairy Sci. 2017;100(2):1223–31.

9. Wang L, Jia P, Wolfinger RD, Chen X, Zhao Z. Gene set analysis of genome-wide association studies: methodological issues and perspectives. Genomics. 2011;98(1):1–8.

10. Han Y, Peñagaricano F. Unravelling the genomic architecture of bull fertility in Holstein cattle. BMC Genet. 2016;17(1):143.

11. Neupane M, Kiser JN, Team the BRDCCAPR, Neibergs HL. Gene set enrichment analysis of SNP data in dairy and beef cattle with bovine respiratory disease. Anim Genet [Internet]. 2018 Dec 1;49(6):527–38. Available from: https://doi.org/10.1111/age.12718

12. Moradband F, Rahimi G, Gholizadeh M. Association of polymorphisms in fecundity genes of GDF9, BMP15 and BMP15-1B with litter size in Iranian Baluchi sheep. Asian-Australasian J Anim Sci. 2011;24(9):1179–83.

13. Matsumoto H, Daikoku T, Wang H, Sato E, Dey SK. Differential Expression of Ezrin/Radixin/Moesin (ERM) and ERM-Associated Adhesion Molecules in the Blastocyst and Uterus Suggests Their Functions During Implantation1. Biol Reprod [Internet]. 2004 Mar 1;70(3):729–36. Available from: https://doi.org/10.1095/biolreprod.103.022764

14. Smits K, Willems S, Van Steendam K, Van De Velde M, De Lange V, Ververs C, et al. Proteins involved in embryo-maternal interaction around the signalling of maternal recognition of pregnancy in the horse. Sci Rep. 2018;8(1):5249.

15. Swegen A, Grupen CG, Gibb Z, Baker MA, de Ruijter-Villani M, Smith ND, et al. From peptide masses to pregnancy maintenance: a comprehensive proteomic analysis of the early equine embryo secretome, blastocoel fluid, and capsule. Proteomics. 2017;17(17–18):1600433.

16. Sheftel AD, Stehling O, Pierik AJ, Elsässer H-P, Mühlenhoff U, Webert H, et al. Humans possess two mitochondrial ferredoxins, Fdx1 and Fdx2, with distinct roles in steroidogenesis, heme, and Fe/S cluster biosynthesis. Proc Natl Acad Sci. 2010;107(26):11775–80.

17. Kalla C, Nentwich H, Schlotter M, Mertens D, Wildenberger K, Döhner H, et al. Translocation t (X; 11)(q13; q23) in B-cell chronic lymphocytic leukemia disrupts two novel genes. Genes, Chromosom Cancer. 2005;42(2):128–43.

18. Liu L, Zhou Z, Huang S, Guo Y, Fan Y, Zhang J, et al. Zc3h12c inhibits vascular inflammation by repressing NF-κB activation and pro-inflammatory gene expression in endothelial cells. Biochem J. 2013;451(1):55–60.

19. Ostankova Y V, Klimovskaya YS, Gorskaya OA, Kolobov A V, Kvetnoi IM, Selkov SA, et al. Expression of thrombospondin-1 gene mRNA and protein in the placenta in gestosis. Bull Exp Biol Med. 2011;151(2):215–8.

20. Goetz SC, Liem Jr KF, Anderson K V. The spinocerebellar ataxia-associated gene Tau tubulin kinase 2 controls the initiation of ciliogenesis. Cell. 2012;151(4):847–58.

21. Tra T, Gong L, Kao L-P, Li X-L, Grandela C, Devenish RJ, et al. Autophagy in human embryonic stem cells. PLoS One. 2011;6(11):e27485.

22. Cohen L, Mizrahi A. Plasticity during motherhood: changes in excitatory and inhibitory layer 2/3 neurons in auditory cortex. J Neurosci. 2015;35(4):1806–15.

23. Bridges RS. Long-term alterations in neural and endocrine processes induced by motherhood in mammals. Horm Behav. 2016;77:193–203.

24. Strid H, Powell TL. ATP-dependent Ca 2+ transport is up-regulated during third trimester in human syncytiotrophoblast basal membranes. Pediatr Res. 2000;48(1):58.

25. Thompson WD, Tyrrell J, Borges M-C, Beaumont RN, Knight BA, Wood AR, et al. Association of maternal circulating 25 (OH) D and calcium with birth weight: A mendelian randomisation analysis. PLoS Med. 2019;16(6):e1002828.

26. Kamity R, Sharma S, Hanna N. MicroRNA-Mediated Control of Inflammation and Tolerance in Pregnancy. Front Immunol. 2019;10.

27. Li Q. Transforming growth factor β signaling in uterine development and function. J Anim Sci Biotechnol. 2014;5(1):52.

28. Rodriguez A, Tripurani SK, Burton JC, Clementi C, Larina I, Pangas SA. SMAD signaling is required for structural integrity of the female reproductive tract and uterine function during early pregnancy in mice. Biol Reprod. 2016;95(2):41–4.

29. Zhao W-X, Lin J-H. Notch signaling pathway and human placenta. Int J Med Sci. 2012;9(6):447.

30. Baldwin Jr AS. The NF-κB and IκB proteins: new discoveries and insights. Annu Rev Immunol. 1996;14(1):649–81.

31. Cookson VJ, Chapman R. NF-kB function in the human myometrium during pregnancy and parturition. Histol Histopathol. 2010;

32. Shamir ER, Ewald AJ. Adhesion in mammary development: novel roles for E-cadherin in individual and collective cell migration. In: Current topics in developmental biology. Elsevier; 2015. p. 353–82.

33. To B, Andrechek ER. Transcription factor compensation during mammary gland development in E2F knockout mice. PLoS One. 2018;13(4):e0194937.

34. Burroughs AM, Aravind L. Analysis of two domains with novel RNA-processing activities throws light on the complex evolution of ribosomal RNA biogenesis [Internet]. Vol. 5, Frontiers in Genetics . 2014. p. 424. Available from: https://www.frontiersin.org/article/10.3389/fgene.2014.00424

35. Salleh MS, Mazzoni G, Höglund JK, Olijhoek DW, Lund P, Løvendahl P, et al. RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high-and low-residual feed intake in Nordic dairy cattle. BMC Genomics. 2017;18(1):258.

36. Johansen M, Lund P, Weisbjerg MR. Feed intake and milk production in dairy cows fed different grass and legume species: A meta-analysis. animal. 2018;12(1):66–75.

37. Lee NK, kook Kim M, Choi JH, Kim EB, Lee HG, Kang SK, et al. Identification of a peptide sequence targeting mammary vasculature via RPLP0 during lactation. Peptides. 2010;31(12):2247–54.

38. Palombo V, Milanesi M, Sgorlon S, Capomaccio S, Mele M, Nicolazzi E, et al. Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using single nucleotide polymorphism arrays. J Dairy Sci. 2018 Dec 1;101(12):11004–19.

39. Raven L-A, Cocks BG, Pryce JE, Cottrell JJ, Hayes BJ. Genes of the RNASE5 pathway contain SNP associated with milk production traits in dairy cattle. Genet Sel Evol. 2013;45(1):25.

40. Farhadian M, Rafat SA, Hasanpur K, Ebrahimi M, Ebrahimie E. Cross-species meta-analysis of transcriptomic data in combination with supervised machine learning models identifies the common gene signature of lactation process. Front Genet. 2018;9:235.

41. Rutten MJM, Bouwman AC, Sprong RC, van Arendonk JAM, Visker MHPW. Genetic variation in vitamin B-12 content of bovine milk and its association with SNP along the bovine genome. PLoS One. 2013;8(4):e62382.

42. Peuhu E, Virtakoivu R, Mai A, Wärri A, Ivaska J. Epithelial vimentin plays a functional role in mammary gland development. Development. 2017;144(22):4103–13.

43. Michalczyk A, Ackland ML, Brown RW, Collins JP. Lactation affects expression of intermediate filaments in human breast epithelium. Differentiation. 2001;67(1–2):41–9.

44. Paten AM. Maternal nutritional programming in the sheep: effects on post-natal growth, mammogenesis and lactation in adult-ewe offspring: a thesis submitted for the degree of Doctor of Philosophy in Animal Science at Massey University, Palmerston North, Aotearoa Ne. Massey University; 2014.

45. Tiezzi F, Parker-Gaddis KL, Cole JB, Clay JS, Maltecca C. A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. PLoS One. 2015;10(2):e0114919.

46. Zhang J, Zan L, Fang P, Zhang F, Shen G, Tian W. Genetic variation of PRLR gene and association with milk performance traits in dairy cattle. Can J Anim Sci. 2008;88(1):33–9.

47. Arora R, Sharma A, Sharma U, Girdhar Y, Kaur M, Kapoor P, et al. Buffalo milk transcriptome: A

598 comparative analysis of early, mid and late lactation. Sci Rep. 2019;9(1):5993.

599 48. Larson BL, Smith VR. Lactation. A comprehensive treatise. Vol. II. Biosynthesis and secretion of
600 milk/Diseases. Academic Press.; 1974.

601 49. Abdollahi-Arpanahi R, Carvalho MR, Ribeiro ES, Peñagaricano F. Association of lipid-related genes
602 implicated in conceptus elongation with female fertility traits in dairy cattle. J Dairy Sci.
603 2019;102(11):10020–9.

604 50. Bhat SA, Ahmad SM, Ibeagha-Awemu EM, Bhat BA, Dar MA, Mumtaz PT, et al. Comparative
605 transcriptome analysis of mammary epithelial cells at different stages of lactation reveals wide differences in
606 gene expression and pathways regulating milk synthesis between Jersey and Kashmiri cattle. PLoS One.
607 2019;14(2):e0211773.

608 51. Bionaz M, Hurley W, Loor J. Milk protein synthesis in the lactating mammary gland: Insights from
609 transcriptomics analyses. Milk protein. 2012;285–324.

610 52. Morrison B, Cutler M Lou. The contribution of adhesion signaling to lactogenesis. J Cell Commun Signal.
611 2010;4(3):131–9.

612 53. Davis SR. TRIENNIAL LACTATION SYMPOSIUM/BOLFA: Mammary growth during pregnancy and
613 lactation and its relationship with milk yield. J Anim Sci. 2017;95(12):5675–88.

614 54. Gholizadeh M, Rahimi-Mianji G, Nejati-Javaremi A, De Koning DJ, Jonas E. Genomewide association
615 study to detect QTL for twinning rate in Baluchi sheep. J Genet. 2014 Aug;93(2):489–93.

616 55. Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM. GenABEL: an R library for genome-wide association
617 analysis. Bioinformatics. 2007;23(10):1294–6.

618 56. Rönnegård L, McFarlane SE, Husby A, Kawakami T, Ellegren H, Qvarnström A. Increasing the power of
619 genome wide association studies in natural populations using repeated measures–evaluation and
620 implementation. Methods Ecol Evol. 2016;7(7):792–9.

621 57. Misztal I, Tsuruta S, Lourenco DAL, Masuda Y, Aguilar I, Legarra A, et al. Manual for BLUPF90 family of

programs. 2018.

58. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005;21(16):3439–40.

59. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184.

60. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25(1):25.

61. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11(2):R14.

Figures

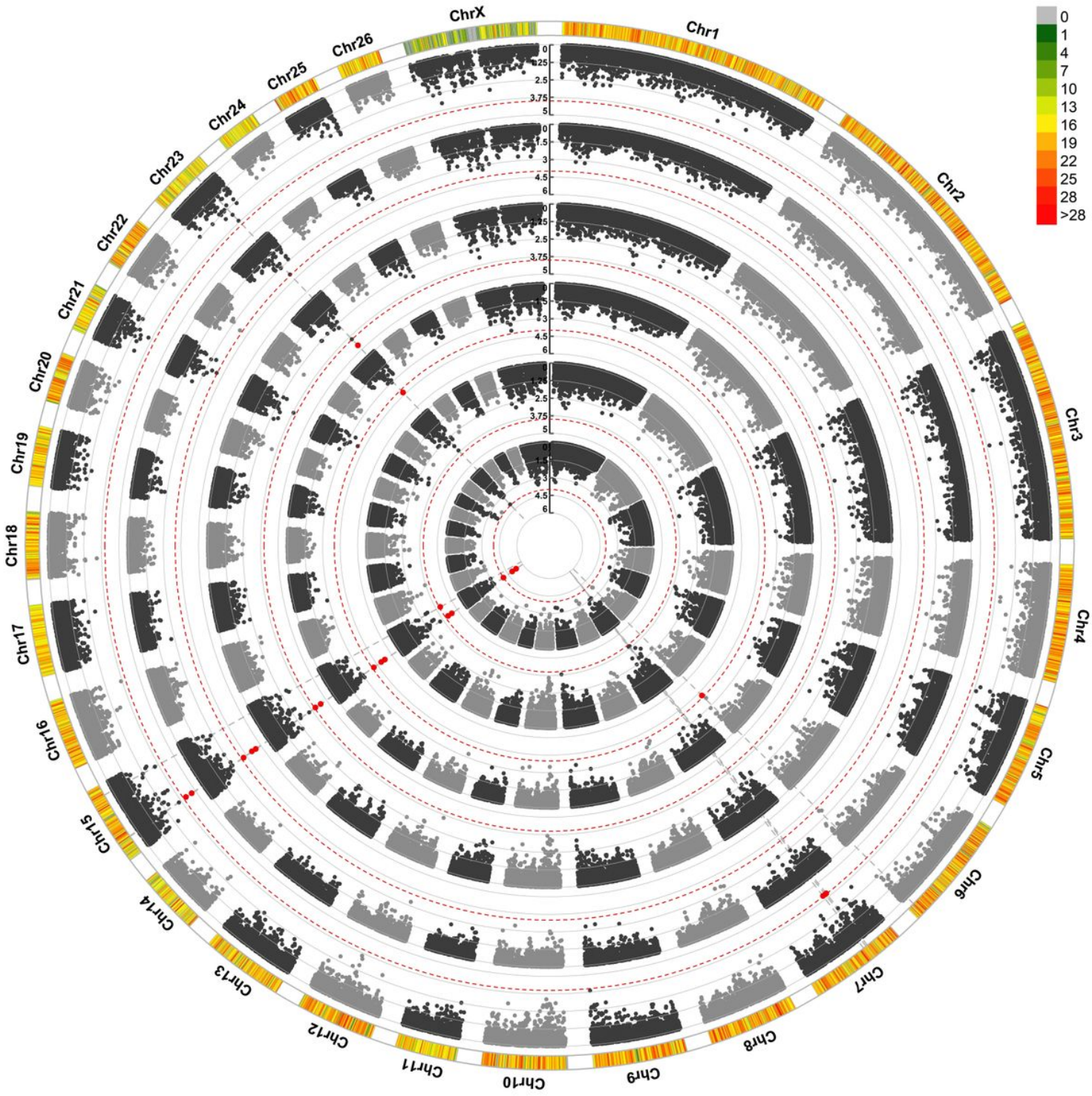


Figure 1

Circular Manhattan plot for associations of SNPs with ewe composite traits at birth by two GWAS approaches. The 6 circles from outside to inside represent Progeny Birth Weight (PBW): pGWAS and eGWAS; Total Litter Weight at Birth (TLWB): pGWAS and eGWAS; Litter Mean Weight at Birth (LMWB): pGWAS and eGWAS. X-axis: SNPs positions on chromosomes, Y-axis: -Log10 P-value. The dashed lines

indicate the threshold for statistical significance (P<10-4). The outermost circle show SNPs density in 1 Mb window for each chromosome. pGWAS: GWAS using phenotypes as response variable; eGWAS: GWAS using EBVs as response variable.

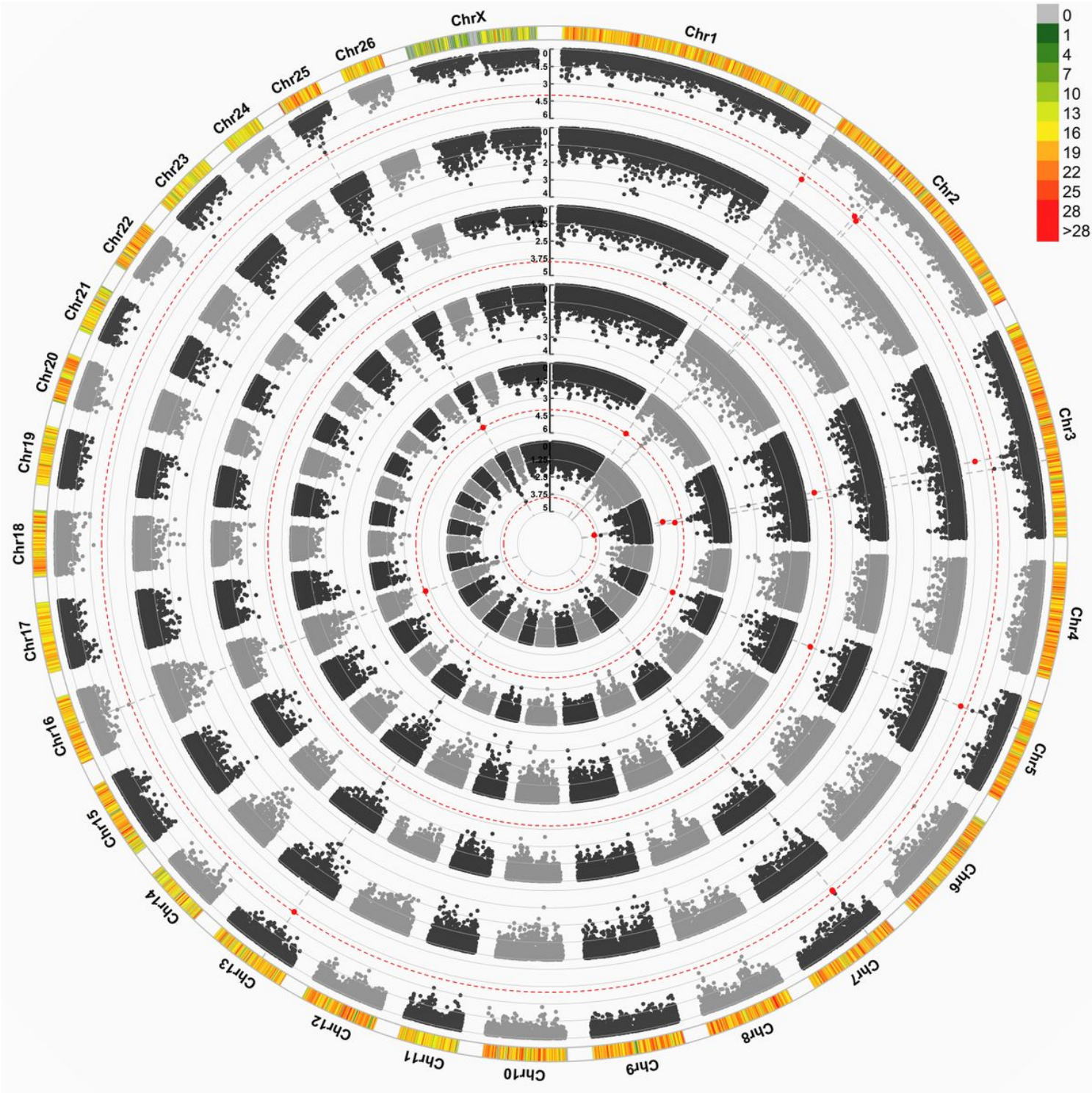


Figure 2

Circular Manhattan plot for associations of SNPs with ewe composite traits at weaning by two GWAS approaches. The 4 circles, from outside to inside, represent Progeny Weaning Weight (PWW): pGWAS and eGWAS; Total Litter Weight at Weaning (TLWW): pGWAS and eGWAS; Litter Mean Weight at Weaning

(LMWW): pGWAS and eGWAS. X-axis: SNPs positions on chromosomes, Y-axis: -Log10 P-value. The dashed lines indicate the threshold for statistical significance (P<10-4). The outermost circle show SNPs density in 1 Mb window for each chromosome. pGWAS: GWAS using phenotypes as response variable; eGWAS: GWAS using EBVs as response variable.

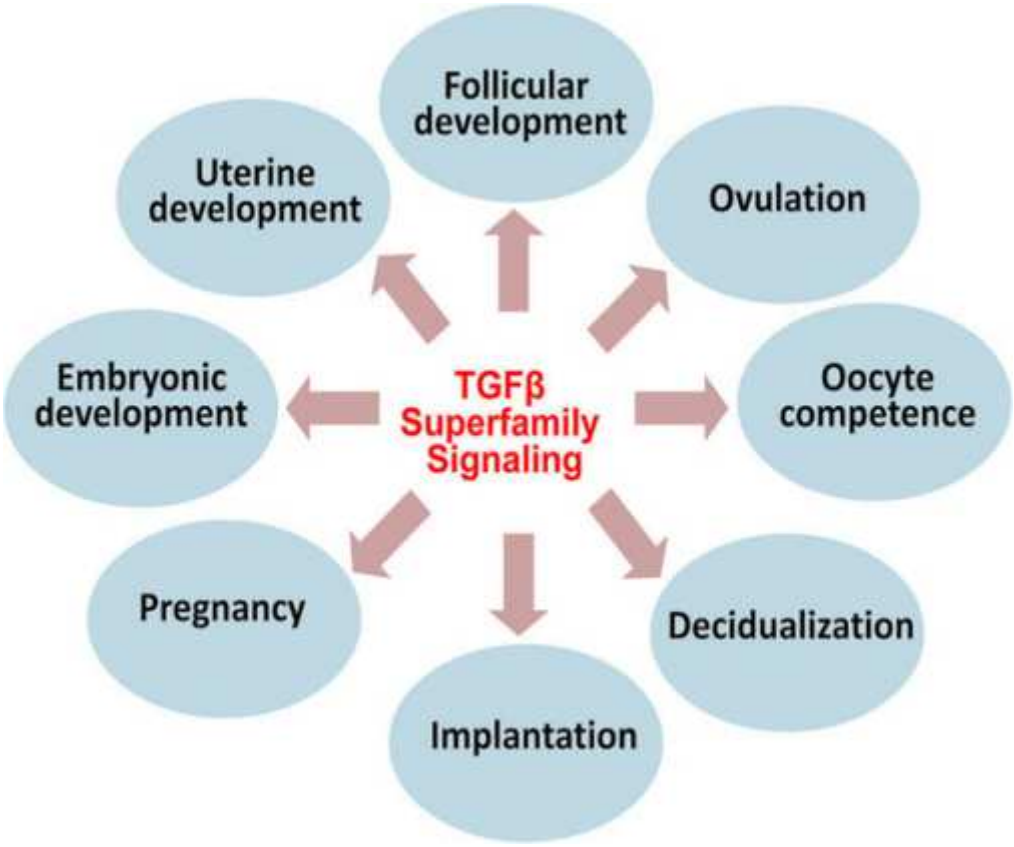


Figure 3

Main functions of TGF-β family signaling in female reproduction [27]. SMAD proteins transduce signals from TGF-β superfamily ligands.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.xlsx