ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi: 10.1111/age.12575

A genome-wide association study reveals candidate genes for the supernumerary nipple phenotype in sheep (*Ovis aries*)

W.-F. Peng*†[1], S.-S. Xu*†[1], X. Ren*‡[1], F.-H. Lv*, X.-L. Xie*†, Y.-X. Zhao*†, M. Zhang*§, Z.-Q. Shen¶, Y.-L. Ren¶, L. Gao††, M. Shen**††, J. Kantanen‡‡§§ and M.-H. Li*[a]**

*CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences (CAS), Beijing, 100101, China. †University of Chinese Academy of Sciences (UCAS), Beijing, 100049, China. ‡Annoroad Gene Technology Co. Ltd, Beijing, 100176, China. §School of Life Sciences, University of Science and Technology of China, Hefei, 230027, China. ¶Shandong Binzhou Academy of Animal Science and Veterinary Medicine, Binzhou, 256600, China. **Institute of Animal Husbandry and Veterinary Medicine, Xinjiang Academy of Agricultural and Reclamation Sciences, Shihezi, 832000, China. ††State Key Laboratory of Sheep Genetic Improvement and Healthy Breeding, Xinjiang Academy of Agricultural and Reclamation Sciences, Shihezi, 832000, China. ‡‡Green Technology, Natural Resources Institute Finland (Luke), Jokioinen, 31600, Finland. §§Department of Environmental and Biological Sciences, University of Eastern Finland, Kuopio, 70211, Finland.

Summary

Genome-wide association studies (GWASs) have been widely applied in livestock to identify genes associated with traits of economic interest. Here, we conducted the first GWAS of the supernumerary nipple phenotype in Wadi sheep, a native Chinese sheep breed, based on Ovine Infinium HD SNP BeadChip genotypes in a total of 144 ewes (75 cases with four teats, including two normal and two supernumerary teats, and 69 control cases with two teats). We detected 63 significant SNPs at the chromosome-wise threshold. Additionally, one candidate region (chr1: 170.723–170.734 Mb) was identified by haplotype-based association tests, with one SNP (rs413490006) surrounding functional genes *BBX* and *CD47* on chromosome 1 being commonly identified as significant by the two mentioned analyses. Moreover, Gene Ontology enrichment for the significant SNPs identified by the GWAS analysis was functionally clustered into the categories of receptor activity and synaptic membrane. In addition, pathway mapping revealed four promising pathways (Wnt, oxytocin, MAPK and axon guidance) involved in the development of the supernumerary nipple phenotype. Our results provide novel and important insights into the genetic mechanisms underlying the phenotype of supernumerary nipples in mammals, including humans. These findings may be useful for future breeding and genetics in sheep and other livestock.

Keywords breast cancer, genome-wide association studies, GO enrichment, pathway mapping, SNPs, Wadi sheep

Introduction

Teats (or nipples) are epidermal appendages on the udders or breasts of mammals (Pumfrey *et al.* 1980) and play an important role in female reproduction and offspring growth. However, the number of teats (NT) varies among and within species. For example, in pigs, NT ranges from 10 to 20 (Duijvesteijn *et al.* 2014; Arakawa *et al.* 2015), whereas sheep normally have only two teats. Earlier studies indicated that NT is correlated with reproductive traits, e.g. litter size, in mammals such as *Antechinus agilis* (Shimmin *et al.* 2000) and pigs (Arakawa *et al.* 2015), whereas in other mammals, e.g. golden hamsters and domestic rabbits, it has no influence on reproductive performance (Anderson & Sinha 1972; Fayeye & Ayorinde 2010).

In addition to different numbers of normal (primary) teats, some mammals, such as humans (Schmidt 1998) and cows (Yapp & St. Clair 1951), also have supernumerary teats (SNTs) beyond the quota. Supernumerary nipples are minor congenital malformations located along or beyond embryonic milk lines and vary from one to eight in humans

Address for correspondence

M.-H. Li, Institute of Zoology, Chinese Academy of Sciences, Beichen West Road No. 1-5, Chaoyang District, Beijing 100101, China.
E-mail: menghua.li@ioz.ac.cn

[1]These three authors contributed equally to this work and should be considered co-first authors.

Accepted for publication 13 May 2017

(Brown & Schwartz 2003). Although SNTs are benign abnormalities, they can be involved in processes such as renal and urinary tract malformations and are associated with diseases that affect normal breast tissues (Mehes 1979). In cattle, SNTs, which include caudal (at the rear of normal teats), intercalary (between normal teats) and ramal (ramification of normal teats) teats (Skjervold 1960), are undesirable due to their negative effect on machine milking and because they can act as a bacterial reservoir (Brka *et al.* 2002; Pausch *et al.* 2012). Normally sheep (and goats) have two teats, whereas a small number of ewes have one to four additional SNTs (Palacios & Abecia 2014), which are heritable and usually result in the removal of these animals from the dairy industry. Nevertheless, a recent investigation found that there were no differences in milk production between dairy sheep with two and four teats and suggested avoiding the removal of animals with SNTs (Palacios & Abecia 2014).

Earlier studies on teats have involved various traits such as length, size, inverted teat defects (Jonas *et al.* 2008), total number (Ding *et al.* 2009), and number on the left or right side (Toro *et al.* 1986). In livestock, the genetic mechanism underlying NT has been extensively studied in various pig breeds (Martínez-Giner *et al.* 2007; Martínez-Giner *et al.* 2011; Arakawa *et al.* 2015). Thus far, a total of 167 quantitative trait loci (QTL; see the pig QTL database: http://www.animalgenome.org) have been found to be associated with NT across the whole genome. Furthermore, a number of strong candidate genes for NT, e.g. *VRTN*, *Prox2*, *MPP7*, *ARMC4* and *MKX*, have been identified. All these genes have been shown to have effects on vertebral development (Ren *et al.* 2012; Duijvesteijn *et al.* 2014). Moreover, a dominant effect has been shown to play an important role in the genetic architecture determining NT in pigs (Lopes *et al.* 2014).

In humans, SNTs are thought to be atavistic structures that follow an autosomal dominant pattern of inheritance (Leung *et al.* 1988; Cellini & Offidani 1992). However, research on human SNTs has been relatively limited, possibly due to its minor clinical significance (Willman *et al.* 2003). In cattle, previous genetic studies have mostly focused on frequency and heritability (15–60%) of the SNT trait (Skjervold 1960; Brka *et al.* 2002). Only recently, a study uncovered several genes of the Wnt signalling pathway in four QTL regions, suggesting both an oligogenic and a polygenic inheritance mode for the trait in cattle (Pausch *et al.* 2012). The earliest investigation on SNT traits in sheep was conducted as early as 1981; in that study, no association was found between SNTs and multiple births (Oppong & Gumedze 1981). Later, a heritability of 14–26% was reported for the SNT trait, and it was concluded that dominant genes controlled the four-teat phenotype but a modifier caused the occurrence of more than four teats (Ritzman 1933). However, until now, no further genetic studies have been carried out on SNTs in sheep.

Wadi sheep, a descendant of Mongolian sheep, is a prolific native breed from the Shandong Province of China. In ewes of the Wadi sheep, NT ranges from two to six, including two normal teats and up to four SNTs. Two and four teats are the common phenotypes with an occurrence of 70–80% and 20–30% respectively (Fig. 1 & Table S2). For four-teated ewes, the two independent SNTs are smaller than the two normal teats, and some can produce milk as well.

In this study, for the first time, we conducted a GWAS to map candidate genes and genomic regions associated with the four-teated phenotype in sheep using the Ovine Infinium HD SNP BeadChip. Our results provide new insight into the genetic mechanisms of SNTs in sheep and other mammals, including humans, and will be useful for sheep breeding in the future.

Materials and methods

Sample collection and DNA extraction

A total of 144 Wadi sheep ewes (75 with two normal teats and two supernumerary nipples, hereafter referred to as 'four-teated', and 69 with two normal teats, hereafter referred to as 'two-teated') were sampled from the Binzhou Academy of Animal Sciences and Veterinary Medicine, Shandong Province, China. Ear marginal tissues were collected and stored in 2-ml microcentrifuge tubes containing 95% ethanol. We used the standard phenol–chloroform protocol to extract DNA from all tissue samples (Sambrook & Russell 2006). In all the cases, particular efforts were

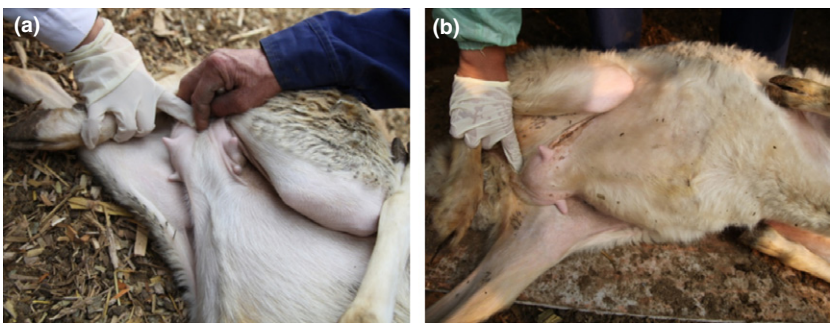

Figure 1 Wadi sheep with the phenotypes of (a) four teats including two normal teats and two supernumerary teats and (b) two normal teats.

made, based on both pedigree information and the knowledge of local herdsmen, to ensure that the animals were as distantly related as possible.

SNP genotyping and quality control

Genotyping for all the samples was implemented with the Illumina Ovine Infinium HD SNP BeadChip and yielded a dataset of 606 006 SNPs based on the manufacturer's protocol (Anderson *et al.* 2014; genotype and phenotype datasets No. ZOCN2017.0511, available from animalgenome.org data repository at https://www.animalgenome.org). Quality control of this SNP dataset was carried out with PLINK v.1.07 software (Purcell *et al.* 2007). The filtering criteria were as follows: (i) SNPs without chromosomal and physical locations, (ii) SNPs with missing genotypes greater than 0.05, (iii) SNPs of minor allele frequency less than 0.05, (iv) individuals with a genotyping rate less than 95%, and (v) a *P*-value of Fisher's exact test (Raymond & Rousset 1995a,b) for Hardy–Weinberg equilibrium less than 0.001. We removed the SNPs and individuals who met any of these criteria. Moreover, pairwise relatedness among individuals was examined using the KING v1.4 program (Manichaikul *et al.* 2010), and one of the pairwise individuals with a kinship coefficient $\Phi > 0.25$ (e.g. parent–offspring or full-sibs from two unrelated parents) was removed from further analyses. After filtering, 509 454 SNPs and 124 individuals (64 four-teated and 60 two-teated sheep) were retained in the dataset for within-population stratification analysis.

Within-breed stratification assessment

Within-breed genetic differentiation was measured by the global F_{ST} estimate (Weir & Cockerham 1984), which was calculated between the two- and four-teated groups using the GENEPOP v.4.4 program (Raymond & Rousset 1995a,b). Moreover, the dataset of 509 454 SNPs was pruned by implementing the command line of 'indep-pairwise 50 5 0.01' in PLINK v.1.07 software. In this procedure, if the linkage disequilibrium (LD) estimate $r^2 > 0.01$, one of the SNPs from the pair would be removed. A total of 10 985 independent SNPs retained by the LD criteria were used to assess within-breed stratification by multidimensional scaling (MDS) analysis using the GENABEL package for R v.3.2.3 (Aulchenko *et al.* 2007). In the subsequent GWAS analysis, the genomic inflation factor lambda (λ) was calculated as an indicator of population substructure and/or technical artefacts before and after adjustment (Price *et al.* 2006; Aulchenko *et al.* 2007).

Genome-wide association analysis

A genome-wide association analysis was conducted with the GENABEL package based on the case–control model (Aulchenko *et al.* 2007). We tested for the association and calculated the significance *P*-value by implementing the 'qtscore' function and adjusting for population substructure using the first two MDS dimensions (Fig. S1). Given the estimated number of independent SNPs ($n = 10\,985$) based on the LD analysis performed with PLINK v.1.07 software, the thresholds of genome-wide ($P_{genome} < 0.05/10\,985 = 4.55 \times 10^{-6}$) and chromosome-wise significance ($P_{chromosome} < 0.05$/number of independent SNPs at each chromosome; Table S1) were determined by the Bonferroni correction at the empirical level of 0.05 (Yang *et al.* 2005). In addition, the false positive rate (FPR) was calculated by a no-replacement resampling test (1000 times), as detailed in Ren *et al.* (2016).

Linkage disequilibrium and haplotype-based association test

Based on the r^2 statistic (Hill & Weir 1994) estimated using the HAPLOVIEW v.4.2 program (Barrett *et al.* 2005), a study of the extent of LD was performed on the candidate region identified by the GWAS. In this case, we ignored pairwise comparison of markers more than 500 kb apart and excluded individuals with more than 50 missing genotypes (Barrett *et al.* 2005). The LD blocks were determined using the algorithm of confidence intervals suggested by Gabriel *et al.* (2002). In addition, we tested the pairwise tests of LD for the most significant SNP (see Results) in the region of 1 Mb upstream and downstream as identified by the GWAS analysis. The LD values were estimated using PLINK v1.07 software. The LD values and related genes were represented graphically using R v3.2.2 (Fig. 2b). Furthermore, we conducted a haplotype-based association test to detect the target haplotype significantly associated with the four-teated phenotype on chromosome 1. This method was detailed in Ren *et al.* (2016). For this analysis, the threshold for the chromosome-wise was obtained after 100 000 permutations ($P_{100\,000} < 0.05$) using HAPLOVIEW v.4.2 (Barrett *et al.* 2005).

Bioinformatics analysis

Functional genes were annotated using the *Ovis aries* assembly Oar_v.4.0 (http://www.ncbi.nlm.nih.gov/genome?term=ovis%20aries) and other databases such as UCSC Genome Bioinformatics (http://genome.ucsc.edu/) and UniProt (http://www.uniprot.org/). Additionally, QTL information from sheep, cattle and pig were searched in AnimalQTLdb (http://www.animalgenome.org/cgi-bin/QTLdb/index). Furthermore, Gene Ontology (GO) enrichment was implemented to retrieve functional information of the functional genes identified by the GWAS using Panther (http://pantherdb.org/). The default options of significance threshold of 0.05 for adjusted *P*-value and at least two genes from the input gene list in the enriched category were applied. Also, we mapped the common pathways for the

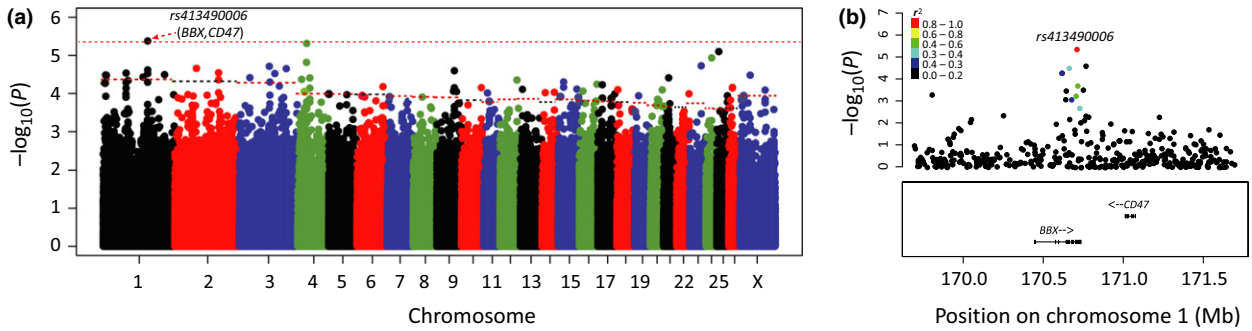

Figure 2 Genome-wide association test for supernumerary nipple phenotypes in Wadi sheep. (a) Manhattan plot based on the GWAS method of the candidate genes listed in Table 1. The 5% genome-wide significant threshold value ($P_{genome} = 4.551 \times 10^{-6}$) is shown as a red dashed line; also, the 5% chromosome-wise significant threshold value is presented on each chromosome. (b) Plot of regional association results for the most significant SNP (rs413490006) is shown as a red dot. Different colours represent the r^2 values of pair-wise LD estimates. Functional genes in this region are plotted in the box.

candidate genes at the chromosome-wise level as identified by the GWAS analysis.

Results

Within-breed population stratification and candidate SNPs by GWAS

We observed a global F_{ST} value of 0.0012, indicating a very low level of genetic differentiation between the two- and four-teated ewes. No individual was removed based on the population stratification analysis, and finally, a total of 509 454 SNPs and 124 individuals (64 four-teated and 60 two-teated ewes) were kept in the working dataset for the GWAS. We obtained a raw inflation factor of 1.15 and a corrected inflation factor of 1.06 after the correction using the first two dimensions (Fig. S2). Only the most significant SNP, rs413490006, located at 170 729 111 bp on OAR1 ($P = 4.18 \times 10^{-6}$), exceeded the genome-wide significance threshold, whereas a total of 63 SNPs, located across 26 different chromosomes, reached the chromosome-wise significance level (Table 1, Table S1 & Fig. 2a). All SNPs had low FPR values (FPR < 0.001) after the bootstrapping test.

Linkage disequilibrium and haplotype-based association test

In the haplotype-based association test, a total of 1403 blocks (3588 haplotypes) were defined in the genomic region (chr 1:170.723–170.734 Mb) containing the most significant SNP, rs413490006. After 100 000 permutations for the chi-square test, only one haplotype [block 682 on chromosome 1 (AGA)] was statistically significant ($P_{100\,000} = 0.0051$) at the chromosome-wise level. The most significant SNP (rs413490006) was included in the significant haplotype, which had a frequency of 0.977 in the case population (four-teated) versus 0.775 in the control population (two-teated). In addition, two genes, *BBX* and *CD47*, were found to be located near the rs413490006 SNP, the most significant SNP on chromosome 1 (Fig. 2b).

Gene annotation

A set of functional genes identified by the GWAS, such as *BBX*, *SEMA3D*, *IRF2BP2*, *SETBP1*, *NAV3*, *CD47*, *LRP1B*, *CERS5*, *CASK* and *CADM2* (Table 1), are involved in the process of cell proliferation and breast cancer and may be potential candidate genes for the supernumerary nipple phenotype. The most significant SNP, rs413490006, is located within the *BBX* gene, which encodes a transcription factor containing high-mobility group (HMG)-box domain and a non-acidic C-terminus and functions as alternative transcripts in human breast cancer (Wen *et al.* 2015). Moreover, CD47, encoded by a gene (*CD47*) that neighbours *BBX*, has been shown to be an active signalling receptor in triple-negative human breast carcinoma cells and has a key role in inhibiting breast cancer cell growth in combination with the antibody B6H12 (Kaur *et al.* 2016). The *SEMA3D* gene is a signalling molecule and is related to thymic development (Berndt & Halloran 2006; Takahashi *et al.* 2008). *IRF2BP2* is a transcriptional repressor and functions in the apoptosis of breast cancer cells by interacting with *EAP1*, *FASTKD2* and *NRIF3* (Yeung *et al.* 2011). *SETBP1* has been identified as a target gene for breast cancer using genome-wide association analysis in humans and has the ability to bind nuclear oncogenes to regulate the development of breast cancer (Michailidou *et al.* 2015). *NAV3* has been identified as a suppressor of breast cancer progression (Cohen-Dvashi *et al.* 2015). *LRP1B* is a gene that functions to reduce cell differentiation of malignant tumours in breast cancer cell lines (Kadota *et al.* 2010). *CERS5* belongs to the ceramide synthase (CerS) family, the genes of which encode important enzymes that control the lengths of ceramides; the overexpression of *CERS5* shows an important role in mediating cell

Table 1 Bonferroni-corrected genome-wide and chromosome-wise significant SNPs and their nearest gene based on the genome-wide association study method.

SNP	Chr.	Position	*P*-value	Gene
rs413490006	1	170 729 111	4.18E-06	*BBX*[1], *CD47*
rs426598066	1	170 784 617	2.40E-05	*BBX*, *CD47*
rs419135687	1	87 817 922	2.89E-05	*OVGP1*[1]
rs430157497	1	170 681 957	3.00E-05	*BBX*[1]
rs406719290	1	236 121 673	3.19E-05	*RNF13*[1]
rs421959057	1	8 932 317	3.26E-05	*C1H1orf94*, *GJB5*
rs429495552	1	152 774 943	3.70E-05	*CADM2*, *VGLL3*
rs402332677	2	83 476 507	2.18E-05	*PSIP1*, *CCDC171*
rs402628430	2	168 692 818	2.85E-05	*LRP1B*[1]
rs419995263	2	168 697 208	4.35E-05	*LRP1B*[1]
rs411465082	3	114 588 115	1.93E-05	*NAV3*, *SYT1*
rs425157991	3	179 817 024	2.21E-05	*PVALB7*[1]
rs413534072	3	135 823 364	2.99E-05	*CERS5*[1]
rs421554889	3	39 874 992	3.82E-05	*CNRIP1*, *PPP3R1*
rs424780352	3	116 922 437	5.00E-05	*ACSS3*[1]
rs418633727	4	34 203 218	4.85E-06	*GRM3*, *SEMA3D*
rs401595294	4	33 560 821	1.51E-05	*GRM3*[1]
rs159766665	4	46 318 235	3.88E-05	*SRPK2*[1]
rs418160769	4	15 010 202	4.24E-05	*ASNS*[1]
rs159766635	4	46 311 840	7.22E-05	*SRPK2*[1]
rs405311330	4	25 385 189	8.87E-05	*AGR2*[1]
rs417066429	5	794 298	1.01E-04	*RASGEF1C*[1]
rs417070058	6	100 512 609	6.58E-05	MAPK10[1]
rs429010450	8	34 762 269	1.23 E-04	*HACE1*, *GRIK2*
rs418336206	9	67 419 429	2.52E-05	*PKHD1L1*[1]
rs160661467	9	67 419 502	7.03E-05	*PKHD1L1*[1]
rs414156085	9	64 120 141	8.65E-05	*CSMD3*[1]
rs422047056	9	75 493 735	8.89E-05	*GRHL2*, *ZNF706*
rs424954193	10	76 605 939	7.00E-05	*TMTC4*, *NALCN*
rs413243007	11	13 838 117	9.65E-05	*RPL13*, *WFDC18*
rs160810084	12	65 020 161	4.42E-05	*HMCN1*[1]
rs417649440	13	7 701 871	7.75E-05	*MACROD2*[1]
rs424744512	14	47 390 262	9.26E-05	*RYR1*[1]
rs427349812	14	12 438 420	9.48E-05	*ZCCHC14*[1]
rs428830693	15	20 439 653	4.97E-05	*ARHGAP20*
rs429974933	15	20 441 216	6.83E-05	*ARHGAP20*
rs424216758	15	71 269 497	7.66E-05	*LRRC4C*, *API5*
rs421389192	15	7 486 466	1.18E-04	*PGR*[1]
rs423920180	15	77 935 764	1.24E-04	*MED19A*[1]
rs404578142	16	67 802 754	5.68E-05	*ADAMTS16*[1]
rs401718234	16	13 753 410	8.66E-05	*NLN*[1]
rs417662713	17	16 615 839	5.86E-05	*RNF150*, *TBC1D9*
rs400698889	17	64 669 244	1.02E-04	*WSCD2*, *PIWIL3*
rs415997036	17	51 773 214	1.31E-04	*RILPL2*, *KMT5A*
rs414657709	17	17 090 321	1.33E-04	*SCOC-AS1*[1]
rs423860641	18	4 166 451	7.92E-05	*GABRG3*[1]
rs409202295	19	15 627 676	1.14E-04	*ABHD5*[1]
rs400264921	19	15 628 123	1.14E-04	*ABHD5*[1]
rs409023706	20	45 926 523	8.02E-05	*LOC100506207*, *SLC35B3*
rs404024164	20	28 012 918	1.41E-04	*UBD*, *OR2H1*
rs418795725	20	32 403 879	1.51E-04	*NRSN1*, *PRP2*
rs405017021	21	19 650 722	3.87E-05	*LUZP2*[1]
rs428508308	22	47 754 858	1.01E-04	*MKI67*, *MGMT*
rs428451278	23	43 938 141	1.86E-05	*MC2R*, *SETBP1*
rs412397177	24	21 939 056	1.15E-05	*CACNG3*[1]
rs400876130	25	7 174 894	7.93E-06	*SLC35F3*, *IRF2BP2*
rs399750484	25	40 192 777	1.14 E-04	*GRID1*[1]
rs414353864	25	40 366 494	1.67E-04	*GRID1*[1]
rs424985719	26	14 510 418	6.90E-05	*SORBS2*[1]
rs418141809	26	14 541 179	7.12E-05	*SORBS2*[1]

Table 1 (Continued)

SNP	Chr.	Position	*P*-value	Gene
rs417322179	26	6 922 551	1.48E-04	*VEGFC*, *NEIL3*
rs413478104	X	38 618 849	3.29E-05	*NYX*, *CASK*
rs420032354	X	95 998 688	8.10E-05	*GPC3*[1]

Chr., chromosome. The *P*-value represents the corrected significance.
The false positive rate for all genes was less than 0.001 (FPR < 0.001).
[1]Genes for which the SNP is intragenic, otherwise the gene is the nearest gene upstream and downstream of the tested SNPs.

proliferation and tumour development in breast cancer cells (Wegner *et al.* 2014). *CASK* has been found to be an important gene of docetaxel resistance by reducing MCF-7 resistant cells in breast cancer cells (Hansen *et al.* 2016). *CADM2*, which has been identified to be associated with breast cancer, is a cell adhesion molecule and plays an important role in the maintenance of cell polarity and tumour suppression (He *et al.* 2013). In addition, *GRM3*, which is relevant to the rs418633727 and rs401595294 SNPs, was found to be related to thyroid cancer (Murugan *et al.* 2013).

Gene Ontology enrichment and pathway mapping

We performed GO enrichment analysis of the functional genes surrounding the significant SNPs at the chromosome-wise level. GO enrichment revealed that these genes were involved in six categories of molecular function, 11 biological process categories and four cellular components (Fig. 3 & Table S3). Significant categories with the number of genes were receptor activity ($n = 3$), synaptic membrane ($n = 7$) in the two structured ontologies (Fig. 3 & Table S3). Three genes (*GRM3*, *GRID1* and *GRIK2*) identified as significant by the GWAS analysis were involved in the two enriched categories (Table S3).

Based on the set of functional genes identified by the GWAS at the chromosome-wise significance level, we also identified four signalling pathways (Wnt, oxytocin, MAPK and axon guidance) that may be related to the supernumerary nipple phenotype through the KEGG pathway mapping (Fig. 4). Six genes—*CACNG3*, *PPP3R1* (also known as *WDR92*), *RYR1*, *MAPK10*, *SEMA3D* and *LRRC4C*—were shown to be involved in one or several of the four pathways (Fig. 4). *PPP3R1* was common in all four pathways, whereas *MAPK10* and *CACNG3* were involved in two of the four pathways. *PPP3R1* encodes calmodulin-regulated protein phosphatase, which has a key role in many signalling pathways (Wang *et al.* 1996). *CACNG3*, which encodes a voltage-dependent calcium channel subunit, was identified as a new cancer target gene in humans (Kumar *et al.* 2015). *MAPK10* was identified to be related to the development of breast cancer by regulating the activity

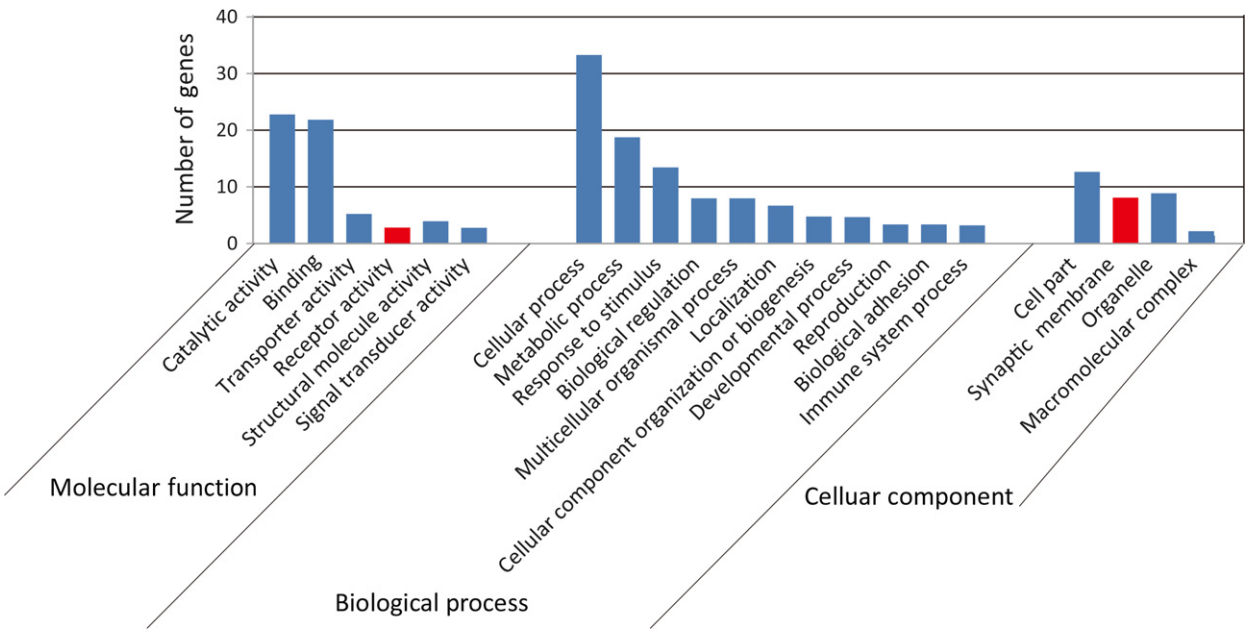

Figure 3 Gene Ontology enrichment analysis based on the candidate functional genes identified by the significant SNPs at the chromosome-wise levels. The significant (adjusted *P*-value <0.05) categories are highlighted in red.

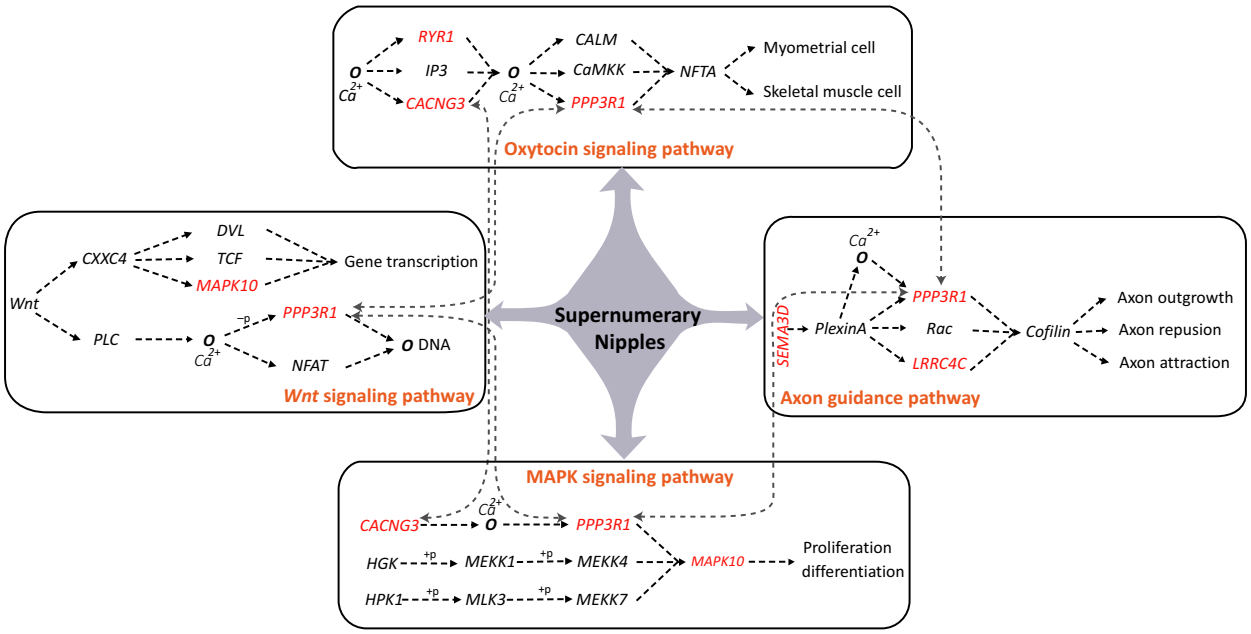

Figure 4 Underlying biological pathways for the supernumerary phenotype in Wadi sheep based on candidate genes. Grey arrows point to the four functional KEGG pathways, whose names are shown in orange. The candidate genes identified by the GWAS are shown in red.

of the mammalian target of rapamycin (mTOR) inhibitor everolimus (Hurvitz *et al.* 2015) and is included in the Wnt and MAPK signalling pathways.

Discussion

In this study, we performed the first GWAS of the SNT trait in sheep to date. By combining a complementary method, i.e. a haplotype-based association test, we herein found some functional genes containing significant SNPs at the chromosome-wise level and four signalling pathways accounting for the variation between the two- and four-teated phenotypes (Figs S3 & 4). Additionally, our results suggest several additional functional genes surrounding the significant SNPs, which may play marginal roles in determining phenotypes.

We detected only one SNP (rs413490006) on chromosome 1 that surpassed the threshold of the genome-wide level determined by the Bonferroni method. Moreover, 63 SNPs, including the most significant SNP rs413490006, were significant ($P_{\text{chromosome}} < 0.05$) at the chromosome-wise level after Bonferroni correction. The candidate genomic region on chromosome 1 (OAR1:170.723–170.734 Mb) most probably contained the causative mutations for the four-teated phenotype. Several possible reasons for many SNPs undetected at a genome-wide significance threshold ($P_{\text{genome}} < 0.05$) include the following: (i) as a complex trait, the SNT trait may be dominated by variants undetected at the stringent genome-wide significance level (Chatterjee *et al.* 2013); (ii) the case–control model may be less powerful for identifying candidate genes for the SNT trait, that have not undergone strong artificial selection (Palacios & Abecia 2014); and (iii) the sample size in this study was not large enough to provide strong enough power to identify numerous candidate genes associated with the SNT trait (Kemper *et al.* 2014). We observed small estimates of the inflation factors lambda (λ), indicating that within-breed stratification was properly adjusted and that small, if any, impact of population substructure on the detection of false positive association would be expected.

In this study, we discovered a variety of functional genes and signalling pathways associated with the SNT trait in sheep. The *BBX* gene on chromosome 1, which was simultaneously identified by the GWAS and haplotype-based association test, has been found to be involved in the development of breast cancer (Wen *et al.* 2015). GO enrichments revealed that a series of functional genes may have a role in the development of the four-teated phenotype in sheep. For example, *CASK* has been reported to be a risk factor associated with breast cancer development (Hansen *et al.* 2016). In addition, four signalling pathways (Wnt, oxytocin, MAPK and axon guidance) were identified to be associated with the SNT trait in sheep. The Wnt signalling pathway has been identified to be related to mammary development and breast cancer (Yu *et al.* 2016), the oxytocin signalling pathway is an important biological pathway in cancer cell proliferation (Cassoni *et al.* 2004), the MAPK signalling pathway plays a key role in regulating breast carcinomas (Chen *et al.* 2015) and the axon guidance signalling pathway is an important pathway involved in the morphogenesis of several tissues and the development of diverse organs (Hinck 2004).

Our results revealed that the SNT trait in sheep is a complex polygenic trait and is controlled by a set of complicated regulated networks. However, a number of genes associated with NT in cattle and pigs did not match our results (Guo *et al.* 2008; Jonas *et al.* 2008; Pausch *et al.* 2012; Hernandez *et al.* 2014; Joerg *et al.* 2014; Arakawa *et al.* 2015; Verardo *et al.* 2015). In our GWAS, we focused on SNTs, which may differ from NT in cattle and pigs. In addition, different species present other influential factors to consider. Furthermore, polymorphisms in candidate genes may not always show an effect large enough to be captured by a GWAS (Van Ingen *et al.* 2016).

In conclusion, our study represents the first genomic attempt to identify candidate genes for the SNT phenotype in sheep. We revealed that this phenotype in sheep might be mediated through the regulation of cell proliferation and that it shares signalling pathways involved in breast cancer development. Future work including other breeds and functional validation, such as gene knockout of the candidate genes and particularly promising candidate genes, such as *BBX*, *CD47* and *CASK*, are warranted. Our results provide additional insights into the genetics and biology of the SNT phenotype in human women as well as the evolution and maintenance of the polymorphism of complex phenotypic traits in sheep and other mammals.

Acknowledgements

This work was supported by the Taishan Scholars Program of Shandong Province (No. ts201511085), the External Cooperation Program of the Chinese Academy of Sciences (152111KYSB20150010), the National High Technology Research and Development Program of China (863 Program, grant No. 2013AA102506), grants from the National Natural Science Foundation of China (grant nos. 31272413 and U1303284), the National Transgenic Breeding Project of China (2014ZX0800952B) and the Academy of Finland (grant nos. 250633 and 256077).

References

Anderson R.R. & Sinha K.N. (1972) Number of mammary glands and litter size in the golden hamster. *Journal of Mammalogy* **53**, 382–4.

Anderson R., McEwan J., Brauning R. *et al.* (2014) Development of a high density (600K) Illumina Ovine SNP chip and its use to fine map the yellow fat locus. *Proceedings of the Plant and Animal Genome XXII Conference*, San Diego, CA.

Arakawa A., Okumura N., Taniguchi M., Hayashi T., Hirose K., Fukawa K., Ito T., Matsumoto T., Uenishi H. & Mikawa S. (2015) Genome-wide association QTL mapping for teat number in a purebred population of Duroc pigs. *Animal Genetics* **46**, 571–5.

Aulchenko Y.S., Ripke S., Isaacs A. & Van Duijn C.M. (2007) GENABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–6.

Barrett J.C., Fry B., Maller J. & Daly M.J. (2005) HAPLOVIEW: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–5.

Berndt J.D. & Halloran M.C. (2006) Semaphorin 3d promotes cell proliferation and neural crest cell development downstream of TCF in the zebrafish hindbrain. *Development* **133**, 3983–92.

Brka M., Reinsch N. & Kalm E. (2002) Frequency and heritability of supernumerary teats in German Simmental and German Brown Swiss cows. *Journal of Dairy Science* **85**, 1881–6.

Brown J. & Schwartz R.A. (2003) Supernumerary nipples: an overview. *Cutis* **71**, 344–6.

Cassoni P., Sapino A., Marrocco T., Chini B. & Bussolati G. (2004) Oxytocin and oxytocin receptors in cancer cells and proliferation. *Journal of Neuroendocrinology* **16**, 362–4.

Cellini A. & Offidani A. (1992) Familial supernumerary nipples and breasts. *Dermatology* **185**, 56–8.

Chatterjee N., Wheeler B., Sampson J., Hartge P., Chanock S.J. & Park J.H. (2013) Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nature Genetics* **45**, 400–5.

Chen X.Y., Zhou J., Luo L.P., Han B., Li F., Chen J.Y., Zhu Y.F., Chen W. & Yu X.P. (2015) Black rice anthocyanins suppress metastasis of breast cancer cells by targeting RAS/RAF/MAPK pathway. *BioMed Research International* **2015**, 1–11.

Cohen-Dvashi H., Ben-Chetrit N., Russell R. *et al.* (2015) Navigator-3, a modulator of cell migration, may act as a suppressor of breast cancer progression. *EMBO Molecular Medicine*, **7**, 299.

Ding N., Guo Y., Knorr C. *et al.* (2009) Genome-wide QTL mapping for three traits related to teat number in a White Duroc × Erhualian pig resource population. *BMC Genetics* **10**, 6.

Duijvesteijn N., Veltmaat J.M., Knol E.F. & Harlizius B. (2014) High-resolution association mapping of number of teats in pigs reveals regions controlling vertebral development. *BMC Genomics* **15**, 542.

Fayeye T. & Ayorinde K. (2010) Effects of season, generation, number of mating, parity and doe number of teat on doe and litter birth characteristics in domestic rabbit. *Research Journal of Animal and Veterinary Sciences* **5**, 6–9.

Gabriel S.B., Schaffner S.F., Nguyen H. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science* **296**, 2225–9.

Guo Y.M., Lee G.J., Archibald A.L. & Haley C.S. (2008) Quantitative trait loci for production traits in pigs: a combined analysis of two Meishan × Large White populations. *Animal Genetics* **39**, 486–95.

Hansen S.N., Ehlers N.S., Zhu S. *et al.* (2016) The stepwise evolution of the exome during acquisition of docetaxel resistance in breast cancer cells. *BMC Genomics* **17**, 442.

He W., Li X.S., Xu S.P. *et al.* (2013) Aberrant methylation and loss of *CADM2* tumor suppressor expression is associated with human renal cell carcinoma tumor progression. *Biochemical and Biophysical Research Communications* **435**, 526–32.

Hernandez S.C., Finlayson H.A., Ashworth C.J., Haley C.S. & Archibald A.L. (2014) A genome-wide linkage analysis for reproductive traits in F2 Large White × Meishan cross gilts. *Animal Genetics* **45**, 191–7.

Hill W.G. & Weir B.S. (1994) Maximum-likelihood estimation of gene location by linkage disequilibrium. *American Journal of Human Genetics* **54**, 705–14.

Hinck L. (2004) The versatile roles of 'axon guidance' cues in tissue morphogenesis. *Developmental Cell* **7**, 783–93.

Hurvitz S.A., Kalous O., Conklin D. *et al.* (2015) *In vitro* activity of the mTOR inhibitor everolimus, in a large panel of breast cancer cell lines and analysis for predictors of response. *Breast Cancer Research and Treatment* **149**, 669–80.

Joerg H., Meili C., Ruprecht O., Bangerter E., Burren A. & Bigler A. (2014) A genome-wide association study reveals a QTL influencing caudal supernumerary teats in Holstein cattle. *Animal Genetics* **45**, 871–3.

Jonas E., Schreinemachers H.J., Kleinwachter T. *et al.* (2008) QTL for the heritable inverted teat defect in pigs. *Mammalian Genome* **19**, 127–38.

Kadota M., Yang H.H., Gomez B., Sato M., Clifford R.J., Meerzaman D., Dunn B.K., Wakefield L.M. & Lee M.P. (2010) Delineating genetic alterations for tumor progression in the MCF10A series of breast cancer cell lines. *PLoS One* **5**, e9201.

Kaur S., Elkahloun A.G., Singh S.P. *et al.* (2016) A function-blocking CD47 antibody suppresses stem cell and EGF signaling in triple-negative breast cancer. *Oncotarget* **7**, 10133–52.

Kemper K.E., Saxton S.J., Bolormaa S., Hayes B.J. & Goddard M.E. (2014) Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics* **15**, 246.

Kumar R.D., Searleman A.C., Swamidass S.J., Griffith O.L. & Bose R. (2015) Statistically identifying tumor suppressors and oncogenes from pan-cancer genome-sequencing data. *Bioinformatics* **31**, 3561–8.

Leung A.K., Opitz J.M. & Reynolds J.F. (1988) Familial supernumerary nipples. *American Journal of Medical Genetics* **31**, 631–5.

Lopes M.S., Bastiaansen J.W., Harlizius B., Knol E.F. & Bovenhuis H. (2014) A genome-wide association study reveals dominance effects on number of teats in pigs. *PLoS One* **9**, e105867.

Manichaikul A., Mychaleckyj J.C., Rich S.S., Daly K., Sale M. & Chen W.M. (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–73.

Martínez-Giner M., Noguera J., Ramirez O., Alves E. & Pena R. (2007) Positive association between porcine PTHLH gene and teat number in a F-2 Meishan and Iberian crossbreed. *Journal of Dairy Science.* **90**, 14–5.

Martínez-Giner M., Noguera J.L., Balcells I., Alves E., Varona L. & Pena R.N. (2011) Expression study on the porcine PTHLH gene and its relationship with sow teat number. *Journal of Animal Breeding and Genetics.* **128**, 344–53.

Mehes K. (1979) Association of supernumerary nipples with other anomalies. *Journal of Pediatrics* **95**, 274–5.

Michailidou K., Beesley J., Lindstrom S. *et al.* (2015) Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nature Genetics* **47**, 373–81.

Murugan A.K., Yang C.F. & Xing M.Z. (2013) Mutational analysis of the *GNA11*, *MMP27*, *FGD1*, *TRRAP* and *GRM3* genes in thyroid cancer. *Oncology Letters* **6**, 437–41.

Oppong E.N. & Gumedze J.S. (1981) Supernumerary teats in Ghanaian livestock. I. Sheep and goats. *Beitrage zur tropischen Landwirtschaft und Veterinarmedizin* **20**, 63–7.

Palacios C. & Abecia J.A. (2014) Supernumerary teat removal can be avoided in dairy sheep. *Journal of Applied Animal Welfare Science* **17**, 178–82.

Pausch H., Jung S., Edel C., Emmerling R., Krogmeier D., Gotz K.U. & Fries R. (2012) Genome-wide association study uncovers four QTL predisposing to supernumerary teats in cattle. *Animal Genetics* **43**, 689–95.

Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A. & Reich D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904–9.

Pumfrey R.A., Johnson R.K., Cunningham P.J. & Zimmerman D.R. (1980) Inheritance of teat number and its relationship to maternal traits in swine. *Journal of Animal Science* **50**, 1057–60.

Purcell S., Neale B., Todd-Brown K. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559–75.

Raymond M. & Rousset F. (1995a) An exact test for population differentiation. *Evolution* **49**, 1280–3.

Raymond M. & Rousset F. (1995b) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248–9.

Ren D.R., Ren J., Ruan G.F., Guo Y.M., Wu L.H., Yang G.C., Zhou L.H., Li L., Zhang Z.Y. & Huang L.S. (2012) Mapping and fine mapping of quantitative trait loci for the number of vertebrae in a White Duroc × Chinese Erhualian intercross resource population. *Animal Genetics* **43**, 545–51.

Ren X., Yang G.L., Peng W.F., Zhao Y.X., Zhang M., Chen Z.H., Wu F.A., Kantanen J., Shen M. & Li M.H. (2016) A genome-wide association study identifies a genomic region for the polycerate phenotype in sheep (*Ovis aries*). *Scientific Reports* **6**, 21111.

Ritzman E.G. (1933) The multinipple trait in sheep and its inheritance. *New Hampshire Agricultural Experiment Station Technical Bulletins* **53**, 32.

Sambrook J. & Russell D.W. (2006) *The Condensed Protocols from Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York, NY.

Schmidt H. (1998) Supernumerary nipples: prevalence, size, sex and side predilection: a prospective clinical study. *European Journal of Pediatrics* **157**, 821–3.

Shimmin G.A., Taggart D.A. & Temple-Smith P.D. (2000) Variation in reproductive surpluses of the agile antechinus (*Antechinus agilis*) at different teat-number locations. *Australian Journal of Zoology* **48**, 511–7.

Skjervold H. (1960) Supernumerary teats in cattle. *Hereditas* **46**, 71–4.

Takahashi K., Ishida M., Hirokawa K. & Takahashi H. (2008) Expression of the semaphorins *Sema 3D* and *Sema 3F* in the developing parathyroid and thymus. *Developmental Dynamics* **237**, 1699–708.

Toro M.A., Dobao M.T., Rodriganez J. & Silio L. (1986) Heritability of a canalized trait: teat number in Iberian pigs. *Genetics Selection Evolution* **18**, 173–83.

Van Ingen G., Li J., Goedegebure A., Pandey R., Li Y.R., March M.E. & Wei Z. (2016) Genome-wide association study for acute otitis media in children identifies *FNDC1* as disease contributing gene. *Nature Communications* **7**, 12792.

Verardo L.L., Silva F.F., Varona L., Resende M.D.V., Bastiaansen J.W.M., Lopes P.S. & Guimarães S.E.F. (2015) Bayesian GWAS and network analysis revealed new candidate genes for number of teats in pigs. *Journal of Applied Genetics* **56**, 123–32.

Wang M.G., Yi H., Guerini D., Klee C.B. & McBride O.W. (1996) *Calcineurin A alpha* (*PPP3CA*), *calcineurin A beta* (*PPP3CB*) and *calcineurin B* (*PPP3R1*) are located on human chromosomes 4, 10q21→q22 and 2p16→p15 respectively. *Cytogenetics and Cell Genetics* **72**, 236–41.

Wegner M.S., Wanger R.A., Oertel S. *et al.* (2014) Ceramide synthases CerS4 and CerS5 are upregulated by 17 beta-estradiol and GPER1 via AP-1 in human breast cancer cells. *Biochemical Pharmacology* **92**, 577–89.

Weir B.S. & Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–70.

Wen J., Toomer K.H., Chen Z.B. & Cai X.D. (2015) Genome-wide analysis of alternative transcripts in human breast cancer. *Breast Cancer Research and Treatment* **151**, 295–307.

Willman J.H., Golitz L.E. & Fitzpatrick J.E. (2003) Clear cells of Toker in accessory nipples. *Journal of Cutaneous Pathology* **30**, 256–60.

Yang Q., Cui J., Chazaro I., Cupples L.A. & Demissie S. (2005) Power and type I error rate of false discovery rate approaches in genome-wide association studies. *BMC Genetics* **6**, S134.

Yapp W.W. & St. Clair L.E. (1951) A meristic mutation involving teat number in the cow. *Journal of Dairy Science* **34**, 4978–88.

Yeung K.T., Das S., Zhang J., Lomniczi A., Ojeda S.R., Xu C.F., Neubert T.A. & Samuels H.H. (2011) A novel transcription complex that selectively modulates apoptosis of breast cancer cells through regulation of FASTKD2. *Molecular and Cellular Biology* **31**, 2287–98.

Yu Q.C., Verheyen E.M. & Zeng Y.A. (2016) Mammary development and breast cancer: a Wnt perspective. *Cancers* **8**, 1–2.

Supporting information

Additional supporting information may be found online in the supporting information tab for this article:

Figure S1 Multi-dimensional scaling (MDS) plot. The purple dots represent animals in the control group (60 with two teats), and the red squares indicate animals from the case group (64 with four teats including two normal and two supernumerary teats).

Figure S2 Q–Q (quantile–quantile) plot of the genome-wide association study. Grey and black rings represent association statistics before and after correction for population stratification respectively.

Figure S3 Linkage disequilibrium analysis of the candidate genomic regions OAR1: 170.723–170.734 Mb. Significant blocks are shown in red, and the significant SNP identified by the genome-wide association study is in a green box. Average frequencies of the haplotypes in the case and control subpopulations are indicated.

Table S1 Bonferroni-corrected 5% chromosome-wise significance threshold.

Table S2 The percentage of samples with a supernumerary nipple phenotype of different numbers in a Wadi sheep population.

Table S3 GO enrichment analysis of the genes associated with the significant SNPs at the chromosome-wise level as identified by the GWAS.