

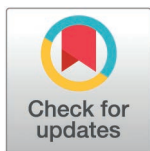
RESEARCH ARTICLE

Integrating sequence-based GWAS and comparative genomic analysis reveals conservation and species-specificity of putative functional variants influencing tail length and tail abnormalities in pigs and sheep

Xuying Zhang^{1*}, Johanna Mainzer¹, Isabella Giambra¹, Tong Yin^{1,2}, Petra Engel¹, Hannah Hümmelchen³, Henrik Wagner³, Axel Wehrend³, Christiane Egerer⁴, Katharina Gerhards⁴, Gerald Reiner⁴, Sven König^{1*}

1 Institute of Animal Breeding and Genetics, Justus-Liebig-University of Gießen, Gießen, Germany, **2** Zhejiang Key Lab of Cow Genetics Improvement and Milk Quality Research, Wenzhou, P.R. China, **3** Veterinary Clinic for Reproductive Medicine and Neonatology, Justus-Liebig-University of Gießen, Gießen, Germany, **4** Department of Veterinary Clinical Sciences, Clinic for Swine, Justus Liebig University Gießen, Gießen, Germany

* Xuying.Zhang@agr.uni-giessen.de (XZ); Sven.Koenig@agr.uni-giessen.de (SK)



OPEN ACCESS

Citation: Zhang X, Mainzer J, Giambra I, Yin T, Engel P, Hümmelchen H, et al. (2026) Integrating sequence-based GWAS and comparative genomic analysis reveals conservation and species-specificity of putative functional variants influencing tail length and tail abnormalities in pigs and sheep. PLoS One 21(3): e0343836. <https://doi.org/10.1371/journal.pone.0343836>

Editor: Amod Kumar, National Bureau of Animal Genetic Resources, INDIA

Received: March 12, 2025

Accepted: February 11, 2026

Published: March 3, 2026

Copyright: © 2026 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Sequence data supporting this study has been uploaded to the NCBI SRA database (PRJNA1230871).

Abstract

Long tails trigger tail biting in pigs and increase the risk of flystrike infections in sheep. Tail docking has been a common management practice in both species for decades, but increasingly conflicts with legal animal welfare guidelines. Sustainable solutions require breeding strategies targeting shorter tails. In consequence, the aims were to conduct whole-genome sequencing (WGS)-based genome-wide association studies (GWAS) and comparative genomic analyses (CGA) to explore functional elements influencing tail traits. Phenotypically divergent experimental populations of pigs and sheep were established through unified selection and mating experiments. Tail traits included tail length (TL) measured at birth, and tail abnormalities (TA) assessed radiographically at 14 weeks of age. WGS-based GWAS identified a significant locus on SSC18 in pigs and suggestive loci for TL in both species, which, together with previously reported loci for TA, were further analyzed by CGA. The genomic windows of the significant locus on SSC18 in pigs and the TL GWAS locus on OAR4 in sheep were found to be conserved, harboring six common genes with predicted functional variants. These variants were jointly associated with TL ($P_{im} < 0.05$) in both species in linear regression models adjusted for sex, age of the dam, body length, and body weight. In other GWAS locus windows (± 1 Mb), species-specific TL candidate genes were identified in sheep (*HOXB13*, *MUC5B*, *EPB41L3*, *MTCL1*, *PIEZO2*, *MPPE1*, and *LOXHD1*) and in pigs (*KNL1*, *DISP2*, *SPRED1*, *TGFB2*, and *HAND1*), each harboring associated putative functional variants. For TA, sheep-specific candidates (*PGM2*, *LRR66*, *CRACD*, *LOC105601916*, and *SH2D4B*) and pig-specific

Funding: The study was carried out as part of the collaborative DFG (German Research Foundation) project titled “A Comparative Functional Genomic Approach to Unravel the Genetic and Genomic Architecture of Tail Length in Pigs and Sheep Based on a Unified Selection Experiment Design” (grant no. KO 3520/12-1, WE 6061/3-1, RE 1026/10-1).

Competing interests: The authors have declared that no competing interests exist.

candidates (*MYOT*, *TMCO6*, and *PCDHAC2*) were revealed using logistic regression models ($P_{glm} < 0.05$). GO analyses of candidate genes predicted shared biological processes between sheep and pigs, whereas pathway analyses indicated that common carbohydrate metabolism pathways, along with species-specific immune and inflammatory signaling, and pig-specific TGF- β signaling and endochondral ossification, may contribute to tail length variation and abnormalities. These findings provided deeper insights into the genetic basis of differential embryonic tail morphogenesis and perinatal tail development across species.

Introduction

In modern animal husbandry systems, long tails pose severe issues, such as the accumulation of dags in the tail area, predisposing sheep to flystrike and infections with associated fertility disorders in sheep, as well as tail biting in pigs [1]. Tail docking directly after birth is a common practice for pigs and sheep, effectively solving these problems [2–4]. However, due to the growing emphasis on animal welfare and the subsequent legal guidelines, such as those defined in EU Directive (Richtlinie 2008/120/EG [5]), tail docking has become controversial [6]. Breeding on shorter tails within breeds or populations might be a sustainable and socially acceptable alternative [7]. Since the wild ancestors of domesticated sheep and pigs naturally had relatively short tails, breeding attempts in this regard reflect a “back-to-the-roots” approach.

Tail length (TL) is a moderately to highly heritable trait in both sheep and pigs. In sheep, heritability estimates ranged from 0.50 in Australian breeds to 0.77 in Finnsheep [8] and 0.60–0.87 in Merinoland sheep [9,10]. In pigs, the heritability for TL was 0.42 [11]. Heritability estimates in other species, such as Toque macaques (0.67) and mice (0.46), further support a strong genetic basis [12,13]. These findings indicate that TL can be effectively reduced through selective breeding within a few generations.

Vertebrate tails originate from a post-anal extension of the embryo known as the tail bud [14]. Its formation, growth and differentiation are closely related to the activity of a cluster of cells from axial progenitors, that build the spinal cord and the trunk musculoskeletal system [15–17]. The tail bud likely results from mechanisms that orderly terminate the molecular and cellular activities constructing vital vertebrate body structures after their key functions are fulfilled [18]. These mechanisms offer plasticity to the tail bud, enabling evolutionary forces to generate tails of different shapes and sizes [18].

The genetic architecture of TL in vertebrates across different species involves both mono- or oligogenic and polygenic components [19], reflecting a combination of qualitative Mendelian and quantitative additive genetic effects. In South Australian Merino sheep, TL inheritance appears to be governed by a small number of interacting genes with large effects, in which short tail genes potentially exhibit dominance [20]. A SNP and an insertion within or near *HOXB13* have been associated with short tail

length in Merino sheep [21]. Two linked SNPs in the *T-box* gene are associated with the tailless phenotype in fat-rumped sheep [22]. Previous studies have reported potential roles of *VRTN* and *BMP2* in determining variation in vertebral number [23], as well as possible involvement of *PDGFD* and *BMP2* in sheep tail formation [24]. A putative functional missense variant in *PCDHA1* (SIFT score = 0.05) was suggestively associated with tail kinks in pigs based on GWAS [25]. Variants in *T-box* were associated with short-tailed phenotypes, the number of cervical and tail vertebrae, and vertebral shape in mice, cats, dogs, and cattle [19,26–28]. However, the *T-box* gene also caused pleiotropic effects, such as spine deformities and embryonic early death [26,27,29]. Furthermore, variants in the *HES7* gene indicated associations with tail lengths in domestic Asian cats [19]. The *Pax1* and *Wnt-3a* genes influenced tail length in mice [30,31]. Phylogenetic studies suggested that tail shortening or loss evolved independently in different animal species, though the evolutionary mechanisms behind tail shortening remain unclear [29]. Despite these insights, the full mechanism and complexity of TL inheritance, particularly when considering the additional evidence for additive genetic effects, are not yet fully understood.

Whole genome sequences (WGS) have been used to study selection signatures in South African Mutton Merino and Northeast China Merino for growth, carcass, and economic traits, as well as morphological traits like tail fat deposition [32–34]. For TL, genome-wide significant regions were capture-sequenced to identify candidate causal genes in the Merinolandschaf breed [21]. WGS data were used to evaluate the frequency of a functional 168 bp deletion of the *HOXB13* gene across different sheep breeds [35].

While most comparative genomic studies have been conducted in model organisms and humans [36–39], similar analyses in livestock species, particularly pigs and sheep, remain scarcely available [40,41]. Leveraging WGS technologies, this study aimed to integrate results from genome wide association studies (GWAS) and ongoing comparative genomic analysis (CGA) into the framework of a multi-species selection experiment. Such approach enables us to investigate and differentiate conserved versus species-specific functional variants for TL and tail abnormalities (TA) across both species of sheep and pigs. The ultimate objective was to elucidate the genetic basis of tail formation and morphogenesis in both species, thereby informing the development of targeted breeding and selection strategies, and improving animal welfare. Additionally, this study will provide valuable insights stimulating ongoing studies with focus on the underlying genetic mechanisms.

Materials and methods

Ethics statement

Trait recording and blood sampling for sheep were approved by the Regional Council of Giessen (V 54—19c 20 15 h 01 GI 18/14 No. G 44/2021). All experiments conducted in pigs were approved by the Animal Welfare Officer of the Ethics Commit-87 tee of the Justus Liebig University, Giessen, Germany, with the reference 88 JLU_kTV_4_2021.

Selection experiment

A selection and mating experiment for the trait TL based on estimated breeding values (EBV) was implemented for both species of sheep and pigs at the University Giessen research herd “Oberer Hardthof”. TL has been routinely recorded in both species at “Oberer Hardthof” since 1988, enabling reliable pedigree-based breeding value estimations, which are routinely carried out in intervals of 6 months. The intention was to use the EBV to create highly diverse TL subgroups, contributing to strong genetic and phenotypic TL group contrasts in offspring. In this regard, in the Merinoland sheep population, 4 rams with most extreme EBV for TL were mated with 142 ewes, following a “EBV short x EBV short” and “EBV long x EBV long” strategy. The generated 254 Merinoland lambs substantially differed genetically and phenotypically for TL. The “short x short” lambs had an average TL at birth of 21.47 cm (SD = 2.26 cm) versus an average TL of 23.51 cm (SD = 2.11 cm) of the “long x long” lamb group, i.e., a difference of 2.04 cm within one generation. The difference in TL EBVs between both lamb groups was 2.14 times the EBV standard deviation.

Accordingly in pigs, genetically and phenotypically diverse subgroups for TL were generated in F1 piglets in the selection experiment, also following an “EBV short x EBV short” and “EBV long x EBV long” mating design, considering 6 Landrace boars and 21 Landrace sows. Average TL in “short x short” offspring comprised $6.26 \text{ cm} \pm 1.91 \text{ cm}$, and $10.80 \text{ cm} \pm 2.11 \text{ cm}$ in the “long x long” piglet’s group. The difference in TL EBVs between both piglet groups comprised 2.39 times the EBV standard deviation.

Phenotypes

The lambs were phenotyped (measurements in cm) for TL at birth. The TL was defined as the distance from the tail base to the tip of the tail. Furthermore, body length at birth was measured, indicating the length from the first thoracic vertebra to the base of the tail. At an age of 14 weeks, the lambs were radiographically examined using a portable X-ray device (Physia GAMMA Light AD 100|20) with a setting of 40 kV and 2.5 mAs to determine the number of vertebrae, and for the detection of TA including wedged vertebrae and axis deviations of the tail. All TA were defined as a binary trait. The trait categories and trait definitions for single traits are outlined in [Table 1](#).

The piglets were measured at birth for TL and body length. The measurement for body length was the distance from the base of the ear to the base of the tail. The TA trait category included kink grades (0°, 30°, 60°, 90°, and 180° deviations), which were subsequently coded as a binary trait (Kinks01). After trait recording, piglet tails were docked.

Genotypes and whole genome sequencing

Blood samples were collected from all lambs within the first week by jugular vein puncture into EDTA tubes and stored at -20°C . For WGS, 140 lambs from different dams were selected out of the 254 offspring, favoring the lambs with most extreme EBV in case of several offspring per dam. Selection of piglets for WGS considered the 70 piglets with the smallest EBV for TL, and the 70 piglets with the largest EBV for TL. In sheep, DNA was extracted from blood using the NucleoSpin® Blood Kit (Macherey Nagel, France) and standardized to 60 ng/μl for sequencing. WGS was performed using the Illumina NovaSeq 6000 (Illumina, San Diego, USA), producing 150 bp paired-end reads with a mean mapped read depth of 15.83 and a coverage of 15x. The *Ovis aries* ARS-UI_Ramb_v2.0 reference genome was used for read alignment and variant calling. After standard quality control (minor allele frequency > 0.05 , non-significant deviation from Hardy Weinberg equilibrium at $p > 0.000001$), 22,603,883 biallelic variants on autosomes were identified using the implemented workflow [42] from the 140 sequenced sheep, with all sheep genomes exhibiting a call rate larger than 0.95.

Tissue from docked piglet tails, stored at -20°C in EDTA tubes, was used as sample material for DNA extraction with the smart 95 DNA prep (m) kit (Analytik Jena, Jena, Germany). DNA concentrations were diluted to 50 ng/μL. In analogy to sheep, WGS was performed on an Illumina NovaSeq 6000, generating 150 bp paired-end reads of 15x coverage. Further sequencing data preparation included demultiplexing of all libraries for each sequencing lane using Illumina bcl2fastq v2.20 software (https://emea.support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html) and clipping of sequencing adapter remainders from all reads. Reads with a final length < 20 bases were discarded. The *Sus scrofa* Sscrofa11.1 reference genome was used for read alignment and variant calling. The ongoing bioinformatics

Table 1. Overview of tail categories in sheep and pigs used in the comparative genomic analyses.

Trait category	Definition of traits considered in sheep	Definition of traits considered in pigs
Tail length (TL)	Absolute TL (in cm)	Absolute TL (in cm)
Tail abnormalities (TA)	Axis deviations (binary trait)	Kink grades (in °)
	Wedged vertebrae (binary trait)	Kinks01(binary trait)

<https://doi.org/10.1371/journal.pone.0343836.t001>

workflow used the infrastructure as implemented in the institute of JLU Gießen [42] and the variant quality criteria as defined for sheep. 14,159,238 variants from 140 piglets were considered for the ongoing genomic analyses.

Genome-wide associations

The WGS-based GWAS in sheep and in pigs for TL was performed by applying the software package GCTA v1.94.1 [43]. The general statistical model was:

$$y = X\beta + x_{\text{Variant}}u_{\text{Variant}} + Zg + e$$

where y = vector for TL observations in cm; β = vector for fixed effects including sex (male, female), age of the dam and linear regressions on body weight (kg) and body length (cm) with incidence matrix X ; x_{Variant} = vector for variant genotypes; u_{Variant} = vector for variant effects; g = vector for polygenic additive-genetic effects following $N(0, G\sigma_g^2)$ with σ_g^2 denoting the additive-genetic variance with the incidence matrix Z ; e = vector for the random residual effects following $N(0, I\sigma_e^2)$ with I denoting an identity matrix and σ_e^2 denoting the residual variance.

The WGS-based GWAS for TA was adopted from previously reported studies in our group [10,25]. For sheep, a linear mixed model (LMM) [44] was implemented in R, whereas for pigs the BLINK model in GAPIT version 3 [45] was applied.

Significantly associated variants were detected using the Bonferroni-corrected threshold, defined as $-\log_{10}(P_{\text{Bonf}}) = -\log_{10}(0.05/\text{number of independent variants})$, which equals $-\log_{10}(0.05/477,627) = 6.98$ in sheep and $-\log_{10}(0.05/373,699) = 6.87$ in pigs. The number of independent variants was estimated in PLINK 1.9 using linkage disequilibrium restrictions ($R^2 \leq 0.5$) within genomic windows of 50,000 variants, shifting by 500 variants after each pruning step [46]. A suggestive threshold was set at $-\log_{10}(P_{\text{Sugg}}) = -\log_{10}(1 \times 10^{-5}) = 5.00$ for both species.

Gene mapping and gene network analyses

Gene mapping was performed using the biomaRt package in R, referencing the NCBI RefSeq assemblies *Ovis aries* ARS-UI_Ramb_v2.0 for sheep and *Sus scrofa* Sscrofa11.1 for pigs. Potential candidate genes were assigned to corresponding significant variants after Bonferroni correction ($pBC = 1.05 \times 10^{-7}$ in sheep, $pBC = 1.34 \times 10^{-7}$ in pigs) or to suggestive variants using a suggestive threshold ($pSC = 1 \times 10^{-5}$). For each locus, gene mapping was conducted within a 1Mb window in both upstream and downstream directions.

Variant functional annotation, including single nucleotide polymorphisms (**SNP**) and small insertion-deletion polymorphisms (**indels**), was carried out using SnpEff [47]. Genes containing predicted high-impact variants (supposed to have deleterious effect) or moderate-impact variants with deleterious SIFT scores (<0.05) were further analysed. Gene ontology (**GO**) and pathway enrichment analyses were conducted using Enrichr [48], PANTHER 19.0 [49] and ShinyGO v0.80 [50] (databases consulted on 10.09.2025), with all protein-coding genes in the respective genome used as the default background for enrichment comparisons. The STRING database [51] was employed to predict protein-protein interactions and functional networks. Identified networks were clustered using the K-means algorithm to delineate functional modules. Additionally, co-expression network analysis was performed with GeneMANIA v3.5.3 [52].

Comparative functional genomic analyses

The comparative functional genomic analyses were based on GWAS results, functional annotations of variants, and gene mapping. Significant or suggestive association signals from GWAS in pigs and sheep were utilized to define candidate functional windows for initial multiple sequence alignment (**MSA**) between both species. These genomic windows encompassed both characterized and novel genes harbouring variants with predicted high or moderate (with deleterious SIFT scores <0.05) functional impact. Additionally, for associated variants located in intergenic regions, genomic windows were defined as the intergenic variant ± 100 kb. MSA was performed using the software packages BLAST version 5 [53] and

Clustal Omega v1.2.4 [54], applying thresholds of a Bit score > 200 and sequence identity > 85%. The alignment analysis evaluated the conservation of identified variants and putative functional elements implicated in the development of tail traits (TL and TA) across pigs and sheep.

Results

Genome-wide associations

For tail length (TL), the WGS-based GWAS identified 320 suggestive chromosomal loci in Merinoland sheep ($\lambda = 1.081$) and 34 in pigs ($\lambda = 1.099$) with additive genetic effects ($pSC = 1 \times 10^{-5}$; Fig 1). The GWAS peaks were located on OAR1, 2, 3, 4, 7, 11, 14, 16, 17, 18, 21, 22 and 23 in sheep, and on SSC1, 2, 5, 6, 7, 10, 11, 13, 14, 15, 16, 17 and 18 in pigs. A summary of these loci is provided in Table 2, with all suggestive loci listed in S1 and S2 Tables. None of the TL-associated loci in sheep reached the genome-wide significance threshold after Bonferroni correction ($pBC = 1.05 \times 10^{-7}$). The GWAS peak at OAR11:40,897,359 is located approximately 3 Mb downstream of a previously reported candidate causal locus at OAR11:37,525,128 [21]. One locus at SSC18:7,836,326 in pigs surpassed the significance threshold ($pBC = 1.34 \times 10^{-7}$). The GWAS peak at SSC6:110,116,231 is located approximately 37 Mb away from a previously reported GWAS locus at SSC6:147,382,689 [25].

For tail abnormalities (TA), the WGS-based GWAS detected 70 suggestive loci for axis deviation (AXISD) and 69 for wedged vertebrae (WDGV) in sheep [10], and 6 suggestive loci for kink in pigs [25] ($pSC = 1 \times 10^{-5}$) with additive genetic effects (Table 3, S3 and S4 Tables). One locus at OAR6:63,347,729 exceeded the Bonferroni-corrected significance threshold for AXISD in sheep ($pBC = 1.05 \times 10^{-7}$), and three loci at SSC1:30,699,535, SSC11:30,485,036, and SSC15:111,206,597 surpassed the threshold in pigs for kink-type TA ($pBC = 1.34 \times 10^{-7}$). The suggestive and significant loci identified in both species provided the basis for subsequent gene mapping and comparative genomic analyses.

Comparative functional genomic analyses

Conserved elements. In sheep, 2,049 genes were mapped within ± 1 Mb of the GWAS loci significantly or suggestively associated with TL or TA. Functional annotation using SnpEff predicted that 778 of these genes carry

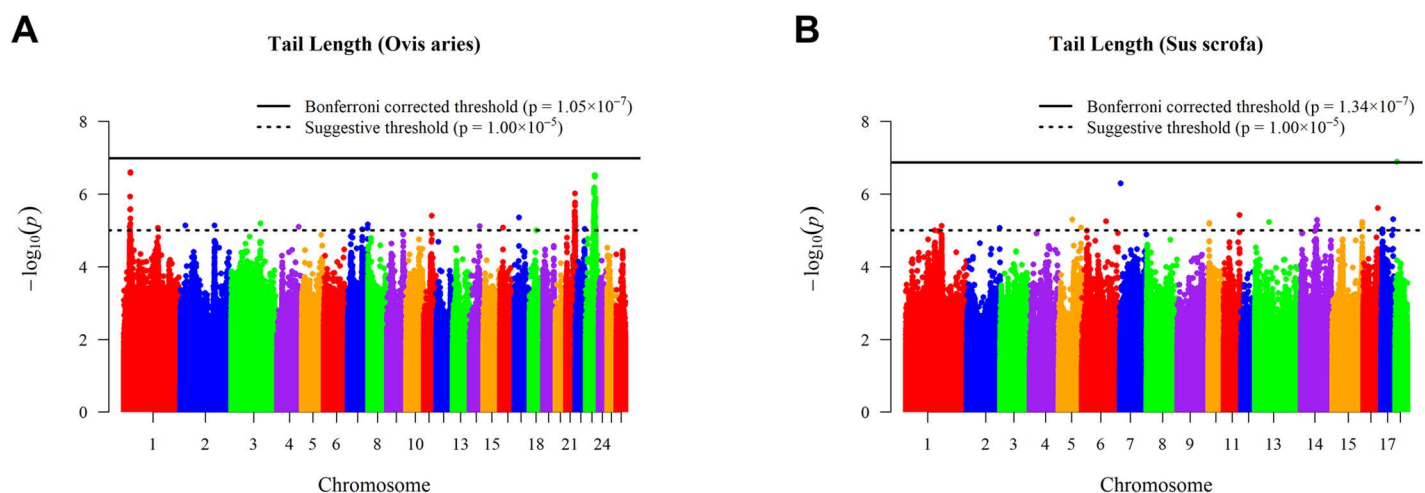


Fig 1. WGS-based GWAS results for tail length (TL) presented as Manhattan plots. (A) Sheep: 320 markers exceeded the suggestive threshold, but none reached the Bonferroni threshold. **(B)** Pigs: 34 markers exceeded the suggestive threshold, and one locus on SSC18:7,836,326 surpassed the Bonferroni threshold.

<https://doi.org/10.1371/journal.pone.0343836.g001>

Table 2. GWAS peak loci associated with tail length in sheep and pigs.

Species	model	Fixed effects	Chr:Pos	Ref>Alt	P# in GWAS	Impact
Sheep	MLMA	sex, age, BL, BW	1:32432765	GA>G	2.50E-07	modifier
			1:168597871	T>C	8.50E-06	modifier
			2:168368498	A>C	7.22E-06	modifier
			2:25067918	A>T	7.30E-06	modifier
			3:145274341	C>A	6.35E-06	modifier
			4:107294668	T>A	7.93E-06	modifier
			7:98709307	T>C	6.84E-06	modifier
			11:40897359	C>T	3.96E-06	modifier
			14:50074477	A>G	7.61E-06	modifier
			16:16227298	G>T	8.41E-06	modifier
			17:23067695	T>C	4.36E-06	modifier
			18:34245224	G>A	9.87E-06	modifier
			21:45702882	A>G	9.59E-07	modifier
			22:44937479	A>G	9.01E-06	modifier
			23:44582760	G>A	3.02E-07	modifier
	cLDLA	sex, age, BH, BW	11:37525128	C>G	5.71E-08	Candidate causal locus [21]
Pigs	MLMA	sex, age, BL, BW	1:161252022	G>T	7.41E-06	modifier
			2:149438280	C>A	8.45E-06	modifier
			5:61839803	C>G	5.01E-06	modifier
			5:100863639	A>G	8.29E-06	modifier
			6:110116231	T>C	5.59E-06	modifier
			7:5520737	C>CT	5.07E-07	modifier
			10:5005427	CCT>C	6.23E-06	modifier
			11:72452748	G>A	3.76E-06	modifier
			13:64707578	A>C	5.87E-06	modifier
			14:73521925	T>*	5.19E-06	modifier
			15:135662299	T>C	5.84E-06	modifier
			16:66174795	CATCACAGAAAGTGCT>C	2.40E-06	modifier
			17:6012151	A>G	9.36E-06	modifier
			17:54799100	A>G	4.87E-06	modifier
			18:7836326	A>C	1.29E-07	modifier
	BLINK	sex, age, LS, LN	6:147382689	C>A	3.85E-12	modifier [25]

Loci refer to the *Ovis aries* ARS-UI_Ramb_v2.0 and the *Sus scrofa* Sscrofa11.1 genome assemblies for sheep and pigs, respectively. MLMA, mixed linear model-based association analysis; cLDLA, combined linkage disequilibrium and linkage analysis; BLINK, bayesian-information and linkage-disequilibrium iteratively nested keyway; BL, body length; BW, body weight; WH, withers height; LS, litter size; LN, litter number (parity). Significance thresholds: for sheep, $\leq 1.05 \times 10^{-7}$ (Bonferroni-corrected) and $\leq 1.00 \times 10^{-5}$ (suggestive); for pigs, $\leq 1.34 \times 10^{-7}$ (Bonferroni-corrected) and $\leq 1.00 \times 10^{-5}$ (suggestive).

<https://doi.org/10.1371/journal.pone.0343836.t002>

high-impact (*hi*) or moderate-impact (*mi*) variants (S5 Table). In pigs, 550 genes were mapped within ± 1 Mb of the GWAS loci significantly or suggestively associated with loci for TL or TA, of which 219 genes harbored *hi* or *mi* variants (S6 Table). A conserved genomic window corresponding to the suggestive TL GWAS locus at OAR4:107,294,668 in sheep and the significant TL GWAS locus at SSC18:7,836,326 in pigs was identified, including 13 overlapping genes from the sheep and pig gene lists (Fig 2). Subsequent SIFT analysis of the *mi* variants identified potentially deleterious missense variants. Six shared genes across both species, *CLEC5A*, *MGAM*, *EPHB6*, *KEL*, *CLCN1*, and *EPHA1*, harbored *hi* or *mi* variants with deleterious SIFT scores (<0.05) (Table 4).

Table 3. GWAS peak loci associated with tail abnormalities, including axis deviation (AXISD) and wedged vertebrae (WDGV) in sheep, and tail kink (Kink) in pigs.

Species	Traits	model	Fixed effects	Chr:Pos	Ref>Alt	P# in GWAS	Impact
Sheep	AXISD	LMM	sex, age, LS, LN	1:88823845	C>T	7.89E-07	modifier [10]
	AXISD			1:164082021	A>G	5.60E-06	modifier [10]
	AXISD			2:22704789	A>G	1.26E-06	modifier [10]
	AXISD			2:136963847	T>G	5.48E-06	modifier [10]
	AXISD			3:70907944	T>A	8.50E-07	modifier [10]
	AXISD			3:193293265	G>A	2.61E-06	modifier [10]
	AXISD			6:63347729	C>G	9.72E-08	modifier [10]
	AXISD			7:66017541	A>C	2.17E-06	modifier [10]
	AXISD			8:84432236	G>A	7.24E-06	modifier [10]
	AXISD			10:217144	T>C	3.00E-06	modifier [10]
	AXISD	LMM	sex, age, LS, LN	12:20417780	T>C	6.34E-06	modifier [10]
	AXISD			16:48358696	C>G	2.44E-06	modifier [10]
	AXISD			20:3013558	T>C	2.65E-06	modifier [10]
	AXISD			21:12623286	C>T	2.27E-06	modifier [10]
	AXISD			25:33098936	T>C	7.14E-07	modifier [10]
	WDGV			1:10081864	G>T	9.68E-07	modifier [10]
	WDGV			2:15855244	C>T	8.84E-06	modifier [10]
	WDGV			3:196834657	A>G	2.61E-06	modifier [10]
	WDGV			5:31830734	T>C	4.81E-06	modifier [10]
	WDGV			6:65064860	C>T	1.92E-06	modifier [10]
Pigs	WDGV	BLINK	sex, age, LS, LN	6:106186919	T>C	7.46E-06	modifier [10]
	WDGV			9:73249875	T>C	2.53E-06	modifier [10]
	WDGV			10:57246689	C>T	1.77E-06	modifier [10]
	WDGV			12:35113679	T>C	5.30E-07	modifier [10]
	WDGV			14:64113935	A>G	1.52E-06	modifier [10]
	WDGV			20:27647344	A>C	5.87E-06	modifier [10]
	WDGV			23:35992193	C>CA	5.90E-06	modifier [10]
	WDGV			25:40843225	C>CT	3.13E-07	modifier [10]
	kinks01			2:142461085	A>G	6.59E-06	tolerated [25]
	kink grades			1:30699535	G>A	5.31E-11	modifier [25]
				11:30485036	T>G	3.69E-13	modifier [25]
				15:111206597	A>G	7.92E-12	modifier [25]

Loci refer to the *Ovis aries* ARS-UI_Ramb_v2.0 and *Sus scrofa* Sscrofa11.1 genome assemblies for sheep and pigs, respectively. LMM, linear mixed model; BLINK, bayesian-information and linkage-disequilibrium iteratively nested keyway; LS, litter size; LN, litter number (parity). Significance thresholds: for sheep, $\leq 1.05 \times 10^{-7}$ (Bonferroni-corrected) and $\leq 1.00 \times 10^{-5}$ (suggestive); for pigs, $\leq 1.34 \times 10^{-7}$ (Bonferroni-corrected) and $\leq 1.00 \times 10^{-5}$ (suggestive).

<https://doi.org/10.1371/journal.pone.0343836.t003>

A linear regression analysis was performed using the *lm* function in R, with sex, age, body weight, and body length included as covariates. We analysed the p-values and the variance additionally explained by the variant(s) beyond the variance explained by the basic model only including the covariates (ΔR^2). Three variants in sheep and four in pigs were significantly associated with TL ($P_{lm} < 0.05$), with two of these commonly located in *EPHB6* and *KEL*. When all variants from the six common genes were included as independent variables along with the covariates, the association became significant in both species (sheep: $P_{lm} = 6.03 \times 10^{-3}$; pigs: $P_{lm} = 7.14 \times 10^{-4}$), with these variants explaining 8.31% and 9.41% of the variance in TL (ΔR^2), respectively.

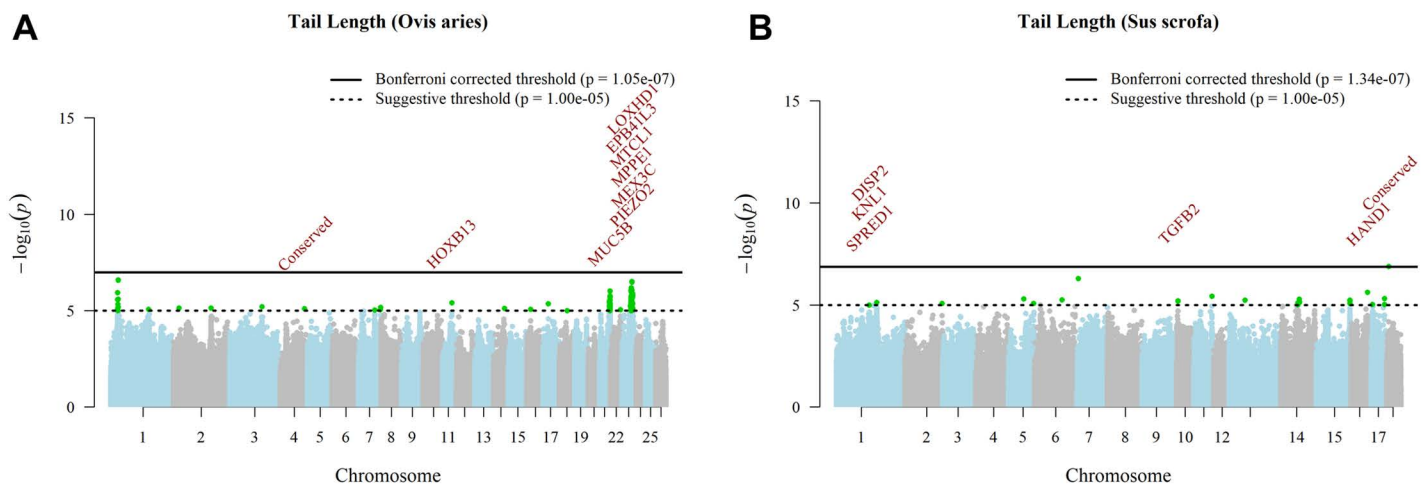


Fig 2. Conserved and species-specific elements in GWAS loci for tail length in sheep and pigs. Manhattan plots labeled with conserved and species-specific elements identified within genomic windows of GWAS loci for tail length in sheep (A) and pigs (B). The genomic window of the suggestive locus on OAR4:107,294,668 in sheep and the significant locus on SSC18:7,836,326 in pigs are conserved, harboring 6 putatively functional common genes. Eight species-specific putative functional candidate genes were identified on OAR11, OAR21, and OAR23 in sheep, and 5 species-specific putative functional candidate genes on SSC1, SSC10, and SSC16 in pigs.

<https://doi.org/10.1371/journal.pone.0343836.g002>

Variants located in intergenic regions, although their functional consequences cannot be predicted by SnpEff, may have the potential to affect regulation of nearby genes [55–57]. Among the variants significantly or suggestively associated with tail traits in the GWAS analyses, 247 in sheep and 20 in pigs were located in intergenic regions. No common genes were mapped within the ± 100 kb genomic windows of these intergenic variants across species. Also, multiple alignment analysis did not reveal highly conserved intergenic sequences across species.

Species-specific elements. Within ± 1 Mb of TL GWAS loci, eight species-specific candidate genes were identified in sheep (*HOXB13*, *MUC5B*, *EPB41L3*, *MTCL1*, *PIEZO2*, *MPPE1*, *LOXHD1*, and *MEX3C*), carrying nine potentially functional variants significantly associated with TL (Table 5, Fig 2A). The 167 bp insertion at OAR11:37,525,005 (*ARS-UI_Ramb_v2.0*) in *HOXB13*, previously reported as a candidate causal variant for TL, was identified by Delly and significantly associated with TL in the *Im* analysis ($P_{lm} = 4.19 \times 10^{-7}$). This variant explained 11.80% of the TL variance (ΔR^2). The putatively functional variants in *EPB41L3*, *MTCL1*, *PIEZO2*, *MPPE1*, *LOXHD1*, and *MEX3C* on OAR23 explained up to 22.54% of TL variance (ΔR^2), with $P_{lm} = 1.77 \times 10^{-13}$. The variant in *MUC5B* on OAR21 accounted for 9.89% of TL variance ($P_{lm} = 4.33 \times 10^{-6}$). Homozygous lambs for the alternate allele at the TL candidate variants in *HOXB13*, *MUC5B*, *EPB41L3*, *MTCL1*, *MPPE1*, *LOXHD1* and *MEX3C* exhibited longer average tail lengths than the lamb group being homozygous for the reference allele, and intermediate lengths for the heterozygous group. The opposite pattern was observed for the variant in *PIEZO2*.

In pigs, five species-specific candidate genes (*KNL1*, *DISP2*, *SPRED1*, *TGFB2*, and *HAND1*) within ± 1 Mb windows of TL GWAS loci, harboring six putatively functional variants, were significantly associated with TL (Fig 2B). Variants in *KNL1*, *DISP2* and *SPRED1* on SSC1 explained up to 10.24% of the TL variance (ΔR^2), with $P_{lm} = 3.34 \times 10^{-7}$. A putative functional variant in an enhancer region of *TGFB2* on SSC10 accounted for 6.12% of the TL variance (ΔR^2 , $P_{lm} = 1.11 \times 10^{-4}$). Additionally, a 9 bp deletion in *HAND1* on SSC16 explained 5.73% of the TL variance ($P_{lm} = 1.90 \times 10^{-4}$). Homozygous and heterozygous piglets for the alternate allele at the TL candidate variants in *KNL1*, *DISP2*, *SPRED1* and *TGFB2* exhibited longer average tail lengths compared to the group being homozygous for the reference allele. The opposite pattern was observed for the variant in *HAND1*.

<https://doi.org/10.1371/journal.pone.0343836.t004>

Table 5. Genotype distributions of short tail (S) and long tail (L) group animals for putative functional candidate variants near TL GWAS loci in sheep and pigs.

Chr:Pos	rs.ID	R>A	Gene	Conseq.	SnEff_SIFT	RR-S(n)	RR-L(n)	RR-TL (M±SD)	RA-S(n)	RA-L(n)	RA-TL (M±SD)	AA-S(n)	AA-L(n)	AA-TL (M±SD)	ΔR ² (%)	P _{im}
Sheep:																
11:37525005	–	SV	HOXB13	insertion	hi	10	0	19.7±1.47	19	4	21.2±2.22	40	66	23.3±2.58	11.80	4.19E-07**
21:46591162	novel	T>G	MUC5B	missense	mi	66	46	22.2±2.57	3	22	24.7±2.17	–	–	–	9.89	4.33E-06**
21:46591665	novel	54 bp [§]	MUC5B	deletion	hi	64	42	22.2±2.62	7	25	24.1±2.39	0	1	26.8±0.00	9.19	9.98E-06**
23:39545427	rs403011667	C>T	EPB41L3	missense	mi_0	52	18	21.2±2.15	17	40	24.0±2.44	0	12	24.8±1.59	13.84	3.18E-08**
23:41727601	rs427246640	G>A	MTCL1	missense	mi_0.03	43	14	21.2±2.17	25	45	23.5±2.65	2	10	24.6±1.77	12.02	3.18E-07**
23:42726665	rs424483114	A>C	PIEZO2	missense	mi_0	15	24	23.4±2.55	44	35	22.6±2.82	13	8	21.8±2.19	2.09	4.05E-02*
23:43248372	rs419737652	C>T	MPPE1	stop gained	hi	41	13	21.2±2.16	29	40	23.3±2.68	1	15	24.9±1.73	11.93	3.55E-07**
23:46757717	rs420232320	C>T	LOXHD1	missense	mi_0.03	25	0	20.1±1.58	38	27	22.2±2.42	7	42	24.7±1.97	22.54	1.77E-13**
23:50986262	novel	A>G	MEX3C	missense	mi	6	1	20.1±1.71	29	17	21.9±2.86	34	52	23.3±2.44	10.37	2.41E-06**
Pigs:																
1:130691649	novel	C>CT	KNL1	frame-shift&splice region	hi	27	20	9.3±1.00	29	30	9.4±1.08	5	27	10.2±0.92	7.99	8.40E-06**
1:130692943	novel	ATGCTTT>A	KNL1	disruptive infr. del.	mi	55	49	9.4±1.04	6	28	10.2±0.97	–	–	–	10.19	3.62E-07**
1:130951854	rs319509181	T>C	DISP2	missense	mi_0.05	53	46	9.4±1.04	8	31	10.1±0.97	–	–	–	7.70	1.25E-05**
1:133016175	novel	G>T	SPRED1	missense	mi	55	50	9.4±1.03	4	29	10.2±0.98	–	–	–	10.24	3.34E-07**
10:8339170	rs690408303	C>G	TGFB2	missense	mi,en-hancer	43	47	9.5±1.06	15	25	9.6±1.14	3	5	10.2±0.92	6.12	1.11E-04**
16:68680269	novel	ACACCAT-CAC>A	HAND1	disruptive infr. del.	mi	52	70	9.6±1.08	9	7	9.3±1.02	–	–	–	5.73	1.90E-04**

Variants include high-impact (hi) or moderate-impact (mi) SNPs with deleterious SIFT scores (<0.05). SNPs are given as Chromosome:Position (Ovis aries ARS-UI_Ramb_v2.0; Sus scrofa Sscrofa11.1) with rs.IDs. Alleles are shown as R>A (reference > alternate). §GGTGGTCACTGTGGAGGGGCTGTCTCACTGGTGTG-GGTGGAGGTCTCTCA>G. Genotypes are coded as RR (homozygous reference), RA (heterozygous), and AA (homozygous alternate). For each genotype group: S (n) = number of short-tail individuals, L (n) = number of long-tail individuals, TL (M±SD) = mean±standard deviation of TL. P_{im} = linear regression p-value (lm in R, adjusted for sex, age, body weight, body length). Significance: P_{im} <0.05 (*), P_{im} <1.00×10⁻³ (**), ΔR² (%) = 100 × [R²(model with tested variant + covariates) – R²(model with covariates only)], where R² is the coefficient of determination.

<https://doi.org/10.1371/journal.pone.0343836.t005>

In our previous studies, the prevalence of TA was significantly associated with the number of vertebrae (nVERT) or TL in sheep and pigs [10,58]. Using sex, age and nVERT as covariates, logistic regression (*glm* in R) was applied to analyze putatively functional variants within ± 1 Mb of previously reported GWAS loci for AXISD and WDGV in sheep. Three variants in *PGM2*, *LRRC66* and *CRACD* on OAR6 were significantly associated with AXISD, while one variant in LOC105601916 on OAR14, and two variants in *SH2D4B* on OAR25, were significantly associated with WDGV ($P_{glm} < 0.05$) (Fig 3AB, Table 6). For variants in *PGM2*, *LRRC66*, *CRACD* and *SH2D4B*, lambs carrying homozygous alternate or heterozygous genotypes showed a higher prevalence of AXISD or WDGV compared to those with homozygous reference genotypes. Notably, in *LRRC66*, no lambs with the homozygous reference genotype exhibited tail abnormalities. In pigs, using sex, age and TL as covariates in the *glm* analysis of putative functional variants within ± 1 Mb of Kink GWAS loci, four putative functional variants in *MYOT*, *TMCO6* and *PCDHAC2* on SSC2 were significantly associated with Kink ($P_{glm} < 0.05$) (Fig 3C). For variants in *MYOT* and *TMCO6*, piglets carrying homozygous alternate or heterozygous genotypes showed a higher prevalence of tail kink compared to those with homozygous reference genotypes. A frameshift variant in *PCDHAC2* exhibited the same pattern, whereas a nearby missense variant (641 bp apart) showed the opposite association.

Gene ontology and pathways

GeneMANIA network analysis in each species revealed interactions among putative functional candidate genes for tail traits (Fig 4), suggesting functional connections underlying the mechanisms of tail development and morphogenesis. GO enrichment analysis of the identified candidate genes for tail traits predicted that the same biological processes enriched in both sheep and pigs, although the relative contribution of each process differed between species (Fig 5). Both sheep and pigs showed enrichment of the galactose metabolism and starch and sucrose metabolism pathways, although the associated p-values and odds ratios differed between species (Table 7). Beyond these shared pathways, sheep exhibited additional enrichment mainly in carbohydrate processing (e.g., glycogen biosynthesis and breakdown, carbohydrate metabolism, O-glycan biosynthesis, glucuronidation, pentose phosphate pathway, amino sugar and nucleotide sugar metabolism), along with immune signaling (interleukin-2 and interleukin-6 pathways) and androgen receptor regulation. In contrast, pigs showed broader enrichment of immune and inflammatory signaling (e.g., TNF-alpha, cytokines, interleukin-1R, interleukin-5, chemokine receptor polarization), growth and developmental pathways (e.g., TGF-beta regulation, endochondral ossification), and additional signaling modules including hedgehog, c-Kit, ATF2 and CTCF. These findings suggest that shared carbohydrate metabolism pathways, together with species-specific immune, developmental and other signaling pathways, may contribute to embryonic tail morphogenesis and perinatal tail development across species.

Discussion

Our GWAS for TL revealed loci overlapping with previously reported signals. In sheep, the locus on OAR11:40,897,359 was located ~ 3 Mb from a previously identified candidate causal locus at OAR11:37,525,128 [21], where we also detected the reported 167 bp insertion in *HOXB13* [21]. In pigs, the GWAS peak on SSC6:110,116,231 was in ~ 37 Mb distance from the previously reported locus SSC6:147,382,689 [25]; however, no putative functional variants were identified at this region. In the conserved genomic windows, only *EPHB6* and *KEL* were significantly associated with TL in both species ($P_{lm} < 0.05$), although no previous functional report linked these genes to tail development. *CLCN1* mutant mice exhibited mild to severe hind limb spasms and abnormal hind limb reflexes, and mice homozygous for an ENU-induced allele displayed spine deformities and reduced body size [59]. *EPHA1* homozygous null mice often presented a kinked tail, while 18% also exhibited vaginal atresia with hydrometrocolpos and infertility [59]. This may suggest a potential link between tail length, tail abnormalities and reproductive traits. In our study, the *CLCN1* variant was significantly associated with TL only in pigs. Regarding *EPHA1*, the association with TL became significant in both species when all six putative functional variants from the common genes were analyzed together with covariates. Further studies are warranted to investigate their

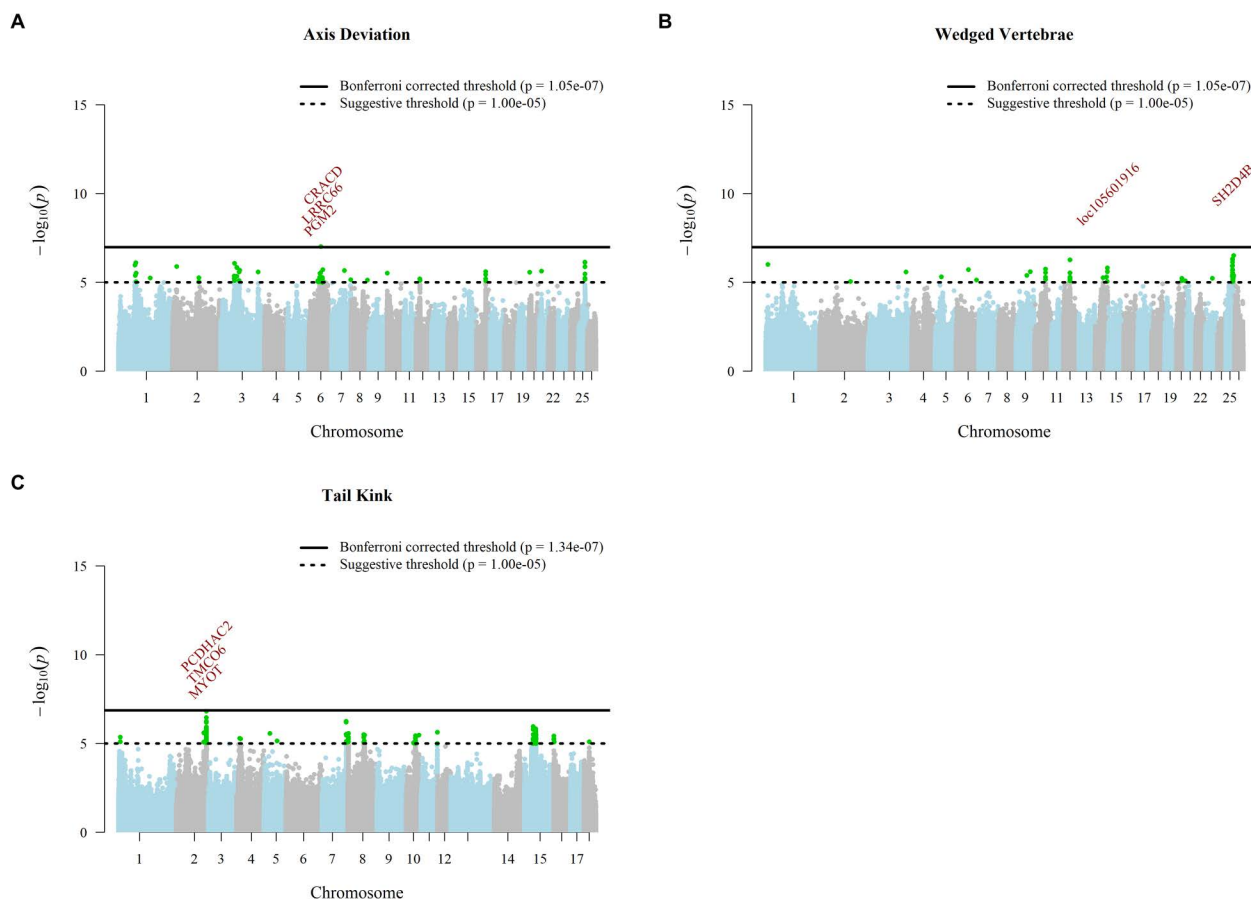


Fig 3. Species-specific candidate genes in GWAS loci for tail abnormalities in sheep and pigs. Manhattan plots labeled with species-specific putative functional candidate genes identified within genomic windows of GWAS loci for axis deviation (A) and wedged vertebrae (B) in sheep, and tail kink (C) in pigs. No conserved genomic window was detected for tail abnormalities across species. Three candidate genes were identified on OAR6 for axis deviation, one on OAR25 for wedged vertebrae in sheep, and three on SSC2 for tail kink in pigs.

<https://doi.org/10.1371/journal.pone.0343836.g003>

potential interactions, as well as interactions with species-specific candidate genes predicted by GeneMANIA, in post-anal tail morphogenesis.

Nine of the 13 sheep-specific candidate genes for tail traits, *EPB41L3*, *MTCL1*, *PIEZO2*, *MPPE1*, *MEX3C*, *PGM2*, *LRRC66*, *CRACD* and *SH2D4B*, have been reported as GWAS loci for (birth) body height in humans, according to GeneCards [60]. In most vertebrate species, tail size is directly linked to the embryo's capacity, e.g., to generate somites after trunk formation [61]. Also, the positive phenotypic correlation of 0.54 between body length and tail length indicated shared genetic regulations [10]. *EPB41L3* is involved in tight junctions, which act as major intercellular regulators of the Hippo pathway [62], a pathway crucial for embryogenesis and tissue elongation [63]. A recessive stop variant in *PIEZO2* caused distal arthrogryposis with distal muscle weakness and scoliosis [64]. Respective phenotype annotations included short necks and short stature in humans, and abnormal embryo size in mice [59]. Mice homozygous for a null allele in *MEX3C* displayed growth retardation and shortened tibiae [59]. *LOXHD1* and *MUC5B* indicated phenotype annotations mainly related to behavior, neurological traits, and hearing [59]. However, the most significant association observed was between a deleterious missense variant (rs420232320) in *LOXHD1* and TL in sheep ($P_{lm} = 1.77 \times 10^{-13}$), explaining 22.54% of the variance in TL (ΔR^2). A functional network linking *MUC5B* with other candidate genes was predicted by GeneMANIA.

Table 6. Genotype distributions of tail abnormality (TA) and normal tail (NT) group animals for putative functional candidate variants near TA GWAS loci in sheep and pigs.

Chr_Pos	rs.ID	R>A	Gene	Con- seq.	SnEff_ SIFT	RR-TA(n)	RR-NT(n)	RR- TA%	RA-TA(n)	RA-NT(n)	RA- TA%	AA- TA(n)	AA- NT(n)	AA- TA%	TA type	P _{gim}
Sheep:																
6:57882091	Novel	C>T	PGM2	mis- sense	Mi	12	89	11.88	11	17	39.29	2	1	66.67	AXISD	3.86E-04**
6:68523704	rs416623867	T>G	LRRRC66	mis- sense	mi_0.04	0	17	0.00	7	47	12.96	19	44	30.16	AXISD	2.79E-03*
6:72450156	rs417947464	G>A	CRACD	mis- sense	mi_0.04	10	84	10.64	16	24	40.00	–	–	–	AXISD	4.73E-04**
14:65770747	rs160996012	T>G	loc105601916	mis- sense	mi_0	6	14	30.00	7	60	10.45	0	47	0.00	WDGV	1.11E-03*
25:34986424	rs596500732	A>G	SH2D4B	mis- sense	mi_0.03	3	93	3.13	9	28	24.32	1	0	100	WDGV	1.74E-04**
25:35000318	rs425695059	C>T	SH2D4B	mis- sense	mi_0.01	3	93	3.13	9	28	24.32	1	0	100	WDGV	1.74E-04**
Pigs:																
2:139914508	rs81212592	C>A	MYOT	mis- sense	mi_0	17	70	19.54	19	23	45.24	5	4	55.56	Kink	1.63E-03*
2:142357467	rs81366497	G>A	TMCO6	mis- sense	mi_0.04	23	82	21.90	18	15	54.55	–	–	–	Kink	4.14E-04**
2:142460474	rs335834834	C>A	PCDHAC2	mis- sense	mi_0.05	5	1	83.33	20	25	44.44	16	71	18.39	Kink	8.11E-05**
2:142461115	Novel	C>CA	PCDHAC2	frame- shift	Hi	16	72	18.18	20	24	45.45	5	1	83.33	Kink	6.21E-05**

Variants include high-impact (hi) or moderate-impact (mi) SNPs with deleterious SIFT scores (<0.05). SNPs are given as Chromosome:Position (Ovis aries ARS-UI_Ramb. v2.0; Sus scrofa Sscrofa11.1) with rs.IDs. Alleles are shown as R>A (reference > alternate); genotypes as RR (homozygous reference), RA (heterozygous), and AA (homozygous alternate). TA types: AXISD (axis deviation), WDG (wedged vertebrae) in sheep; Kink (tail kink) in pigs. TA (n) = number of affected individuals, NT (n) = number of unaffected individuals, TA% = prevalence of affected individuals. P_{gim} = logistic regression p-value (g/m in R, adjusted for sex, age and tail length/vertebrae count). Significance: P_{gim} < 0.05 (*), P_{gim} < 1.00 × 10⁻³ (**).

<https://doi.org/10.1371/journal.pone.0343836.t006>

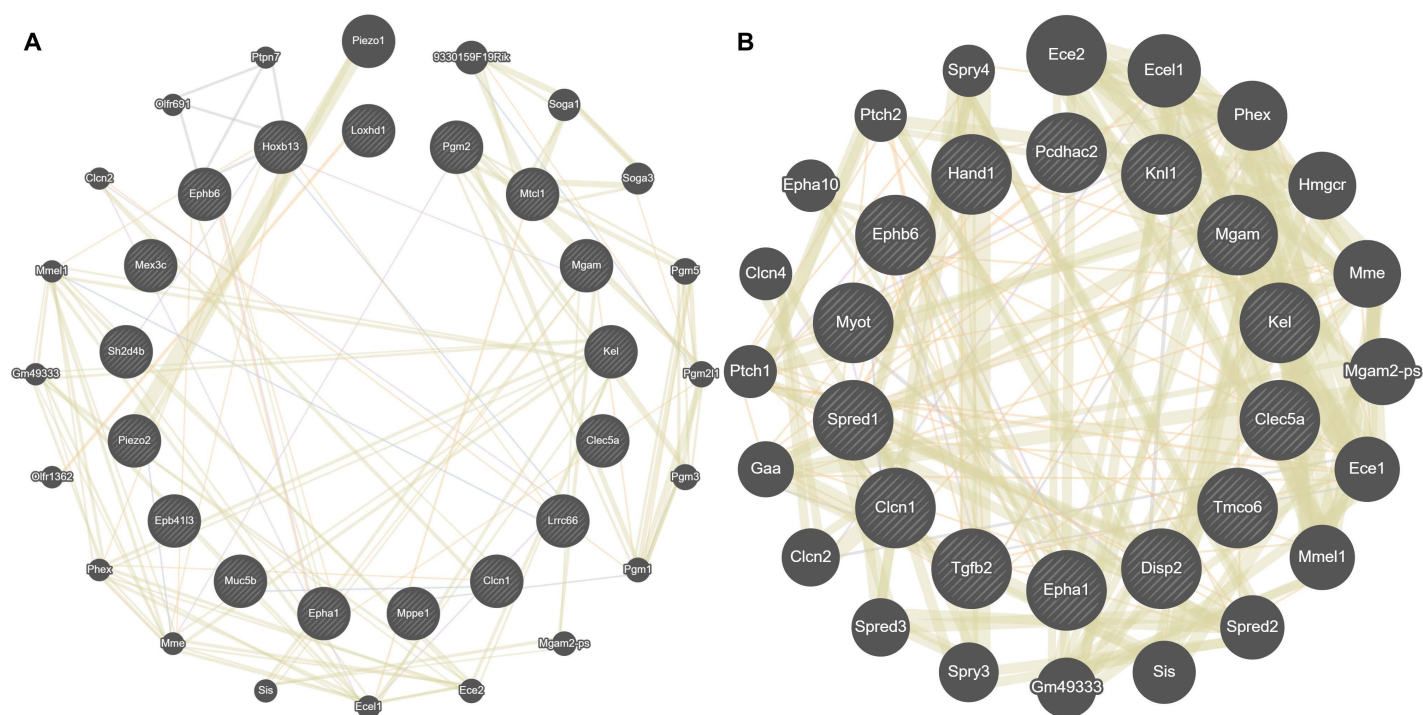


Fig 4. Network analysis. GeneMANIA diagrams illustrating interactions among the putative functional candidate genes (central nodes) and their neighboring genes (outer nodes) for tail traits in sheep (A) and pigs (B).

<https://doi.org/10.1371/journal.pone.0343836.g004>

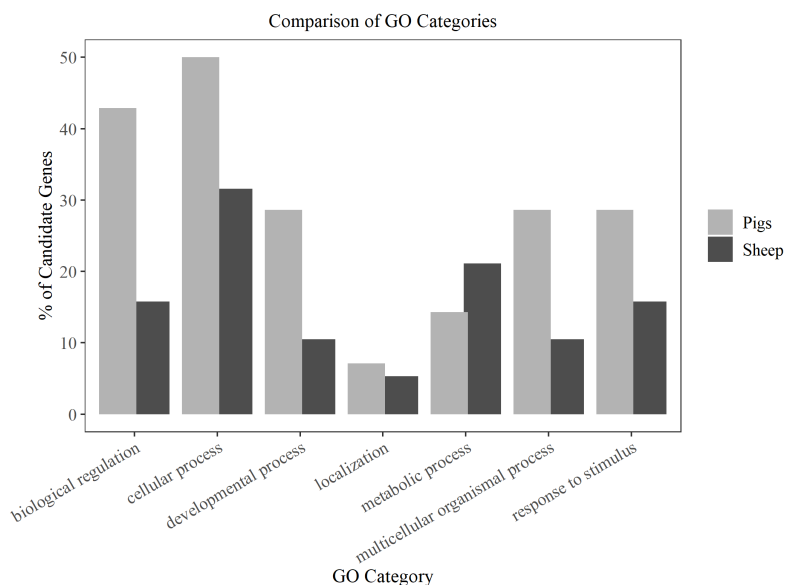


Fig 5. Comparison of GO enrichment analysis of candidate genes for tail traits in sheep and pigs. Bars indicate the percentage of candidate genes mapped to each GO category.

<https://doi.org/10.1371/journal.pone.0343836.g005>

Table 7. Pathways enriched in candidate genes associated with tail traits in sheep and pigs.

Species	Pathway	P-value	Odds Ratio	Combined Score
Sheep	Galactose metabolism	2.45E-04	104	864
	Starch and sucrose metabolism	1.03E-03	49	336
	Glycogen biosynthesis	7.18E-03	168	829
	Glycogen breakdown (glycogenolysis)	1.25E-02	90	396
	Interleukin-6 regulation of target genes	1.25E-02	90	396
	Carbohydrate metabolism	1.86E-02	11	42
	Termination of O-glycan biosynthesis	2.32E-02	47	177
	Glucuronidation	2.40E-02	45	168
	Pentose phosphate pathway	2.40E-02	45	168
	Interleukin-2 signaling pathway	3.84E-02	5	15
	Amino sugar and nucleotide sugar metabolism	3.98E-02	27	86
	Androgen receptor regulation of biosynthesis and transcription	4.67E-02	23	69
Pigs	TNF-alpha effects on cytokine activity, cell motility, and apoptosis	3.90E-03	25	138
	Digestion of dietary carbohydrate	5.59E-03	220	1139
	TGF-beta regulation of extracellular matrix	6.47E-03	9	48
	Signaling events mediated by the hedgehog family	1.60E-02	70	289
	Galactose metabolism	1.81E-02	61	247
	Selective expression of chemokine receptors during T-cell polarization	1.81E-02	61	247
	TGF-beta signaling in development	1.81E-02	61	247
	CTCF pathway	1.87E-02	59	235
	Cytokines and inflammatory response	2.42E-02	45	168
	Signal transduction through IL-1R	2.63E-02	41	151
	Interleukin-5 signaling pathway	3.38E-02	32	108
	Signaling events mediated by stem cell factor receptor (c-Kit)	3.58E-02	30	100
	Starch and sucrose metabolism	3.65E-02	29	98
	ATF2 transcription factor network	4.05E-02	26	85
	Endochondral ossification	4.52E-02	24	73

Pathways shared between sheep and pigs are highlighted in bold.

<https://doi.org/10.1371/journal.pone.0343836.t007>

Seven of the eight pig-specific candidate genes for tail traits, *KNL1*, *DISP2*, *SPRED1*, *TGFB2*, *HAND1*, *TMCO6* and *PCDHAC2*, have also been reported as GWAS loci for body height in humans [60]. *KNL1* was annotated in the Human Phenotype Ontology with abnormal axial skeleton morphology [60]. *DISP2* may be required for normal Hedgehog (Hh) signaling during embryonic development, and together with upstream Wnt signaling, these pathways likely function in a coordinated manner to regulate tail regeneration [65]. Homozygous null mice for *SPRED1* exhibit kinked tails, shortened faces, and altered CNS transmission, while deficiency of *TGF-β2* resulted in severe developmental abnormalities across limbs, digits, the tail and other organs [59]. Variants in *HAND1* were associated with abnormal limb bud morphology, disrupted somite development, and reduced embryo size [59]. *MYOT*, involved in myogenesis, has been implicated in regenerative processes along the developing tail axis [66]. Collectively, these functional insights support the potential roles of pig-specific candidate genes in tail length regulation and morphogenesis.

Our GO and pathway analyses align with previous mechanistic findings on tail development. Convergent extension (CE) was a key mechanism for body axis elongation without additional growth [67], and Wnt/planar cell polarity (PCP) signaling played a crucial role in cell polarization during CE [68]. In both zebrafish and mice, Wnt/PCP signaling has been demonstrated to be indispensable for tail elongation [69]. Fibroblast Growth Factor (FGF) signaling was imperative for

sustaining axial progenitors and promoting posterior body elongation in mouse embryos [70]. In axolotls, TGF- β , Wnt and FGF signaling jointly regulated tail regeneration [71], while a member of the TGF- β superfamily, *Gdf11*, interacted with *Lin28* and *Hox13* genes to control axial progenitor activity in the tail bud [72]. Consistent with these findings, our pathway analyses revealed enrichment of TGF- β signaling among pig-specific candidate genes for tail traits. Furthermore, a gradient of glycolytic activity has been shown to coordinate FGF and Wnt signaling during axis elongation in amniote embryos [73]. Supporting this, we observed enrichment of carbohydrate metabolism pathways, including shared galactose and starch/sucrose metabolism, as well as sheep-specific glycogen biosynthesis. This suggests that energy metabolism and biosynthetic fluxes may be critical for cell proliferation, differentiation, and tissue remodeling during tail elongation. Species-specific enrichment of immune and inflammatory signaling pathways further suggests a role for localized signaling cues in orchestrating morphogenesis. Together, these findings highlight an interplay between metabolic supply and signaling regulation in driving species-specific aspects of embryonic tail development.

Implementation of results from integrative genomic analyses into breeding programs

The challenge for improving selection in farm animals is the combination of classical quantitative-trait loci (QTL) mapping approaches such as complex segregation analysis, high-throughput genomics carried out by regional GWAS and gene expressions combined with expression QTL detection [38]. Results from biological case studies considering extreme subgroups (e.g., between different extreme environmental conditions or lactation stages, or, as in the present study, offspring from a selection experiment) will contribute to a deeper understanding of genomic mechanisms as shown for adipogenesis in cattle [74]. In a simulation study, Teng et al. [75] compared three genomic prediction models considering causal gene information with the GBLUP model that assumes a flat prior. The inclusion of causal genes in a genomic prediction model increased the accuracy for different modelling strategies, i.e., considering the causal genes as fixed effects, as separate random components, or as a weight factor in the genomic relationship matrix. Similar improvements were observed for real data when using the most significant SNP from GWAS (mostly located within potential candidate genes) as weighting factors in building up the genomic relationship matrix [76].

However, a prerequisite for applying such enhanced genomic prediction models is that the causal genes explain a considerable proportion of genetic variance and that potential pleiotropic effects are explored. The classical example in this regard are the variants in the *T-box* gene with favorable effects on TL, but proven pleiotropic effects on vertebral defects [29]. In the present study, a potential role of *KNL1*, *SPRED1*, *TGFB2*, and *PIEZO2* in TL development in pigs or sheep was inferred. However, previous studies in humans and mice have indicated pleiotropic effects of these genes, including distal arthrogryposis, altered CNS transmission, severe developmental abnormalities in other organs, perinatal mortality, and so on [59,60,64].

Study limitations and prospects

When comparing to “classical” 50K based GWAS, the sample size in the present study including 140 lambs and 140 piglets seems to be quite small. Nevertheless, in the present study, we focused on WGS-based GWAS in Merinoland sheep with similar sample sizes in other sheep breeds (e.g., N = 196 sequenced Tibetan sheep [77]). In pigs, most of the WGS-based GWAS considered larger sample sizes, but using imputed WGS data (e.g., [78]). However, in the present study, the genotypes and phenotypes were available from animals kept on one research herd, implying a smaller number of “disturbing” environmental or fixed effects which need to be considered in the genetic-statistical modelling approach. Such specific on-station data generally contributed to smaller residual variances and increased trait heritabilities. As a concern, possible genotype-by-environment interactions due to differences in feeding, housing or climatic conditions indicate losses in selection response in commercial pig herds when basing selection solely on pig or sheep on-station records. Furthermore, our quite small sample size can be justified in the context of the selection experiment. We explicitly focused on the generation of extreme sub-groups for TL (from a genetic and a phenotypic perspective), contributing to increasing contrasts

in case-control studies and more pronounced SNP effects. Such advantages of selective genotyping or sequencing have been highlighted, for example, in studies of clinical mastitis and production traits in Holstein dairy cattle [79].

Based on the putative functional candidate variants identified through genome sequencing, gene tests can be developed in follow-up studies for in-depth validation across other sheep and pig breeds. Some of the candidate genes, such as *PIEZO2*, *CLCN1* and *EPHA1*, have been shown to be expressed in the whole sheep embryo [80,81]. It would be informative to perform transcriptomic analyses to identify differentially expressed genes between short- and long-tailed individuals.

Conclusions

The current comparative genomic study, based on WGS, investigated tail traits across species. Conserved genomic windows were identified in sheep (OAR4:107,294,668) and in pigs (SSC18:7,836,326), harboring predicted functional variants in six common genes associated with TL ($P_{lm} = 6.03 \times 10^{-3}$ in sheep; $P_{lm} = 7.14 \times 10^{-4}$ in pigs). In addition, eight sheep-specific and five pig-specific putatively functional candidate genes were identified for TL. For TA, no conserved elements were detected. Species-specific candidate genes included five variants associated with axis deviation or wedged vertebrae in sheep, and three variants associated with tail kink in pigs. Variants in some candidate genes, including *KNL1*, *SPRED1*, *TGFB2* and *PIEZO2*, have been linked to skeletal abnormalities, perinatal lethality, altered CNS transmission, and other developmental defects in mice and humans, suggesting potential unfavorable pleiotropic effects. Gene ontology analyses of the identified candidate genes revealed similar biological processes enriched in both sheep and pigs, although the relative contribution of each process differed between species. Pathway analysis suggested that shared carbohydrate metabolism pathways, together with species-specific immune and inflammatory signaling, as well as pig-specific TGF- β signaling and endochondral ossification, may contribute to embryonic tail morphogenesis and perinatal tail development. Overall, these findings highlight the complex and multi-layered genetic mechanisms underlying tail development across species.

Supporting information

S1 Table. Significant and suggestive GWAS loci for tail length in sheep.

(XLSX)

S2 Table. Significant and suggestive GWAS loci for tail length in pigs.

(XLSX)

S3 Table. Significant and suggestive GWAS loci for tail abnormalities (axis deviation, wedged vertebrae) in sheep.

(XLSX)

S4 Table. Significant and suggestive GWAS loci for tail abnormalities (kink grades, Kinks01) in pigs.

(XLSX)

S5 Table. Genes mapped within ± 1 Mb windows of GWAS loci for tail traits in sheep, carrying high- or moderate-impact variants.

(XLSX)

S6 Table. Genes mapped within ± 1 Mb windows of GWAS loci for tail traits in pigs, carrying high- or moderate-impact variants.

(XLSX)

Author contributions

Conceptualization: Petra Engel, Christiane Egerer, Sven König.

Data curation: Tong Yin, Katharina Gerhards.

Formal analysis: Xuying Zhang, Tong Yin, Katharina Gerhards.

Funding acquisition: Sven König.

Investigation: Gerald Reiner, Sven König.

Methodology: Hannah Hümmelchen, Henrik Wagner, Axel Wehrend, Christiane Egerer, Gerald Reiner, Sven König.

Project administration: Sven König.

Resources: Johanna Mainzer.

Software: Xuying Zhang.

Supervision: Gerald Reiner, Sven König.

Validation: Isabella Giambra.

Writing – original draft: Xuying Zhang, Sven König.

Writing – review & editing: Xuying Zhang, Isabella Giambra, Tong Yin, Petra Engel, Hannah Hümmelchen, Henrik Wagner, Axel Wehrend, Christiane Egerer, Katharina Gerhards, Gerald Reiner, Sven König.

References

1. Sutherland MA, Tucker CB. The long and short of it: A review of tail docking in farm animals. *Appl Anim Behav Sci*. 2011;135(3):179–91. <https://doi.org/10.1016/j.applanim.2011.10.015>
2. French NP, Wall R, Morgan KL. Lamb tail docking: a controlled field study of the effects of tail amputation on health and productivity. *Vet Rec*. 1994;134(18):463–7. <https://doi.org/10.1136/vr.134.18.463> PMID: 8059511
3. Schrader L, Roth H-R, Winterling C, Brodmann N, Langhans W, Geyer H, et al. The Occurrence of Tail Tip Alterations in Fattening Bulls Kept Under Different Husbandry Conditions. *Anim welf*. 2001;10(2):119–30. <https://doi.org/10.1017/s0962728600023794>
4. Shelton M. Studies on tail length of Rambouillet and Mouflon sheep. *J Hered*. 1977;68(2):128–30. <https://doi.org/10.1093/oxfordjournals.jhered.a108788> PMID: 874307
5. Richtlinie 2008/120/EG des Rates vom 18. Dezember 2008 über Mindestanforderungen für den Schutz von Schweinen. 2008.
6. Bauer B, Ganter M, Strobel H, Hannemann R. Schmerzhaftes Eingriffe beim Schaf – Schwanzkupieren. *Tierarztl Prax Ausg G*. 2017;45(05):302–11. <https://doi.org/10.15653/tpg-170354>
7. Genetic selection as alternative to tail docking in Hampshire Down and Clun Forest. 2010.
8. Oltenacu EA, Boylan WJ. Inheritance of tail length in crossbred Finnsheep. *J Hered*. 1974;65(6):331–4. <https://doi.org/10.1093/oxfordjournals.jhered.a108543> PMID: 4448895
9. Oberpenning J, Bohlouli M, Engel P, Hümmelchen H, Wagner H, Wehrend A, et al. Multiple-trait and structural equation modelling approaches to infer genetic relationships between tail length and weight traits in Merinoland sheep. *J Anim Breed Genet*. 2023;140(2):132–43. <https://doi.org/10.1111/jbg.12752> PMID: 36583443
10. Mainzer J, Yin T, Giambra I, Hümmelchen H, Engel P, Wagner H, et al. Genetic parameters, genome-wide associations and potential candidate genes for additive and dominance effects of tail traits in Merinoland sheep based on whole-genome sequence data in a selection experiment. *Anim Genet*. 2025;56(5):e70041. <https://doi.org/10.1111/age.70041> PMID: 40965274
11. Aikins-Wilson S, Bohlouli M, König S. Maternal and direct genetic parameters for tail length, tail lesions, and growth traits in pigs. *J Anim Sci*. 2021;99(1):skaa398. <https://doi.org/10.1093/jas/skaa398> PMID: 33320242
12. Cheverud JM, Dittus WPJ. Primate population studies at Polonnaruwa. II. Heritability of body measurements in a natural population of toque macaques (*Macaca sinica*). *Am J Primatol*. 1992;27(2):145–54. <https://doi.org/10.1002/ajp.1350270209> PMID: 31948145
13. Kramer MG, Vaughn TT, Pletscher LS, King-Ellison K, Adams E, Erikson C, et al. Genetic variation in body weight gain and composition in the intercross of Large (LG/J) and Small (SM/J) inbred strains of mice. *Genet Mol Biol*. 1998;21(2):211–8. <https://doi.org/10.1590/s1415-47571998000200007>
14. de Holmdahl 14. Experimentelle Untersuchungen über die Lage der Grenze zwischen primärer und sekundärer Körperentwicklung beim Huhn. *Anat Anz*. 1925;59(6):6.
15. Handrigan GR. Concordia discors: duality in the origin of the vertebrate tail. *J Anat*. 2003;202(Pt 3):255–67. <https://doi.org/10.1046/j.1469-7580.2003.00163.x> PMID: 12713266
16. Stern CD, Charité J, Deschamps J, Duboule D, Durston AJ, Kmita M, et al. Head-tail patterning of the vertebrate embryo: one, two or many unresolved problems?. *Int J Dev Biol*. 2006;50(1):3–15. <https://doi.org/10.1387/ijdb.052095cs> PMID: 16323073

17. Wilson V, Olivera-Martinez I, Storey KG. Stem cells, signals and vertebrate body axis extension. *Development*. 2009;136(12):2133. <https://doi.org/10.1242/dev.039172>
18. Mallo M. The vertebrate tail: a gene playground for evolution. *Cell Mol Life Sci*. 2020;77(6):1021–30. <https://doi.org/10.1007/s00018-019-03311-1> PMID: 31559446
19. Xu X, Sun X, Hu X-S, Zhuang Y, Liu Y-C, Meng H, et al. Whole Genome Sequencing Identifies a Missense Mutation in HES7 Associated with Short Tails in Asian Domestic Cats. *Sci Rep*. 2016;6:31583. <https://doi.org/10.1038/srep31583> PMID: 27560986
20. James PJ, Ponzoni RW, Gare, Cockrum KS. Inheritance of short tailedness in South Australian Merinos. 1991.
21. Lagler DK, Hannemann E, Eck K, Klawatsch J, Seichter D, Russ I, et al. Fine-mapping and identification of candidate causal genes for tail length in the Merinolandschaf breed. *Commun Biol*. 2022;5(1):918. <https://doi.org/10.1038/s42003-022-03854-3> PMID: 36068271
22. Han J, Yang M, Guo T, Niu C, Liu J, Yue Y, et al. Two linked TBXT (brachyury) gene polymorphisms are associated with the tailless phenotype in fat-rumped sheep. *Anim Genet*. 2019;50(6):772–7. <https://doi.org/10.1111/age.12852> PMID: 31475743
23. Mastrangelo S, Moiola B, Ahbara A, Latairish S, Portolano B, Pilla F, et al. Genome-wide scan of fat-tail sheep identifies signals of selection for fat deposition and adaptation. *Anim Product Sci*. 2018;59(5):835–48. <https://doi.org/10.1071/an17753>
24. Li R, Gong M, Zhang X, Wang F, Liu Z, Zhang L, et al. The first sheep graph-based pan-genome reveals the spectrum of structural variations and their effects on tail phenotypes. 2021.
25. Gerhards K, Egerer C, Becker S, Willems H, Engel P, Koenig S, et al. Genome-Wide Association Study Reveals Single Nucleotide Polymorphisms Associated with Tail Length and Tail Kinks in Piglets. *Vet Sci*. 2025;12(3):198. <https://doi.org/10.3390/vetsci12030198> PMID: 40266940
26. Hytönen MK, Grall A, Hédan B, Dréano S, Seguin SJ, Delattre D, et al. Ancestral T-box mutation is present in many, but not all, short-tailed dog breeds. *J Hered*. 2009;100(2):236–40. <https://doi.org/10.1093/jhered/esn085> PMID: 18854372
27. Kromik A, Ulrich R, Kusenda M, Tipold A, Stein VM, Hellige M, et al. The mammalian cervical vertebrae blueprint depends on the T (brachyury) gene. *Genetics*. 2015;199(3):873–83. <https://doi.org/10.1534/genetics.114.169680> PMID: 25614605
28. Wu B, Shao Y, Chen B, Liu C, Xue Z, Wu P, et al. Identification of a novel mouse brachyury (T) allele causing a short tail mutation in mice. *Cell Biochem Biophys*. 2010;58(3):129–35. <https://doi.org/10.1007/s12013-010-9097-9> PMID: 20809182
29. Buckingham KJ, McMillin MJ, Brassil MM, Shively KM, Magnaye KM, Cortes A, et al. Multiple mutant T alleles cause haploinsufficiency of Brachyury and short tails in Manx cats. *Mamm Genome*. 2013;24(9–10):400–8. <https://doi.org/10.1007/s00335-013-9471-1> PMID: 23949773
30. Greco TL, Takada S, Newhouse MM, McMahon JA, McMahon AP, Camper SA. Analysis of the vestigial tail mutation demonstrates that Wnt-3a gene dosage regulates mouse axial development. *Genes Dev*. 1996;10(3):313–24. <https://doi.org/10.1101/gad.10.3.313> PMID: 8595882
31. Wilm B, Dahl E, Peters H, Balling R, Imai K. Targeted disruption of Pax1 defines its null phenotype and proves haploinsufficiency. *Proc Natl Acad Sci U S A*. 1998;95(15):8692–7. <https://doi.org/10.1073/pnas.95.15.8692> PMID: 9671740
32. Li X, Yang J, Shen M, Xie X-L, Liu G-J, Xu Y-X, et al. Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat Commun*. 2020;11(1):2815. <https://doi.org/10.1038/s41467-020-16485-1> PMID: 32499537
33. Liu Z, Bai C, Shi L, He Y, Hu M, Sun H, et al. Detection of selection signatures in South African Mutton Merino sheep using whole-genome sequencing data. *Anim Genet*. 2022;53(2):224–9. <https://doi.org/10.1111/age.13173> PMID: 35099062
34. Yi W, Hu M, Shi L, Li T, Bai C, Sun F, et al. Whole genome sequencing identified genomic diversity and candidate genes associated with economic traits in Northeastern Merino in China. *Front Genet*. 2024;15:1302222. <https://doi.org/10.3389/fgene.2024.1302222> PMID: 38333624
35. Zhu Q, Liu P, Zhang M, Kang Y, Lv L, Xu H, et al. The Detection of a Functional 168 bp Deletion of the HOXB13 Gene Determining Short Tail and Its Association with Senior Growth Traits in Sheep Breeds Worldwide. *Animals (Basel)*. 2024;14(11):1617. <https://doi.org/10.3390/ani14111617> PMID: 38891664
36. Alföldi J, Lindblad-Toh K. Comparative genomics as a tool to understand evolution and disease. *Genome Res*. 2013;23(7):1063–8. <https://doi.org/10.1101/gr.157503.113> PMID: 23817047
37. Coan PM, Hummel O, Garcia Diaz A, Barrier M, Alfazema N, Norsworthy PJ, et al. Genetic, physiological and comparative genomic studies of hypertension and insulin resistance in the spontaneously hypertensive rat. *Dis Model Mech*. 2017;10(3):297–306. <https://doi.org/10.1242/dmm.026716> PMID: 28130354
38. de Las Heras-Saldana S, Lopez BI, Moghaddar N, Park W, Park J-E, Chung KY, et al. Use of gene expression and whole-genome sequence information to improve the accuracy of genomic prediction for carcass traits in Hanwoo cattle. *Genet Sel Evol*. 2020;52(1):54. <https://doi.org/10.1186/s12711-020-00574-2> PMID: 32993481
39. Lee J-S, Grisham JW, Thorgeirsson SS. Comparative functional genomics for identifying models of human cancer. *Carcinogenesis*. 2005;26(6):1013–20. <https://doi.org/10.1093/carcin/bgi030> PMID: 15677630
40. Jann OC, King A, Corrales NL, Anderson SI, Jensen K, Ait-Ali T, et al. Comparative genomics of Toll-like receptor signalling in five species. *BMC Genomics*. 2009;10:216. <https://doi.org/10.1186/1471-2164-10-216> PMID: 19432955
41. Rubin C-J, Megens H-J, Martinez Barrio A, Maqbool K, Sayyab S, Schwochow D, et al. Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci U S A*. 2012;109(48):19529–36. <https://doi.org/10.1073/pnas.1217149109> PMID: 23151514
42. Bathke J, Lühken G. OVarFlow: a resource optimized GATK 4 based Open source Variant calling workFlow. 2021.
43. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011> PMID: 21167468

44. Aliloo H, Pryce JE, González-Recio O, Cocks BG, Hayes BJ. Validation of markers with non-additive effects on milk yield and fertility in Holstein and Jersey cows. *BMC Genet.* 2015;16:89. <https://doi.org/10.1186/s12863-015-0241-9> PMID: [26193888](#)
45. Huang M, Liu X, Zhou Y, Summers RM, Zhang Z. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience.* 2019;8(2):giy154. <https://doi.org/10.1093/gigascience/giy154> PMID: [30535326](#)
46. Purcell S. PLINK 1.9. 2007. Available from: <http://pngu.mgh.harvard.edu/purcell/plink/>
47. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin).* 2012;6(2):80–92. <https://doi.org/10.4161/fly.19695> PMID: [22728672](#)
48. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene Set Knowledge Discovery with Enrichr. *Curr Protoc.* 2021;1(3):e90. <https://doi.org/10.1002/cpz1.90> PMID: [33780170](#)
49. Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albu L-P, Mi H. PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci.* 2022;31(1):8–22. <https://doi.org/10.1002/pro.4218> PMID: [34717010](#)
50. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics.* 2020;36(8):2628–9. <https://doi.org/10.1093/bioinformatics/btz931> PMID: [31882993](#)
51. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43(Database issue):D447–52. <https://doi.org/10.1093/nar/gku1003> PMID: [25352553](#)
52. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 2010;38(Web Server issue):W214–20. <https://doi.org/10.1093/nar/gkq537> PMID: [20576703](#)
53. Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 2008;36(Web Server issue):W5–9. <https://doi.org/10.1093/nar/gkn201> PMID: [18440982](#)
54. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011;7:539. <https://doi.org/10.1038/msb.2011.75> PMID: [21988835](#)
55. Aldersey JE, Sonstegard TS, Williams JL, Bottema CDK. Understanding the effects of the bovine POLLED variants. *Anim Genet.* 2020;51(2):166–76. <https://doi.org/10.1111/age.12915> PMID: [31999853](#)
56. Stadhouders R, Aktuna S, Thongjuea S, Aghajanirofeh A, Pourfarzad F, van Ijcken W, et al. HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. *J Clin Invest.* 2014;124(4):1699–710. <https://doi.org/10.1172/JCI71520> PMID: [24614105](#)
57. Visser M, Palstra R-J, Kayser M. Human skin color is influenced by an intergenic DNA polymorphism regulating transcription of the nearby BNC2 pigmentation gene. *Hum Mol Genet.* 2014;23(21):5750–62. <https://doi.org/10.1093/hmg/ddu289> PMID: [24916375](#)
58. Egerer C, Gerhards K, Becker S, Engel P, König S, Reiner G. Pig tail length is associated with the prevalence of tail malformations but not with inflammation of the tail. *BMC Vet Res.* 2025;21(1):111. <https://doi.org/10.1186/s12917-025-04598-y> PMID: [40011910](#)
59. Baldarelli RM, Smith CL, Ringwald M, Richardson JE, Bult CJ, Mouse Genome Informatics Group. Mouse Genome Informatics: an integrated knowledgebase system for the laboratory mouse. *Genetics.* 2024;227(1):iyae031. <https://doi.org/10.1093/genetics/iyae031> PMID: [38531069](#)
60. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics.* 2016;54:1.30.1-1.30.33. <https://doi.org/10.1002/cpbi.5> PMID: [27322403](#)
61. Mallo M. Of Necks, Trunks and Tails: Axial Skeletal Diversity among Vertebrates. *Diversity.* 2021;13(7):289. <https://doi.org/10.3390/d13070289>
62. Fulford A, Tapon N, Ribeiro PS. Upstairs, downstairs: spatial regulation of Hippo signalling. *Curr Opin Cell Biol.* 2018;51:22–32. <https://doi.org/10.1016/j.ceb.2017.10.006> PMID: [29154163](#)
63. Campàs O, Noordstra I, Yap AS. Adherens junctions as molecular regulators of emergent tissue mechanics. *Nat Rev Mol Cell Biol.* 2024;25(4):252–69. <https://doi.org/10.1038/s41580-023-00688-7> PMID: [38093099](#)
64. Haliloglu G, Becker K, Temucin C, Talim B, Küçükşahin N, Pergande M, et al. Recessive PIEZO2 stop mutation causes distal arthrogryposis with distal muscle weakness, scoliosis and proprioception defects. *J Hum Genet.* 2017;62(4):497–501. <https://doi.org/10.1038/jhg.2016.153> PMID: [27974811](#)
65. Singh BN, Weaver CV, Garry MG, Garry DJ. Hedgehog and Wnt Signaling Pathways Regulate Tail Regeneration. *Stem Cells Dev.* 2018;27(20):1426–37. <https://doi.org/10.1089/scd.2018.0049> PMID: [30003832](#)
66. Hutchins ED, Markov GJ, Eckalbar WL, George RM, King JM, Tokuyama MA, et al. Transcriptomic analysis of tail regeneration in the lizard *Anolis carolinensis* reveals activation of conserved vertebrate developmental and repair mechanisms. *PLoS One.* 2014;9(8):e105004. <https://doi.org/10.1371/journal.pone.0105004> PMID: [25140675](#)
67. Tada M, Heisenberg C-P. Convergent extension: using collective cell migration and cell intercalation to shape embryos. *Development.* 2012;139(21):3897–904. <https://doi.org/10.1242/dev.073007> PMID: [23048180](#)
68. Gray RS, Roszko I, Solnica-Krezel L. Planar cell polarity: coordinating morphogenetic cell behaviors with embryonic polarity. *Dev Cell.* 2011;21(1):120–33. <https://doi.org/10.1016/j.devcel.2011.06.011> PMID: [21763613](#)
69. Rauch GJ, Hammerschmidt M, Blader P, Schuete HE, Strähle U, Ingham PW, et al. WNT5 Is Required for Tail Formation in the Zebrafish Embryo. *Cold Spring Harb Symp Quant Biol.* 1997; 62:227–34. <https://doi.org/10.1101/SQB.1997.062.01.028>

70. Boulet AM, Capecchi MR. Signaling by FGF4 and FGF8 is required for axial elongation of the mouse embryo. *Dev Biol.* 2012;371(2):235–45. <https://doi.org/10.1016/j.ydbio.2012.08.017> PMID: [22954964](#)
71. Qiu Q. Tgf- β , Wnt, and Fgf signaling pathways during axolotl tail regeneration and forelimb bud development. University of Kentucky Libraries; 2019.
72. Aires R, de Lemos L, Nóvoa A, Jurberg AD, Mascres B, Duboule D, et al. Tail Bud Progenitor Activity Relies on a Network Comprising Gdf11, Lin28, and Hox13 Genes. *Dev Cell.* 2019;48(3):383–395.e8. <https://doi.org/10.1016/j.devcel.2018.12.004> PMID: [30661984](#)
73. Oginuma M, Moncuquet P, Xiong F, Karoly E, Chal J, Guevorkian K, et al. A Gradient of Glycolytic Activity Coordinates FGF and Wnt Signaling during Elongation of the Body Axis in Amniote Embryos. *Dev Cell.* 2017;40(4):342–353.e10. <https://doi.org/10.1016/j.devcel.2017.02.001> PMID: [28245921](#)
74. Tan SH, Reverter A, Wang Y, Byrne KA, McWilliam SM, Lehnert SA. Gene expression profiling of bovine in vitro adipogenesis using a cDNA microarray. *Funct Integr Genomics.* 2006;6(3):235–49. <https://doi.org/10.1007/s10142-005-0016-x> PMID: [16470362](#)
75. Teng J, Huang S, Chen Z, Gao N, Ye S, Diao S, et al. Optimizing genomic prediction model given causal genes in a dairy cattle population. *J Dairy Sci.* 2020;103(11):10299–310. <https://doi.org/10.3168/jds.2020-18233> PMID: [32952023](#)
76. Li B, Zhang N, Wang Y-G, George AW, Reverter A, Li Y. Genomic Prediction of Breeding Values Using a Subset of SNPs Identified by Three Machine Learning Methods. *Front Genet.* 2018;9:237. <https://doi.org/10.3389/fgene.2018.00237> PMID: [30023001](#)
77. Liu D, Li X, Wang L, Pei Q, Zhao J, Sun D, et al. Genome-wide association studies of body size traits in Tibetan sheep. *BMC Genomics.* 2024;25(1):739. <https://doi.org/10.1186/s12864-024-10633-3> PMID: [39080522](#)
78. Falker-Gieske C, Blaj I, Preuß S, Bennewitz J, Thaller G, Tetens J. GWAS for Meat and Carcass Traits Using Imputed Sequence Level Genotypes in Pooled F2-Designs in Pigs. *G3 (Bethesda).* 2019;9(9):2823–34. <https://doi.org/10.1534/g3.119.400452> PMID: [31296617](#)
79. Bagheri M, Miraie-Ashtiani R, Moradi-Shahrababak M, Nejati-Javaremi A, Pakdel A, von Borstel UU, et al. Selective genotyping and logistic regression analyses to identify favorable SNP-genotypes for clinical mastitis and production traits in Holstein dairy cattle. *Livestock Sci.* 2013;151(2–3):140–51. <https://doi.org/10.1016/j.livsci.2012.11.018>
80. Wu C, Jin X, Tsueng G, Afrasiabi C, Su AI. BioGPS: building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Res.* 2016;44(D1):D313–6. <https://doi.org/10.1093/nar/gkv1104> PMID: [26578587](#)
81. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* 2009;10(11):R130. <https://doi.org/10.1186/gb-2009-10-11-r130> PMID: [19919682](#)