



OPEN

## New genomic insights into the conformation of Lipizzan horses

A. I. Gmel<sup>1,2</sup>, G. Brem<sup>3</sup> & M. Neuditschko<sup>2,3</sup>✉

Conformation traits are important selection criteria in equine breeding, as they describe the exterior aspects of the horse (height, joint angles, shape). However, the genetic architecture of conformation is not well understood, as data of these traits mainly consist of subjective evaluation scores. Here, we performed genome-wide association studies on two-dimensional shape data of Lipizzan horses. Based on this data, we identified significant quantitative trait loci (QTL) associated with cresty neck on equine chromosome (ECA)16 within the *MAG1* gene, and with type, hereby differentiating heavy from light horses on ECA5 within the *POU2F1* gene. Both genes were previously described to affect growth, muscling and fatty deposits in sheep, cattle and pigs. Furthermore, we pin-pointed another suggestive QTL on ECA21, near the *PTGER4* gene, associated with human ankylosing spondylitis, for shape differences in the back and pelvis (roach back vs sway back). Further differences in the shape of the back and abdomen were suggestively associated with the *RYR1* gene, involved in core muscle weakness in humans. Therefore, we demonstrated that horse shape space data enhance the genomic investigations of horse conformation.

Conformation refers to all exterior, visible traits of the overall morphology of an animal (height, segment lengths, joint angles, shape of the head, neck, back or croup). Horses have been selected based on conformation traits for centuries, resulting in diverse breeds with specific use (heavy type draught horses, light type sport horses, small ponies, etc.)<sup>1–3</sup>. Some traits, such as the concave (dished) head shape in the Arabian horse, are breed-specific and essentially aesthetic. Other traits are relevant for health (limb angles) or influence performance (slope of the shoulder or croup)<sup>4</sup>. While some of these traits, such as height length and angle measurements, can be objectively quantified, those describing shapes, proportions or having an aesthetic component are usually scored by experts of a particular breed. Conformation scores have been shown to be subjective and unreliable in many breeds, and are therefore difficult to be included in genetic analyses<sup>5–8</sup>. To date, only one recent genome-wide association analysis (GWAS) has identified a quantitative trait locus (QTL) for the conformation of the croup in the Icelandic horse based on subjective scoring data<sup>9</sup>. Advances in data imagery through the field of geometric morphometrics have led to the development of the horse shape space model, tracing the contour and certain anatomical landmarks on standardized photographs of horses<sup>10</sup>. Shape space data has been shown to be more heritable and reliable than traditional scores for the shoulder joint and fetlock joint angles of the front and hind limb in Franches–Montagnes (FM) horses<sup>11</sup>. Furthermore, a GWAS of joint angles from the horse shape space model applied to FM and Lipizzan horses revealed novel QTL for poll angle and elbow joint angle<sup>12</sup>. To date, two-dimensional (2D) geometric morphometric data have been already used to explore the genetic architecture of morphological differences in mollusks<sup>13</sup>, sorghum<sup>14</sup>, or house flies<sup>15</sup>, while three-dimensional (3D) data has been applied to study facial traits in humans (summarized in<sup>16</sup>), cats<sup>17</sup> and dogs<sup>18</sup>.

Our aim was to explore the genetic architecture of conformation in horses. For this purpose, we used 2D shape data from the Lipizzan horse breed. The Lipizzan breed is considered a baroque riding horse, traced back to a few known Italian and Spanish founder animals born in the eighteenth century. Between 1776 and 1945, Arabian horses were also introgressed to “ameliorate” its type, affecting the conformation of the head, neck, withers and extremities<sup>19</sup>. Currently, Lipizzan horses of a lighter type are associated with higher Arabian admixture proportions and have a more concave head and more salient withers<sup>20</sup>. Consequently, there is a relatively large phenotypic variation in the Lipizzan horse, which makes it an ideal model to explore conformation differences in the horse in a GWAS, while avoiding population stratification present when comparing different breeds.

<sup>1</sup>Equine Department, Vetsuisse Faculty, University of Zurich, Winterthurerstrasse 260, 8057 Zurich, Switzerland. <sup>2</sup>Animal GenoPhenomics, Agroscope, Rte de La Tioleyre 4, 1725 Posieux, Switzerland. <sup>3</sup>Institute of Animal Breeding and Genetics, Veterinary University Vienna, Veterinärplatz 1, 1220 Vienna, Austria. ✉email: markus.neuditschko@agroscope.admin.ch

Here, we report QTL for shape-derived conformation traits representing a gradient from a heavy to light type, from a roach to sway back, as well as for the shape of neck. These findings may have implications for the selection of Lipizzan horses, and potentially, the domesticated horse in general.

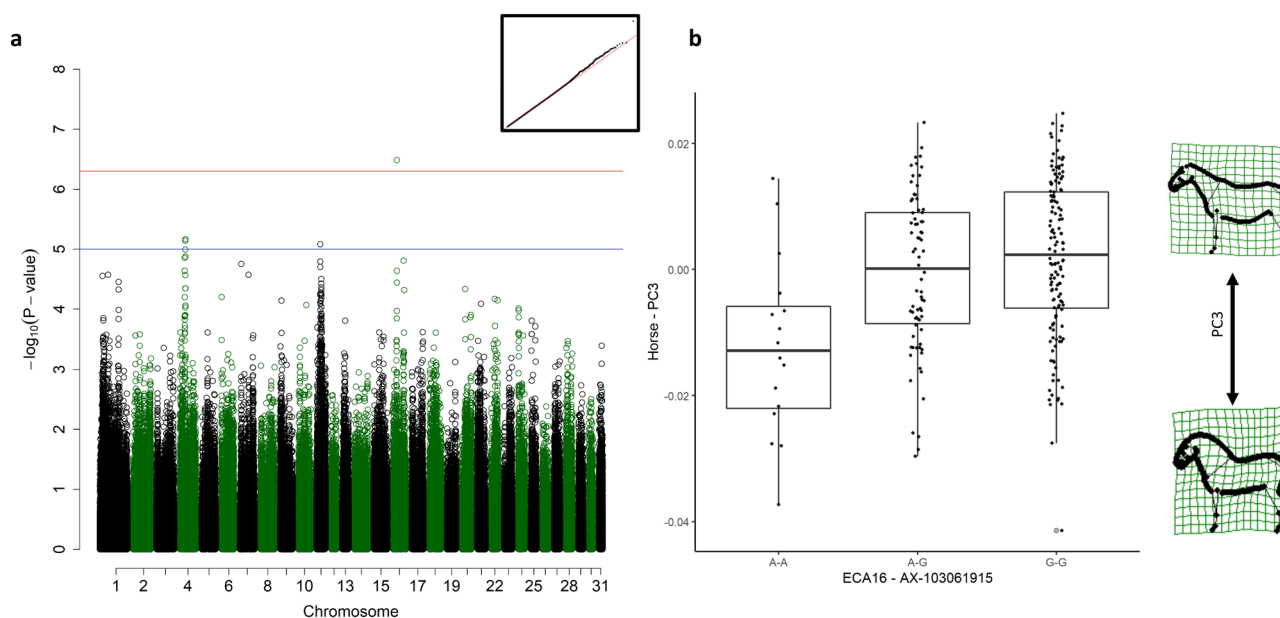
## Results

**Phenotypic shape variation.** The first twelve principal components (PCs) accounting for more than 1% of the shape variance (246 landmarks) were visualised in Supplementary Fig. 1. Based on the warp grids along the twelve PCs, eight PCs (PC1, PC2, PC4, PC5, PC8, PC9, PC10 and PC12) were excluded from the downstream analyses, as they were strongly influenced by the posture (head, neck and or limbs) of the horse (e.g. the first PC, accounting for 40.27% of the variance, mainly described the gradient of the head position). Compared to the discarded PCs, the remaining four PCs were associated with shape differences of the horses. The third PC, accounting for 8.68% of the variance, highlighted the width and shape of the neck, with the most extreme shape reminiscent of a “cresty neck”. The sixth PC showed a gradient between a horse with a shorter sway back to a horse with a longer and straight or slightly roach back and explained 4.50% of the variance. The seventh PC, accounting for 3.59% of the variance, described differences in the shape of the Lipizzan horse previously associated with Arabian genetic contributions (lighter horse type)<sup>20</sup>. The eleventh PC, explaining 1.32% of the variance, described a gradient between a horse with a larger neck, torso and a more sloping croup compared to a horse with a slimer torso and horizontal croup.

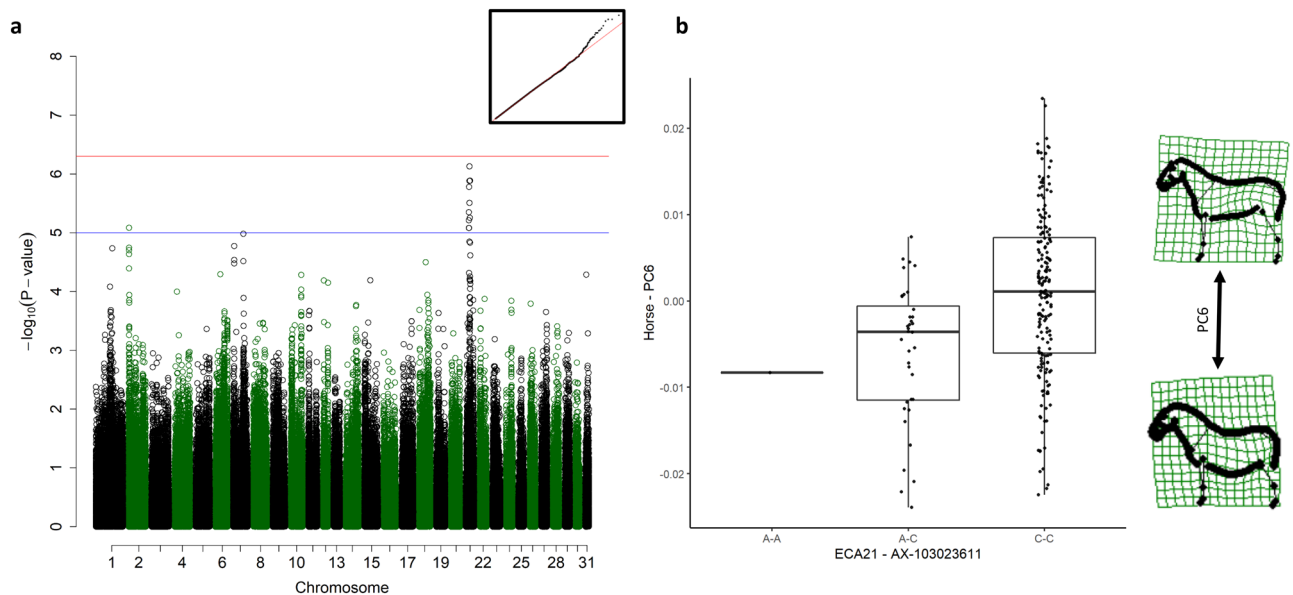
For the shapes of the body parts (torso and neck), we applied the same principle and selected the second and the fifth PC accounting for 21.38% and 3.51% of the torso and neck shape variance, respectively.

**Cresty neck.** The scores from PC3 were significantly associated with a single SNP on ECA16 at 25,245,630 ( $p = 3.25 \times 10^{-07}$ ), within the *membrane-associated guanylate kinase, WW and PDZ containing 1 (MAG1)* gene (ECA16: 24,927,284–25,524,757, Fig. 1a). Horses with the alternative A-A genotype had a much larger (cresty) neck than heterozygous A-G or homozygous G-G horses (Fig. 1b).

**Spinal curvature.** The scores from PC6 were significantly associated with a QTL on ECA21 spanning 10 SNPs from 22,096,555–30,582,615 (Fig. 2a). The best associated SNP ( $p = 7.42 \times 10^{-07}$ ) was located at 26,328,965 near the *Prostaglandin E Receptor 4 (PTGER4)* gene (ECA21: 26,143,605–26,157,288). Horses homozygous for the reference allele (C–C genotypes) tended to have a rounder croup, a hyperkyphotic back (roach back) and a tight abdomen (Fig. 2b), while A-C and A-A genotypes showed a more distended abdomen, swayback and horizontal croup.



**Figure 1.** Genome-wide association study (GWAS) for the scores from the third principal component (PC3). (a) Manhattan plot with the red line representing the significance threshold corrected for the effectively independent single nucleotide polymorphisms (SNPs) ( $p_{\text{ind}} < 4.98 \times 10^{-7}$ ). The inset on the right-hand corner shows the quantile–quantile (Q–Q) plot with the observed  $p$ -value plotted against the expected one. (b) Boxplots representing the genotype effect of the SNP on chromosome 16 on the shape of the horse. The horizontal line shows the median, the box extends from the lower to the upper quartile, and the whiskers to  $1.5 \times$  the interquartile range above the upper quartile or below the lower quartile. On the right-hand side, the extreme shapes for PC3 (shape with the highest PC3 score on top, shape with the lowest PC3 score on the bottom) are visualised in deformation grids.



**Figure 2.** Genome-wide association study (GWAS) for the scores from the sixth principal component (PC6). (a) Manhattan plot with the red line representing the significance threshold corrected for the effectively independent single nucleotide polymorphisms (SNPs) ( $p_{\text{ind}} < 4.98 \times 10^{-7}$ ). The inset on the right-hand corner shows the quantile–quantile (Q–Q) plot with the observed  $p$ -value plotted against the expected one. (b) Boxplots representing the genotype effect of the SNP on chromosome 21 on the shape. The horizontal line shows the median, the box extends from the lower to the upper quartile, and the whiskers to  $1.5 \times$  the interquartile range above the upper quartile or below the lower quartile. On the right-hand side, the extreme shapes for PC6 (shape of the horse with the highest PC6 score on top, shape of the horse with the lowest PC6 score on the bottom) are visualised in deformation grids.

**Type.** The scores from PC11 were significantly associated with a large QTL on ECA5 spanning 22 SNPs from 1,552,569–4,094,935 (Fig. 3a). The best-associated SNP ( $p = 4.15 \times 10^{-07}$ ) was located at 3,879,741 in the *POU Class 2 Homeobox 1 (POU2F1)*, also known as *OCT1* or *OTF1* gene (ECA5: 3,876,009–4,050,835). The horses with an A-A genotype (reference allele) were of a lighter type, with less neck and hindquarter musculature and a more horizontal croup, while G-G horses were heavier built, with a sloping, muscular croup (Fig. 3b). The heterozygous horses were of medium size. For the scores from PC7, which was also associated with type, no significant nor suggestive association was determined.

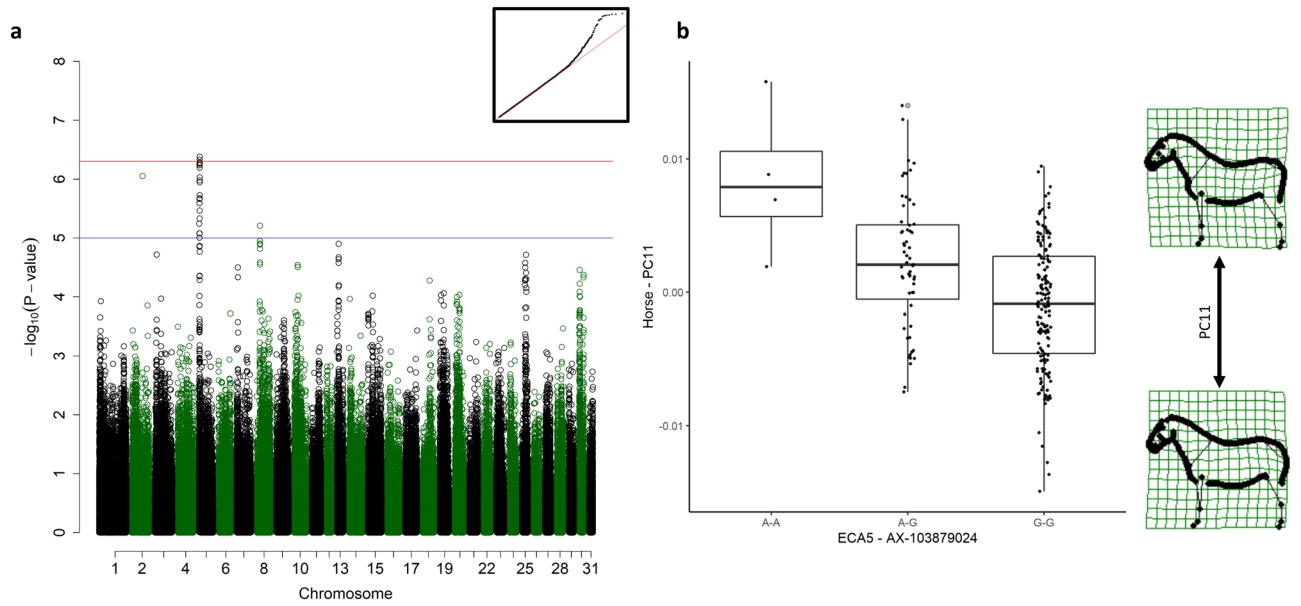
**Withers' salience.** The scores from PC5 of neck landmarks were significantly associated with a single SNP on ECA22 at 20,846,113 ( $p = 2.88 \times 10^{-07}$ ), near the Transglutaminase 3 (TGM3) gene (ECA22: 20,818,610–20,858,852) (Supplementary Fig. 2a). Heterozygous T-C horses had little withers' salience and a larger underside of the neck, while homozygous C-C horses had a tendency for a swan neck and clearly defined withers (Supplementary Fig. 2b). There were no homozygous T-T horses in the dataset.

**Abdominal shape.** The scores from PC2 of torso landmarks were suggestively associated with a QTL on ECA10 including four SNPs spanning from 8,390,393 to 9,703,042, with the best-associated SNP ( $p = 2.34 \times 10^{-06}$ ) at 9,664,363, embedded in the *Ryanodine receptor 1 (RYR1)* gene (ECA10: 20,818,610–20,858,852) (Fig. 4a). Homozygous T-T horses had a sagging abdomen and a tendency towards a sway back (Fig. 4b). G-T and G-G horses had a straighter back and abdominal line.

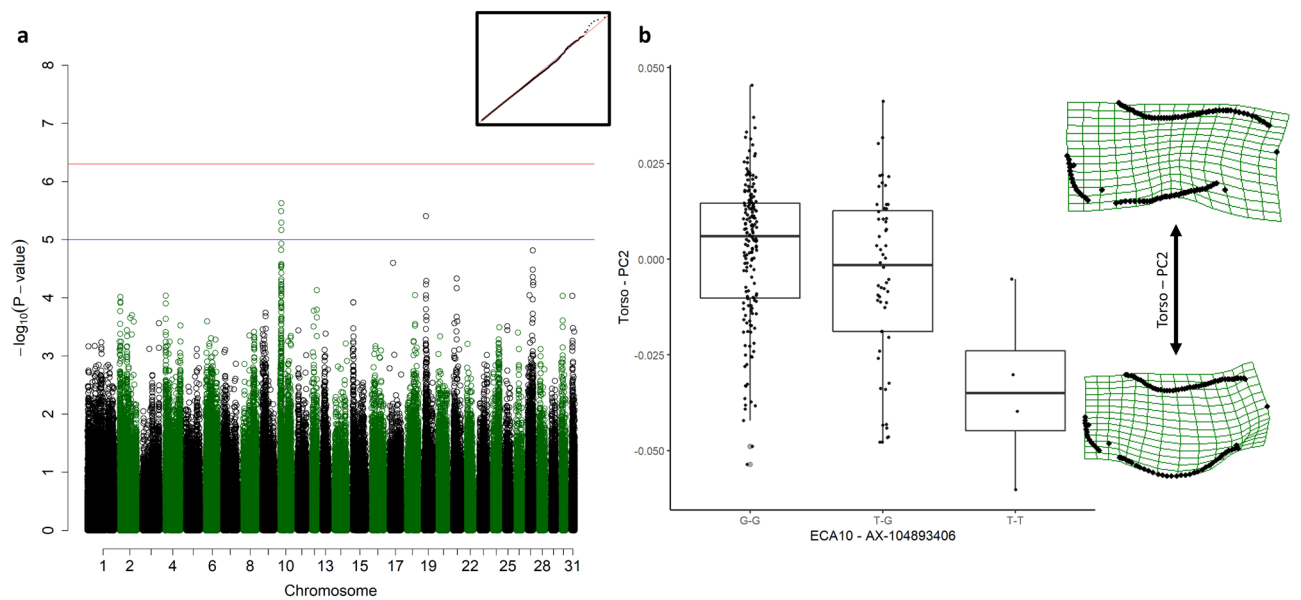
**Heritability of shape-derived conformation traits.** The genome-wide heritability ( $h^2_{\text{SNP}}$ ) of the five shape-derived conformation traits ranged from  $0.09 \pm 0.12$  (PC11) to  $0.63 \pm 0.14$  (PC3, Table 1). The standard errors were therefore fairly high in comparison, especially for traits with a lower  $h^2_{\text{SNP}}$  (PC11, PC5 of neck landmarks, PC2 of torso landmarks). The aforementioned QTL, including surrounding SNPs  $\pm 2.5$  Mb, explained 33% up to 99% of  $h^2_{\text{SNP}}$  while the standard errors were smaller.

## Discussion

**Shape variation.** The first two PCs, accounting for more than 50% of the shape variation, were consistent with previous results applying the horse shape space model on either Lipizzan<sup>6,10,20,21</sup> or FM<sup>5,11</sup> horses. The posture of the head systematically has the greatest effect on shape variation overall and cannot be entirely standardised in the field. Despite explaining most of the shape variance, PC1 (head height) and PC2 (poll flexion–extension) were not as heritable as PC3 describing the width of the neck in the FM horse<sup>11</sup>. Therefore, it is important to notice, that posture effects might account for the majority of shape variance and need to be considered in the selection of PC associated with conformation of the horse.



**Figure 3.** Genome-wide association study (GWAS) for the scores from the eleventh principal component (PC11). **(a)** Manhattan plot with the red line representing the significance threshold corrected for the effectively independent single nucleotide polymorphisms (SNPs) ( $p_{ind} < 4.98 \times 10^{-7}$ ). The inset on the right-hand corner shows the quantile–quantile (Q–Q) plot with the observed  $p$ -value plotted against the expected one. **(b)** Boxplots representing the genotype effect of the SNP on ECA5 on the shape. The horizontal line shows the median, the box extends from the lower to the upper quartile, and the whiskers to  $1.5 \times$  the interquartile range above the upper quartile or below the lower quartile. On the right-hand side, the extreme shapes for PC11 (shape of the horse with the highest PC11 score on top, shape of the horse with the lowest PC11 score on the bottom) are visualised in deformation grids.



**Figure 4.** Genome-wide association study (GWAS) for the scores from the second principal component of the torso landmarks (Torso PC2). **(a)** Manhattan plot with the red line representing the significance threshold corrected for the effectively independent single nucleotide polymorphisms (SNPs) ( $p_{ind} < 4.98 \times 10^{-7}$ ). The inset on the right-hand corner shows the quantile–quantile (Q–Q) plot with the observed  $p$ -value plotted against the expected one. **(b)** Boxplots representing the genotype effect of the SNP on ECA10 on the shape of the torso. The horizontal line shows the median, the box extends from the lower to the upper quartile, and the whiskers to  $1.5 \times$  the interquartile range above the upper quartile or below the lower quartile. On the right-hand side, the extreme shapes for PC2 of the torso (shape of the torso with the highest PC2 score on top, shape of the torso with the lowest PC2 score on the bottom) are visualised in deformation grids.

Trait	$h^2_{\text{SNP}} \pm \text{SE}$	SNP with lowest $p$ -value in GWAS	$P$ -value from GWAS	ECA	Position of redefined QTL (EquCab 3.0)	$h^2_{\text{QTL}} \pm \text{SE}$	Proportion $h^2_{\text{SNP}}$ explained by $h^2_{\text{QTL}}$ (%)
PC11	0.09 $\pm$ 0.12	AX-103879024	4.15E-07	5	1,379,741–6,379,741	0.09 $\pm$ 0.05	99
PC3	0.62 $\pm$ 0.14	AX-103061915	3.25E-07	16	22,745,630–27,745,630	0.20 $\pm$ 0.10	33
PC6	0.54 $\pm$ 0.14	AX-103023611	7.42E-07	21	23,828,965–28,828,965	0.34 $\pm$ 0.11	63
PC5 (neck)	0.19 $\pm$ 0.14	AX-104245250	2.88E-07	22	18,346,113–23,346,113	0.10 $\pm$ 0.07	54
PC2 (torso)	0.21 $\pm$ 0.14	AX-104893406	2.34E-06	10	7,164,363–12,164,363	0.12 $\pm$ 0.06	59

**Table 1.** Heritability of phenotypic traits across all SNPs ( $h^2_{\text{SNP}}$ ) and the 5-Mb QTL windows ( $h^2_{\text{QTL}}$ ), including their standard error (SE).

**Cresty neck.** Only a single SNP within the *MAG11* gene was significantly associated with the shape and width of the neck. This gene has not been described previously in equine studies, while in sheep, *MAG11* was associated with average daily gain and Kleiber ratio (average daily gain/end weight<sup>0.75</sup>) in 6- to 9-month-old sheep<sup>22</sup>. Furthermore, this gene was reported in a selection signature analysis differentiating two sheep breeds, the Sarda and Sardinian Ancestral Black, the latter having a much smaller body size<sup>23</sup>. In the horse, a cresty neck is the consequence of localised fat deposits in the dorsal neck region delineated by the nuchal ligament<sup>24,25</sup> and is highly prevalent in Baroque breeds such as Pura Raza Español (PRE, or Andalusian), Lusitano, Morgan or Paso Fino horses<sup>26</sup>. Cresty neck is an undesirable trait in the PRE, and horses with high cresty neck scores are greatly penalised during breeding shows<sup>26</sup>. Furthermore, there are some health concerns associated with cresty neck, as it has been associated with insulin dysregulation, one of the risk factors for equine metabolic syndrome, especially when measuring the neck circumference near the base of the neck<sup>27,28</sup>. Interestingly, *MAG11* variants have previously been associated with glucose response<sup>29</sup> and insulin resistance<sup>30,31</sup> in human studies. We suggest that the identified QTL in the *MAG11* gene might be associated with cresty neck in Lipizzan horses by affecting insulin regulation. Further studies including additional data from other Baroque breeds including horses with or without cresty neck, in combination with metabolic data such as blood insulin and glucose levels, are required to confirm this QTL.

**Type.** The QTL associated with type is located within the *POU2F1* gene, a transcription factor expressed in many different tissues<sup>32</sup>. In the Brazilian Nellore cattle, a QTL containing the *POU2F1* gene explained 41% of genetic variance in conformation and 29% in muscling scores<sup>33</sup> and was also identified as one of the major transcription factors regulating growth and fatty deposits in Iberian  $\times$  Landrace pigs<sup>34</sup>. In humans, certain *Pou2f1* isoforms were either heavily up- or downregulated by repulsive guidance molecule a (RGMa), important in skeletal cell differentiation, cell fusion and hypertrophy<sup>35</sup>. *POU2F1* variants were also associated with human Type 2 Diabetes Mellitus<sup>36,37</sup>, and it is common that adults affected by Type 2 Diabetes lose skeletal muscle mass<sup>38</sup>. Therefore, the identified QTL is likely associated with both muscle mass and fatty deposits.

**Spinal curvature: skeletal stiffness or muscle weakness?** The scores from PC6, representing the shape variation between a sway and a roach back, were suggestively associated with a QTL near the *PTGER4* gene. In humans, this gene has previously been associated with the severity of ankylosing spondylitis (also known as Morbus Bechterew)<sup>39,40</sup>. Main clinical signs include lower back pain, spinal stiffness and loss of mobility<sup>41,42</sup>. Hyperkyphosis (roach back) is particularly common in young men<sup>43</sup>. The hip and shoulders may also become affected<sup>44,45</sup>, with an observed shift between movements of the shoulder girdle and the pelvis<sup>46</sup>. Although we are comparing phenotypes between bipedal humans and quadrupedal horses, the shape variance explained by PC6 was still consistent with some of the aforementioned symptoms. The higher extreme showed a horse shape with a roach back and sloping croup, as opposed to the lowest extreme of a sway backed horse (hyperlordosis) with a horizontal croup, although we have no information on actual clinical signs in this dataset of horse photographs. Follow-up studies should include radiographs of genotyped horses with extreme roach back or sway back, as well as an objective movement analysis of homozygous horses from both genotype groups to determine whether the conformation would affect the gait quality of the horse, as stride length and over-tracking distance are highly dependent on the movement of the pelvis at walk, and to a lesser extent, at trot<sup>47,48</sup>. Previous genomic research investigated juvenile onset lordosis in the American Saddlebred horse identifying a large haplotype on ECA20 potentially involving the *TRERF1*, *TAF8* and *C6orf132* genes<sup>49</sup>. The causal variant could not be identified yet, which the authors attributed to the phenotyping method and the complexity of the trait<sup>50</sup>. The new QTL on ECA21 may be one of the contributing genetic factors influencing the severity of lordosis. However, this should be confirmed by multi-breed studies to better understand whether this QTL is only present in Lipizzan horses. There have been no genetic studies on roach back in the horse to date, although it is also an undesirable trait in horse breeding, similarly to sway back.

The shape of the torso was suggestively associated with a QTL containing *RYR1*. *Ryanodine receptor 1 gene* mutations have been associated with malignant hyperthermia in many mammals including pigs<sup>51</sup>, dogs<sup>52</sup> and horses<sup>53</sup>, but also with myopathies such as central core disease (CCD), Multiminicore Disease (MmD), Centronuclear myopathy (CNM) and axial hereditary myopathy<sup>54–56</sup>. Symptoms vary, but commonly include proximal muscle weakness, scoliosis and hip girdle weakness. In the case of the axial hereditary myopathy, pronounced lumbar hyperlordosis is frequent<sup>56</sup>. Joint hypermobility may lead to hip dislocations, although for MmD cases the spine is very rigid<sup>55</sup>. In the horse, the gravitational pull of the viscera will over time extend the back towards

lordosis, especially in the case of pregnant mares. This effect can be counteracted by well-developed ventral musculature, which can flex the back<sup>57</sup>. The extreme phenotype of the equine torso indicates a weakness of the abdominal muscles, which may cause the back to sag towards the ground.

For the shape of the back and abdomen, we have discovered two genes with different actions, but similar phenotypes. While *PTGER4* could influence the shape of the spine due to osseous fusion by increasing spinal rigidity, *RYR1* could weaken the abdominal musculature, causing the spine to bend as a consequence. These results underline the complex phenotype and genetic architecture of conformation traits.

**Neck shape and mane.** The scores from PC5 of the neck landmarks were significantly associated with a single SNP near the *TGM3* gene. *TGM3* is involved in skin and hair physiology. Mutations in this gene were associated with uncombable hair syndrome in humans<sup>58</sup> and wavy hair (“wellhaarig” phenotype) in mice<sup>59</sup>. *TGM3* was identified as one of the cis-regulatory elements involved in the adaptation of the Yakutian horse to the Arctic environment, probably by increasing the thickness of the coat<sup>60</sup>. It is unclear how this gene might affect the shape of the neck and withers. It is possible that if the structure of the mane hair is more rigid, the outline of the neck would be overestimated during the digitising process. *TGM3* was also associated with some forms of head and neck cancers, which would suggest a regulatory function in this anatomical region<sup>61</sup>.

**Heritability of shape-derived conformation traits.** The  $h^2_{\text{SNP}}$  of the five shape-derived morphological traits were low (<0.10) to high (>0.40) and exceeded high standard errors, most likely due to the small sample size. Interestingly, the redefined QTL accounted for 33% up to 99% of  $h^2_{\text{SNP}}$ . The scores from PC3 associated with the neck width variation had the highest  $h^2_{\text{SNP}}$ . This was consistent with findings from a study of the shape space in FM horses, where the PC describing the neck width also had a high heritability, albeit pedigree-based<sup>11</sup>.

**Current limitations and future implications.** This study has revealed several novel QTL associated with conformation in the Lipizzan horse. However, the sample size was small, and the applied SNP data did not include whole-genome sequence information. The horse shape space model is an efficient high-throughput phenotyping method to infer objective information on different aspects of equine conformation (PCs and joint angles<sup>5,11,12</sup>). Based on this method it becomes feasible to derive objective conformation traits of hundreds or thousands of horses independently from breeding judges, which simultaneously improves the data quality and quantity for genetic analyses. Furthermore, shape-derived conformation traits can be pooled across breeds and therefore support, multi-breed comparisons. However, there are currently two major limitations: the posture of the horses when analysing a single photograph, and the association of the PCs with conformation, which remains subjective. Excluding the effect of posture, by standardising the position of the head and limbs during photo taking, or by using the mean of several shapes of the same horse photographed in different postures, should result in PCs associated with differences in conformation.

In conclusion, with the application of the horse shape space model, it became feasible to identify several novel QTL associated with complex conformation traits in the Lipizzan horse. Additional studies including larger sample sizes, whole-genome sequencing and different breeds will improve our understanding of the functional genomics of the reported QTL.

## Material and methods

**Animals.** In total, 229 horses from the Lipizzan horse breed were included in the analyses (102 females and 127 males), born between 1987 and 2013; median of 2005). Samples were collected in the studs of Piber (Austria), Đakovo (Croatia), Topolčianky (Slovakia) and Szilvasvárad (Hungary) between 2014 and 2017, following national rules and regulations. All horses were genotyped on the commercial Axiom™ Equine genotyping Array containing 670,795 evenly distributed single nucleotide polymorphisms (SNP)<sup>62</sup>. After quality control, 226 horses remained for analysis.

**Phenotyping.** Each horse was phenotyped based on one unique photograph from the side. The horses were placed in open posture, as previously described<sup>5,10,11</sup>. The photographs were digitised with tpsDig2<sup>63</sup> following the outline horse model described in Druml et al.<sup>10</sup>: 18 anatomical and 15 somatometric landmarks combined with eight outlines (head upper side, neck, back, hindquarter, belly, chest and neck lower side, jaw and head lower side), for a total of 246 landmarks (Supplementary Fig. 3). After a generalised Procrustes analysis (GPA) which scales, rotates and centres every shape according to the mean configuration of the sample, the 213 semi-landmarks were slid to minimize the amount of bending energy between each configuration and the average of all specimen in an iterative process with the software tpsRelw v1.70<sup>64</sup>. The shape variation in the 246 landmarks was then summarised in principal components (PCs), where each score on a PC corresponded to one specific horse, and therefore to a specific shape pattern. The PCs explaining more than 1% of the variance for the outline horse model were considered for GWAS. From the 246 landmarks of the full horse model, we also extracted landmark configurations for the neck and torso (Supplementary Fig. 3), each summarised in PCs. The PCs were visualised using warp grids, and the main variation of each PC was described.

**Statistics and reproducibility.** The reproducibility of the horse shape space model has been extensively studied elsewhere: The digitisation process of single horse photographs was highly repeatable (same image and digitiser) and reproducible (same image, different digitisers)<sup>5,10</sup>. The consistency of the shape (same animal, different images) was influenced by age and posture<sup>5</sup>. Therefore, we corrected for age in our statistical analyses and excluded PCs affected by posture.

**Genome-wide association studies.** After removing SNPs not located on the autosomes, with a minor allele frequency below 5%, a genotyping rate below 90% or departing from Hardy–Weinberg equilibrium (HWE) at  $p < 0.0001$  using PLINK<sup>65,66</sup>, 361,411 genome-wide SNPs were available for analysis. Genome-wide association studies (GWAS) were performed using a polygenic model (*polygenic\_hgml*) in the R-package GenABEL<sup>67</sup> with the R version 3.4.1<sup>68</sup>, taking into account sex (male or female) and age by categories (three to four, five to eight, nine to sixteen and over sixteen years old) as covariates in addition to the genome-derived kinship matrix. The age categories were previously defined by Druml et al.<sup>21</sup> for the Lipizzan horse. The significance of each SNP was extracted using *mmscore*.

The results were visualised with Manhattan plots, with a suggestive significance threshold at  $p < 10^{-5}$ , and the significance threshold as the effective number of independent loci ( $p_{\text{ind}}$ ) by pruning the 361,411 SNPs for linkage disequilibrium (LD). Using a 50-kb sliding window size, a 5-kb window step size and an  $r^2$  exclusion threshold of 0.5, resulting in an LD-independent  $p$ -value threshold of  $p_{\text{ind}} < 4.98 \times 10^{-7}$ . We also evaluated the quantile–quantile (Q–Q) plots for the inflation of small  $p$ -values, hinting at false positive association signals. The effect of the SNP on the phenotype was evaluated graphically with boxplots to visualize differences in phenotypic values. We investigated which genes were located near significant SNPs using the NCBI Genome Data Viewer, based on the EquCab 3.0 reference genome assembly<sup>69</sup>. Genes functions in horses, humans and other livestock species were investigated in current published peer-reviewed literature (last research: 03.05.2022).

**Heritability estimates of shape-derived conformation traits.** For each reported QTL, we estimated the genome-wide heritability ( $h^2_{\text{SNP}}$ ) using the GCTA software<sup>70</sup> by including all available SNPs simultaneously into a genetic relationship matrix (grm), followed by a restricted maximum likelihood estimation (REML) with the grm including the aforementioned covariates age categories and sex as predictors and each conformation trait as the response variable<sup>71</sup>. To estimate the proportion of the heritability explained by the SNP with the highest-log<sub>10</sub>  $p$ -value, we defined a QTL window of  $\pm 2.5$  Mb surrounding the most significant SNP in order to establish an informative grm<sup>72</sup>. Thereafter, we estimated the heritability of the 5-Mb QTL ( $h^2_{\text{QTL}}$ ) with the covariates age categories and sex, and calculated the proportion of the heritability explained by the 5-Mb QTL compared to all genome-wide SNPs ( $h^2_{\text{QTL}}/h^2_{\text{SNP}}$ ).

**Ethical approval.** This study was discussed and approved by the institutional Commission for Ethics and Animal Welfare, University of Veterinary Medicine, Vienna, protocol number: ETK-06/05/2015, in accordance with GSP guidelines, national legislation and ARRIVE guidelines.

### Data availability

The data that support the findings of this study are available from project consortium FFG project number 843464; Veterinary University Vienna, Xenogenetik and five European state stud farms, but restrictions apply to the availability of these data. Data are available from the corresponding author upon reasonable request and with permission of project consortium, FFG project number 843464; Veterinary University Vienna, Xenogenetik and five European state stud farms.

Received: 23 February 2023; Accepted: 31 May 2023

Published online: 02 June 2023

### References

1. Szmatoła, T. et al. Assessment and distribution of runs of homozygosity in horse breeds representing different utility types. *Animals* **12**, 3293 (2022).
2. Grilz-Seger, G. et al. Genome-wide homozygosity patterns and evidence for selection in a set of European and near eastern horse breeds. *Genes* **10**, 491 (2019).
3. Grilz-Seger, G. et al. Analysis of ROH patterns in the Noriker horse breed reveals signatures of selection for coat color and body size. *Anim. Genet.* **50**, 334–346 (2019).
4. Holmström, M. & Back, W. The effects of conformation. In *Equine Locomotion* 2nd edn (eds Back, W. & Clayton, H.) (Saunders Elsevier, 2013).
5. Gmel, A. I., Druml, T., Portele, K., von Niederhäusern, R. & Neuditschko, M. Repeatability, reproducibility and consistency of horse shape data and its association with linearly described conformation traits in Franches–Montagnes stallions. *PLoS ONE* **13**, e0202931 (2018).
6. Druml, T., Dobretsberger, M. & Brem, G. Ratings of equine conformation—new insights provided by shape analysis using the example of Lipizzan stallions. *Arch. Anim. Breed.* **59**, 309–317 (2016).
7. Schroderus, E. & Ojala, M. Estimates of genetic parameters for conformation measures and scores in Finnhorse and Standardbred foals. *J. Anim. Breed. Genet.* **127**, 395–403 (2010).
8. Samoré, A., Pagnacco, G. & Miglior, F. Genetic parameters and breeding values for linear type traits in the Haflinger horse. *Livest. Prod. Sci.* **52**, 105–111 (1997).
9. Rosengren, M. K. et al. A QTL for conformation of back and croup influences lateral gait quality in Icelandic horses. *BMC Genom.* **22**, 1–13 (2021).
10. Druml, T., Dobretsberger, M. & Brem, G. The use of novel phenotyping methods for validation of equine conformation scoring results. *Animal* **9**, 928–937 (2015).
11. Gmel, A. I., Burren, A. & Neuditschko, M. Estimates of genetic parameters for shape space data in Franches–Montagnes horses. *Animals* **12**, 2186 (2022).
12. Gmel, A. I., Druml, T., von Niederhäusern, R., Leeb, T. & Neuditschko, M. Genome-wide association studies based on equine joint angle measurements reveal new QTL affecting the conformation of horses. *Genes* **10**, 370 (2019).
13. Kess, T. & Boulding, E. G. Genome-wide association analyses reveal polygenic genomic architecture underlying divergent shell morphology in Spanish *Littorina saxatilis* ecotypes. *Ecol. Evol.* **9**, 9427–9441 (2019).
14. Sakamoto, L. et al. Comparison of shape quantification methods for genomic prediction, and genome-wide association study of sorghum seed morphology. *PLoS ONE* **14**, e0224695 (2019).

15. Pitchers, W. *et al.* A multivariate genome-wide association study of wing shape in *Drosophila melanogaster*. *Genetics* **211**, 1429–1447 (2019).
16. Claes, P. & Shriver, M. D. New entries in the lottery of facial GWAS discovery. *PLoS Genet.* **12**, e1006250 (2016).
17. Gordon, C. R., Marchant, T. W., Lodzinska, J., Schoenebeck, J. J. & Schwarz, T. Morphological variation of the caudal fossa of domestic cat skulls assessed with CT and geometric morphometrics analysis. *J. Feline Med. Surg.* **20**, 752–758 (2018).
18. Marchant, T. W. *et al.* Canine brachycephaly is associated with a retrotransposon-mediated missplicing of SMO2. *Curr. Biol.* **27**, 1573–1584 (2017).
19. Zechner, P. *et al.* Analysis of diversity and population structure in the Lipizzan horse breed based on pedigree information. *Livest. Prod. Sci.* **77**, 137–146 (2002).
20. Druml, T., Horna, M., Grilz-Seger, G., Dobretsberger, M. & Brem, G. Association of body shape with amount of Arabian genetic contribution in the Lipizzan horse. *Arch. Anim. Breed.* **61**, 79–85 (2018).
21. Druml, T., Dobretsberger, M. & Brem, G. The interplay of performing level and conformation—a characterization study of the Lipizzan riding stallions from the Spanish riding school in Vienna. *J. Equine Vet. Sci.* **60**, 74–82 (2018).
22. Pasandideh, M., Rahimi-Mianji, G. & Gholizadeh, M. A genome scan for quantitative trait loci affecting average daily gain and Kleiber ratio in Baluchi Sheep. *J. Genet.* **97**, 493–503 (2018).
23. Cesarani, A. *et al.* Investigation of genetic diversity and selection signatures between Sarda and Sardinian Ancestral black, two related sheep breeds with evident morphological differences. *Small Rumin. Res.* **177**, 68–75 (2019).
24. Geor, R. J. Metabolic predispositions to laminitis in horses and ponies: Obesity, insulin resistance and metabolic syndromes. *J. Equine Vet.* **28**, 753–759 (2008).
25. Carter, R. A., Treiber, K. H., Geor, R. J., Douglass, L. & Harris, P. A. Prediction of incipient pasture-associated laminitis from hyperinsulinaemia, hyperleptinaemia and generalised and localised obesity in a cohort of ponies. *Equine Vet. J.* **41**, 171–178 (2009).
26. Sánchez, M. *et al.* Prevalence, risk factors and genetic parameters of cresty neck in Pura Raza Español horses. *Equine Vet. J.* **49**, 196–200 (2017).
27. Martín-Gimenez, T., Aguirre-Pascasio, C. & De Blas, I. Beyond scoring systems: Usefulness of morphometry considering demographic variables, to evaluate neck and overall obesity in Andalusian horses. *Animal* **12**, 597–605 (2018).
28. Martín-Gimenez, T., De Blas, I., Aguilera-Tejero, E., de Castro, E. D. & Aguirre-Pascasio, C. Endocrine, morphometric, and ultrasonographic characterization of neck adiposity in Andalusian horses. *Domest. Anim. Endocrinol.* **56**, 57–62 (2016).
29. Ellis, K. *et al.* Genetic modifiers of response to glucose–insulin–potassium (GIK) infusion in acute coronary syndromes and associations with clinical outcomes in the IMMEDIATE trial. *Pharmacogenomics J.* **15**, 488–495 (2015).
30. Palmer, N. *et al.* Candidate loci for insulin sensitivity and disposition index from a genome-wide association analysis of Hispanic participants in the Insulin Resistance Atherosclerosis (IRAS) Family Study. *Diabetologia* **53**, 281–289 (2010).
31. Norris, J. M. & Rich, S. S. Genetics of glucose homeostasis: Implications for insulin resistance and metabolic syndrome. *Arterioscler. Thromb. Vasc. Biol.* **32**, 2091–2096 (2012).
32. Dey, A., Sen, S., Uversky, V. N. & Maulik, U. Structural facets of POU2F1 in light of the functional annotations and sequence-structure patterns. *J. Biomol. Struct. Dyn.* **39**, 1093–1105 (2021).
33. Carreño, L. O. D. *et al.* Genome association study for visual scores in Nellore cattle measured at weaning. *BMC Genom.* **20**, 1–9 (2019).
34. Pérez-Montarelo, D. *et al.* Identification of genes regulating growth and fatness traits in pig through hypothalamic transcriptome analysis. *Physiol. Genomics* **46**, 195–206 (2014).
35. Copola, A. G. L. *et al.* Transcriptomic characterization of the molecular mechanisms induced by RGMa during skeletal muscle nuclei accretion and hypertrophy. *BMC Genomics* **23**, 1–16 (2022).
36. Ng, M. *et al.* Association of the POU class 2 homeobox 1 gene (POU2F1) with susceptibility to Type 2 diabetes in Chinese populations. *Diabet. Med.* **27**, 1443–1449 (2010).
37. Batool, A., Jahan, N., Sun, Y., Hanif, A. & Xue, H. Genetic association of IDE, POU2F1, PON1, IL1 $\alpha$  and IL1 $\beta$  with type 2 diabetes in Pakistani population. *Mol. Biol. Rep.* **41**, 3063–3069 (2014).
38. Park, S. W. *et al.* Excessive loss of skeletal muscle mass in older adults with type 2 diabetes. *Diabetes Care* **32**, 1993–1997 (2009).
39. Cortes, A. *et al.* Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat. Genet.* **45**, 730–738 (2013).
40. Chai, W. *et al.* JARID1A, JMY, and PTGER4 polymorphisms are related to ankylosing spondylitis in Chinese Han patients: A case-control study. *PLoS ONE* **8**, e74794 (2013).
41. Braun, J. & Sieper, J. Ankylosing spondylitis. *The Lancet* **369**, 1379–1390 (2007).
42. Dakwar, E., Reddy, J., Vale, F. L. & Uribe, J. S. A review of the pathogenesis of ankylosing spondylitis. *Neurosurg. Focus* **24**, E2 (2008).
43. Vosse, D. *et al.* Determinants of hyperkyphosis in patients with ankylosing spondylitis. *Ann. Rheum. Dis.* **65**, 770–774 (2006).
44. Vander Cruyssen, B., Vastesaegeer, N. & Collantes-Estévez, E. Hip disease in ankylosing spondylitis. *Curr. Opin. Rheumatol.* **25**, 448–454 (2013).
45. Will, R. *et al.* Ankylosing spondylitis and the shoulder: Commonly involved but infrequently disabling. *J. Rheumatol.* **27**, 177–182 (2000).
46. Mangone, M. *et al.* Pelvis–shoulder coordination during level walking in patients with ankylosing spondylitis. *Gait Posture* **34**, 1–5 (2011).
47. Gmel, A. I. *et al.* Determining objective parameters to assess gait quality in Franches–Montagnes horses for ground coverage and over-tracking—Part 1: At walk. *J. Equine Vet.* **115**, 104024 (2022).
48. Gmel, A. I. *et al.* Determining objective parameters to assess gait quality in franchises-montagnes horses for ground coverage and over-tracking—part 2: At trot. *J. Equine Vet.* **120**, 104166 (2023).
49. Cook, D., Gallagher, P. & Bailey, E. Genetics of swayback in American Saddlebred horses. *Anim. Genet.* **41**, 64–71 (2010).
50. YousefiMashouf, N., Graves, K., Kalbfleisch, T., Bailey, E. Genetic basis for juvenile onset lordosis in saddlebred horses. In *13th International Havemeyer Foundation Horse Genome Workshop (July 24–28, 2022)* (2022).
51. Fujii, J. *et al.* Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* **253**, 448–451 (1991).
52. Roberts, M. C. *et al.* Autosomal dominant canine malignant hyperthermia is caused by a mutation in the gene encoding the skeletal muscle calcium release channel (RYR1). *J. Am. Soc. Anesthesiol.* **95**, 716–725 (2001).
53. Aleman, M., Nieto, J. & Magdesian, K. Malignant hyperthermia associated with ryanodine receptor 1 (C7360G) mutation in Quarter Horses. *J. Vet. Intern. Med.* **23**, 329–334 (2009).
54. Jungbluth, H., Sewry, C. A. & Muntoni, F. What's new in neuromuscular disorders? The congenital myopathies. *Eur. J. Paediatr. Neurol.* **7**, 23–30 (2003).
55. Zhou, H. *et al.* Molecular mechanisms and phenotypic variation in RYR1-related congenital myopathies. *Brain* **130**, 2024–2036 (2007).
56. Witting, N., Andersen, L. K. & Vissing, J. Axial myopathy: An overlooked feature of muscle diseases. *Brain* **139**, 13–22 (2016).
57. van Weeren, R. Kinematics. In *Equine neck and back pathology: Diagnosis and treatment* 2nd edn (ed. Henson, F. M. D.) (Wiley, 2018).



58. Basmanav, F. B. Ü. *et al.* Mutations in three genes encoding proteins involved in hair shaft formation cause uncombable hair syndrome. *Am. J. Hum. Genet.* **99**, 1292–1304 (2016).
59. Brennan, B. M. *et al.* The mouse wellhaartig (we) mutations result from defects in epidermal-type transglutaminase 3 (Tgm3). *Mol. Genet. Metab.* **116**, 187–191 (2015).
60. Librado, P. *et al.* Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proc. Natl. Acad. Sci.* **112**, E6889–E6897 (2015).
61. Wu, X. *et al.* TGM3, a candidate tumor suppressor gene, contributes to human head and neck cancer. *Mol. Cancer* **12**, 1–13 (2013).
62. Schaefer, R. J. *et al.* Developing a 670k genotyping array to tag~ 2M SNPs across 24 horse breeds. *BMC Genomics* **18**, 1–18 (2017).
63. RohlF. tpsDig2 version 1.78.) (2001).
64. RohlF. tpsRelw. Version 1.70.) (2003).
65. Purcell S. PLINK Version 1.07.) (2017).
66. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
67. Aulchenko, Y. S., Ripke, S., Isaacs, A. & Van Duijn, C. M. GenABEL: An R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007).
68. R Core Team. R: A language and environment for statistical computing. (2013).
69. Kalbfleisch, T. S. *et al.* Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun. Biol.* **1**, 197 (2018).
70. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
71. Norton, E. M. *et al.* Heritability of metabolic traits associated with equine metabolic syndrome in Welsh ponies and Morgan horses. *Equine Vet. J.* **51**, 475–480 (2019).
72. Signer-Hasler, H. *et al.* A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PLoS ONE* **7**, e37282 (2012).

## Acknowledgements

We are grateful to Dr. Thomas Druml for providing the photographs and shape data for this study. For assistance and cooperation regarding the shape data and for providing the hair samples, we thank the stud farms of Piber (Austria), Đakovo (Croatia), Topolčianky (Slovakia) and Szilvasvárad (Hungary) and the Spanish Riding School in Vienna (Austria).

## Author contributions

G.B. and M.N. designed the study. G.B. was responsible for funding acquisition. A.I.G. and M.N. performed the analyses and the visualisations. A.I.G. and M.N. wrote the initial manuscript. All authors read and approved the final version of the manuscript.

## Funding

This work was financially supported by the Austrian Research Promotion Agency (FFG) under contract number 843464 and by Xenogenetik.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-36272-4>.

**Correspondence** and requests for materials should be addressed to M.N.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023