Accepted Manuscript

A Genome-wide association study for morphometric traits in Quarter Horse

Camila Tangari Meira , Michel Marques Farah , Marina R.S. Fortes , Stephen S. Moore , Guilherme L. Pereira , Josineudson Augusto II. V. Silva , Marcílio Dias S. da Mota , Rogério Abdallah Curi

PII: S0737-0806(14)00204-4

DOI: 10.1016/j.jevs.2014.05.011

Reference: YJEVS 1732

To appear in: *Journal of Equine Veterinary Science*

Received Date: 11 March 2014

Revised Date: 13 May 2014

Accepted Date: 21 May 2014

Please cite this article as: Meira CT, Farah MM, Fortes MRS, Moore SS, Pereira GL, Silva JAIV, da Mota MDS, Curi RA, A Genome-wide association study for morphometric traits in Quarter Horse, *Journal of Equine Veterinary Science* (2014), doi: 10.1016/j.jevs.2014.05.011.

A Genome-wide association study for morphometric traits in Quarter Horse

Camila Tangari Meira[a], Michel Marques Farah[a], Marina R.S Fortes[b], Stephen S. Moore[b], Guilherme L. Pereira[a], Josineudson Augusto II V. Silva[c], Marcílio Dias S. da Mota[c], Rogério Abdallah Curi[c]

[a] Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, SP 14884-900, Brazil.

[b] Queensland Alliance for Agriculture and Food Innovation, Centre for Animal Science, The University of Queensland, Brisbane, QLD, Australia.

[c] Faculdade de Medicina Veterinária e Zootecnia, Universidade Estadual Paulista, Botucatu, SP 18618-000, Brazil.

Corresponding author:

Camila Tangari Meira

Postal address: Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, SP 14884-900, Brazil.

E-mail: camilatangari@zootecnista.com.br

Phone number: 55 14 38117187

Fax number: 55 14 38117187

Abstract

A genome-wide association study for morphometric traits was conducted in 184 Quarter Horses, 120 from a racing population and 64 from a cutting population, which were genotyped using the Illumina EquineSNP50 chip. Association analysis was performed with 42,058 SNPs (after quality control) using Qxpak5 software. The following traits were measured: weight, rump and body length. These morphometric traits are important for the best performance in race and cutting events. For weight, 3 SNPs associated ($P < 0.0001$) were found on chromosomes (ECA) 2 and 3. For rump were found 8 associated ($P < 0.0001$) on ECA 2, 3, 6, 7, 9, 21, and 26. On ECA 3 and ECA 8, 2 SNPs were associated ($P < 0.0001$) with body length. So, a total of 13 important chromosomal regions were identified with Q-values of 0.53 (SNPs for W), 0.40 (SNPs for RL) and 0.99 (SNPs for BL). Positional and functional candidate genes emerging from this study were *WWOX* and *AAVPR1A*. Further studies are required to confirm these associations in other populations.

Keywords: *Equus caballus*, genetics, quantitative traits, single-nucleotide polymorphism

1. Introduction

Breed of great economic expression, especially due to its versatility in different equestrian events, the Quarter Horse is subdivided into different populations according to skills resulting from distinct selection objectives. The racing population explores the sprinting ability of the animals over shorts distances, whereas the cutting population is used in functional tests, exploring abilities such as agility, obedience and cow sense.

The morphology of horses presents functional relationship with the performance of the animals in their various activities, such as racing, reining, barrel racing, cutting, etc. Many times the differences in performance exist owing to adequacy or inadequacy morphometric traits. According to Meira et al. [1], racing horses are heavier and taller and present greater body lengths and perimeters than cutting horses.

The availability of an equine SNP array gave rise to genome wide association studies (GWAS). Recent studies applying genomic approaches reported to conformation traits have been conducted. Signer-Hasler et al. [2], Tetens et al. [3] and Meira et al. [4] reported the identification of QTL for height at withers on *Equus caballus* autosomes (ECA) 3 and ECA 9 in Franches-Montagnes horses, on ECA 3 in German Warmblood horses and on ECA 9 in Quarter Horse, respectively. Studies with body size also had been conducted. Makvandi-Nejad et al. [5] identify four loci on ECA 3, 6, 9 and 11 explain the vast majority of variation in horse size. Metzger et al. [6] investigated associations of polymorphisms of the candidate gene *LCORL* with the development of body size in horses. In this study we aimed to perform GWAS to identify chromosomal regions and positional candidate genes associated with other important morphometric traits in Quarter Horse.

2. Methods and Materials

For the analyses, a total of 188 Quarter Horses of both sexes born between 1985 and 2009, including 120 racing horses and 68 cutting horses, registered at the Brazilian Association of Quarter Horse Breeders (ABQM), were studied. All experimental procedures were conducted in accordance with the Brazilian legislation on animal welfare.

The following physical traits were measured according to Torres and Jardim [7]: weight (W), rump length (RL) and body length (BL). The measurements were performed by the same person with a tape measure and measuring stick, always on the right side of the animal, with the horse standing with front and rear legs perpendicular to the ground. For genotyping, a 5-mL sample of whole blood was collected from each animal by puncture of the left jugular vein in the neck region into vacuum tubes containing 7.5 mg ethylenediaminetetraacetic acid.

The racing horses (18 males and 102 females), born to 48 stallions and 107 mares, belonged to five farms in the countryside of the State of São Paulo, Brazil. The cutting horses (26 males and 42 females), born to 44 stallions and 64 mares, belonged to three other farms in the countryside of the State of São Paulo. In both populations, full sibs were avoided.

Genomic DNA was extracted from the blood samples of Quarter Horses using the Illustra Blood Genomicprep Mini Spin kit (GE Healthcare, Little Chalfont, Buckinghamshire, UK) according to manufacturer instructions. DNA integrity was analyzed by 0.8% agarose gel electrophoresis and DNA was quantified with a NanoDrop ND-2000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The DNA concentration in the samples was adjusted to 40-60 ng/μL.

SNPs were genotyped on the HiScan system (Illumina Inc.) using the Illumina Equine SNP50 BeadChip at the Faculty of Agricultural and Veterinary Sciences, UNESP, Jaboticabal, São Paulo, Brazil. The chip contains 54,602 SNPs uniformly distributed across the entire genome of 15 horse breeds. The SNPs are distributed across the 31 autosomes and X chromosome. The mean interval between SNPs is 43,200 bp. This content is derived from the EquCab2.0 SNP Collection compiled by the Broad Institute's Equine Genome Sequencing Project, which identified more than 940,000 SNPs in Arabian, Andalusian, Akhal-Teke, Icelandic Pony, Standardbred, Thoroughbred, and Quarter Horses.

The quality of individual and SNP genotype data was investigated using the Genome Studio program, version 2011.1 (Illumina Inc.). For individuals, call rate, heterozygosity, and gender estimation were determined. Animals with a call rate < 0.95, heterozygosity of ± 3

standard deviations from the mean, and errors in gender estimation were excluded from the sample. In addition, agreement between four replicates and parentage concordance (allele sharing) between four stallion/progeny and three stallion/mare/progeny pairs were evaluated. With respect to the quality of SNP genotypes in the whole population, SNPs located on the X chromosome, with low genotyping quality (cluster separation < 0.3), a call frequency < 0.9, a minor allele frequency (MAF) < 0.05 (including fixed alleles), and a Hardy-Weinberg *P*< 0.001were excluded. Quality control led to the exclusion of four cutting animals from the sample and 12,544 SNPs. Thus 42,058 autosomal SNPs for 184 animals were used in the analysis.

Genome wide association studies (GWAS) were performed with the 42,058 SNPs for each of the three traits (W, RL and BL) separately using Qxpak5 and fitting one SNP at a time [8]. Qxpak5 relies on the well-known theory of mixed models. It performs a likelihood ratio test with every SNP in turn, testing the model with the SNP *versus* the model without the SNP, against a chi-square distribution with 1 degree of freedom.

Heritability estimates and the effect of each SNP were estimated using the mixed model (Equation 1), which included sex (2 levels), population (2 levels) and stud (8 levels) as fixed effects, age at the time of trait measurement as a linear covariate and animal and SNP as a random effect.

$$y_{ij} = X\beta + Zu + S_k s_{jk} + e_{ij} \quad \text{(Eq.1)},$$

where y_{ij} represents the vector of phenotypic observations from the *i*-th horse (*i* = 1 to 184) at the *j*-th trait (*j* = W, RL and BL, respectively); X is the incidence matrix relating fixed effects in *β* with observations in y_{ij}; Z is the incidence matrix relating random additive polygenic effects in *u* with observations in y_{ij}; S_k is the vector of genotypes for the *k*-th SNP across all animals; s_{jk} represents the additive effect of the *k*-th SNP on the *j*-th trait, and e_{ij} is the vector of random residual effects.

The genetic relationship between animals was estimated from genotypes, used to build a genomic relationship matrix (G), using the same methodology describe by VanRanden [9] (Equation 2):

$$G = \frac{(M-P)(M-P)'}{2\Sigma_{j=1}^{m} p_j(1-p_j)} \quad \text{(Eq. 2)},$$

where M is an allele-sharing matrix with m columns (m=42,058 SNPs) and n rows (n=184 individuals), and P is a matrix containing the observed frequency of the second allele (p_j), expressed as $2p_j$. M_{ij} was 0 if the genotype of individual i for SNP j was homozygous for the first allele, was 1 if heterozygous, or 2 if the genotype was the other homozygous state.

To account for multiple tests, Q-value was calculated with the package for R (Version 2.10) [10]. The percentage of the genetic variance accounted by the each SNP was estimated according to Equation 3:

$$\%V_i = 100\left(\frac{2p_i q_i \hat{\alpha}_i^2}{\sigma_g^2}\right) \text{ (Eq. 3)},$$

where p_i and q_i are the allele frequencies for the i-th SNP estimated across the entire population, α_i is the estimated additive effect of the i-th SNP on the trait in question, and σ_g^2 is the REML estimate of the (poly)genetic variance for the trait.

According to McCue et al. [11] the linkage disequilibrium (LD) of the Quarter Horse decline within the first 50 - 100 Kb. Therefore, the window considered for gene identification was 100Kb upstream and downstream to each significant SNP (*P*-value < 0.0001 for any of the traits). Genes with biologically interesting findings were highlighted. Positions of SNPs and genes were according to NCBI Reference Assembly [12], based on the latest assembly of the horse genome sequence (EquCab2.0).

Descriptive statistics of the traits are reported in Table 1. Conformation traits are reported in the literature to be of moderate to high heritability [13]. In this study, the heritabilities estimated ranged from 0.17 to 0.51, indicating additive genetic variability and the potential for improvement through genetic selection of these morphometric traits in Quarter Horse.

We performed a GWAS for conformation traits (W, RL, BL) using a mixed-model, fitting genomic relationship data, fixed effects and SNP alleles. The genomic relationship matrix was used to obtain more accurate relationship between animals, using identical by state (IBS) information. Fitting a G matrix allowed to estimate relationships between animals of the 2 different populations that were unrelated according to pedigree records.

4. Results and Discussion

A total of thirteen significant SNP associations ($P < 0.0001$) were found for the traits, being 3 SNPs on ECA 2 and 3 associated with weight, 8 SNPs on ECA 2, 3, 6, 7, 9, 21 and 26 associated with rump length and 2 SNPs on ECA 3 and 8 associated with body length (Figure 1). This *P*-value corresponds to Q-values of 0.53 (SNPs for W), 0.40 (SNPs for RL) and 0.99 (SNPs for BL). These association values are indicative of a possible true association, given the small sample size and the fact that each of these SNP accounted for an important proportion of the genetic variance, which ranged from 17.8 to 54.4% for all the traits (Table 2). Among significant results, the SNP rs68488737 (ECA3) that was associated with W and RL and the SNP rs68714893 (ECA6) associated with RL point to interesting candidate genes: *WWOX* and *AVPR1A*. These genes are biologically interesting findings, because they can be functionally related to important traits in the Quarter Horses as detailed below. The SNP rs68488737 (ECA3) is positioned in non-coding region of the *WWOX*, however it can affect the gene expression or be in LD with causal SNPs.

In our study, two SNP associations at ECA 3 were observed implicating that the gene *WWOX* (WW domain containing oxidoreductase) was associated with W and RL. This gene, according to Aqeilan et al. [14] contributes to bone formation through regulation of *RUNX2* activity in osteoblasts, which encodes a nuclear protein essential for osteoblastic differentiation and skeletal morphogenesis [12]. In a previous study, a different region of ECA 3 was implicated. Tetens et al. [3] found a single major QTL for height at withers explaining ~18% of the phenotypic variance mapped to ECA3 in German Warmblood horses, indicating

the *LCORL*/*NCAPG* locus as a strong candidate underlying this QTL. Differences in studied breeds and traits may account for these different results. A fine mapping exercise of these regions of ECA3 would contribute to resolving this QTL.

The association with RL on ECA6 implicated a new candidate gene: *AVPR1A* (arginine vasopressin receptor 1A). The protein encoded by *AVPR1A* acts as receptor for arginine vasopressin (AVP). Vasopressin is a powerful vasoconstrictor and an important component in the control of blood pressure during exercise in horses [15]. According to McKeever et al. [16], exercise causes an increase in plasma concentration AVP that is correlated with the duration and intensity of the exercise in horses and man. What would be the direct effect of *AVPR1A* in RL is difficult to speculate. From the above described gene function, it is logical to assume that *AVPR1A* could affect exercise performance. Further investigation of the association reported here could provide insight for the molecular links between RL and exercise performance.

In conclusion, genomic regions on ECA 2, 3, 6, 7, 8, 9, 21 and 26 were associated with morphometric traits in Quarter Horse. Genes mapped to these regions and thus positional candidates for weight, rump and body length emerged from this GWAS. However, most of these genes have no obvious function related to the studied traits. Positional and functional candidate genes from this study are *WWOX* and *AVPR1A*. Further studies are required to confirm these SNP associations and candidate genes in other populations and horse breeds.

Acknowledgements

The authors thank the horse farms for providing the Quarter Horses and the financial support of Fapesp and Capes. They also thank the invaluable contribution of the late Professor Dr. Marcílio Dias Silveira da Mota.

References

[1] Meira CT, Curi RA, Silva JAV, Corrêa MJM, Oliveira HN, Mota MDS. Morphological and genomic differences between cutting and racing lines of Quarter Horses. J Equine Vet Sci 2013;33:244-49.

[2] Signer-Hasler H, Flury C, Haase B, Burger D, Simianer H, Leeb T, et al. A genome-wide association study reveals loci influencing height and other conformation traits in horses. PLoS One 2012;7(5):e37282.

[3] Tetens J, Widmann P, Kühn C, Thaller G. A genome-wide association study indicates *LCORL*/*NCAPG* as a candidate locus for withers height in German Warmblood horses. Anim Genet 2013;44:467–71.

[4] Meira CT, Fortes MRS, Farah MM, Porto-Neto LR, Curi RA, Moore SS, et al. A genome-wide association study for height at withers in racing quarter horse. Proc Assoc Advmt Anim Breed Genet 2013;20:420-23.

[5] Makvandi-Nejad S, Hoffman GE, Allen JJ, Chu E, Gu E, Chandler AM, et al. Four loci explain 83% of size variation in the horse. PLoS One 2012;7(7):e39929.

[6] Metzger J, Schrimpf R, Philipp U, Distl O. Expression levels of *LCORL* are associated with body size in horses. PLoS One 2013;8(2):e56497.

[7] Torres AP, Jardim WR. Criação do cavalo e de outros equinos. 3rd ed. São Paulo: Nobel; 1992.

[8] Perez-Enciso M, Misztal I. Qxpak.5: Old mixed model solutions for new genomics problems. BMC Bioinformatics 2011;12:202.

[9] VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci 2008;91:4414-23.

[10] Storey JD, Tibshirani R. Statistical significance for genome wide studies. Proc Natl Acad Sci USA 2003;100:9440-5.

[11] McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM, et al. A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. PLoS Genet 2012;8(1):e1002451.

[12] NCBI. National Center for Biotechnology Information. Available at: http://www.ncbi.nlm.nih.gov. Accessed February 7, 2014.

[13] Zamborlini LC, Bergmann JAG, Pereira CS, Fonseca AG, Carneiro ASR. Estudo genético quantitativo das medidas lineares da raça Mangalarga Marchador –I. Estimativas dos fatores de ambiente e parâmetros genéticos. Rev Bras Ci Vet 1996;3:33-7.

[14] Aqeilan RI, Hassan MQ, Bruin A, Hagan JP, Volinia S, Palumbo T, et al. The WWOX tumor suppressor is essential for postnatal survival and normal bone metabolism. J Biol Chem 2008;283:21629-39.

[15] McKeever KH, Hinchcliff KW. Neuroendocrine control of blood volume, blood pressure and cardiovascular function in horses. Equine Vet J 1995;18:77-81.

[16] McKeever KH, Hinchcliff KW, Schmall LM, Lamb DR, Muir WW. Changes in plasma renin activity, aldosterone, and vasopressin, during incremental exercise in horses. Am J Vet Res1992;53:1290-3.

Table 1. Summary statistics and heritability (h^2) estimates for weight (kilogram), rump length (centimeter) and body length (meter) of the Quarter Horse.

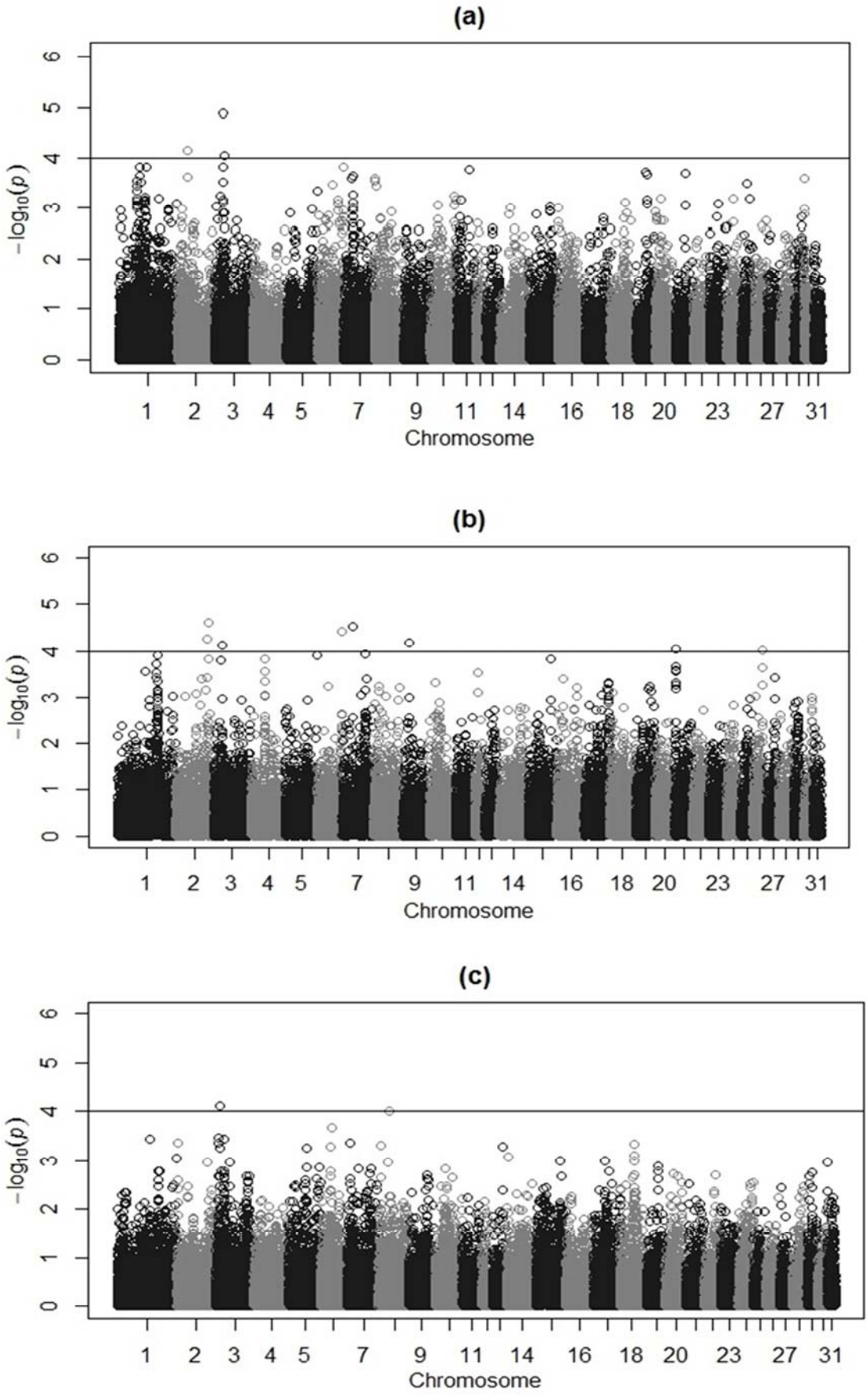
Parameter	Weight (Kg)	Rump length (cm)	Body length (m)
N (animals)	184	184	184
Mean (± std) (racing animals)	538.97 (± 42.05)	62.03 (±0.03)	1.80 (± 0.06)
Mean (± std) (cutting animals)	450.69 (± 46.59)	55.17 (±0.02)	1.63 (± 0.07)
σ^2_g	823.2840	0.0004	0.0006
h^2	0.4975	0.5164	0.1710

std = standard deviation; σ^2_g = REML estimate of the (poly)genetic variance for the trait

Table 2. Minor allele frequency (MAF), effect, proportion of the additive genetic variance explained (%V) and genes for SNPs associated ($P < 0.0001$) with weight, rump and body length of the Quarter Horse.

SNP	ECA: Position (Mb)	MAF	Weight		Rump length		Body length		Genes
			Effect	%V	Effect	%V	Effect	%V	
rs68626382	2:33.5	0.17	24.30	19.9	-	-	-	-	*EIF4G3, HP1BP3, SH2D5, KIF17, LOC102148467, DDOST, PINK1, LOC102148800, CDA,LOC102148657*
rs68488737	3:26.7	0.36	-19.60	21.7	-	-	-	-	*WWOX*
rs68559987	3:29.1	0.42	-17.70	18.6	-	-	-	-	*BCMO1, GAN, LOC102150591, CMIP*
rs68614112	2:103.7	0.06	-	-	-0.03	22.9	-	-	*ANKRD50*
rs68562545	2:98.9	0.10	-	-	0.02	21.6	-	-	-
rs68488737	3:26.7	0.36	-	-	-0.01	19.7	-	-	*WWOX*
rs68714893	6:79.4	0.38	-	-	0.01	19.9	-	-	*AVPR1A*
rs68762127	7:34.1	0.09	-	-	-0.02	24.1	-	-	*LOC100072341, LOC100630076, LOC100630107, LOC102148118*
rs68802762	9:18.5	0.24	-	-	0.01	17.8	-	-	*MCMDC2, SGK3, LOC102147850, C9H8orf44, VCPIP1*
rs69315059	21:4.1	0.26	-	-	-0.01	19.1	-	-	*LOC102148385, LOC100054881, LOC102148426*
rs69402054	26:29.9	0.42	-	-	-0.01	20.5	-	-	*IFNAR2, IL10RB, LOC102150829, IFNAR1*
rs68625621	3:15.0	0.40	-	-	-	-	0.02	54.4	-
rs68795337	8:32.8	0.36	-	-	-	-	-0.02	52.3	*PTPRM*

Figure 1. Manhattan plot of *P*-value for (a) weight, (b) rump length and (c) body length. The log inverse *P*-values estimated for each polymorphism is plotted in the y-axis. Chromosome number is plotted in the x-axis. Horizontal line indicates the threshold $P < 0.0001$


(a)
$-\log_{10}(p)$
0 1 2 3 4 5 6
1 2 3 4 5 7 9 11 14 16 18 20 23 27 31
Chromosome
(b)
$-\log_{10}(p)$
0 1 2 3 4 5 6
1 2 3 4 5 7 9 11 14 16 18 20 23 27 31
Chromosome
(c)
$-\log_{10}(p)$
0 1 2 3 4 5 6
1 2 3 4 5 7 9 11 14 16 18 20 23 27 31
Chromosome