A single-step genome wide association study on Body Size Traits using imputation-based whole-genome sequence data in Yorkshire pigs

Huatao Liu, Hailiang Song, Yifan Jiang, Yao Jiang, Fengxia Zhang, Yibing Liu, Yong Shi, Xiangdong Ding*, Chuduan Wang*

National Engineering Laboratory for Animal Breeding, Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture, College of Animal Science and Technology, China Agricultural University, Beijing 100193, P.R. China;

Email addresses:
Huatao Liu, 925457179@qq.com
Hailiang Song, 530638625@qq.com
Yifan Jiang, jiangyf2020_2020@163.com
Yao Jiang, jiangyao133996@126.com
Fengxia Zhang, 895315294@qq.com
Yibing Liu, 279547013@qq.com
Yong Shi, 2273190789@qq.com
Xiangdong Ding, xding@cau.edu.cn
Chuduan Wang, cdwang@cau.edu.cn

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.
Abstract

Background: The body shape of pig is the most direct production index of pig, which can fully reflect the growth status of pig and is closely related to some important economic traits. In this study, genome-wide association study on seven body size traits, the body length (BL), height (BH), chest circumference (CC), abdominal circumference (AC), cannon bone circumference (CBC), rump width (RW) and chest width (CW) were conducted in Yorkshire pigs.

Methods: Illumina Porcine 80K SNP chip were used to genotype 589 of 5,572 Yorkshire pigs with body size records, and then the chip data was imputed to sequencing data. After quality control of imputed sequencing data, 784,267 SNPs were obtained, and the averaged linkage disequilibrium ($r^2$) was 0.191. We used the single-trait model and the two-trait model to conduct single-step genome wide association study (ssGWAS) on seven body size traits.

Results: A total of 198 significant SNPs were finally identified according to the $P$ value and the contribution to the genetic variance of individual SNP. 11 candidate genes (CDH13, SIL2, CDC14A, TMRPSS15, TRAPPC9, CTNND2, KDM6B, CHD3, MUC13, MAPK4 and HMGA1) were found to be associated with body size traits in pigs, KDM6B and CHD3 jointly affect AC and CC, and MUC13 jointly affect RW and CW. These genes are involved in the regulation of bone growth and development as well as the absorption of nutrients and are associated with obesity. HMGA1 is proposed as strong candidate gene for body size traits because of its
important function and high consistency with other studies regarding the regulation of body size traits. Our results could provide valuable information for pig breeding based on molecular breeding.

Keywords: pigs, body size traits, ssGWAS

Introduction

Pork is widely used as an important animal protein resource and has become one of the main sources of human protein. Commercial pigs (e.g. Duroc, Yorkshire and Landrace pigs) have the characteristics of fast growth, high feed utilization rate, high lean meat rate and obvious economic benefits. Therefore, it is not only a large number of breeding production, but also the focus of research. The body size trait is one kind of important phenotypic trait that can reflect the overall appearance of animals. Compared with the description of physical appearance, body size traits can objectively reflect the response of pigs to environment and other aspects[35]. In pig breeding, the body shape character index is often used as the most direct production index of pig. Body size is a typical quantitative (or complex) trait, understanding the genetic mechanism of body size differences among individuals can effectively help control the growth and production of animals[34]. At present, there are many researches on genetic parameters of pig external traits, which accelerate the process of genetic improvement of related traits. With the development of molecular
biotechnology, many studies have been carried out to clarify the genetic basis of pig body size traits.

By far, 1172 QTLs have been found related to body size traits in pigs according to PigQTLdb database (http://www.genome.iastate.edu/cgibin/QTLdb/ss/index). Although a range of researches have been done in QTL mapping, wide confidence intervals (covering more than 20 CM) for the positions of QTL remain that have rarely been replicated[39; 42]. A new research era was initiated with advances in single nucleotide polymorphism (SNP) chip and sequencing technology, and genome wide association study (GWAS) has become one of the most efficient methods to detect genetic variation in livestock[30]. Compared with traditional QTL localization, GWAS has more advantages in mining the intensity of medium-potency variation sites and defining the accuracy of genome segments containing variation sites[19; 26; 38; 41]. Although a large number of genome-wide association studies have been carried out in pigs, only few GWAS focused on identifying genes related to external traits. In particular, the investigation on body height, cannon bone circumference, rump width and other important body size traits are still lacking.

Marker density is one key factor affecting the efficiency of GWAS as gene mapping mainly relies on the linkage disequilibrium between causal mutation and markers[9]. Whole genome sequence data can definitely meet such requirements. In recent years, with the rapid development of the new generation of sequencing technology, the cost
of sequencing has been reduced rapidly, on one hand, a large number of samples and
the subsequent processing of sequence data are still time-consuming and costly,
limiting its utilization in genetic analysis. On the other hand, genotype imputation
provides one efficient tool to improve the marker density of SNP chip based on
sequence data. It can accurately predict the genotypes of polymorphic sites not
covered by the widely used SNP chip, allowing more genetic loci to be applied to
association analysis and improving the possibility of discovering new pathogenic
genes [32; 45]. In this study, we used imputation-based whole genome sequence data
to carry out GWAS on seven body size traits in pigs.

Materials and methods

Ethics statement

The whole recording procedure of ear tissue samples was carried out in strict
accordance with the protocol approved by the Institutional Animal Care and Use
Committee (IACUC) at the China Agricultural University. The IACUC of the China
Agricultural University approved this study (permit number DK996).

Animals and phenotypes

Yorkshire pigs born 2013-2016 from one pig breeding farm in Beijing were collected
in this study. Performance test on seven body size traits were carried out at the body
weight of about 100 kg for pigs. In total, 5,572 Yorkshire pigs with phenotypic
records and pedigree information were selected. The seven body size traits included body length (BL), body height (BH), chest circumference (CC), abdominal circumference (AC), cannon bone circumference (CBC), chest width (CW) and rump width (RW). Table 1 presents the descriptive statistics of body weight and seven body size traits. There were 4898 records for AC and 5572 records for the other six body size traits and body weight. Normal test showed all the traits followed normal distribution, and the body weight had the largest standard deviation of 12.59 and coefficient of variation of (12.43%), it was used as a covariate considering its influence on the body size traits in further analysis.

Genotype data and imputation

In this study, 589 out of 5572 Yorkshire pigs with body size records were genotyped using the PorcineSNP80 Bead Chip (Illumina, San Diego, CA), which includes 68,528 SNPs across the whole pig genome. In order to improve the marker density, the genotyped animals with another 6103 pigs genotyped with PorcineSNP80 [43] were imputed to whole genome sequence data using Beagle 4.1[10]. A wide collection of 289 sequenced pigs all with average sequencing depth of ~25X from 6 different pig breeds were used as reference data for imputation and each breed contained 24 to 94 pigs. The composition of reference data and the SNP calling of these individuals were described by Yan et al.[54]. After SNP calling, 46,766,110 SNPs were retained as the reference panel for imputation. On average, the genotype concordance rate across all variants was 92%, which is sufficient for further
After imputation, in this study, the following genotype quality control procedure was carried out using the PLINK software (v1.90) [36]. (1) SNPs with minor allele frequency (MAF) lower than 0.01 and deviated from Hardy - Weinberg equilibrium (P < 10^{-6}) were excluded and only variants located on autosomes were used for further analysis; (2) the SNP with call rate less than 0.95 were removed; (3) individuals with call rate less than 0.90 were excluded. In addition, in order to decrease the influence of the dependence of adjacent markers on the high false positive of GWAS analysis, the SNP were further pruned, the SNP with linkage disequilibrium ($r^2$) in slide window of 50 SNPs less than 0.9 were selected. Finally, all the genotyped animals and 784267 SNPs were retained.

**Statistical models**

**genetic correlation**

According to the information of 5,572 pigs in this study, the restricted maximum likelihood method (AI-REML) in DMU v6.0 software [31] was used to estimate the genetic correlations of seven body size traits.

The animal model was used to estimate the genetic parameters:

$$ y = \mu + Xb + Z_1a + Z_2t + e,$$

with
where, \( y \) is the vector of phenotypic values of each body size trait; \( \mu \) is the population mean; \( b \) is the fixed effect of herd-year-season; \( a \) is the vector of additive genetic effects; \( t \) is the covariate vector of body weight effects; \( e \) is a vector of residual effects. \( X, Z_1 \) and \( Z_2 \) are incidence matrices associating \( b, a \) and \( t \) with \( y \), respectively. \( A \) is the genetic relationship matrix, five generations of pedigree were traced back to construct \( A \), and \( \sigma_a^2 \) is the additive genetic variance. \( I \) is the identity matrix of appropriate dimension, \( \sigma_t^2 \) is the variance of body weight effect and \( \sigma_e^2 \) is the residual variance.

Subsequently, genetic correlations were calculated based on the variance components as follows:

\[
\rho_A = \frac{\text{cov}(a_1, a_2)}{\sigma_{a_1}\sigma_{a_2}}
\]

where, \( \rho_A \) is the genetic correlation between trait 1 and trait 2, \( a_1 \) and \( a_2 \) represent the additive genetic values of trait 1 and trait 2 for same individuals, \( \text{cov} (a_1, a_2) \) and \( \sigma_{a_1}, \sigma_{a_2} \) refer to the genetic covariance of two traits and the genetic standard deviation of trait 1 and trait 2, respectively.
Genome-wide Association Study

In this study, single-step GWAS (ssGWAS), which can simultaneously use all the SNP information and utilize the ungenotyped animals with phenotypic records[47], was implemented to identify significant SNPs associated with body size traits. Considering the genetic correlations between body size traits, two-trait ssGWAS model was also conducted on traits with high genetic correlations.

Single-trait ssGWAS

Single-trait ssGWAS model was used for three body size traits BL, BH and CBC.

\[ y = Xb + \gamma W + Zg + e \]

where \( y \) is the vector of phenotypic values, \( b \) is the vector of fixed effects including herd-year-season-sex, \( W \) is the covariate of body weight, \( g \) is the vector of additive genetic effects, following a normal distribution of \( N(0, H\sigma_g^2) \), in which \( H \) is the matrix of additive genetic relationships incorporating both pedigree and genomic information, \( \sigma_g^2 \) is the additive genetic variance, \( e \) is the vector of random residuals with distribution of \( N(0, I\sigma_e^2) \), in which \( I \) is the identity and \( \sigma_e^2 \) is the residual variance. \( X, W \) and \( Z \) is the incidence matrix associating \( b, w, g \) with \( y \), respectively.

The genotyped and ungenotyped animals were considered simultaneously based on a H matrix[4]. The inverse of the H matrix was written as follows:
\[ H^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & G_w^{-1} - A_{22}^{-1} \end{bmatrix} + A^{-1} \]

where \( A^{-1} \) is the inverse of the numerator relationship matrix, \( A_{22}^{-1} \) is only the inverse of the pedigree-based relationship matrix for the genotyped animals, and \( G_w^{-1} \) is the inverse of the genomic relationship matrix; G weight markers were obtained by reciprocals of expected marker variance[46].

The SNP effects could be estimated by ssGWAS. The proportion of genetic variance explained by single SNP was calculated as follows:

\[
\frac{\text{Var}(Z_j \hat{u}_j)}{\sigma_a^2} \times 100\%
\]

where \( \sigma_a^2 \) is the total genetic variance, \( Z_j \) is a vector of the gene content of the jth SNP for all animals, and \( \hat{u}_j \) is the estimated marker effect of the jth SNP.

**Two-trait ssGWAS**

According to the genetic correlation estimations, four body size traits with high genetic correlations (CC and AC, RW and CW) were carried out using two-trait ssGWAS model.

\[
\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Y_1 \\ W_1 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix},
\]

where \( \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \) is the vector of observation values of trait I and II, \( b_1 \) and \( b_2 \) are the vector of fixed effects of herd-year-season-sex of trait I and II, \( X_1 \) and \( X_2 \) are the incidence matrix associating \( b_1 \) and \( b_2 \) with \( y_1 \) and \( y_2 \), \( \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \) is the vector of
covariate of body weight of trait I and II, $y_1$ and $y_2$ are the regression coefficient associating $W_1$ and $W_2$, $[g_1; g_2]$ is the vector of additive genetic effects of the two traits, following a normal distribution of $N(0, HM)$, where $M = \begin{bmatrix} \sigma_{g1}^2 & \sigma_{g12}^2 \\ \sigma_{g12}^2 & \sigma_{g2}^2 \end{bmatrix}$ is the additive genetic variance and covariance matrix of the two traits. $Z_1$ and $Z_2$ are the incidence matrix associating $g_1$ and $g_2$ with $y_1$ and $y_2$, $[e_1; e_2]$ is the vector of random errors with distribution of $N(0, IR)$, where $I$ is the identity matrix and $R = \begin{bmatrix} \sigma_{e1}^2 & \sigma_{e12}^2 \\ \sigma_{e12}^2 & \sigma_{e2}^2 \end{bmatrix}$ is the residual variance and covariance matrix of the two traits.

In this study, for both single trait model and two-trait model of ssGWAS, blupf90 [5] was implemented to estimate genomic breeding values (GEBV), and afterwards, based on GEBV, SNP effects and P-values were estimated via postGSf90. The P value of each marker was calculated as follows [3]:

$$P_i = P_t(\frac{\hat{u}_i}{\sqrt{\sigma_{e1}^2/n}}, n - 1),$$

where $P_i$ is the distribution function of t distribution, $\hat{u}_i$ is ith SNP effect, $\sigma_{e1}^2$ is the genetic variance of ith SNP, n is the number of animals with ith SNP. In addition, the proportion of genetic variance explained by the ith SNP could also be calculated as $\frac{\sigma_{i}^2}{\sigma_g^2}$. Manhattan plots of SNP variance were obtained by the “qqman” R package [14].

In order to control false positives, the False Discovery Rate (FDR) [6; 52] method for multiple testing was used as follow:
where \( m \) is the number of times to be tested, \( n \) is the number of significant SNPs at assigned FDR level, e.g. 0.05. \( P_{\text{Max}} \) is the genome-wide significance level empirical P-value of FDR adjusted. Based on the P-values of SNPs obtained by ssGWAS, the empirical P-value of FDR adjusted at the genome-wide significance level of 0.05 was calculated on each trait in this study.

**Identification of candidate genes**

After identifying significant SNPs by ssGWAS, the genes located in the 50Kb downstream and 50 Kb upstream region of the significant SNPs were determined using BedTools[37] and pig reference gene annotation (http://www.ensembl.org/Sus_scrofa/Info/Index/; Sus scrofa 11.1 genome version).

Using the R package bioconductor (http://www.bioconductor.org/) to identify the related pathways and functional annotation. QTLdb (http://www.animalgenome.org/cgi-bin/QTLdb/SS/download?file=gbsSS_11.1) was used to annotate significant SNPs located in previously mapped QTLs in pigs. R package 'Cluster Profiler'[55] was used to carry out Gene Ontology (GO) and Kyoto research on annotated candidate genes Encyclopedia of Genes and Genomes (KEGG) enrichment analysis.

**Results**
Table 2 shows the genetic correlations of seven body size traits. The genetic correlations ranged from -0.286 to 0.840 with standard errors ranging from 0.028 to 0.106. Among the seven body size traits, chest circumference (CC) and abdominal circumference (AC), chest width (CW) and rump width (RW) had the higher genetic correlations of 0.747 and 0.840 with standard errors of 0.055 and 0.028, respectively. The genetic correlations between other traits were lower than 0.3, and some traits were almost not genetic correlated with other traits, e.g. body length (BL) had very low genetic correlation of -0.010, 0.03, -0.01, 0.01 with body height (BH), CC, AC, CW, respectively.

Identification of significant SNPs associated with body size traits

Two criteria of P value and SNP effect were respectively used to determine the SNPs associated with body size traits. As to the P value, after the 0.05 significance level of the whole genome was adjusted, the $P_{\text{Max}}$ values of FDR-based multiple tests were 9.26E-06 for BL, 1.08E-05 for BH, 1.02E-05 for CBC, 9.74E-06 for AC, 1.05E-05 for CC, 9.60E-06 for RW, and 1.01E-05 for CW. As shown in Table 3, a total of 88 significant SNPs was identified for seven body size traits. The Manhattan plots of the three traits BL, BH and CBC using the single trait model are shown in Figure 1. For BL, a total of 9 significant SNPs reached the genome-wide significance level, totally accounting for 0.0085% of the genetic variance. These significant SNPs were located
on SSC1, SSC6, SSC8, SSC13, SSC14, SSC16, and SSC17. The SNP at SSC17:33632497 explained the largest genetic variance (0.0029%). For BH, only 6 SNPs were genome-wide significant, accounting for a total of 0.0123% of genetic variance. They were located on SSC3, SSC5, SSC14, and SSC16. The interpretation of scc16: 886074 has the largest genetic variance (0.0082%). For CBC, there were 15 significant SNPs at the genome-wide level, which explained 0.0267% of the genetic variance, and the most significant SNPs were closely located on SSC1. For the two pairs of genetic correlated traits using the two-trait model, the Manhattan plots of AC and CC, RW and CW are shown in Figure 2. In total, 8, 17, 9, and 24 SNPs were identified associated with AC, CC, RW, and CW, respectively, and these SNPs explained 0.0109%, 0.0242%, 0.0099% and 0.0281% of genetic variances for the corresponding traits. For each trait, the genetic variance explained by a single significant SNP was very small, the largest of which for each trait were 0.0051% (SSC5:15137502), 0.0067% (SSC4:64552365), 0.0038% (SSC9:2330339) and 0.0065% (SSC7:115471416), respectively. Although the genetic correlations existed among seven body size traits, no common significant SNPs were found.

Considering the small contribution of above significant SNPs to the genetic variance, the proportion of genetic variance explained by each SNP were also illustrated as shown in Figure 3 in this study. Top 20 SNPs with the largest genetic variance were selected for each trait (Table 3), SNPs for BL were located on SSC17, BH on SSC2, SSC5 and SSC16, CBC on SSC7 and SSC4, SNPs with largest genetic variance for
AC and CC are located on SSC12, those for RW and CW were on SCC6 SCC7, SCC13 and SCC17. For each body size trait, BL, BH, CBC, AC, CC, RW and CW, the top 20 SNPs explained 2.01%, 1.56%, 1.63%, 2.39%, 2.32%, 1.54% and 1.23% of the genetic variance, respectively. Interestingly, the top 20 SNPs for AC and CC were same, RW and CW shared half of the 20 SNPs. In total, 110 SNPs with larger proportion of explanatory genetic variance were retained for further analysis (Table 1).

Identification of candidate genes

All the significant SNPs identified by the two methods were annotated within the 50 Kb downstream and upstream region with reference to the Sus scrofa 11.1 genome assembly. According to the two methods of SNP significance and explained genetic variance, 88 and 110 SNPs were identified without overlapping, and 64 and 40 genes were found near these SNPs and only two of them were common, respectively (Table 3 and Stable 1). Six and seven genes were found to be related to the corresponding body size traits by the two methods. The biological processes and pathways involved in these genes include calcium channel proteins, lipid metabolism, and cell proliferation.

Discussion

The superiority of imputation-based WGS data
Genotype marker density is one important factor affecting the efficiency of GWAS [9]. With the increase of marker density, the linkage disequilibrium between markers and the target trait QTL is increased, it is helpful for QTL detection. In previous studies, the advantages of whole genome sequencing data have been demonstrated[49]. However, its high cost hampered the widely application of sequencing data. Genotype imputation was proved efficiently to impute the SNP chip data to sequencing data with high accuracy[20]. Our results indicated that imputation-based WGS data dramatically improved the power of GWAS, among the significant SNPs identified in this study, only 3 out of the 88 significant SNPs were located in the PorcineSNP80 SNP chip, the remaining 85 loci were identified in the sequencing data. Moreover, among the 110 non-repeating loci screened by interpretation variance, 101 are new loci after imputation, which indicates that imputed WGS data adds a lot of useful information.

Increasing marker density could lead to high linkage disequilibrium (LD) to improve the resolution of gene mapping, while it may also be a burden[24]. Too high LD between markers will cause noise and increase false positive[50]. One of the strategies to deal with such dilemma is to pre-select SNP, which can be done via SNP selection to only keep a set of SNPs that are mutually uncorrelated[11; 18]. Therefore, we pruned SNPs according to the genome-wide sequence data to reduce the LD degree between SNPs, and retained the loci in the original 80K chip. In this study, 44003 out of the qualified 50179 SNPs in PorcineSNP80 chip according to the
genotype quality control were retained, and the average linkage disequilibrium of the finally used 784,267 SNPs is similar to that of the chip data, the average $r^2$ was 0.191 and 0.195, respectively. This not only retains the original SNPs but also increases a large number of SNPS, and does not cause the increase of LD.

**The advantage of ssGWAS**

Single SNP regression model is widely used in GWAS to identify the association of SNP with traits of interest, whereas it usually yields a high false-positive rate due to ignoring the linkage disequilibrium between adjacent SNPs. Wang et al. [47] proposed Single-step GWAS (ssGWAS) that combines all the data (genotype, phenotype and pedigree information) in one step. It can simultaneously utilize all the markers compared with the single-marker regression genome-wide association analysis, resulting in higher power and accuracy [48]. In addition, ssGWAS is able to use sliding windows to simultaneously analyze multiple SNPs to reduce errors [8; 17]. Wang et al. (2012) reported that ssGWAS achieved accuracy of 0.81 ± 0.02 using 1500 genotype animals, which was more accurate than single SNP regression model [47]. Moreover, ssGWAS can utilize more individuals, the sample size in this study is not very large, but has a large number of phenotypic data of ungenotyped animals. Compared with traditional GWAS, ssGWAS can make full use of this part of information, expand the sample size to a certain extent, improve the accuracy of SNP effect estimation, and further improve the efficiency of SNP identification.
The determination of significant SNP using P value or SNP effect

Theoretically, the SNPs with smallest p values were supposed to explain relatively high proportion of genetic variance. Likewise, the SNPs with large effects should be significantly associated with the trait of interest. However, our results indicated that the SNPs with smallest P values did not have large effects, there was no overlap between the top 20 SNPs with smallest P values and with largest SNP effects for each trait. Therefore, in order to locate QTLs related to traits more accurately and comprehensively, this study identified significant SNPs from both P value and SNP effect. The proportion of genetic variance explained by most the significant SNPs was small (0.00004%-0.00653%) for all traits, and the maximum genetic variance of all SNPs was also not large (0.0557%-0.1205%), perhaps because too many SNPs were used in the sequencing data in this study, leading to small effect of each related SNP for each trait. It also indicates that SNPs controlling body size traits are widely distributed on the genome, fitting well the infinitesimal model. It was reported that for complex traits such as height, action sites are widely distributed across the entire genome, indicating that almost all genes are involved in the regulation of height[7].

Pleiotropic effects can lead to genetic correlation between traits. From the aspect of P value, no overlap of significant SNPs associated with two genetic related trait pairs AC and CC, RW and CW were detected in this study. However, more common SNPs with largest effects (not statistically significant) were found in each pair of
genetic related traits, e.g. the top 20 SNPs with largest effects for AC and CC were completely overlapped, these SNPs were adjacent to each other and located near SCC12:53132997. Therefore, it is speculated that these SNPs constitute an important QTL and jointly affect AC and CC. Similarly, there may be QTLs associated with RW and CW around SSC6:39553559 and SCC13:135373704. In addition, we took 20 SNPs as a sliding window, and found that the top 20 windows with largest genetic effects respective for AC and CC were overlapped, as well for RW and CW. The above results further reflect 'one factor produces multiple effects', suggesting that highly genetic related traits are probably regulated by the same QTL.

Potential Candidate Genes for BL, BH and CBC

The body length (BL) is an important index to investigate the breeding performance of animals. According to bioinformatics analysis, CDH13 near SCC6:5671575 could be used as a candidate gene affecting body length. CDH13 is a unique cadherin that regulates cell adhesion, signal transduction and cell growth, and plays an important role in the formation of tissues and organs. The ingestion and transfer of Ca will affect the bone development of body for a long time, therefore, CDH13 has a certain influence on the growth and development of the body. For BH, SIL2 was found to be associated with this trait near SCC2:46827557, Proteomic studies showed that SIL1 elevation alters the expression of proteins including crucial players in neurodegeneration, abnormal expression of SIL1 has an impact on the morphology of
the body, which can reduce the body size[28]. CBC reflects the physical quality of the
animal, whether it is strong or not. There are three candidate genes associated with
CBC, CDC14A, TMPRSS15 and TRAPPC9. CDC14A is widely expressed in
eukaryotic cell biology of a special kind of highly conservative dual specificity
phosphatase, a variety of studies from yeast to human somatic cells have shown that
CDC14 involves extensive roles, including embryonic development and body size[1].
TMPRSS15 has an impact on the digestive efficiency of animals, and has been found
to be associated with the formation of cholesterol in humans and has been shown to
be associated with the development of fat and body weight in mice[51]. TMPRSS15
has also a higher variance ranking based on the SNP effect. The gene mutation of
transporter particle complex 9 (TRAPPC9), a protein subunit of transporter particle II
(TRAPPII), can lead to abnormal embryonic development, abnormal dietary behavior,
and is associated with body mass index[2; 33].

Potential Candidate Genes for AC and CC

AC and CC are a pair of highly genetically-related body size traits, which determine
the body size of animals and are indicators to fatness and thinness. CTNND2 was
closely related to AC according to the P value. Studies have shown that CTNND2
participates in the regulation of cell proliferation and affects the body node number of
zebrafish[57]. It is found that KDM6B and CHD3 jointly affect AC and CC. KDM
subfamily 6 enzymes B (KAM6B) plays an important role in repression of
developmental genes[25], and has a regulatory effect on chondrocyte differentiation, thus affecting bone growth and development[15]. CDH3 is a calcium-binding protein that is involved in calcium ion binding and protein binding and is associated with diseases such as malnutrition and developmental malformations. Studies have shown that CHD3 regulates the developmental morphology of zebrafish heart, thereby affecting the abdominal circumference and body shape of zebrafish[12].

Potential Candidate Genes for RW and CW

For RW and CW, MUC13 was detected to affect both RW and CW. MUC13 promotes cell proliferation and migration, inhibits apoptosis, and reduces adhesion through a number of signaling pathways[40], and has a certain effect on the absorption of intestinal nutrients, thus affecting the growth and development of bone and the organism. It was found that MAPK4 and HMGA1 affect RW and CW of pigs respectively. MAPK4 is mitogen-activated protein kinase 4, which is involved in the absorption and decomposition of sugars and the formation of fat, so it is related to obesity traits[53]. HMGA1 affects the expression of two IGFBP(insulin-like growth factor binding protein) protein species and plays an important role in cell growth and differentiation[13; 21]. Studies have shown that the deletion of HMGA1 gene results in a significant decrease in the body size of mice[16]. Moreover, a large number of studies have shown that HMGA1 is related to the body size character of pigs. Ji et al.[23] found HMGA1 was a candidate gene affecting body size of pig through
genome-wide association analysis. Zhang et al[56] found that HMGA1 is expressed in pig limb cells and affects the growth and differentiation of chondrocytes. Because of the functional importance of HMGA1 and several studies have shown that it is highly associated with body size traits, it is worth being verified in the future.

**Conclusion**

In this study, among seven body size traits in pigs, CC and AC, CW and RW were highly genetic correlated with correlation of 0.747 and 0.840, respectively. We implemented ssGWAS to identify SNPs associated with body size traits based on two aspects of P value and the proportion of explanatory genetic variance of SNP. In total, 198 SNPs were identified associated with seven body size traits in Yorkshire, correspondingly, 11 genes were related to body size traits, among which HMGA1 could be worth being validated in further study.

**Data availability**

The ped and the map are not publicly available because the genotyped animals belong to commercial breeding companies, but are available from the corresponding author on reasonable request.

**Conflict of interest**

The authors declare that they have no competing interest.
Author contributions

XD and CD conceived and supervised the study. HT, HL and LF helped complete the imputation of the chip data and provide technical guidance. HT, JY, HL, YF collected the samples and recorded the phenotypes. JY and LF extracted the DNA for genotyping. HT, FX, YB and SY contributed to the visualization of data. HT and XD wrote and revised the manuscript. All authors read and approved the manuscript.

Funding

The authors are grateful to all the funding agencies. This work was supported by grants for the National Key Research and Development Project (2019YFE0106800, 2018YFD0501000), China Agriculture Research System (CARS-35), Modern Agriculture Science and Technology Key Project of Hebei Province (19226376D), the National Natural Science Foundation of China (31671327).

References:

1. Research progress in function of cell division cycle protein Cdc14A. Journal of Jilin University. Medicine edition. 2018;44(1671-587X(2018)44:5<1105:XBFLZQ>2.0.TX;2-55): 1105-8.

2. Abbasi AA, Blaesius K, Hu H, Latif Z, Picker-Minh S, Khan MN et al. Identification of a novel homozygous TRAPPC9 gene mutation causing non-syndromic intellectual disability, speech disorder, and secondary microcephaly. Am J Med Genet B Neuropsychiatr Genet. 2017;174(8): 839-45. doi:10.1002/ajmg.b.32602.
3. Aguilar I, Legarra A, Cardoso F, Masuda Y, Lourenc D, Misztal I. Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. GENETICS SELECTION EVOLUTION. 2019;51(28). doi:10.1186/s12711-019-0469-3.

4. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J Dairy Sci. 2010;93(2): 743-52. doi:10.3168/jds.2009-2730.

5. Aguilar I, Misztal I, Legarra A, Tsuruta S. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. JOURNAL OF ANIMAL BREEDING AND GENETICS. 2011;128(6SI): 422-8. doi:10.1111/j.1439-0388.2010.00912.x.

6. BENJAMINI Y, HOCHBERG Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES B-STATISTICAL METHODOLOGY. 1995;57(1): 289-300. doi:10.1111/j.2517-6161.1995.tb02031.x.

7. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: From polygenic to omnigenic. Cell. 2017;169(7): 1177-86. doi:10.1016/j.cell.2017.05.038.

8. Braz CU, Taylor JF, Bresolin T, Espigolan R, Feitosa F, Carvalheiro R et al. Sliding window haplotype approaches overcome single SNP analysis limitations in identifying genes for meat tenderness in Nelore cattle. BMC Genet. 2019;20(1): 8. doi:10.1186/s12863-019-0713-4.
9. Brondum RF, Ma P, Lund MS, Su G. Short communication: Genotype imputation within and across Nordic cattle breeds. J Dairy Sci. 2012;95(11): 6795-800. doi:10.3168/jds.2012-5585.

10. Browning BL, Browning SR. A unified approach to genotype imputation and Haplotype-Phase inference for large data sets of trios and unrelated individuals. The American Journal of Human Genetics. 2009;84(2): 210-23. doi:10.1016/j.ajhg.2009.01.005.

11. Calus M, Vandenplas J. SNPrune: An efficient algorithm to prune large SNP array and sequence datasets based on high linkage disequilibrium. Genet Sel Evol. 2018;50(1): 34. doi:10.1186/s12711-018-0404-z.

12. Cho SH, Lee CH, Gi E, Yim Y, Koh HJ, Kang K et al. The rice rolled fine striped (RFS) CHD3/Mi-2 chromatin remodeling factor epigenetically regulates genes involved in oxidative stress responses during leaf development. Front Plant Sci. 2018;9(364. doi:10.3389/fpls.2018.00364.

13. Cleynen I, Van de Ven WJ. The HMGA proteins: A myriad of functions (Review). Int J Oncol. 2008;32(2): 289-305.

14. D. Turner S. Qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. Journal of open source software. 2018;3(25): 731. doi:10.21105/joss.00731.

15. Dai J, Yu D, Wang Y, Chen Y, Sun H, Zhang X et al. Kdm6b regulates cartilage development and homeostasis through anabolic metabolism. Ann Rheum Dis. 2017;76(7): 1295-303. doi:10.1136/annrheumdis-2016-210407.

16. Federico A, Forzati F, Esposito F, Arra C, Palma G, Barbieri A et al. Hmgal/Hmga2
double knock-out mice display a "superpygmy" phenotype. BIOLOGY OPEN. 2014;3(5):372-8. doi:10.1242/bio.20146759.

17. Guerra FP, Suren H, Holliday J, Richards JH, Fiehn O, Famula R et al. Exome resequencing and GWAS for growth, ecophysiology, and chemical and metabolomic composition of wood of Populus trichocarpa. BMC Genomics. 2019;20(1): 875. doi:10.1186/s12864-019-6160-9.

18. Hazelett DJ, Conti DV, Han Y, Al OA, Easton D, Eeles RA et al. Reducing GWAS complexity. Cell Cycle. 2016;15(1): 22-4. doi:10.1080/15384101.2015.1120928.

19. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nature Reviews Genetics. 2005;6(2): 95-108. doi:10.1038/nrg1521.

20. Hoze C, Fouilloux MN, Venot E, Guillaume F, Dassonneville R, Fritz S et al. High-density marker imputation accuracy in sixteen French cattle breeds. Genet Sel Evol. 2013;45(3). doi:10.1186/1297-9686-45-33.

21. Hristov AC, Cope L, Di Cello F, Reyes MD, Singh M, Hillion JA et al. HMGA1 correlates with advanced tumor grade and decreased survival in pancreatic ductal adenocarcinoma. Modern Pathology. 2010;23(1): 98-104. doi:10.1038/modpathol.2009.139.

22. Iotzova-Weiss G, Freiberger SN, Johansen P, Kamarachev J, Guenova E, Dziunycz PJ et al. TLR4 as a negative regulator of keratinocyte proliferation. PLoS One. 2017;12(10): e0185668. doi:10.1371/journal.pone.0185668.

23. Ji J, Zhou L, Guo Y, Huang L, Ma J. Genome-wide association study identifies 22 new loci
for body dimension and body weight traits in a White DurocxErhualian F2 intercross population. Asian-Australas J Anim Sci. 2017;30(8): 1066-73. doi:10.5713/ajas.16.0679.

24. Joiret M, Mahachie JJ, Gusareva ES, Van Steen K. Confounding of linkage disequilibrium patterns in large scale DNA based gene-gene interaction studies. BioData Min. 2019;12(11. doi:10.1186/s13040-019-0199-7.

25. Jones SE, Olsen L, Gajhede M. Structural Basis of Histone Demethylase KDM6B Histone 3 Lysine 27 Specificity. Biochemistry. 2018;57(5): 585-92. doi:10.1021/acs.biochem.7b01152.

26. Klein RJHJ. Complement factor h polymorphism in Age-Related macular degeneration. Science. 2005;308): 385-9.

27. Kovacs CS. Maternal mineral and bone metabolism during pregnancy, lactation, and Post-Weaning recovery. Physiol Rev. 2016;96(2): 449-547. doi:10.1152/physrev.00027.2015.

28. Labisch T, Buchkremer S, Phan V, Kollipara L, Gatz C, Lentz C et al. Tracking effects of SIL1 increase: Taking a closer look beyond the consequences of elevated expression level. Mol Neurobiol. 2018;55(3): 2524-46. doi:10.1007/s12035-017-0494-6.

29. Liu B, Song J, Luan J, Sun X, Bai J, Wang H et al. Promoter methylation status of tumor suppressor genes and inhibition of expression of DNA methyltransferase 1 in non-small cell lung cancer. Exp Biol Med (Maywood). 2016;241(14): 1531-9. doi:10.1177/1535370216645211.

30. Mackay TFC, Stone EA, Ayroles JF. The genetics of quantitative traits: Challenges and
prospects. Nature Reviews Genetics. 2009;10(8): 565-77. doi:10.1038/nrg2612.

31. Madsen PAJ. A user’s guide to DMU: A package for analysing multivariate mixed models.. 2013.

32. Marchini J, Howie B. Genotype imputation for genome-wide association studies. NATURE REVIEWS GENETICS. 2010;11(7): 499-511. doi:10.1038/nrg2796.

33. Mbimba T, Hussein NJ, Najeed A, Safadi FF. TRAPPC9: Novel insights into its trafficking and signaling pathways in health and disease (Review). Int J Mol Med. 2018;42(6): 2991-7. doi:10.3892/ijmm.2018.3889.

34. Niu P, Kim S, Choi B, Kim T, Kim J, Kim K. Porcine insulin-like growth factor 1 (IGF1) gene polymorphisms are associated with body size variation. GENES & GENOMICS. 2013;35(4): 523-8. doi:10.1007/s13258-013-0098-0.

35. Ohnishi C, Satoh M. Estimation of genetic parameters for performance and body measurement traits in Duroc pigs selected for average daily gain, loin muscle area, and backfat thickness. Livestock Science. 2018;214(161-6. doi:10.1016/j.livsci.2018.05.022.

36. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D et al. PLINK: A tool set for Whole-Genome association and Population-Based linkage analyses. The American Journal of Human Genetics. 2007;81(3): 559-75. doi:10.1086/519795.

37. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6): 841-2. doi:10.1093/bioinformatics/btq033.

38. Risch NMK. The future of genetic studies of complex human diseases. Epidemiology. 1996;9): 350-4.
39. Schreiweis MA, Hester PY, Moody DE. Identification of quantitative trait loci associated with bone traits and body weight in an F2 resource population of chickens. Genetics and Selection Evolution. 2005;37(6): 677-98. doi:10.1051/gse:2005023.

40. Sheng YH, Triyana S, Wang R, Das I, Gerloff K, Florin TH et al. MUC1 and MUC13 differentially regulate epithelial inflammation in response to inflammatory and infectious stimuli. Mucosal Immunol. 2013;6(3): 557-68. doi:10.1038/mi.2012.98.

41. Simon-Sanchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat Genet. 2009;41(12): 1308-12. doi:10.1038/ng.487.

42. Soller M, Weigend S, Romanov MN, Dekkers JCM, Lamont SJ. Strategies to assess structural variation in the chicken genome and its associations with biodiversity and biological performance. Poultry Science. 2006;85(12): 2061-78. doi:10.1093/ps/85.12.2061.

43. Song H, Ye S, Jiang Y, Zhang Z, Zhang Q, Ding X. Using imputation-based whole-genome sequencing data to improve the accuracy of genomic prediction for combined populations in pigs. Genet Sel Evol. 2019;51(1): 58. doi:10.1186/s12711-019-0500-8.

44. Takeuchi T, Liang SB, Matsuyoshi N, Zhou S, Miyachi Y, Sonobe H et al. Loss of T-cadherin (CDH13, H-cadherin) expression in cutaneous squamous cell carcinoma. Lab Invest. 2002;82(8): 1023-9. doi:10.1097/01.lab.0000025391.35798.f1.

45. van Leeuwen EM, Kanterakis A, Deelen P, Kattenberg MV, Slagboom PE, de Bakker PIW et al. Population-specific genotype imputations using minimac or IMPUTE2. NATURE
46. VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91(11):4414-23. doi:10.3168/jds.2007-0980.

47. Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet Res (Camb). 2012;94(2):73-83. doi:10.1017/S0016672312000274.

48. Wang H, Misztal I, Aguilar I, Legarra A, Fernando RL, Vitezica Z et al. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. Front Genet. 2014;5(134). doi:10.3389/fgene.2014.00134.

49. Wang L, Zhang Y, Zhang T, Zhang L, Yan H, Liu X et al. Genotyping by sequencing reveals a new locus for pig teat number. Anim Genet. 2017;48(4):470-2. doi:10.1111/age.12547.

50. Wang X, Elston RC, Zhu X. The meaning of interaction. Hum Hered. 2010;70(4):269-77. doi:10.1159/000321967.

51. Wang Z, Xu S, Du K, Huang F, Chen Z, Zhou K et al. Evolution of digestive enzymes and RNASE1 provides insights into dietary switch of cetaceans. Mol Biol Evol. 2016;33(12):3144-57. doi:10.1093/molbev/msw191.

52. Weller JI, Song JZ, Heyen DW, Lewin HA, Ron M. A new approach to the problem of multiple comparisons in the genetic dissection of complex traits. GENETICS. 1998;150(4):1699-706.
53. Wu YS, Chen YT, Bao YT, Li ZM, Zhou XJ, He JN et al. Identification and verification of potential therapeutic target genes in Berberine-Treated zucker diabetic fatty rats through bioinformatics analysis. PLoS One. 2016;11(11): e166378. doi:10.1371/journal.pone.0166378.

54. Yan G, Qiao R, Zhang F, Xin W, Xiao S, Huang T et al. Imputation-Based Whole-Genome sequence association study rediscovered the missing QTL for lumbar number in sutai pigs. Sci Rep. 2017;7(1): 615. doi:10.1038/s41598-017-00729-0.

55. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R package for comparing biological themes among gene clusters. OMICS. 2012;16(5): 284-7. doi:10.1089/omi.2011.0118.

56. Zhang LC, Li N, Liu X, Liang J, Yan H, Zhao KB et al. A genome-wide association study of limb bone length using a Large White x Minzhu intercross population. Genet Sel Evol. 2014;46(56. doi:10.1186/s12711-014-0056-6.

57. Zhang Y, Zhang R, Ding X, Ai K. EFNB2 acts as the target of miR-557 to facilitate cell proliferation, migration and invasion in pancreatic ductal adenocarcinoma by bioinformatics analysis and verification. Am J Transl Res. 2018;10(11): 3514-28.
**Figures**

**Figure 1. Manhattan plot of the genome-wide association study on three body size traits by using single-trait model ssGWAS.** BL, Body length; BH, Body height; CBC, Cannon bone circumference. In the Manhattan plots, negative log10 P-values of the quantified SNPs were plotted against their genomic positions. The x-axis represents the chromosomes, and the y-axis represents the observed \(-\log_{10}(P\text{-value})\). Different colors indicate various chromosomes. Each trait has a significant threshold of FDR adjusted, for (A) BL, it was $9.26 \times 10^{-6}$. Similarly, (B) BH was $1.08 \times 10^{-5}$, and (D) CBC was $1.02 \times 10^{-5}$.

**Figure 2. Manhattan plot of the genome-wide association study on four body size traits by using two-trait model ssGWAS.** AC, Abdominal circumference; CC, Chest circumference; RW, Rump width; CW, Chest width. AC and CC are a pair of traits, RW and CW are a pair of traits. In the Manhattan plots, negative log10 P-values of the quantified SNPs were plotted against their genomic positions. The x-axis represents the chromosomes, and the y-axis represents the observed \(-\log_{10}(P\text{-value})\). Different colors indicate various chromosomes. Each trait has a significant threshold of FDR adjusted, for (A) AC, it was $9.74 \times 10^{-6}$. Similarly, (B) CC was $1.05 \times 10^{-5}$, (C) RW was $9.60 \times 10^{-6}$, and (D) CW was $1.01 \times 10^{-5}$.

**Figure 3. Manhattan plot of the genome-wide association study on seven body size traits and Venn plot of SNPs according to the contribution of SNP to genetic variance by using ssGWAS.** BL, Body length; BH, Body height; CBC, Cannon bone circumference; AC, Abdominal circumference; CC, Chest circumference; RW, Rump width; CW, Chest width. BL, BH and CBC were single-trait models, AC, CC, RW and CW were two-trait models. AC and CC are a pair of traits, RW and CW are a pair of traits. In the Manhattan plots (A-G), the proportion of genetic variance of the quantified SNPs were plotted against their genomic positions. The x-axis represents the chromosomes, and the y-axis represents the percentage of SNP explaining the genetic variance. Different colors indicate different chromosomes. Venn plot (H) of SNPs for the two pairs of body size traits, AC and CC, RW and CW are a pair of traits, respectively.
### Table 1 Descriptive statistics for body weight and seven body size traits

| Trait | N-obs | Mean | S.D. | CV(%) | Min value | Max value |
|-------|-------|------|------|-------|-----------|-----------|
| BL(cm) | 5573 | 108.89 | 6.18 | 5.67 | 88 | 134 |
| BH(cm) | 5573 | 62.87 | 2.92 | 4.64 | 51 | 75 |
| CC(cm) | 5573 | 104.58 | 5.75 | 5.50 | 85 | 126 |
| AC(cm) | 4898 | 113.52 | 6.31 | 5.56 | 94 | 137 |
| CW(cm) | 5572 | 29.75 | 2.31 | 7.76 | 19 | 38 |
| RW(cm) | 5573 | 31.64 | 2.13 | 6.73 | 22 | 40 |
| CBC(cm) | 5573 | 17.98 | 1.03 | 5.73 | 13 | 23 |
| BW(kg) | 5573 | 101.31 | 12.59 | 12.43 | 61 | 150 |

Note: 1 BL = body length, BH = body height, CC = chest circumference, AC = abdominal circumference, CBC = cannon bone circumference, RW = rump width, CW = chest width, 2 N-obs = number of observations

### Table 2 Genetic correlations between seven body size traits

| Trait | BL | BH | CC | AC | CW | RW | CBC |
|-------|----|----|----|----|----|----|-----|
| BL | -0.010(0.088) | 0.033(0.092) | -0.014(0.092) | 0.014(0.078) | -0.286(0.078) | 0.206(0.078) |
| BH | 0.171(0.104) | 0.071(0.106) | -0.221(0.091) | -0.217(0.090) | -0.105(0.096) |
| CC | | **0.747(0.055)** | 0.255(0.093) | 0.127(0.095) | 0.197(0.096) |
| AC | | | 0.153(0.096) | 0.204(0.095) | 0.202(0.096) |
| CW | | | | **0.840(0.028)** | 0.015(0.086) |
| RW | | | | | -0.032(0.085) |
| CBC | | | | | | |

Note: 1 BL = body length, BH = body height, CC = chest circumference, AC = abdominal circumference, CBC = cannon bone circumference, RW = rump width, CW = chest width, SE of estimates are in parentheses
| Trait | Chromosome | Position (bp) | P_value  | SNP effect (%) | Gene         | Distance | Gene function                                      |
|-------|------------|---------------|----------|----------------|--------------|----------|---------------------------------------------------|
| BL    | 6          | 5671575       | 2.35E-07 | 0.00186        | CDH13        | +13217   | cadherin 13                                       |
|       | 1          | 6435744       | 1.5E-06  | 0.00017        | NA           | NA       | NA                                                |
|       | 1          | 6472959       | 1.5E-06  | 0.00080        | PRKN         | +38323   | parkin RBR E3 ubiquitin protein ligase            |
|       | 17         | 33632497      | 2.62E-06 | 0.00289        | ENSSSCG00000028 | -47405   | signal regulatory protein alpha                    |
|       | 13         | 25520933      | 4.45E-06 | 0.00004        | ULK4         | -8396    | unc-51 like kinase 4                              |
|       | 16         | 1276330       | 4.57E-06 | 0.00031        | NA           | NA       | NA                                                |
|       | 14         | 137476010     | 6.47E-06 | 0.00054        | NA           | NA       | NA                                                |
|       | 8          | 28933773      | 7.46E-06 | 0.00144        | NWD2         | -23316   | NACHT and WD repeat domain containing 2           |
|       | 13         | 166328893     | 8.39E-06 | 0.00039        | NA           | NA       | NA                                                |
| BH    | 16         | 886074        | 2.84E-06 | 0.00817        | CTNND2       | +28239   | alpha-2-macroglobulin like 1                       |
|       | 8          | 7942460       | 3.01E-06 | 0.00083        | NA           | NA       | NA                                                |
|       | 3          | 26586077      | 4.62E-06 | 0.00117        | CLEC19A      | -45911   | C-type lectin domain containing 19A               |
|       | 5          | 62690298      | 6.5E-06  | 0.00004        | A2ML1        | -42827   | alpha-2-macroglobulin like 1                       |
|       | 14         | 33580513      | 9.85E-06 | 0.00060        | HSPB8        | +45615   | heat shock protein family B (small) member 8      |
| CBC   | 4          | 117759672     | 2.16E-07 | 0.00279        | CDC14A       | -34935   | cell division cycle 14A                           |
|       | 13         | 182971424     | 1.83E-06 | 0.00420        | TMPRSS15     | -29625   | transmembrane serine protease 15                  |
|       | 17         | 12868538      | 1.85E-06 | 0.00635        | PSD3         | -43049   | pleckstrin and Sec7 domain containing 3           |
|       | 1          | 1201299       | 2.3E-06  | 0.00025        | ENSSSCG00000041      | -47914   | NA                                                |
|       | 1          | 1205821       | 2.3E-06  | 0.00018        | ENSSSCG00000050      | -42855   | NA                                                |
|       | 1          | 1220233       | 2.3E-06  | 0.00039        | ENSSSCG00000045      | -18409   | NA                                                |
|       | 1          | 1367723       | 2.3E-06  | 0.00064        | ENSSSCG00000043      | +5537    | NA                                                |
|       | 18         | 21663467      | 0.00003  | 0.00400        | GRM8         | -14659   | glutamate metabotropic receptor 8                 |
|       | 14         | 9698552       | 3.19E-06 | 0.00026        | ENSSSCG00000049      | 9436     | NA                                                |
|       | 5          | 7020488       | 3.46E-06 | 0.00023        | PMM1         | -49963   | phosphomannomutase 1                              |
|       | 12         | 50490164      | 4.08E-06 | 0.00233        | SPNS3        | -47230   | sphingolipid transporter 3 (putative)             |
|       | 3          | 12869355      | 4.1E-06  | 0.00259        | ENSSSCG00000036      | +18272   | NA                                                |
|       | 4          | 10221008      | 5.3E-06  | 0.00076        | ASAP1        | -37722   | ArfGAP with SH3 domain, ankyrin repeat and ...    |
| Trait | Chromosome | Position (bp) | P_value | SNP effect (%) | Gene | Distance | Gene function |
|-------|------------|---------------|---------|----------------|------|----------|---------------|
| AC    | 8          | 3249196       | 1.88E-06| 0.00134        | AFAP1| -46660   | actin filament associated protein 1 |
|       | 9          | 14578071      | 2.31E-06| 0.00128        | NA   | NA       | NA            |
|       | 14         | 13670622      | 2.71E-06| 0.00048        | PRSS55| -4581    | serine protease 55 |
|       | 5          | 15137502      | 2.96E-06| 0.00513        | RHEBL1| -39837   | RHEB like 1    |
|       | 4          | 5362087       | 4.48E-06| 0.00139        | ENSSSCG00000004493| +36176  | NA            |
|       | 7          | 26363076      | 5.4E-06 | 0.00093        | NA   | NA       | NA            |
|       | 14         | 43227411      | 5.77E-06| 0.00024        | ENSSSCG00000003338| -49062  | KIAA1671 ortholog |
|       | 16         | 522752        | 6.96E-06| 0.00014        | CTNND2| -3796    | catenin delta 2 |
| CC    | 3          | 63528527      | 1.32E-07| 0.00015        | ENSSSCG0000000825| -41861  | catenin alpha 2 |
|       | 6          | 19429624      | 3.27E-07| 0.00022        | Metazoa_SRP| -49801  | Metazoan signal recognition particle RNA |
|       | 1          | 3149903       | 7.78E-07| 0.00047        | PDE10A| -28771   | phosphodiesterase 10A |
|       | 6          | 120477523     | 1.95E-06| 0.00173        | FHOD3 | -40874   | formin homology 2 domain containing |
|       | 10         | 56219300      | 2.01E-06| 0.00203        | ITGB1 | -46293   | integrin subunit beta 1 |
|       | 17         | 18990746      | 2.63E-06| 0.00004        | ANKEF1| -32351   | ankyrin repeat and EF-hand domain containing 1 |
|       | 17         | 18997949      | 2.63E-06| 0.00008        | ANKEF1| -37701   | ankyrin repeat and EF-hand domain containing 1 |
|       | 2          | 122228151     | 3.03E-06| 0.00105        | ENSSSCG00000005134| -14167  | NA            |
|       | 2          | 122235537     | 3.03E-06| 0.00074        | ENSSSCG00000005134| -21553  | NA            |
|       | 16         | 33630686      | 3.58E-06| 0.00018        | NA   | NA       | NA            |
|       | 16         | 33638300      | 3.58E-06| 0.00079        | NA   | NA       | NA            |
|       | 16         | 5533970       | 4.64E-06| 0.00394        | ENSSSCG00000001679| +16579  | NA            |
|       | 12         | 5297390       | 5.41E-06| 0.00092        | RNF157| -48707   | ring finger protein 157 |
|       | 4          | 64552365      | 5.7E-06 | 0.00666        | ENSSSCG00000004202| -24706  | NA            |
|       | 10         | 43341283      | 7.22E-06| 0.00064        | CUBN  | -39687   | cubilin        |
|       | 8          | 21799389      | 8.84E-06| 0.00030        | ENSSSCG00000005098| -18261  | NA            |
| RW | 10 | 60737384 | 9.42E-06 | 0.00449 | ENSSCG0000001112 | -24200 | CUGBP Elav-like family member 2 |
| --- | --- | --------- | --------- | -------- | --------------- | ------ | --------------------------- |
| RW | 9  | 137165913 | 5.64E-07 | 0.00049 | NA             | NA    | NA                         |
| RW | 1  | 2330339  | 2.74E-06 | 0.00381 | SYT9           | -11533 | synaptotagmin 9           |
| RW | 1  | 38033383 | 4.19E-06 | 0.00149 | NKAIN2         | -12912 | sodium/potassium transporting ATPase 1 | |
| RW | 3  | 63682227 | 5.03E-06 | 0.00010 | NA             | NA    | NA                         |
| RW | 11 | 32555905 | 6.78E-06 | 0.00015 | DIAPH3         | +46764 | diaphanous related formin 3 |
| RW | 16 | 48600234 | 7.1E-06  | 0.00118 | ENSSCG0000004608 | -23005 | NA                         |
| CW | 8  | 132277288 | 8.17E-07 | 0.00009 | PTPN13         | -27877 | protein tyrosine phosphatase non-receptor type 13 |
| CW | 7  | 115471416 | 9.52E-07 | 0.00653 | MAPK10         | -27281 | mitogen-activated protein kinase 10 |
| CW | 14 | 37118119 | 1.09E-06 | 0.00289 | ENSSCG0000005178 | -18233 | protein phosphatase 4 regulatory subunit 11 |
| CW | 14 | 37165658 | 1.09E-06 | 0.00051 | ENSSCG0000005178 | -2275  | NA                         |
| CW | 14 | 37230969 | 1.09E-06 | 0.00039 | ENSSCG0000005178 | -49874 | NA                         |
| CW | 2  | 80016213 | 2.07E-06 | 0.00192 | COL23A1        | -9755  | NA                         |
| CW | 14 | 139878474 | 2.34E-06 | 0.00437 | TCERG1L        | -9755  | NA                         |
| CW | 14 | 49725382 | 2.67E-06 | 0.00112 | TRPV1          | -46102 | DEAD-box helicase 4 |
| CW | 16 | 35012960 | 3.46E-06 | 0.00029 | DDX4           | -31787 | proline rich membrane anchor 1 |
| CW | 16 | 115132809 | 4.65E-06 | 0.00069 | ENSSCG000000246 | -31787 | proline rich membrane anchor 1 |
| CW | 16 | 73800572 | 4.82E-06 | 0.00153 | U6             | +41611 | U6 spliceosomal RNA |
| CW | 16 | 73812833 | 4.82E-06 | 0.00172 | U6             | +29350 | U6 spliceosomal RNA |
| CW | 16 | 73816240 | 4.82E-06 | 0.00175 | U6             | +25343 | U6 spliceosomal RNA |
| CW | 9  | 107845695 | 5.74E-06 | 0.00014 | ENSSCG0000003290 | -7136  | NA                         |
| CW | 8  | 76456715 | 6.42E-06 | 0.00057 | ENSSCG0000004227 | -45304 | NA                         |
| CW | 3  | 131731345 | 6.59E-06 | 0.00001 | ENSSCG0000004975 | -23407 | NA                         |
| CW | 3  | 131738702 | 6.59E-06 | 0.00016 | ENSSCG0000004975 | -30764 | NA                         |
| Trait | 20 SNPs distributions of maximum effect | SNPs of the maximum effect | Top 20 SNPs effect (%) | Number of nearest gene | Candidate gene |
|-------|----------------------------------------|---------------------------|------------------------|------------------------|----------------|
| BL    | SSC17                                  | 17_7477978                | 0.117                  | 4                      | NA             |
| BH    | SSC2, SSC5, SSC16                       | 2_46827557                | 0.087                  | 8                      | SIL1           |
| CBC   | SSC7, SSC4                             | 7_55099416                | 0.101                  | 17                     | TRAPPC9        |
| AC    | SSC12                                  | 12_53181656               | 0.128                  | 8                      | KDM6B, CHD3    |
| CC    | SSC12                                  | 12_53169477               | 0.129                  | 8                      | KDM6B, CHD3    |
| RW    | SSC6, SSC7, SSC12, SSC13, SSC17        | 17_13175224               | 0.099                  | 15                     | MUC13          |
| CW    | SSC6, SSC7, SSC13, SSC17               | 6_39554872                | 0.070                  | 28                     | MUC13          |

Note: 1BL=body length, BH= body height, CC=chest circumference, AC=abdominal circumference, CBC= cannon bone circumference, RW= rump width, CW=chest width, gene effect= proportion of genetic variance explained
Additional file Table S1

Format: DOCX

Title: Each trait explains the 20 SNPs with the greatest genetic variance.