



**HAL**  
open science

## **Genome-wide association study and accuracy of genomic prediction for teat number in Duroc pigs using genotyping-by-sequencing**

Cheng Tan, Zhenfang Wu, Jiangli Ren, Zhuolin Huang, Dewu Liu, Xiaoyan He, Dzianis Prakapenka, Ran Zhang, Ning Li, Yang Da, et al.

### ► To cite this version:

Cheng Tan, Zhenfang Wu, Jiangli Ren, Zhuolin Huang, Dewu Liu, et al.. Genome-wide association study and accuracy of genomic prediction for teat number in Duroc pigs using genotyping-by-sequencing. *Genetics Selection Evolution*, 2017, 49 (1), pp.35. <10.1186/s12711-017-0311-8>. <hal-01499291>

**HAL Id: hal-01499291**

**<https://hal.science/hal-01499291v1>**

Submitted on 31 Mar 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

RESEARCH ARTICLE

Open Access



# Genome-wide association study and accuracy of genomic prediction for teat number in Duroc pigs using genotyping-by-sequencing

Cheng Tan<sup>1,2</sup>, Zhenfang Wu<sup>3</sup>, Jiangli Ren<sup>1</sup>, Zhuolin Huang<sup>1</sup>, Dewu Liu<sup>3</sup>, Xiaoyan He<sup>3</sup>, Dzianis Prakapenka<sup>2</sup>, Ran Zhang<sup>1</sup>, Ning Li<sup>1</sup>, Yang Da<sup>2\*</sup> and Xiaoxiang Hu<sup>1\*</sup>

## Abstract

**Background:** The number of teats in pigs is related to a sow's ability to rear piglets to weaning age. Several studies have identified genes and genomic regions that affect teat number in swine but few common results were reported. The objective of this study was to identify genetic factors that affect teat number in pigs, evaluate the accuracy of genomic prediction, and evaluate the contribution of significant genes and genomic regions to genomic broad-sense heritability and prediction accuracy using 41,108 autosomal single nucleotide polymorphisms (SNPs) from genotyping-by-sequencing on 2936 Duroc boars.

**Results:** Narrow-sense heritability and dominance heritability of teat number estimated by genomic restricted maximum likelihood were  $0.365 \pm 0.030$  and  $0.035 \pm 0.019$ , respectively. The accuracy of genomic predictions, calculated as the average correlation between the genomic best linear unbiased prediction and phenotype in a tenfold validation study, was  $0.437 \pm 0.064$  for the model with additive and dominance effects and  $0.435 \pm 0.064$  for the model with additive effects only. Genome-wide association studies (GWAS) using three methods of analysis identified 85 significant SNP effects for teat number on chromosomes 1, 6, 7, 10, 11, 12 and 14. The region between 102.9 and 106.0 Mb on chromosome 7, which was reported in several studies, had the most significant SNP effects in or near the *PTGR2*, *FAM161B*, *LIN52*, *VRTN*, *FCF1*, *AREL1* and *LRRC74A* genes. This region accounted for 10.0% of the genomic additive heritability and 8.0% of the accuracy of prediction. The second most significant chromosome region not reported by previous GWAS was the region between 77.7 and 79.7 Mb on chromosome 11, where SNPs in the *FGF14* gene had the most significant effect and accounted for 5.1% of the genomic additive heritability and 5.2% of the accuracy of prediction. The 85 significant SNPs accounted for 28.5 to 28.8% of the genomic additive heritability and 35.8 to 36.8% of the accuracy of prediction.

**Conclusions:** The three methods used for the GWAS identified 85 significant SNPs with additive effects on teat number, including SNPs in a previously reported chromosomal region and SNPs in novel chromosomal regions. Most significant SNPs with larger estimated effects also had larger contributions to the total genomic heritability and accuracy of prediction than other SNPs.

\*Correspondence: yda@umn.edu; huxx@cau.edu.cn

<sup>1</sup> State Key Laboratory for Agrobiotechnology, China Agricultural University, Beijing 100193, China

<sup>2</sup> Department of Animal Science, University of Minnesota, Saint Paul, MN 55108, USA

Full list of author information is available at the end of the article

## Background

A sufficient number of teats is necessary for a sow to rear its piglets to weaning age. Many putative QTL (quantitative trait loci) for teat number have been reported on most of the porcine chromosomes, but most of these were detected using microsatellite markers and lacked specific gene targets [1]. Genome-wide association studies (GWAS) using single nucleotide polymorphisms (SNPs) and analyses of candidate genes have identified several specific gene targets that affect teat number in swine. A GWAS using 42,654 SNPs on 936 Large White pigs reported 39 QTL with 211 significant SNP effects on teat number [2]. Among those SNP effects, the region between 102.0 and 105.2 Mb on chromosome 7 had the most significant effects and the percentage of the genetic variance explained by SNPs in this region ranged from 0.04 to 2.51%. Within this region, the *VRTN* and *PROX2* genes were identified as the most convincing candidate genes. The chromosomal locations of the significant SNPs that were detected in this GWAS differed from all previously reported QTL for teat number that have been compiled in the animal QTL database [1]. Another GWAS, using 32,911 SNPs on 1550 Large White pigs, reported 21 QTL with additive effects on chromosomes 6, 7 and 12, one QTL with a dominant effect on chromosome 4, and identified *VRTN* as the most promising candidate gene for teat number [3]. A third GWAS using 41,647 SNPs on 1657 Large White pigs found 65 significant SNPs on chromosomes 1, 2, 7, 8, 12 and 14, including SNPs in the region 102.9 between 105.2 Mb on chromosome 7 [4]. A fourth GWAS using 39,778 SNPs identified the *VRTN* gene with pleiotropic and desirable effects on thoracic vertebral number, teat number and carcass (body) length across four pig populations, and showed that, of all SNPs on chromosome 7, a SNP within the *VRTN* gene had the most significant effect on teat number in Duroc pigs [5]. Among all significant SNPs that have been detected for teat number by GWAS, the significance of the *VRTN* gene on chromosome 7 achieved the widest consensus and has been identified as a strong candidate gene for teat number [2–5]. However, in the literature some discrepancies regarding the most significant location and many SNP effects in other genomic regions have been reported. A GWAS using the porcine 60 K SNP chip on a F2 population from a cross between Landrace and Korean pigs identified highly significant SNPs on chromosome 7 that were more than 40 Mb away from the *VRTN* gene [6], and in another GWAS using 36,588 SNPs and 1024 Duroc pigs, the most significant SNPs on chromosome 7 were found 2 to 3 Mb downstream of the *VRTN* region [7]. However, other than for the *VRTN* region, there is little consensus among the GWAS results on genomic regions that affect teat number [2–4, 6, 7]. Therefore,

additional studies are needed to identify the genetic factors that affect swine teat number. Furthermore, it is unclear what the impact of the highly significant SNPs is on the accuracy of genomic prediction for teat number.

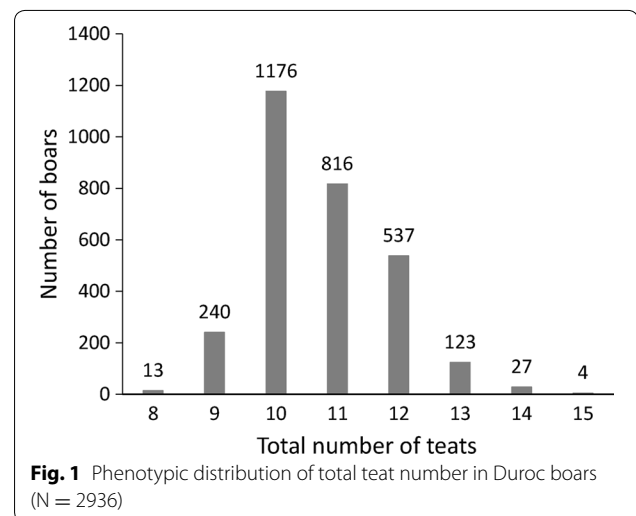
The objective of this study was to identify genetic factors that affect teat number in pigs, evaluate the accuracy of genomic prediction, and evaluate the contribution of significant genes and genomic regions to the heritability and accuracy of genomic prediction using 41,108 autosomal SNPs from genotyping-by-sequencing (GBS) on 2936 Duroc boars.

## Methods

### Animals, phenotyping, and genotyping-by-sequencing

Animal and phenotype data used for this study were provided by Guangdong Wen's Foodstuff Group (Guangdong, China). The study population included 2936 Duroc boars born from September 2011 to September 2013 in 1456 litters from 79 sires with one to three piglets per litter, and all pigs were managed at a single nucleus farm. The left and right teats were counted separately within 48 h after birth and only normal teats were recorded. In this study, the phenotype used for 'teat number' was the total number of teats that was equal to the sum of the left and right normal teats. The 'mean  $\pm$  (standard deviation)' of teat number was  $10.72 \pm 1.72$ . The phenotypic values followed a near bell-shaped distribution (Fig. 1), which was similar to the leptokurtic distribution of teat number that is observed for Landrace and Large White pigs [8], with most animals (1992 out of 2936) having 10 or 11 teats.

Genomic DNA was extracted from ear tissue of all 2936 Duroc boars and quantified using a Qubit 2.0 Fluorometer. DNA concentrations were normalized to 50 ng/ml in 96-well plates. A two-enzyme i.e. *EcoRI* and *MspI*



genotyping-by-sequencing (GBS) was used. A set of 96 forward barcoded adapters with an *EcoRI* overhang were designed by the GBS Barcode Generator (<http://www.deenabio.com/services/gbs-adapters>), and the reverse adapter with a *MspI* overhang was designed according to [9]. DNA samples (150 ng each) were digested with *EcoRI* and *MspI*, then ligated to the designed adapters. Following adapter ligation, samples were pooled in 96-plex and size-selected using two cycles of purification with Agencourt AMPure XP Beads (Beckman Coulter, Pasadena, CA). The purified libraries were amplified by PCR and then sequenced on Illumina NextSeq500 by a 90-bp single-end sequencing. SNP genotypes were called according to the pipeline implemented in Tassel 5.0 with default parameters [10], and Beagle 4.0 was used to impute missing SNP genotypes. A total of 90,051 SNPs were identified for the population used in this study. SNP filtering was based on the following criteria: only SNPs that had a minor allele frequency higher than 5%, for which the frequency of the least frequent homozygous genotype was at least 0.01, and passed the Hardy–Weinberg equilibrium test ( $p \geq 10^{-6}$ ) were retained. Among the autosomal SNPs, 41,108 SNPs satisfied these requirements and were used for analyses.

#### GWAS analysis

Three single-SNP methods were used for the GWAS analysis: a *t* test of additive and dominance SNP effects using a generalized least squares (GLS) analysis that takes intraclass correlation of sibs into account and is implemented in the EPISNP2 program [11, 12], and the least squares (LS) analyses of additive effects by PLINK [13] and EPISNP1 [11] with population stratification correction using the first 50 dimensions from multidimensional scaling (MDS) as covariates. We report the PLINK and EPISNP1 results using the first 35 MDS dimensions for stratification correction because the genomic inflation factor [14] and the patterns of Manhattan plots of SNP significance stabilized when fitting the first 35 MDS dimensions (Fig. 2).

The statistical model for the EPISNP2 analysis was:

$$\mathbf{y} = \mathbf{X}_b \mathbf{b} + \mathbf{X}_g \mathbf{g} + \mathbf{Z} \mathbf{f} + \mathbf{e},$$

where  $\mathbf{y}$  is the vector of phenotypic values,  $\mathbf{b}$  is the vector of fixed year-month effects,  $\mathbf{X}_b$  is the incidence matrix for  $\mathbf{b}$ ,  $\mathbf{g}$  is the vector of the effects of SNP genotypes,  $\mathbf{X}$  is the incidence matrix of  $\mathbf{g}$ ,  $\mathbf{f}$  is the vector of random family effects with a common variance  $\sigma_f^2$  for sibs in the same family, and  $\mathbf{Z}$  is the incidence matrix of  $\mathbf{f}$ . The variance–covariance matrix of the family effects was assumed to be  $\mathbf{G} = \text{Var}(\mathbf{f}) = \mathbf{I}\sigma_f^2$ , where  $\mathbf{I}$  is an identity matrix, and the phenotypic variance–covariance matrix is  $\text{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{I}\sigma_e^2$  [12].

The statistical model for the PLINK analysis was:

$$\mathbf{y} = \mathbf{X}_b \mathbf{b} + \mathbf{X}_1 \mathbf{b}_1 + \alpha \mathbf{x} + \mathbf{e},$$

where  $\mathbf{b}_1$  is the vector of fixed effect(s) of the MDS dimension(s),  $\mathbf{X}_1$  is the matrix of the MDS dimension(s) as calculated by PLINK from the SNP matrix of identity-by-state [13],  $\alpha$  is the additive SNP effect, and  $\mathbf{x}$  is a column vector of genotype codes for  $\alpha$  created by PLINK.

The statistical model for EPISNP1 analysis was:

$$\mathbf{y} = \mathbf{X}_b \mathbf{b} + \mathbf{X}_1 \mathbf{b}_1 + \mathbf{X}_g \mathbf{g} + \mathbf{e},$$

where the matrices have the same definitions as in the previous two models. Significance tests for additive and dominance SNP effects by EPISNP1 and EPISNP2 were implemented by *t* tests for the additive and dominance contrasts of the estimated SNP genotypic values [11, 12, 15].

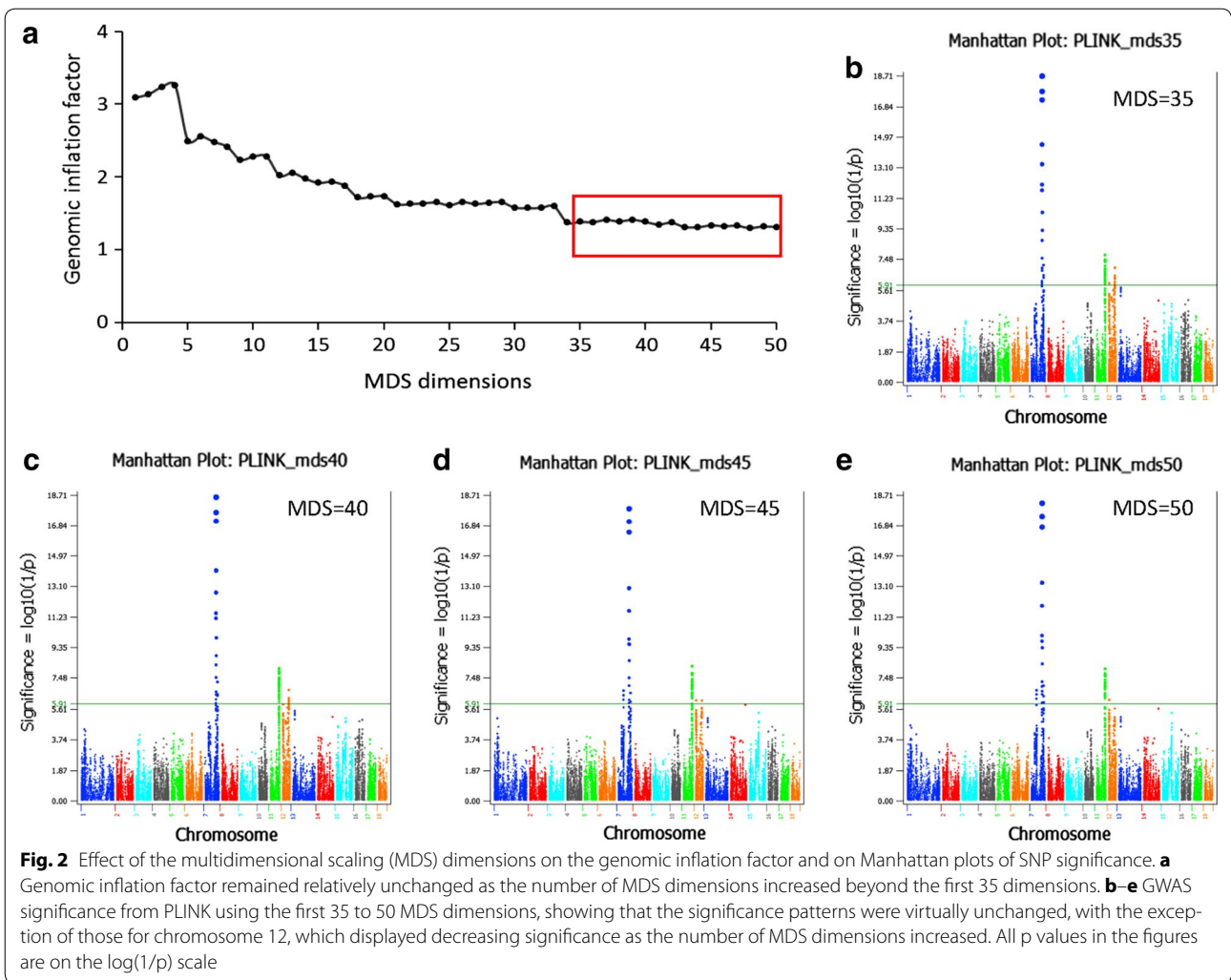
#### Genomic heritability and accuracy of genomic prediction

Genomic heritability and genomic prediction were estimated by using a mixed model with additive and dominance effects as described previously [16–18]. Briefly, the mixed model for heritability estimation and genomic prediction was:

$$\mathbf{y} = \mathbf{X}_b \mathbf{b} + \mathbf{Z} \mathbf{a} + \mathbf{Z} \mathbf{d} + \mathbf{e},$$

with  $\text{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{A}_g\mathbf{Z}'\sigma_\alpha^2 + \mathbf{Z}\mathbf{D}_g\mathbf{Z}'\sigma_\delta^2 + \mathbf{I}\sigma_e^2$ , where  $\mathbf{Z}$  is an incidence matrix allocating phenotypic observations to each individual,  $\mathbf{a}$  is the vector of genomic additive (breeding) values,  $\mathbf{d}$  is the vector of genomic dominance values or dominance deviations,  $\mathbf{A}_g$  is a genomic additive relationship matrix calculated from the SNPs,  $\mathbf{D}_g$  is a genomic dominance relationship matrix calculated from the SNPs,  $\sigma_\alpha^2$  is the additive variance,  $\sigma_\delta^2$  is the dominance variance, and  $\sigma_e^2$  is the residual variance. The  $\mathbf{A}_g$  and  $\mathbf{D}_g$  matrices were calculated using Definition II of genomic relationships implemented by the GVCBLUP package, and variance components of additive, dominance and random residual values were estimated by genomic restricted maximum likelihood estimation (GREML) using the GREML\_CE program in the GVCBLUP package [18]. The genomic heritability was defined as:  $h_\alpha^2 = \sigma_\alpha^2/\sigma_y^2$ , i.e. the narrow-sense heritability,  $h_\delta^2 = \sigma_\delta^2/\sigma_y^2$ , i.e. the dominance heritability, and  $h_t^2 = h_\alpha^2/h_\delta^2$ , i.e. the broad-sense heritability, where  $\sigma_y^2 = \sigma_\alpha^2 + \sigma_\delta^2 + \sigma_e^2$  is the phenotypic variance. The genomic best linear unbiased prediction (GBLUP) of additive, dominance and genetic values of individuals in the training and validation samples were calculated at the last iteration of the GREML.

A tenfold validation study was conducted to evaluate the prediction accuracy. The 2936 Duroc boars were randomly divided into 10 validation datasets of 293 individuals except the 10th sample, which included



299 individuals. For each of the 10 validation analyses, phenotypic observations in the validation dataset were omitted in the GBLUP calculation. Three measures of prediction accuracy were calculated and compared:  $R_{0jp} = \text{corr}(\hat{g}_{0j}, y_0)$ , which is the observed accuracy of predicting the phenotypic values in the validation population and is calculated as the correlation between the estimated genetic values ( $\hat{g}_{0j}$ ) and the phenotypic observations ( $y_0$ ) of validation individuals, averaged across all validation datasets;  $R_{0j} = \text{corr}(\hat{g}_{0j}, g_{0j})$ , which is the expected accuracy of predicting the true genetic values ( $g_{0j}$ ) of individuals in the validation population and is calculated as the square root of the reliability estimate for each individual from the GVCBLUP package [18], where ‘j =  $\alpha$ ’ indicates additive prediction, ‘j =  $\delta$ ’ indicates dominance prediction, and ‘j = t’ indicates prediction of total genetic value; and  $R_{0jp} = R_{0j} \sqrt{h_j^2}$ , which is the expected accuracy of predicting the phenotypic values, where  $h_j^2$  is the genomic narrow-sense

(j =  $\alpha$ ), dominance (j =  $\delta$ ), or broad-sense (j = t) heritability. The accuracy of predicting phenotypic values was previously termed as ‘predictive ability’ [19] to distinguish it from ‘expected prediction accuracy’ of predicting genetic values ( $R_{0j}$  in this study). The formula of the expected accuracy of predicting phenotypic values,  $R_{0jp} = R_{0j} \sqrt{h_j^2}$ , is a slightly different form of the relationship between ‘predictive ability’ and ‘prediction accuracy’ [19]. The mathematical difference between  $R_{0jp}$  and  $R_{0j}$  is in the denominators of these two measures: the denominator of  $R_{0jp}$  is the phenotypic standard deviation, whereas the denominator of  $R_{0j}$  is the genetic standard deviation, which is necessarily smaller than the phenotypic standard deviation in the presence of non-zero residual variance. Therefore,  $R_{0j}$  is the upper limit of  $R_{0jp}$  but this upper limit may not hold for the observed accuracy of predicting phenotypic values ( $\hat{R}_{0jp}$ ) due to unknown variations in the data or genetic mechanisms that are not explained by the statistical model. The

observed accuracy of predicting genetic values can only be defined when the true genetic values such as simulated genetic values are known [16] but they could not be defined in this study because the true genetic values were unknown.

**Contribution of significant SNPs and genomic regions to total heritability and prediction accuracy**

The contributions of each SNP to additive, dominance and total heritability can be estimated [18]. However, as shown by the results of this study, the contribution of each SNP is affected by the number of SNPs in the mixed model: the larger is the number of SNPs, the smaller is the contribution of each SNP. To avoid this dependency in the mixed model for GREML and to estimate each SNP's independent contribution to genomic heritability and prediction accuracy, we used the approach of 'partial heritability' and 'partial accuracy' based on differences in heritability and prediction accuracy between a full model and a reduced model. The full model fits all SNPs as random effects and the reduced model fits the target SNP or SNPs as fixed effects to remove their effects from the phenotypic values. The reduced model was as follows:

$$y = X_f b + X_s s + Z a + Z d + e,$$

where **s** is a column vector of fixed SNP effects and **X** is the incidence matrix of **s**. The phenotypic variance-covariance matrix was assumed to be the same as in the full model, i.e.,  $Var(y) = V = Z A_g Z' \sigma_a^2 + Z D_g Z' \sigma_d^2 + I \sigma_e^2$ , but variance components were estimated under the reduced model. Let  $\hat{h}_i^2$  ( $\hat{h}_i$ ) be the estimated heritability from the full (reduced) model, and  $R_0$  ( $R_{0i}$ ) a measure of prediction accuracy for the full (reduced) model. Then, the relative contribution of the *i*th SNP or the *i*th set of SNPs to the total heritability was calculated as  $c_{hi}^2 = 1 - \hat{h}_i / \hat{h}$ , and

the relative contribution of the *i*th SNP or the *i*th set of SNPs to the prediction accuracy as  $c_{ri} = 1 - R_{0i} / R_0$ .

**Results**

**Genomic heritability and prediction accuracy**

The estimate of genomic narrow-sense heritability ( $\hat{h}_a^2$ ) was  $0.365 \pm 0.030$ , of dominance heritability ( $\hat{h}_d^2$ ) was  $0.035 \pm 0.019$ , of broad-sense heritability ( $\hat{h}_t^2$ ) was  $0.400 \pm 0.034$ , and the estimate of narrow-sense heritability for the mixed model with additive effects only, was  $0.368 \pm 0.030$ , which is slightly higher than the corresponding estimate for the mixed model with additive and dominance effects (Table 1). The observed accuracy of predicting phenotypic values from the tenfold validation study was  $0.437 \pm 0.064$  for the mixed model with additive and dominance SNP effects ( $\hat{R}_{0tp}$ , Model 1A in Table 2), and was  $0.435 \pm 0.064$  for the mixed model with additive effects only ( $\hat{R}_{0ap}$ , Model 2A in Table 2), which is only 0.46% lower than that from the mixed model with additive and dominance effects. These slight differences in both heritability and accuracy of prediction between the additive model and the model with additive and dominance effects indicates that additive SNP effects were the primary genetic effects that affect teat number and that dominance SNP effects only had a negligible contribution to the prediction accuracy for teat number.

**GWAS results**

The GWAS that was done with EPISNP2, which accounted for the sib intraclass correlation and was implemented by a GLS analysis [11, 12], identified 73 SNPs on chromosomes 1, 6, 7, 10, 11, 12 and 14 with additive effects but no SNP with dominance effects reached genome-wide significance with the Bonferroni multiple testing correction ( $p < 10^{-5.91}$ ) (Fig. 3a, b; Table 3; Additional file 1: Table S1). LS analysis of PLINK [13] and EPISNP1 [11] with stratification correction using the first

**Table 1 Estimates of genomic heritabilities for teat number using 41,108 autosomal SNPs on 2936 Duroc boars**

Model	All SNPs as random effects	85 significant SNPs removed	85 significant SNPs as fixed effects
Additive and dominance effects	$\hat{h}_a^2 = 0.365 \pm 0.030$	$\hat{h}_a^2 = 0.346 \pm 0.030$ $-c_{hi}^2 = -5.20\%$	$\hat{h}_a^2 = 0.260 \pm 0.030$ $-c_{hi}^2 = -28.77\%$
	$\hat{h}_d^2 = 0.035 \pm 0.019$	$\hat{h}_d^2 = 0.036 \pm 0.020$ $-c_{hi}^2 = +2.86\%$	$\hat{h}_d^2 = 0.037 \pm 0.022$ $-c_{hi}^2 = +5.71\%$
	$\hat{h}_t^2 = 0.400 \pm 0.034$	$\hat{h}_t^2 = 0.382 \pm 0.034$ $-c_{hi}^2 = -4.50\%$	$\hat{h}_t^2 = 0.297 \pm 0.036$ $-c_{hi}^2 = -25.75\%$
Additive effects only	$\hat{h}_a^2 = 0.368 \pm 0.030$	$\hat{h}_a^2 = 0.350 \pm 0.030$ $-c_{hi}^2 = -4.89\%$	$\hat{h}_a^2 = 0.263 \pm 0.030$ $-c_{hi}^2 = -28.53\%$

$\hat{h}_a^2$  = narrow-sense heritability,  $\hat{h}_d^2$  = dominance heritability,  $\hat{h}_t^2$  = broad-sense heritability =  $\hat{h}_a^2 + \hat{h}_d^2$ ,  $-c_{hi}^2$  = decrease in heritability relative to the heritability estimated by using all SNPs fitted as random effects

**Table 2 Accuracies of genomic prediction for the phenotypic values and true genetic values of teat number using 41,108 autosomal SNPs on 2936 Duroc boars in a tenfold validation study**

Model and accuracy change	$\widehat{R}_{0tp} = \text{corr}(\widehat{g}_{0j}, y_0)$	$R_{0jp} = R_{0j}\sqrt{h_j^2}$	$R_{0j} = \text{corr}(\widehat{g}_{0j}, g_{0j})$
Model 1A	$\widehat{R}_{0tp} = 0.437 \pm 0.064$	$R_{0tp} = 0.460$	$R_{0t} = 0.728 \pm 0.004$
Model 1B	$\widehat{R}_{0tp} = 0.279 \pm 0.076$	$R_{0tp} = 0.360$	$R_{0t} = 0.661 \pm 0.007$
– $c_{ri}$ of 1B relative to 1A	–36.16%	–21.74%	–9.20%
Model 2A	$\widehat{R}_{0\alpha p} = 0.435 \pm 0.064$	$R_{0\alpha p} = 0.425$	$R_{0\alpha} = 0.700 \pm 0.007$
Model 2B	$\widehat{R}_{0\alpha p} = 0.275 \pm 0.074$	$R_{0\alpha p} = 0.320$	$R_{0\alpha} = 0.624 \pm 0.009$
– $c_{ri}$ of 2B relative to 2A	–36.78%	–24.70%	–10.86%
Model 3A	$\widehat{R}_{0tp} = 0.426 \pm 0.066$	$R_{0tp} = 0.446$	$R_{0t} = 0.721 \pm 0.004$
– $c_{ri}$ of 3A relative to 1A	–2.52%	–3.04%	–0.96%
Model 4A	$\widehat{R}_{0\alpha p} = 0.424 \pm 0.066$	$R_{0\alpha p} = 0.409$	$R_{0\alpha} = 0.691 \pm 0.007$
– $c_{ri}$ of 4A relative to 2A	–2.53%	–3.76%	–1.28%

Model 1A has additive and dominance effects and uses all 41,108 autosome SNPs. Model 1B is a modification of Model 1A by using the 85 significant SNPs as fixed non-genetic effects. Model 2A has additive effects only and uses all 41,108 autosome SNPs. Model 2B is a modification of Model 2A by using the 85 significant SNPs as fixed non-genetic effects. Model 3A has additive and dominance effects and uses 41,023 autosomal SNPs after removing the 85 significant SNPs. Model 4A has additive effects only and uses 41,023 autosomal SNPs after removing the 85 significant SNPs.  $\widehat{R}_{0jp}$  is the observed accuracy of predicting phenotypic values from tenfold validations.  $R_{0jp}$  is the expected accuracy of predicting phenotypic values.  $R_{0j}$  is the expected accuracy of predicting genetic values calculated by GVCBLUP from tenfold validations,  $j = t$  or  $\alpha$ .  $h_t^2 = 0.400$  for Model 1A,  $= 0.297$  for Model 1B,  $= 0.382$  for Model 3.  $h_{\alpha}^2 = 0.368$  for Model 2A,  $0.263$  for Model 2B,  $= 0.350$  for Model 4. – $c_{ri}$  is the decrease in accuracy

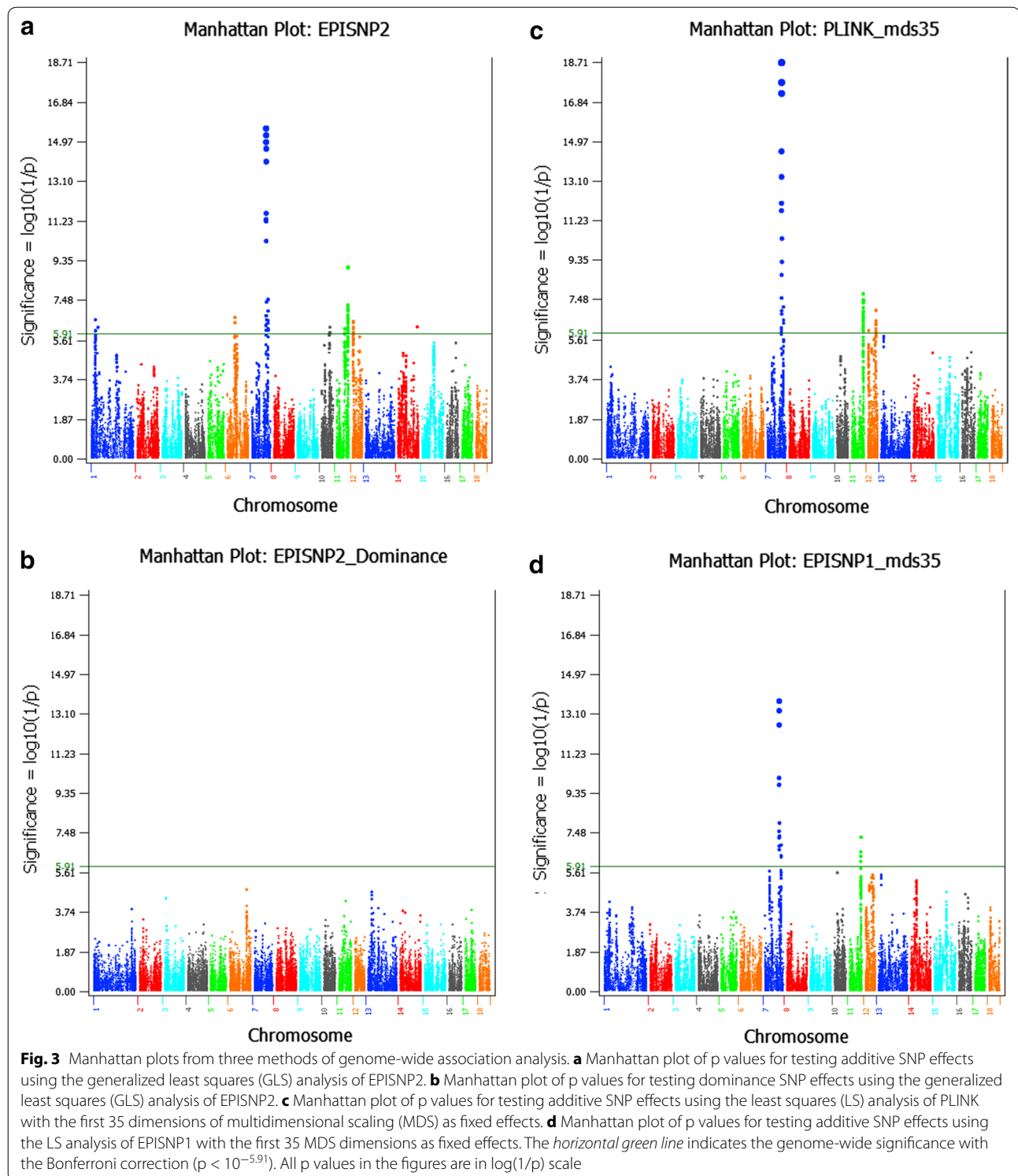
35 dimensions of MDS as fixed covariates, identified 54 and 21 significant SNPs, respectively (Fig. 3c, d; Additional file 1: Table S1). Twelve SNPs detected by PLINK and two SNPs detected by EPISNP1 did not overlap with the SNPs detected by EPISNP2. Eighteen SNPs detected by EPISNP1 overlapped with those detected by EPISNP2 and PLINK. For this dataset, EPISNP1 was the most conservative for declaring significance. We report SNPs detected by EPISNP2 because they all had a substantial contribution to the broad-sense genomic heritability (see Additional file 1: Table S1). A graphical view of the GWAS results obtained by EPISNP2, PLINK and EPISNP1 for all autosomes is in Additional file 2: Figure S1.

To evaluate the impact of the significant SNPs on the phenotypic variance, we estimated the decreases in observed genomic narrow-sense heritability and prediction accuracy when the phenotypic values were adjusted for the estimated genotypic values of the significant SNPs. The results showed that the 85 significant SNPs identified by the three methods, i.e. EPISNP2, PLINK and EPISNP1, accounted for 28.5 to 28.8% of the genomic narrow-sense heritability (Table 1) and for 36.2 to 36.8% of the observed prediction accuracy (Model 1A and Model 2A in Table 2). These results show that many SNPs that were deemed insignificant by the GWAS analysis were relevant for genomic prediction of teat number. Each of the 85 SNPs had a relatively large contribution to the genomic heritability and prediction accuracy, with the contribution of each SNP to the observed genomic narrow-sense

heritability ranging from 0.7 to 7.3% and relative contribution of each SNP to the observed prediction accuracy ranging from 0.5 to 5.6% (see Additional file 1: Table S1).

#### Analysis of the region between 102.9 and 106.0 Mb on chromosome 7

A cluster of 14 SNPs within or near the *PTGR2*, *FAM161B*, *LIN52*, *VRTN*, *FCF1*, *AREL1* and *LRRC74A* genes in the region between 102.9 and 106.0 Mb on chromosome 7 had the most significant effects on teat number with genome-wide significance (Fig. 4a). Based on the GLS analysis of EPISNP2, the two SNPs upstream of *PTGR2* had the most significant additive effects, followed by the three SNPs within and upstream of *AREL1*, whereas the LS analysis of PLINK and EPISNP1 with stratification correction ranked the three *AREL1* SNPs as the most significant and the two SNPs upstream of *PTGR2* in the 6th and 7th positions (see Additional file 1: Table S1). The six most significant SNPs in the region between 102.9 and 103.8 Mb on chromosome 7 accounted for 7.4% of the genomic additive heritability and 7.0% of the observed prediction accuracy in the tenfold validation study (Table 3), and all the 14 SNPs in this region with genome-wide significance accounted for 10.0% of the genomic narrow-sense heritability and 8.0% of the observed prediction accuracy. Removal of the genotypic effects of the 14 SNPs by fitting these SNPs as fixed effects in the model for EPISNP2 removed all significant effects in the region between 102.9 and 103.8 Mb on chromosome 7 and also removed the significant effects of seven SNPs



in the region between 116.1 and 117.4 Mb on chromosome 7 (Fig. 4b). SNPs within or near the *AREL1*, *PTGR2*, *FMA161B*, *LIN52* and *LRRC74A* genes also had the largest contributions to the genomic narrow-sense heritability

and prediction accuracy (Fig. 4c). We did not detect SNPs within the *VRTN* gene but a SNP upstream of and nearest to *VRTN* (*S7\_103355294*) was highly significant, ranking 6th based on EPISNP2 and 5th based on PLINK and

**Table 3** Chromosome regions with significant SNP effects on teat number

Chr	Region (Mb)	Size (Mb)	Most significant SNP				Contribution % of $\hat{R}_0$	Gene region
			Name	MAF	p value	% of $h_a^2$		
1	29.63–30.18	0.55	S1_29635241	0.499	2.68 ( $10^{-07}$ )	2.25	1.51	<i>TNFAIP3, OLIG3</i>
7	102.91–103.80	0.89	S7_102911357	0.438	2.49 ( $10^{-16}$ )	7.35	6.98	<i>PTGR2</i> (U), <i>FAM161B, LIN52, VRTN</i> (U, D), <i>FCF1, AREL1</i> (16 genes)
7	116.07–117.43	1.36	S7_116899295	0.342	2.97 ( $10^{-08}$ )	3.07	2.13	<i>GALC, KCNK10, SPATA7, PTPN21, ZC3H14, EML5, TTC8</i>
11	56.56–58.58*	2.21	S11_58558301	0.422	7.14 ( $10^{-7}$ )	1.64	2.27	<i>SPRY2, SNORA70</i>
11	77.75–79.69	1.94	S11_79009219	0.226	9.16 ( $10^{-10}$ )	5.07	5.23	<i>FGF14, BIVM</i> (U, D), <i>LOC102167592, LOC102167785</i> (U)
12	4.53–6.26*	1.73	S12_5615207	0.335	3.16 ( $10^{-07}$ )	2.59	2.63	<i>MFSD11-KCTD2</i> (49 genes)
12	50.54–51.74	1.20	S12_51574540	0.138	1.06 ( $10^{-07}$ )	0.64	1.16	<i>OR3A2, ASPA, TRPV3, TRPV1, CTNS, TAX1BP3, EMC6, CAMKK1</i>

Chr = chromosome; MAF = minor allele frequency.  $h_a^2$  is the additive heritability.  $\hat{R}_0$  is the observed accuracy of prediction. U indicate the significant SNP is located upstream of the gene. D indicates the significant SNP is located downstream of the gene. Contribution was calculated for the six most significant SNPs in each region. \* This region has three significant SNPs

EPISNP1. Linkage disequilibrium (LD) analysis using Haploview [20] showed that the two SNPs flanking *VRTN* were in strong LD ( $D' = 0.90$ , Fig. 4d), implying that either of these two SNPs could also be in strong LD with *VRTN*. Therefore, assuming that *VRTN* is a causal gene, the significant effect of S7\_103355294 could be a linked effect of *VRTN*. The LD analysis showed that the significant effects of the region between 116.1 and 117.4 Mb on chromosome 7 could also be due to LD with the region between 102.9 and 106.0 Mb since three of the seven SNPs in the former region were in low LD with five significant SNPs in the latter region ( $D' = 0.13$  to  $0.24$ , Fig. 4d). This did not consider the possibility of multilocus LD between the two regions. The low LD was the only known reason that could explain the disappearance of the significant QTL effects of the region between 116.1 and 117.4 Mb when the 14 SNPs in the region between 102.9 and 106.0 Mb were fitted as fixed effects.

#### Other chromosomes with significant SNPs

In addition to chromosome 7, SNPs on chromosomes 1, 6, 10, 11, 12 and 14 also had significant effects with genome-wide significance (Table 3; Additional file 1: Table S1). Among these chromosomes, the region between 77.3 and 79.7 Mb on chromosome 11 had the most significant additive effects. The top six SNPs within or near the *FGF14*, *BIVM*, *LOC10216759*, and *LOC102167785* genes accounted for 5.1% of the genomic narrow-sense heritability and 5.2% of the observed prediction accuracy (Table 3). The remaining regions on chromosomes 1, 7, 11 and 12 each with at least six SNPs accounted for 0.6 to 3.1% of the genomic narrow-sense heritability and accounted for 1.2 to 2.6% of the observed prediction accuracy (Table 3).

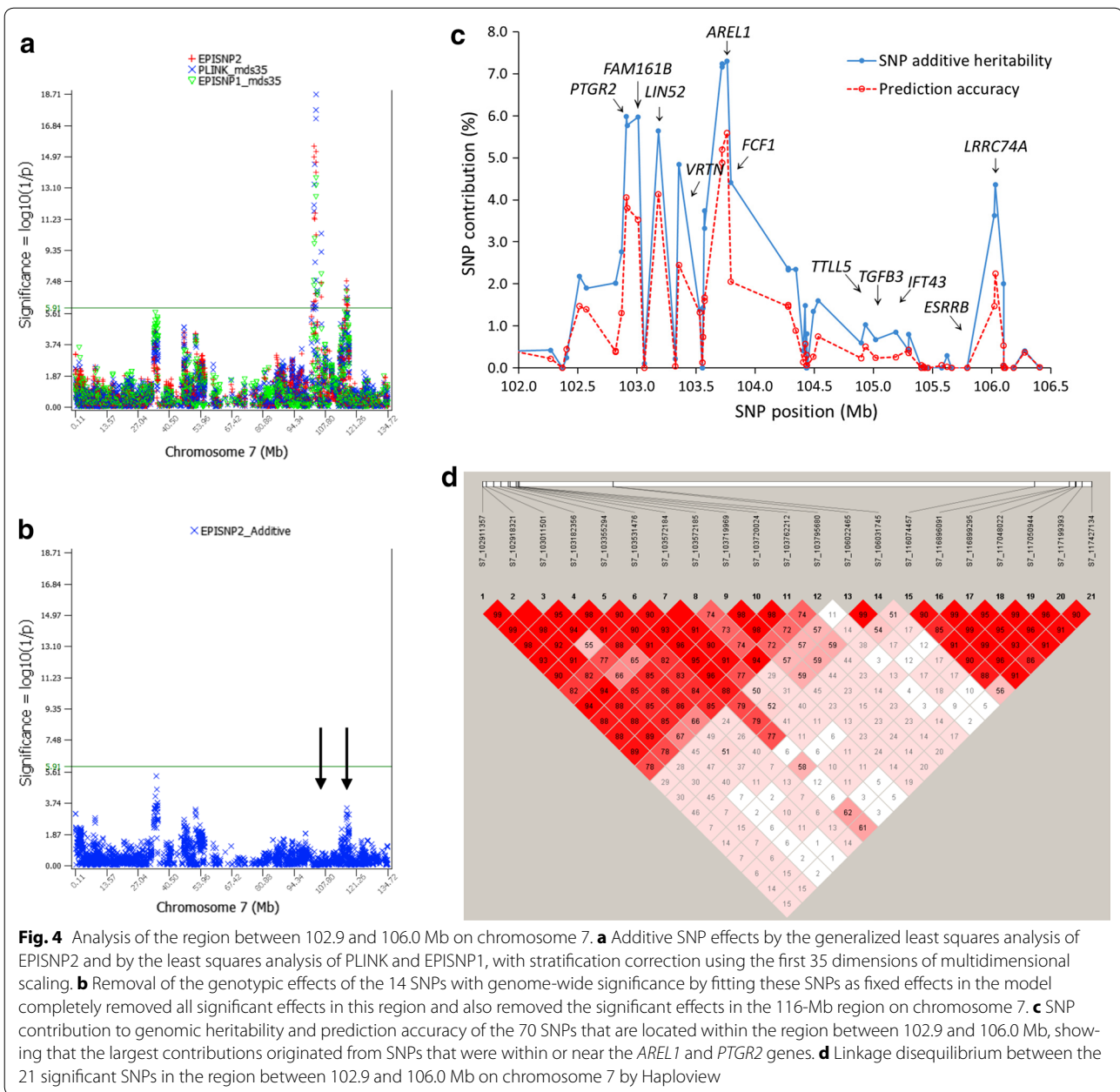
## Discussion

### Comparison with previous GWAS results

The region between 102.9 and 106.0 Mb on chromosome 7 that was identified in the current study was also reported in several previous GWAS but with varying lengths, e.g., between 102.1 and 105.2 Mb [2], 102.9 and 105.2 Mb [4], 103.0 and 103.6 Mb [3], the *VRTN* gene [5], the *VRTN-PROX2-FOS* region that is equivalent to the region between 103.4 and 104.3 Mb based on a microsatellite study [21], and the region between 106.7 and 106.9 Mb [7]. Five significant SNPs on chromosome 10 were located close to some previously reported significant regions on this chromosome [2], and a significant SNP at 51.6 Mb on chromosome 12 was close to the previously reported chromosome 12 region between 52.9 and 52.6 Mb [2, 4]. In the current study, the region between 77.7 and 79.7 Mb on chromosome 11 was the second most significant chromosome region, which, to our knowledge, has not been reported in previous GWAS. However, the Animal QTLdb [1] has one entry for pig teat number in the region between 79.2 and 85.7 Mb on chromosome 11, which partially overlaps the region between 77.7 and 79.7 Mb that we detected in this study, and this Animal QTLdb result was based on a QTL mapping study using 137 microsatellite markers on 573 F2 females and 530 F2 males from a Meishan  $\times$  Large White cross [22].

### Heritabilities and factors that affect teat number

The genomic narrow-sense heritability estimates reported in the current study are the only ones available for teat number. Estimated narrow-sense heritabilities ranged from 0.346 to 0.350 (Table 1) and were within the range of recently published heritability estimates based on pedigree relationships, e.g., 0.39 in a study using



57,000 Yorkshire pigs [23], and 0.37 in a study using 1550 Landrace pigs [3]. The above genomic and pedigree-based heritability estimates indicate that swine teat number has a strong genetic component, and also that a large portion of the phenotypic variation is not explained by additive genetic effects. Only one GWAS reported dominance effects on chromosome 4 in Landrace pigs [3]. Our GWAS results differ from the reported dominance effects on chromosome 4 because we found that many other chromosomes had more significant dominance effects than chromosome 4, although none of the dominance effects that we observed reached the genome-wide

significance threshold of  $p < 10^{-5.91}$  (Fig. 3b). The estimated dominance heritability was low (0.036) and inclusion of dominance effects in the prediction model only had negligible effects on the observed prediction accuracy. Removing dominance effects from the mixed model resulted only in a 0.5% reduction in the observed accuracy of predicting phenotypic values, although the reductions in the expected accuracy of predicting phenotypic values (7.6%) and in the expected accuracy of predicting genetic values (3.9%) were considerably larger for unknown reasons (Table 2). Previously, a maternal effect on teat number was reported [24] but has not been

studied by GWAS or genomic prediction. Data error is a source of phenotypic variation, but teat number is easy to measure and any data errors that might have occurred would only be minor, given that our GWAS results on chromosome 7 agreed with those of other studies [2, 3] and that our genomic heritability estimates are consistent with those based on pedigree data [3, 23].

**SNP contributions to genomic heritability and prediction accuracy**

We analyzed three methods (Methods I, II and III) of estimating the contribution of a set of SNPs to genomic heritability and compared Methods II and III for estimating the contributions of SNPs to both genomic heritability and prediction accuracy. Method I for estimating the contribution of SNPs to genomic heritability consisted of summing together the heritability estimates of the target set of SNPs because the narrow-sense, dominance and broad-sense heritabilities for each SNP can be estimated individually by the GVCBLUP package [18]. However, this method was inappropriate for estimating SNP contributions to the phenotypic variance because the heritability estimate for each SNP decreases as the number of SNPs in the mixed model increases and is approximately proportional to 1/m, where m is the number of SNPs in the mixed model (see “Appendix” for an approximate proof).

The results in Table 4 show the dependency of the size of the heritability estimate on the number of SNPs fitted as random effects in the mixed model. When dropping every other SNP from the model (i.e. reducing from 41,108 to 20,554 SNPs), the average  $\hat{h}_{\alpha i}^2$  (narrow-sense heritability of the *i*th SNP) of each SNP for the same 20,554 SNPs when all 41,108 SNPs were fitted in the model nearly doubled ( $\hat{h}_{\alpha 1}^2/\hat{h}_{\alpha 2}^2 = 1.960$ ), while the total narrow-sense heritability was nearly unaffected ( $\hat{h}_{\alpha 1}^2 = 0.360$  and  $\hat{h}_{\alpha 3}^2 = 0.368$ ), showing that the size of the heritability estimate for each SNP was approximately divided by 2 when the model had twice as many SNPs. Therefore, the heritabilities of the significant SNPs are

not suitable for measuring their contributions to the phenotypic variance due to the dependency of the size of the heritability estimate for each SNP on the number of SNPs fitted as random effects in the mixed model.

Method II for estimating the contribution of a set of target SNPs to genomic heritability and prediction accuracy consisted in calculating the difference between the model with all SNPs and the model without the target SNPs. However, removing the target SNPs from the statistical model may not completely remove their effects because some of them could be explained by other SNPs in LD with the target SNPs. The results in Table 4 supported this expectation, i.e., the total narrow-sense heritability when halving the number of SNPs fitted was nearly unaffected. Using Method II, the 85 significant SNPs accounted for 4.5% of the total genomic heritability (Table 1) and 2.5% of the observed prediction accuracy (Model 3A, Table 2). Due to the partial effects of the removed SNPs that could have been explained by other SNPs, the contributions of SNPs to genomic heritability and observed prediction accuracy using Method II can be considered as the lower bound of the SNP contributions.

Method III for estimating SNP contribution to genomic heritability and to prediction accuracy calculated the difference between the model with all SNPs fitted as random effects and the model with the target SNPs fitted as fixed effects, an approach that we refer to as ‘partial heritability’ and ‘partial accuracy’. The example of the region between 102.9 and 106.3 Mb on chromosome 7 showed that fitting significant SNPs as fixed effects completely removed the significant effects of those SNPs and also removed the effects of the SNPs that are still fitted in the statistical model as random effects but are in LD with the SNPs fitted as fixed effects (Fig. 4b). Figure 5 is a graphical view of a specific chromosome region, showing that the contributions of SNPs to the total narrow-sense heritability from the two models with 41,108 SNPs and 20,554 SNPs estimated using Method III were nearly the same for the region between 102.9 and 106.0 Mb on chromosome 7, i.e., partial heritability estimates were

**Table 4 Estimates of SNP additive heritabilities of teat number when using 41,108 autosomal SNPs or every other (20,554) of the 41,108 SNPs**

SNP set	Average $\hat{h}_{k\alpha}^2$ per SNP	Ratio	Total $\hat{h}_{\alpha i}^2$
20,554 SNPs	$\bar{h}_{\alpha 1}^2 = 1.75 (10^{-5})$	$\bar{h}_{\alpha 1}^2/\bar{h}_{\alpha 2}^2 = 1.960$	$\hat{h}_{\alpha 1}^2 = 0.360$
20,554 SNPs with all 41,108 SNPs in the mixed model	$\bar{h}_{\alpha 2}^2 = 8.93 (10^{-6})$	$\bar{h}_{\alpha 2}^2/\bar{h}_{\alpha 3}^2 = 0.998$	$\hat{h}_{\alpha 2}^2 = 0.184$
41,108 SNPs	$\bar{h}_{\alpha 3}^2 = 8.95 (10^{-6})$	$\bar{h}_{\alpha 1}^2/\bar{h}_{\alpha 3}^2 = 1.955$	$\hat{h}_{\alpha 3}^2 = 0.368$

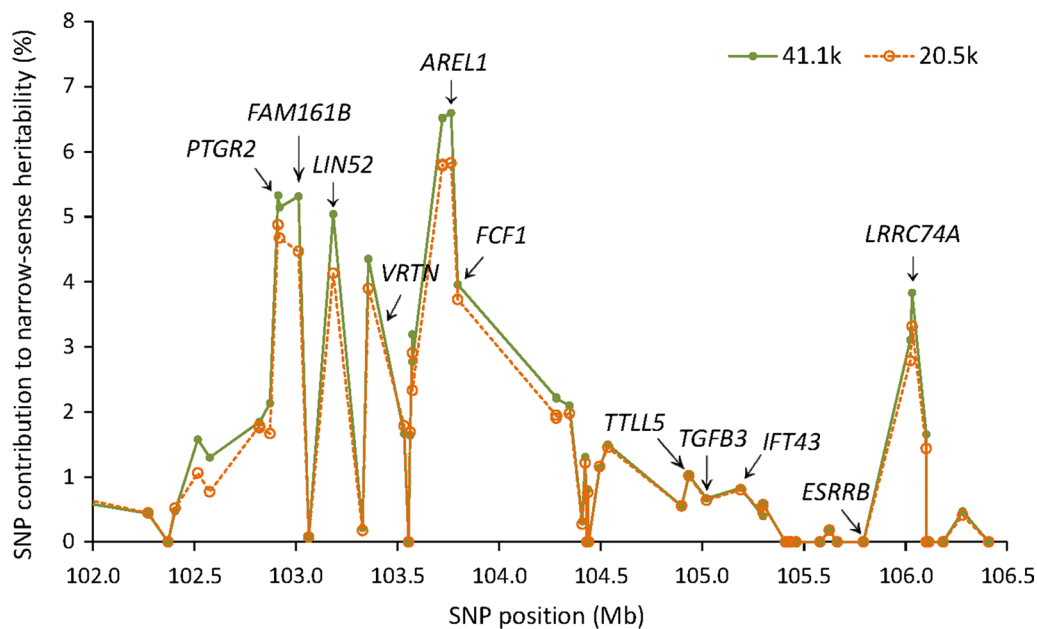
$\hat{h}_{k\alpha}^2$  is the heritability of the *k*th SNP.  $\bar{h}_{\alpha i}^2 = \sum_{k=1}^{m_i} \hat{h}_{k\alpha}^2/m_i$  is the average of SNP additive heritability of the *i*th SNP set,  $\hat{h}_{\alpha i}^2 = \sum_{k=1}^{m_i} \hat{h}_{k\alpha}^2$  is the total additive heritability of all SNPs in the *i*th SNP set, where *m<sub>i</sub>* is the number of SNPs in the *i*th SNP set

nearly unaffected by the number of SNPs in the model. Using this method, the 85 significant SNPs accounted for 28.5 to 28.8% of the genomic narrow-sense heritability (Table 1) and for 36.2 to 36.8% of the observed prediction accuracy (Table 2). In general, contributions of SNPs to genomic heritability and to the observed prediction accuracy were consistent, i.e., most SNPs with higher contributions to heritability also had greater contributions to prediction accuracies. On average, the contributions of SNPs obtained with Method III were larger than those with Method II by 24.0% (4.6 to 28.6%) for the genomic additive heritability and by 33.3% (2.3 to 35.6%) for the observed prediction accuracy (Table 2). Such large differences could be due to two reasons: overestimation by Method III and underestimation by Method II. Overestimation by Method III is expected since some effects that do not come from the target SNPs that are fitted as fixed effects could also be removed, in addition to removing the effects of the target SNPs. e.g., fitting the 14 SNPs in the region between 102.9 and 106.0 Mb on chromosome 7 as fixed effects also removed the significant effects of SNPs in the region between 116.1 and 117.4 Mb on chromosome 7 (Fig. 4b). Therefore, the contributions of SNPs estimated by Method III could be considered as the upper bound of the true SNP contributions. However, underestimation of Method II is likely the main reason for the large differences between Methods II and III, because a large percentage of the effects of the removed SNPs could have been explained by other SNPs in the

model. For the sample in Table 4, the effects of half of the 41,108 SNPs were almost completely explained by the remaining half of the 41,108 SNPs because estimates of genomic heritabilities from those two sets of SNPs were almost the same, as we discussed above. Based on this analysis, we report contributions of SNPs by Method III in the abstract but also show the results obtained with Method II in the main body of the article (Tables 1, 2).

**Observed and expected prediction accuracies**

In the current study, we compared three measures of prediction accuracy: the observed prediction accuracy of predicting phenotypic values based on the correlation between predictions from GBLUP and phenotypic observations of the validation individuals ( $\widehat{R}_{0jp}$ ,  $j = \alpha$  or  $t$ ), the expected accuracy of predicting phenotypic values ( $R_{0jp}$ ), and the expected accuracy of predicting genetic values ( $R_{0j}$ ). For the models using all 41,108 SNPs fitted as random effects (Model 1A and Model 2A in Table 2), we found excellent consistency between the observed accuracies of predicting phenotypic values ( $\widehat{R}_{0tp} = 0.437$  for Model 1A and  $\widehat{R}_{0tp} = 0.435$  for Model 2A) and the expected accuracies of predicting phenotypic values ( $R_{0tp} = 0.460$  for Model 1A and  $R_{0tp} = 0.425$  for Model 2A). For the models using the 85 SNPs as fixed effects to remove the genetic values of those SNPs from the phenotypic values (Model 1B and Model 2B in Table 2), some differences between the observed accuracies of predicting phenotypic values ( $\widehat{R}_{0tp} = 0.279$  for Model 1B and



**Fig. 5** SNP partial heritability in the region between 102.9 and 106.0 Mb on chromosome 7 from two models with 20.5 and 41 K SNPs. The results show that partial heritability estimates were nearly unaffected by the number of SNPs in the model

$\widehat{R}_{0\alpha p} = 0.275$  for Model 2B) and between the expected accuracies of predicting phenotypic values ( $R_{0tp} = 0.360$  for Model 1B and  $R_{0\alpha p} = 0.320$  for Model 2B) were larger (Table 2), but those differences were mostly within one standard deviation of the observed accuracies and should be considered as acceptable. As expected, both observed and expected accuracies of predicting phenotypic values ( $\widehat{R}_{0jp}$  and  $R_{0jp}$ ) were lower than the expected accuracies of predicting genetic values ( $R_{0j}$ ).

## Conclusions

Swine teat number has a strong genetic component with narrow-sense heritability estimates of about 0.365. The GWAS results confirmed the previously reported region on chromosome 7 and identified several new regions associated with swine teat number; they also indicated that the additive effects are the primary genetic effects for teat number and indicated consistency between statistical significance of SNP effects and SNP contribution to the genomic heritability. Most SNPs with higher statistical significance also had greater contributions to the genomic broad-sense heritability and prediction accuracy. The 85 significant SNPs accounted for about 28% of the genomic heritability and 36% of the prediction accuracy.

## Additional files

**Additional file 1: Table S1.** Significant SNP effects with Bonferroni significance ( $p < 10^{-5.91}$ ) by three methods of GWAS analysis.

**Additional file 2: Figure S1.** Manhattan plots of additive SNP effects of all 18 autosomes by three methods of GWAS analysis. All p-values in the figures are in  $\log(1/p)$  scale.

## Authors' contributions

XH, YD, NL and CT designed the experiments. ZW, DL, RZ and XH performed data collection. CT, JR and ZH performed the experiments. CT and YD analyzed the data. DP developed some of the analysis software. YD, CT and XHU wrote the manuscript. All authors read and approved the final manuscript.

## Author details

<sup>1</sup> State Key Laboratory for Agrobiotechnology, China Agricultural University, Beijing 100193, China. <sup>2</sup> Department of Animal Science, University of Minnesota, Saint Paul, MN 55108, USA. <sup>3</sup> National Engineering Research Center for Breeding Swine Industry, South China Agricultural University, Guangdong 510642, China.

## Funding

This project was supported by the National Basic Research Program of China (2014CB138501), the 948 Program of the Ministry of Agriculture of China (2012-G1(4)), and the National High Technology Research and Development Program of China (2011AA100301).

## Competing interests

The authors declare that they have no competing interests.

## Availability of data and materials

The datasets supporting the results of this article are included within the article and its additional files. The phenotypic data and original SNP data are private property of Guangdong Wens Foodstuffs Group and are not for public distribution at the present time.

## Ethics statement

All animal work was conducted according to the guidelines for the care and use of experimental animals established by the Ministry of Science and Technology of the People's Republic of China (Approval Number: 2006-398). Ear tissues were collected using the standard method approved by the Animal Welfare Committee of China Agricultural University (Permit Number: SKLAB-2014-04-02).

## Appendix: Approximate proof for the decrease in SNP heritability as the number of SNPs increases

The heritability estimate for each SNP decreases as the number of SNPs in the mixed model increases and is approximately proportional to  $1/m$ , where  $m$  is the number of SNPs in the mixed model. An approximate mathematical proof for this result can be derived based on the invariance property of GBLUP and GREML to duplicating SNPs. Assuming a set of  $m$  SNPs is duplicated  $r$  times in the mixed model, GBLUP of genetic values (additive, dominance and genotypic values) of individuals and SNP genetic variance components, as well as the associated variance estimates by GREML are invariant to the duplication of SNPs, and GBLUP of SNP additive, dominance and genotypic effects differ from those without duplicate SNPs by the square root of  $r$  [25]. In the example of additive SNP effects,  $\widehat{\alpha}_{ri} = \widehat{\alpha}_i / \sqrt{r}$ , where  $\widehat{\alpha}_{ri}$  is the additive GBLUP estimate of the  $i$ th SNP from the mixed model with  $m$  SNPs repeated  $r$  times,  $\widehat{\alpha}_i$  is the additive GBLUP estimate of the  $i$ th SNP from the model with  $m$  SNPs. The additive heritability for the  $i$ th SNP from the mixed model with  $m$  SNPs ( $h_{\alpha i}^2$ ) is:  $h_{\alpha i}^2 = (\widehat{\alpha}_i^2 / \sum_{i=1}^m \widehat{\alpha}_i^2) h_{\alpha}^2$  [18], where  $\widehat{\alpha}_i$  is the additive GBLUP of the  $i$ th SNP and  $h_{\alpha}^2$  is the total additive heritability based on all SNPs. Since  $h_{\alpha}^2$  is unaffected by repeated SNPs, the additive heritability for the  $i$ th SNP from the mixed model with  $m$  SNPs repeated  $r$  times ( $h_{\alpha ri}^2$ ) is:

$$h_{\alpha ri}^2 = \left( \widehat{\alpha}_{ri}^2 / \sum_{i=1}^{rm} \widehat{\alpha}_{ri}^2 \right) h_{\alpha}^2 = \left( \widehat{\alpha}_i^2 / r \right) / \left( \sum_{i=1}^{rm} \widehat{\alpha}_i^2 / r \right) h_{\alpha}^2 = h_{\alpha i}^2 / r,$$

i.e., the heritability for the  $i$ th SNP from the model with  $r$  times of the  $m$  SNPs is  $1/r$  of the SNP heritability with  $m$  SNPs in the model without repeat. This theoretical result under the simple assumption of repeated SNPs was almost the same as the results obtained with the real data in Table 4, i.e., when the number of SNPs is reduced by half, the heritability of each SNP as an average of all SNP heritability estimates nearly doubled.

Received: 6 September 2016 Accepted: 22 March 2017

Published online: 29 March 2017

## References

1. Hu ZL, Park CA, Wu XL, Reecy JM. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res.* 2013;41:D871–9.

2. Duijvesteijn N, Veltmaat JM, Knol EF, Harlizius B. High-resolution association mapping of number of teats in pigs reveals regions controlling vertebral development. *BMC Genomics*. 2014;15:542.
3. Lopes MS, Bastiaansen JW, Harlizius B, Knol EF, Bovenhuis H. A genome-wide association study reveals dominance effects on number of teats in pigs. *PLoS One*. 2014;9:e105867.
4. Verardo LL, Silva FF, Lopes MS, Madsen O, Bastiaansen JW, Knol EF, et al. Revealing new candidate genes for reproductive traits in pigs: combining Bayesian GWAS and functional pathways. *Genet Sel Evol*. 2016;48:9.
5. Yang J, Huang L, Yang M, Fan Y, Li L, Fang S, et al. Possible introgression of the *VRTN* mutation increasing vertebral number, carcass length and teat number from Chinese pigs into European pigs. *Sci Rep*. 2016;6:19240.
6. Lee JB, Jung EJ, Park HB, Jin S, Seo DW, Ko MS, et al. Genome-wide association analysis to identify SNP markers affecting teat numbers in an F2 intercross population between Landrace and Korean native pigs. *Mol Biol Rep*. 2014;41:7167–73.
7. Arakawa A, Okumura N, Taniguchi M, Hayashi T, Hirose K, Fukawa K, et al. Genome-wide association QTL mapping for teat number in a purebred population of Duroc pigs. *Anim Genet*. 2015;46:571–5.
8. Clayton GA, Powell JC, Hiley PG. Inheritance of teat number and teat inversion in pigs. *Anim Prod*. 1981;33:299–304.
9. Poland JA, Brown PJ, Sorrells ME, Jannink JL. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*. 2012;7:e32253.
10. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One*. 2014;9:e90346.
11. Ma L, Runesha HB, Dvorkin D, Garbe J, Da Y. Parallel and serial computing tools for testing single-locus and epistatic SNP effects of quantitative traits in genome-wide association studies. *BMC Bioinformatics*. 2008;9:315.
12. Ma L, Wiggans GR, Wang S, Sonstegard TS, Yang J, Crooker BA, et al. Effect of sample stratification on dairy GWAS results. *BMC Genomics*. 2012;13:536.
13. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
14. Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, et al. Assessing the impact of population stratification on genetic association studies. *Nat Genet*. 2004;36:388–93.
15. Mao Y, London NR, Ma L, Dvorkin D, Da Y. Detection of SNP epistasis effects of quantitative traits using an extended Kempthorne model. *Physiol Genomics*. 2006;28:46–52.
16. Da Y, Wang C, Wang S, Hu G. Mixed model methods for genomic prediction and variance component estimation of additive and dominance effects using SNP markers. *PLoS One*. 2014;9:e87666.
17. Wang C, Da Y. Quantitative genetics model as the unifying model for defining genomic relationship and inbreeding coefficient. *PLoS One*. 2014;9:e114484.
18. Wang C, Prakapenka D, Wang S, Pulugurta S, Runesha HB, Da Y. GVCBLUP: a computer package for genomic prediction and variance component estimation of additive and dominance effects. *BMC Bioinformatics*. 2014;15:270.
19. Legarra A, Robert-Granié C, Manfredi E, Elsen JM. Performance of genomic selection in mice. *Genetics*. 2008;180:611–8.
20. Barrett JC, Fry B, Maller J, Daly M. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21:263–5.
21. Ren DR, Ren J, Ruan GF, Guo YM, Wu LH, Yang GC, et al. Mapping and fine mapping of quantitative trait loci for the number of vertebrae in a White Duroc × Chinese Erhualian intercross resource population. *Anim Genet*. 2012;43:545–51.
22. Bidanel J, Rosendo A, Iannuccelli N, Riquet J, Gilbert H, Caritez J, Billon Y, Amigues Y, Prunier A, Milan D. Detection of quantitative trait loci for teat number and female reproductive traits in Meishan × Large White F2 pigs. *Animal*. 2008;2:813–20.
23. Lundeheim N, Chalkias H, Rydhmer L. Genetic analysis of teat number and litter traits in pigs. *Acta Agric Scand A*. 2013;63:121–5.
24. Pumfrey RA, Johnson RK, Cunningham PJ, Zimmerman DR. Inheritance of teat number and its relationship to maternal traits in swine. *J Anim Sci*. 1980;50:1057–60.
25. Da Y, Tan C, Parakapenka D. Joint SNP-haplotype analysis for genomic selection based on the invariance property of GBLUP and GREML to duplicate SNPs. *J Anim Sci*. 2016;94:161–2.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

