

RESEARCH ARTICLE

Open Access



# Evaluation of the phenotypic and genomic background of variability based on litter size of Large White pigs

Ewa Sell-Kubiak<sup>1\*</sup> , Egbert F. Knol<sup>2</sup> and Marcos Lopes<sup>2,3</sup>

## Abstract

**Background:** The genetic background of trait variability has captured the interest of ecologists and animal breeders because the genes that control it could be involved in buffering various environmental effects. Phenotypic variability of a given trait can be assessed by studying the heterogeneity of the residual variance, and the quantitative trait loci (QTL) that are involved in the control of this variability are described as variance QTL (vQTL). This study focuses on litter size (total number born, TNB) and its variability in a Large White pig population. The variability of TNB was evaluated either using a simple method, i.e. analysis of the log-transformed variance of residuals (LnVar), or the more complex double hierarchical generalized linear model (DHGLM). We also performed a single-SNP (single nucleotide polymorphism) genome-wide association study (GWAS). To our knowledge, this is only the second study that reports vQTL for litter size in pigs and the first one that shows GWAS results when using two methods to evaluate variability of TNB: LnVar and DHGLM.

**Results:** Based on LnVar, three candidate vQTL regions were detected, on *Sus scrofa* chromosomes (SSC) 1, 7, and 18, which comprised 18 SNPs. Based on the DHGLM, three candidate vQTL regions were detected, i.e. two on SSC7 and one on SSC11, which comprised 32 SNPs. Only one candidate vQTL region overlapped between the two methods, on SSC7, which also contained the most significant SNP. Within this vQTL region, two candidate genes were identified, *ADGRF1*, which is involved in neurodevelopment of the brain, and *ADGRF5*, which is involved in the function of the respiratory system and in vascularization. The correlation between estimated breeding values based on the two methods was 0.86. Three-fold cross-validation indicated that DHGLM yielded EBV that were much more accurate and had better prediction of missing observations than LnVar.

**Conclusions:** The results indicated that the LnVar and DHGLM methods resulted in genetically different traits. Based on their validation, we recommend the use of DHGLM over the simpler method of log-transformed variance of residuals. These conclusions can be useful for future studies on the evaluation of the variability of any trait in any species.

## Background

Living organisms are under the constant influence of unpredictable changes in the environment. Thus, in ecology as well as in animal and plant breeding, genes that

can buffer the negative effects of unpredictable (e.g., diseases) or difficult to avoid (e.g., temperature changes) environmental factors are highly desirable [1]. It is assumed that such genes can control the variation of a trait (at either the population or the individual level) and maintain it at an optimal level [2]. One of the most well-studied genes that is involved in buffering the effects of genetic and environmental factors is the *heat-shock protein 90 (Hsp90)* gene. In *Drosophila* and *Arabidopsis*,

\*Correspondence: ewa.sell-kubiak@puls.edu.pl

<sup>1</sup> Department of Genetics and Animal Breeding, Poznan University of Life Sciences, Poznań, Poland

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



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

*Hsp90* is described as a stabilizer of developmental and morphological traits [3–5]. This suggests that traditionally applied methods that focus on the genetic control of the mean of traits could be extended by also accounting for the variability around that mean. This is possible since it has been observed that not only the mean of the trait is under genetic control, but also the variation around the mean, which is described in the literature as “variance heterogeneity” or “phenotypic variability”.

Phenotypic variability can be assessed by studying the heterogeneity of residual variance across observations [6]. Empirical evidence that residual variance has a genetic component has been reported for different traits in many animal species [7–15] and in humans [16, 17]. Recently, residual variance has even been linked with filling part of the gap of “missing heritability” in genome-wide association studies (GWAS) in humans [17]. One of the most common methods used to obtain variability phenotype is the double hierarchical generalized linear model (DHGLM) [18]. This rather complex method requires substantial computation time. Thus, to verify the need to use DHGLM, it was compared with simpler approaches: log-transformed variance of a trait [14], log-transformed squared estimated residuals [19], and log-transformed variance of residuals (LnVar) [15]. However, only Sell-Kubiak et al. [14] and Lung et al. [19] reported comprehensive comparative studies. Thus, the extended evaluation would also be needed for LnVar and DHGLM. Furthermore, many studies have reported quantitative trait loci (QTL) that are associated with phenotypic variability, the so-called variance QTL (vQTL) [20]. Detection of vQTL in a population can indicate the presence

of an unmodeled interaction associated with the locus [1, 2, 20, 21] or the presence of QTL that directly control the variance of a trait [22, 23]. An overview of selected vQTL that have been detected to date is in Table 1. Still, so far no study has compared the genomic background of variability phenotypes for the same trait obtained with different methods.

In the current study, we continue to focus on the variability of litter size in a Large White population. Litter size is a trait of high economic relevance for pig breeding and has been under intense selection in the past decades. Many reports have shown that litter size has increased from an average of 11.7 live piglets in 2000 to 17.5 in 2019 [24] and this increase has also led to an increase in its variability, which is due to the positive genetic correlation between the mean litter size and its variability [25]. Although litter size might be one of the most studied traits in pigs in terms of genetics, with more than 255 associated single nucleotide polymorphisms (SNPs) reported between 2011 and 2021 [26], to our knowledge, only the Sell-Kubiak et al. [25] study has detected vQTL for litter size in pigs. One of the most promising candidate genes for the variability of litter size is *HSPCB*, which is better known as the aforementioned *Hsp90* [27]. Although our previous study [25] reported the first SNPs associated with phenotypic variability in litter size in pigs, a follow-up study is necessary to confirm more precisely the genomic regions that affect variability in litter size, especially because now we have access to high-density SNP-chip data (660 K instead of 60 K) and the number of genotyped animals has increased exponentially since the publication of our previous paper.

**Table 1** Selected examples of vQTL that affect the variability of quantitative traits across species

Authors	Species	Trait
Mackay and Lyman [53]	<i>Drosophila melanogaster</i>	Bristle number
Ordas et al. [54]	Maize	Days to flowering, ear height, and tassel length
Paré et al. [55]	Human	Levels of inflammatory biomarkers
Ayroles et al. [56]	<i>Drosophila melanogaster</i>	Locomotor handedness
Perry et al. [57]	F2 cross of the genetically hypercalciuric <i>Rattus norvegicus</i> with normocalciuric Wistar-Kyoto	Urinary calcium levels
Wolc et al. [58]	Laying hens	Egg weight
Jimenez-Gomez et al. [59]; Shen et al. [2]	<i>Arabidopsis thaliana</i>	Flowering time
Yang et al. [16]	Human	Body mass index (BMI)
Mulder et al. [60]	Dairy cows	Somatic cell score (SCS)
Nelson et al. [20]	<i>S. cerevisiae</i> strains	Expression traits in different treatments
Sell-Kubiak et al. [25]	Pigs	Litter size
Ek et al. [61]	Human	Variation in DNA methylation levels
Wang et al. [62]; Wang et al. [63]	Pigs	Birth weight
lung et al. [64]	Nellore cattle	Yearling weight
Hussain et al. [65]	Bread wheat	Cadmium levels

Here, we compare two methods to estimate the phenotypic variability of litter size and to evaluate its genetic background and detect new genomic regions associated with it, by performing a GWAS using genotypes from a high-density SNP-chip. Thus, our study has two main aims: (1) to compare two approaches for obtaining variability phenotypes for litter size, and for analyzing their genetic background, i.e. a simpler log-transformed variance of residuals (LnVar) and a more complex double hierarchical generalized linear model (DHGLM); and (2) to perform a single-SNP GWAS to detect novel genomic regions associated with litter size variability using high-density SNP-chip data on ~12,000 Large White pigs. The data used in this study is an updated version of the litter size records and genotypes used in the previous GWAS performed by Sell-Kubiak et al. [25], where only DHGLM was used to obtain variability phenotypes.

**Methods**

**Phenotypes**

The phenotypic data on Large White pigs used in this study were collected on multiplication farms of Topigs Norsvin (Vught, the Netherlands) between March 1982 and January 2019. In total, litter size records (total number born, TNB) for over 640,000 litters were available before data editing. Records were removed if TNB was equal to 3 or less, or if only one record per sow was available. Parities 10+ were treated as parity 10 and litters with a TNB larger than 25 were recoded to 25. This allowed us to keep such records in the analysis, rather than removing the most extreme values. After data editing, 607,553 litter records from 121,088 sows were available for further analysis. The average TNB was 13.76

(±3.64) and the distribution of TNB is presented in Fig. 1. The pedigree contained 168,230 records and was on average five generations deep.

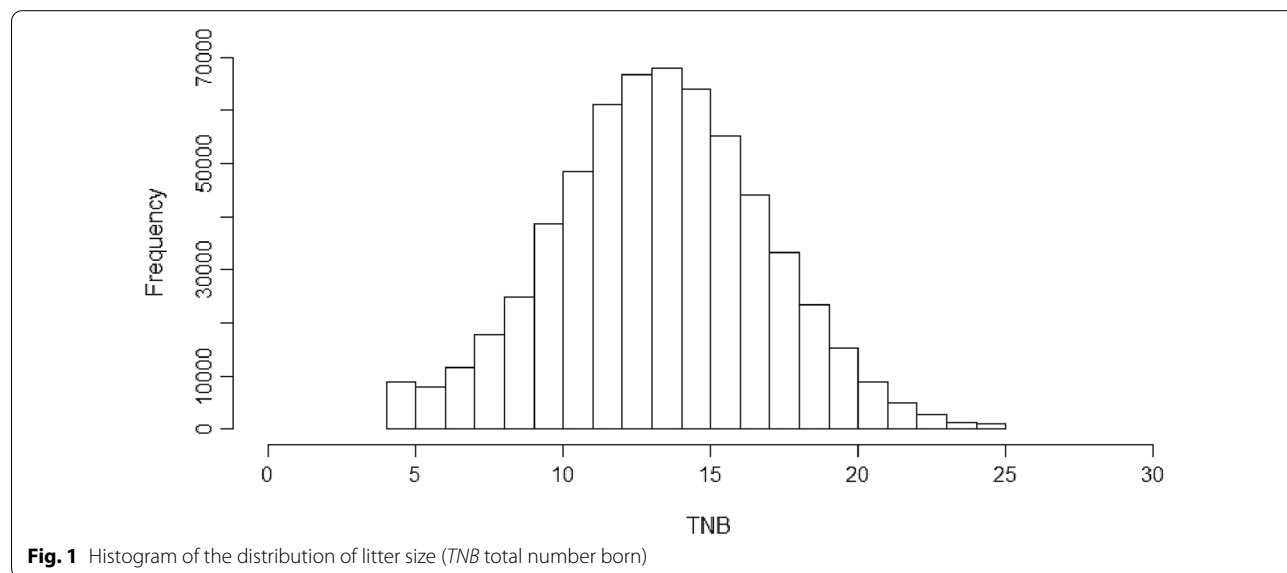
**Estimation and analysis of variability phenotypes using the log-transformed variance of residuals**

Variability phenotype was obtained as the log-transformed variance of residuals (LnVar) of TNB in two steps. First, to obtain estimates of variance components, the TNB data were analyzed with the following traditional animal model, as previously described and tested by Dobrzański et al. [28], using the ASReml 3.0 package [29]:

$$y = Xb + Za + Upe + e, \tag{1}$$

where **y** is a vector of phenotypes for TNB; **X**, **Z**, and **U** are incidence matrices relating phenotypes to effects; **b** is a vector of the fixed effects of herd-year-season and parity; **a** is a vector of additive genetic effects, with  $a \sim N(0, A\sigma_a^2)$ , where **A** is the augmented numerator relationship matrix and  $\sigma_a^2$  is the additive genetic variance; **pe** is a vector of permanent environmental sow effects, which accounts for repeated observations per sow, with  $pe \sim N(0, I_{pe}\sigma_{pe}^2)$ , where **I<sub>pe</sub>** is the appropriate identity matrix and  $\sigma_{pe}^2$  is the permanent environmental sow variance; and **e** is a vector of residuals, with  $e \sim N(0, I_e\sigma_e^2)$ , where **I<sub>e</sub>** is the appropriate identity matrix and  $\sigma_e^2$  is the residual variance.

In the second step, estimates of the residuals from the first step were used to estimate the phenotype for the log of variability of TNB (LnVarTNB) for each sow as the within-individual variance of residuals and log-transforming the result. Log transformation was used to



**Fig. 1** Histogram of the distribution of litter size (TNB total number born)

normalize the distribution of the variability phenotypes and to assume an exponential model for the environmental variance [8, 30–32], which in general is described as:  $y_i = \mu + u + \exp\left(\frac{1}{2}(\eta + \nu)\right)\epsilon_j$  for  $j = 1, \dots, n$ , where  $y_i$  are the phenotypic observations for LnVarTNB,  $\mu$  is the population mean,  $\eta$  is the log environmental variance mean;  $u$  and  $\nu$  are the genetic values of the mean and environmental variance, respectively, following  $\begin{bmatrix} u \\ \nu \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \sigma_u^2 & \sigma_{u,\nu} \\ \sigma_{u,\nu} & \sigma_\nu^2 \end{bmatrix}\right)$ , where  $\sigma_u^2$  and  $\sigma_\nu^2$  are the genetic variances and  $\sigma_{u,\nu}$  is the covariance between the genetic effects; and  $\epsilon_j$  refers to independent, identically and normally distributed variables, that are independent from  $u$  and  $\nu$ .

The variability phenotypes from LnVar were analyzed with the following model, as previously used by Berghof et al. [15] and Dobrzański et al. [28]:

$$\mathbf{y}_{\text{var}} = \mathbf{X}_{\text{var}}\mathbf{b}_{\text{var}} + \mathbf{Z}_{\text{var}}\mathbf{a}_{\text{var}} + \mathbf{e}_{\text{var}}, \tag{2}$$

where  $\mathbf{y}_{\text{var}}$  is a vector of LnVarTNB phenotypes;  $\mathbf{X}_{\text{var}}$  and  $\mathbf{Z}_{\text{var}}$  are incidence matrices relating phenotypes to model effects;  $\mathbf{b}_{\text{var}}$  is a vector of the fixed effects of farm-year-season of the first farrowing;  $\mathbf{a}_{\text{var}}$  is a vector of additive genetic effects, with  $\mathbf{a}_{\text{var}} \sim N(\mathbf{0}, \mathbf{A}\sigma_{\text{avar}}^2)$ , where  $\sigma_{\text{avar}}^2$  is the additive genetic variance; and  $\mathbf{e}_{\text{var}}$  is a vector of residuals, with  $\mathbf{e}_{\text{var}} \sim N(\mathbf{0}, \mathbf{I}_{\text{evar}}\sigma_{\text{evar}}^2)$ , where  $\mathbf{I}_{\text{evar}}$  is the appropriate incidence matrix and  $\sigma_{\text{evar}}^2$  is the residual variance. To account for differences in residual variance due to the varying numbers of TNB records available per sow, we estimated a unique residual variance for each of the nine groups of sows based on numbers of records (Group 1: 19,670 sows with two litters; Group 2: 18,173 sows with three litters; Group 3: 17,601 sows with four litters; Group 4: 16,868 sows with five litters; Group 5: 15,661 sows with six litters; Group 6: 13,365 sows with seven litters; Group 7: 9516 sows with eight litters; Group 8: 5581 sows with nine litters; and Group 9: 4188 sows with ten or more litters).

**Estimation and analysis of variability using a double hierarchical generalized linear model**

We also analyzed the variability of TNB with a double hierarchical generalized linear model (DHGLM) [18, 33], which also allows estimation of variance components for residual variance in ASReml 3.0. The method is based on a bivariate model that requires several iterations until convergence and was used in an earlier study to obtain variability phenotypes for a GWAS for litter size variability [25]. The applied DHGLM was:

$$\begin{bmatrix} \mathbf{y} \\ \boldsymbol{\psi} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_v \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{b}_v \end{bmatrix} + \begin{bmatrix} \mathbf{Z} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_v \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{a}_v \end{bmatrix} + \begin{bmatrix} \mathbf{U} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_v \end{bmatrix} \begin{bmatrix} \mathbf{pe} \\ \mathbf{pe}_v \end{bmatrix} + \begin{bmatrix} \mathbf{e} \\ \mathbf{e}_v \end{bmatrix}, \tag{3}$$

where  $\mathbf{y}$  is the vector of phenotypes TNB and  $\boldsymbol{\psi}$  is the vector of response variables for the variance part of the DHGLM; the residuals  $\mathbf{e}$  and  $\mathbf{e}_v$  are assumed to be independent and normally distributed but with heterogeneous variances across phenotypes;  $\mathbf{b}$  and  $\mathbf{b}_v$  are vectors of the fixed effects of parity of the sow and farm-year-season of the farrowing for  $\mathbf{y}$  and  $\boldsymbol{\psi}$ , respectively;  $\mathbf{a}$  and  $\mathbf{a}_v$  are vectors of random additive genetic effects for  $\mathbf{y}$  and  $\boldsymbol{\psi}$ , with  $\begin{bmatrix} \mathbf{a} \\ \mathbf{a}_v \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \sigma_a^2 & \sigma_{a,\text{av}} \\ \sigma_{a,\text{av}} & \sigma_{\text{av}}^2 \end{bmatrix} \otimes \mathbf{A}\right)$ , where  $\sigma_{\text{av}}^2$  is the additive genetic variance,  $\sigma_{a,\text{av}}$  is the covariance between the genetic effects and  $\otimes$  is the Kronecker product;  $\mathbf{pe}$  and  $\mathbf{pe}_v$  are vectors of random non-genetic permanent sow effects for  $\mathbf{y}$  and  $\boldsymbol{\psi}$ , with  $\begin{bmatrix} \mathbf{pe} \\ \mathbf{pe}_v \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \sigma_{\text{pe}}^2 & \sigma_{\text{pe,pe}_v} \\ \sigma_{\text{pe,pe}_v} & \sigma_{\text{pe}_v}^2 \end{bmatrix} \otimes \mathbf{I}\right)$ , where  $\sigma_{\text{pe}_v}^2$  is the permanent sow variance,  $\sigma_{\text{pe,pe}_v}$  is the covariance between the permanent sow effects; and  $\mathbf{e}$  and  $\mathbf{e}_v$  are vectors of residuals, with  $\begin{bmatrix} \mathbf{e} \\ \mathbf{e}_v \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{W}^{-1}\sigma_e^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_v^{-1}\sigma_{e_v}^2 \end{bmatrix}\right)$ , where  $\mathbf{W} = \text{diag}\left(\exp\left(\hat{\boldsymbol{\psi}}\right)^{-1}\right)$  and  $\mathbf{W}_v = \text{diag}\left(\frac{1-h}{2}\right)$  are expected reciprocals of the residual variance from the previous iteration, and  $\sigma_e^2$  and  $\sigma_{e_v}^2$  are scaling variances and are expected to be equal to 1 [31]. The predicted residual variance for each TNB phenotype  $\exp\left(\hat{\boldsymbol{\psi}}\right)$  is based on the estimated fixed and random effects for  $\boldsymbol{\psi}$  in the previous iteration of the algorithm. This methodology produced the second variability phenotype—varTNB and estimated its variance components.

Note: to fully evaluate the genetic variation within litter size variability based on LnVar and DHGLM, the genetic coefficient of variation on the standard deviation (SD) level ( $\text{GCV}_{\text{SDe}}$ ) was used, which can be approximated as:  $\text{GCV}_{\text{SDe}} = \frac{\sigma_{\text{addv}}(\sigma_e)}{\sigma_e} \approx \frac{1}{2}\sigma_{\text{avar}}$  or  $\approx \frac{1}{2}\sigma_{\text{av}}$ , where  $\sigma_{\text{addv}}$  is the genetic SD in residual variance.

**Comparison of the LnVar and DHGLM**

The two methods used to estimate variability phenotypes for litter size, i.e. LnVar and DHGLM, were evaluated by:

- (1) Comparing the estimates of variance components and the estimated breeding values (EBV) of LnVarTNB and varTNB and their theoretical accuracies which were estimated based on the following equation:  $r_{A,I} = \sqrt{1 - \frac{s_i}{\sigma_{add}^2}}$ , where  $s_i$  is the standard error for the EBV of the  $i$ th individual and  $\sigma_{add}^2$  is the additive genetic variance for the trait.
- (2) Three-fold cross-validation used to predict EBV for LnVarTNB and varTNB. For this, we selected paternal families with at least three half-sisters with litter size records. Then, phenotypes for one of the paternal half-sisters were set to missing, which resulted in removing 3650 sows and their records in three subsets of data. This cross-validation imitates a situation where a breeding program aims at predicting the phenotype of a new-born sow that already has paternal half-sisters with litter size records. The predicted EBV were then correlated with the log-transformed variance of litter size per sow. A similar validation approach was described by Sell-Kubiak et al. [14].

### SNP genotypes for GWAS

In total, 12,232 genotyped sows (N=11,451) and boars (N=781) were available for the GWAS. Genotyping was performed at GeneSeek (Lincoln, NE, USA) using three medium-density SNP chips. Most animals (N=7079) were genotyped using the (Illumina) GeneSeek custom 50K SNP chip (Lincoln, NE, USA), while 3276 and 1877 animals were genotyped using the (Illumina) GeneSeek custom 80K SNP chip (Lincoln, NE, USA) and the Illumina Porcine SNP60 Beadchip (Illumina, San Diego, CA, USA), respectively. In total, 499 animals that had already been genotyped with the medium-density chips were also genotyped with the Axiom porcine 660K array from Affymetrix (Affymetrix Inc., Santa Clara, CA, USA) at GeneSeek (Lincoln, NE, USA). These animals were the sires with the largest number of offspring in the genotyped dataset.

Quality control of the genotype data included exclusion of SNPs with a GenCall score < 0.15 (Illumina Inc., 2005), a call rate < 0.95, and a minor allele frequency < 0.01, and of the SNPs that deviated strongly from the Hardy–Weinberg equilibrium ( $\chi^2 > 600$ ), that were located on sex chromosomes, and that were unmapped. The positions of the SNPs were based on the Sscrofa11.1 assembly of the reference genome. Since all animals had a frequency of missing genotypes  $\leq 0.05$ , none were removed based on that criterion.

After quality control, missing genotypes of the animals genotyped with the (Illumina) GeneSeek custom 50K

SNP chip were imputed within the population using the Fimpute v2.2 software [34], while animals genotyped with the other two chips had their genotypes imputed to the set of SNPs on the (Illumina) GeneSeek custom 50K SNP chip that passed the quality control. After quality control and imputation, genotypes on 50,717 SNPs were available and were used in imputation to the 660K genotypes using Fimpute v2.2 [34]. After quality control and imputation, genotypes on 526,505 SNPs were available for the GWAS.

### Genome-wide association study

For use in GWAS, the EBV for variability obtained with the two methods were deregressed using the methodology of Garrick et al. [35]. Deregression of EBV avoids double-counting of parental information due to various information sources and complex family structures in a population, and provides phenotypes for boars that do not have own litter size records. This also allows more genotyped animals to be included in the GWAS and thus increase its power to detect QTL.

A single-SNP GWAS was performed using the imputed 660K genotypes and applying the following linear animal model in the GCTA software [36, 37]:

$$y_k^* = \mu + X^* \hat{\beta} + u_k^* + e_k^*, \tag{4}$$

where  $y_k^*$  is the deregressed EBV of the  $k$ th animal for LnVarTNB or varTNB;  $\mu$  is the average of the deregressed EBV;  $X^*$  is the genotype (0, 1, 2) of the  $k$ th animal for the evaluated SNP;  $\hat{\beta}$  is the unknown allele substitution effect of the evaluated SNP;  $u_k^*$  is the random additive genetic effect, assumed to be distributed as  $\sim N(\mathbf{0}, \mathbf{G}\sigma_a^{2*})$ , which accounts for the (co)variances between animals through the formation of a genomic numerator relationship matrix ( $\mathbf{G}$ ), constructed using the imputed 660 K genotypes,  $\sigma_a^{2*}$  is the additive genetic variance; and  $e_k^*$  is the random residual effect, which was assumed to be distributed as  $\sim N(\mathbf{0}, \mathbf{I}\sigma_e^{2*})$ .

The genetic variance explained by a SNP ( $\sigma_{snp}^2 = 2pq\hat{\beta}^2$ ) was estimated based on the allele frequencies ( $p$  and  $q$ ) and the estimated allele substitution effect ( $\hat{\beta}$ ). The proportion of phenotypic variance explained by the SNP was defined as  $\sigma_{snp}^2/\sigma_p^2$ , where  $\sigma_p^2$  is the total phenotypic variance (sum of the additive genetic and residual variances), which was estimated based on the model described above without a SNP effect. Significant SNPs were declared based on a p-value <  $10^{-6}$ , while suggestive SNPs were declared based on a p-value <  $10^{-4}$  and linkage disequilibrium (LD) with the significant SNP. Each region with the significant and suggestive SNPs was defined as a separate vQTL region.

Following the GWAS based on the imputed 660 K data, we investigated the linkage disequilibrium (LD) in the

QTL region on SSC7 that overlapped between the two variability phenotypes. LD, as measured by  $r^2$ , was calculated between SNPs using the Plink 1.9 software [38].

**Search for candidate genes**

Based on the GWAS results, we used the location of the significant SNPs to search for candidate genes with the software BIOMART available in Ensembl Sscrofa 11.1 [39] by entering the position of a SNP and  $\pm 50$  kb if the SNP was not located within a known gene. Furthermore, the PigQTL database of the Animal Genome project [26] was used to find previously detected associations and QTL related to pig reproduction within the most promising regions. This was done by entering the start and end position of the identified QTL regions for litter size variability.

**Results**

**Comparison of the variability phenotypes**

Estimates of the variance components and heritability for total number born (TNB) and its variability are in Table 2. The variance components obtained for the variability of litter size differed between the two estimation methods (LnVar and DHGLM), which resulted in differences between heritability estimates. The genetic coefficient of variation on the SD level ( $GCV_{SDe}$ ) indicated a possible change of 9.1 and 9.6% per generation in the genetic SD of LnVarTNB and of varTNB, respectively. Table 3 shows estimates of the correlations between random effects in the DHGLM, with a correlation of 0.43 between additive effects and a correlation of  $-0.87$  between permanent sow effects.

**Table 2** Estimates of variance components, heritability, and the genetic coefficient of variation on the standard deviation level ( $GCV_{SDe}$ )

Estimate	TNB	LnVarTNB	varTNB
Permanent sow variance	0.86 (0.01)	–	0.15 (0.003)
Additive genetic variance	1.33 (0.2)	0.033 (0.001)	0.037 (0.002)
Residual variance	7.67 (0.8)	1.33 (0.4)	0.84
Heritability	0.135 (0.02)	0.024 (0.002) <sup>a</sup>	0.036 <sup>a</sup>
$GCV_{SDe}$ <sup>b</sup>	–	0.091	0.096

TNB: litter size as total number born

LnVarTNB: phenotypic variability of litter size estimated by the log-transformed variance of residuals

varTNB: phenotypic variability estimated with double hierarchical generalized linear model measured in 121,088 sows from the Large White pig population

<sup>a</sup> Heritability is a measure of the reliability of EBV for LnVarTNB and varTNB; it does not reflect the magnitude of the genetic variance in varTNB

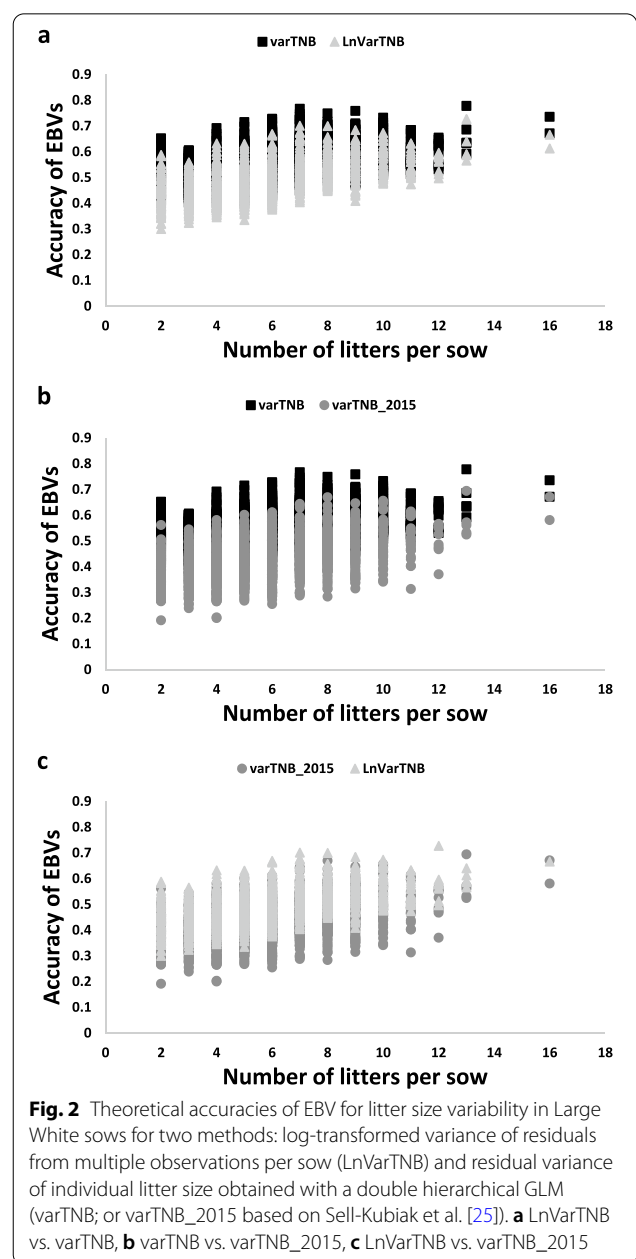
<sup>b</sup> Measure of the genetic variation in residual variance

$$GCV_{SDe} = \frac{\sigma_{add}(\sigma_e)}{\sigma_e} \approx \frac{1}{2}\sigma_{avar} \text{ OR } \approx \frac{1}{2}\sigma_{av}$$

**Table 3** Estimates of correlations<sup>a</sup> (with SE) between the random effects on litter size and its variance

Effect	Correlation
Additive genetic	0.43 (0.03)
Permanent sow	$-0.87$ (0.02)

<sup>a</sup> Obtained with double hierarchical GLM



**Table 4** Correlations between log-transformed variance of litter size and EBV for LnVar<sup>c</sup> and DHGLM<sup>c</sup> obtained from three-fold cross-validation<sup>a</sup>

Validation run	Correlations of ln(var(TNB)) <sup>b</sup> with	
	EBV from LnVar <sup>c</sup>	EBV from DHGLM <sup>c</sup>
1	0.47	0.55
2	0.42	0.49
3	0.40	0.44
Average	0.43	0.49

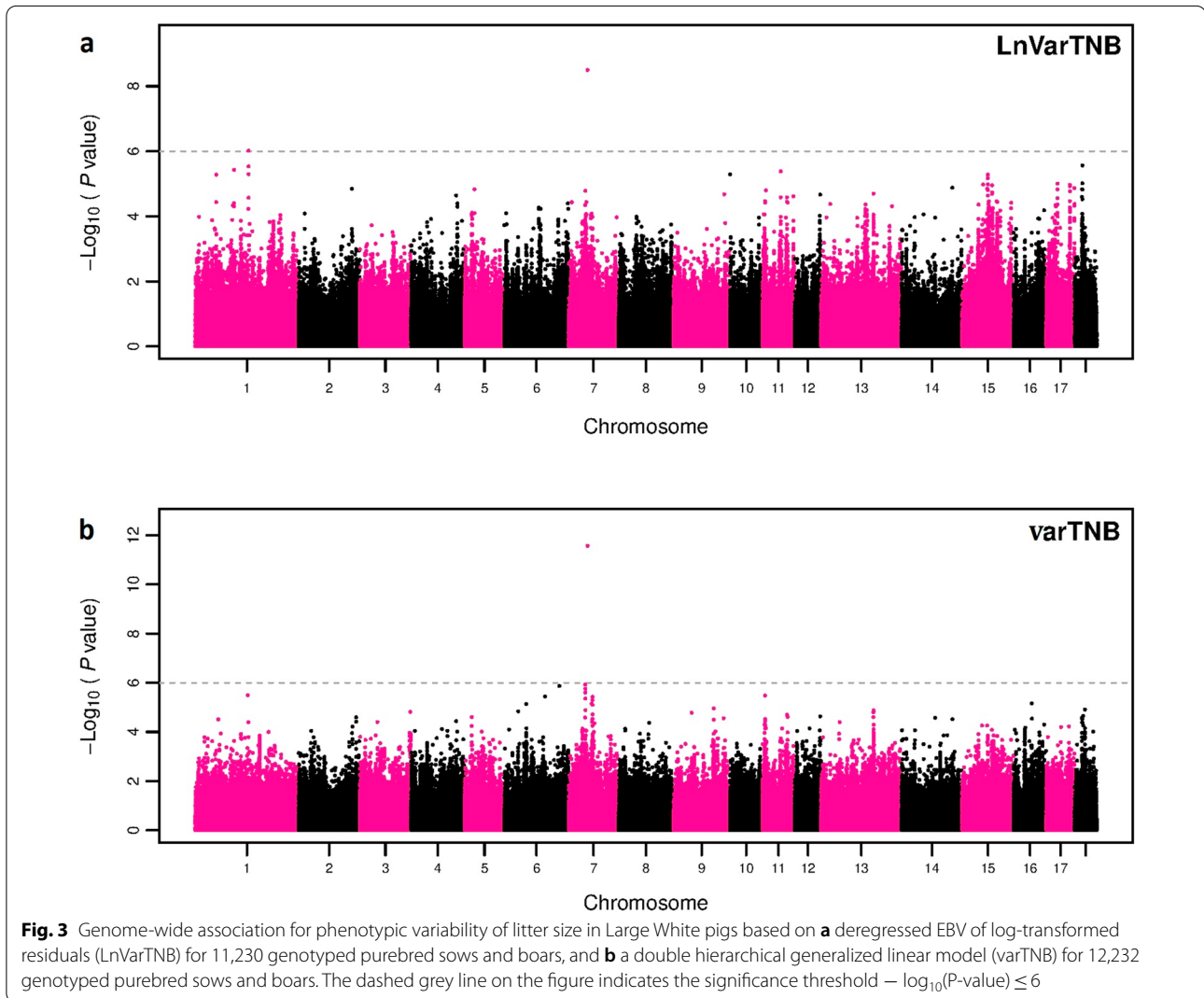
<sup>a</sup> Three-fold cross-validation was performed to compare the two methods used to obtain phenotypic variability of litter size: log-transformed variance of residuals (LnVar) and double hierarchical GLM (DHGLM)

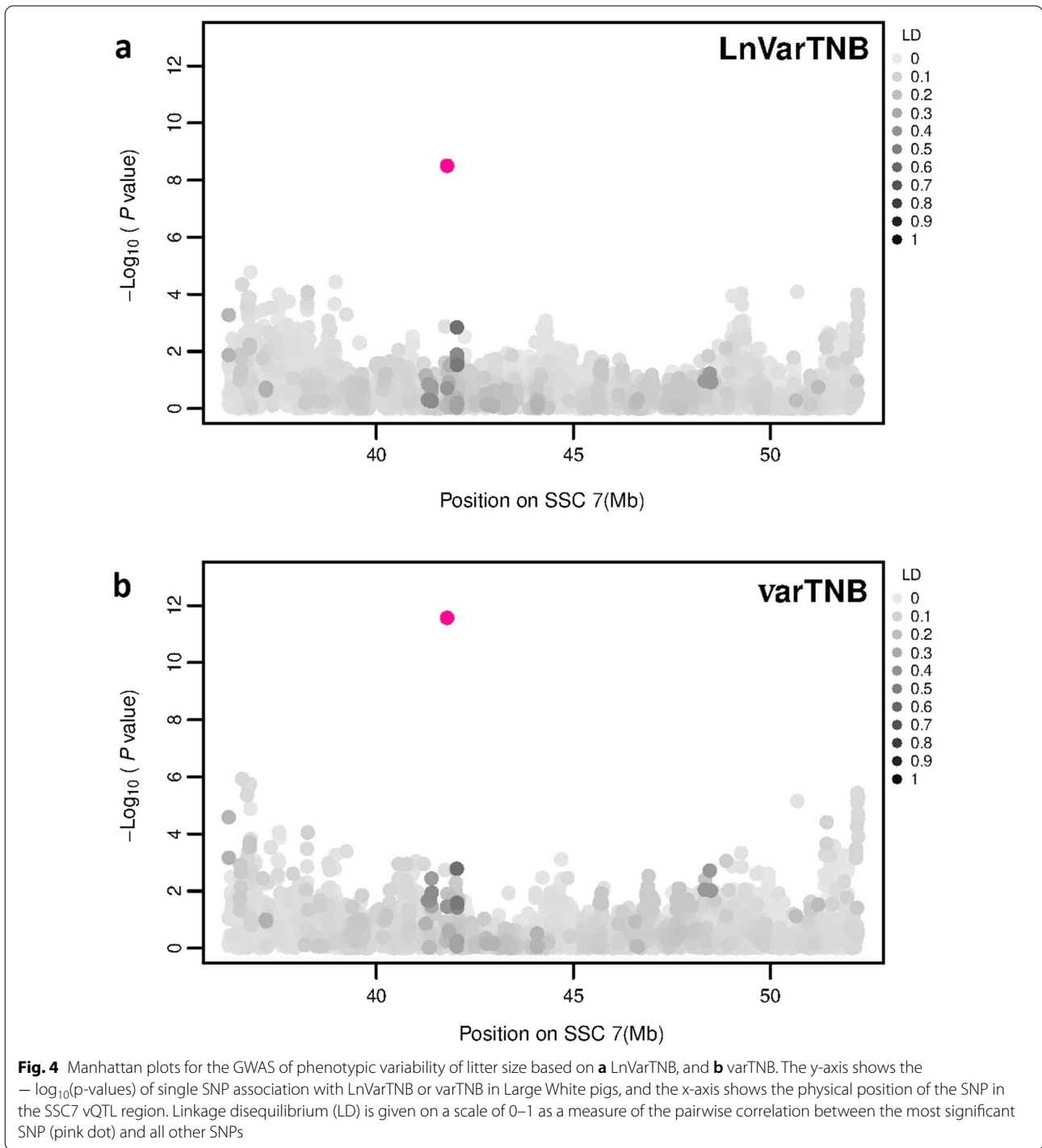
<sup>b</sup> Log-transformed variance of litter size per sow

<sup>c</sup> Estimated in each run for 3650 sows with their records set to missing

The correlation between the EBV obtained with the two methods was 0.86, which indicates some reranking of the animals. In the validation, we also compared the EBV from the two methods with those obtained for the variability of litter size by Sell-Kubiak et al. [25], which will be referred to as varTNB\_2015. Estimates of the correlation of the EBV for varTNB\_2015 with those for LnVarTNB and varTNB were 0.73 and 0.83, respectively. These results suggest that the variability phenotypes obtained with the two methods are not as similar as expected.

The two methods were also compared by evaluating the theoretical accuracies of the EBV and by cross-validation. The theoretical accuracies presented in Fig. 2 indicate that the highest accuracy was reached for varTNB, while LnVarTNB and varTNB\_2015 resulted in very similar





accuracies, although the older dataset contained fewer records per sow. The results of the three-fold cross-validation indicated that the EBV from DHGLM had a significantly ( $P=0.038$  based on a t-test) higher precision than the EBV based on LnVar (Table 4).

**Significant SNPs and candidate genes**

Figures 3 and 4 present the GWAS results of this study and Tables 5 and 6 show the significant and suggestive SNPs for each identified vQTL region for LnVarTNB and varTNB, respectively. The estimates of the



**Table 5** Most significant (*in italics*) and suggestive SNPs in identified vQTL for litter size variability<sup>a</sup>

SSC	Name	Position (Mb)	MAF	SNP effect (with SE)	Genetic variance explained	Phenotypic variance explained	− log(p-value)	SNP location	Candidate genes	Previously detected QTL
1	AX-116127462	129.07	0.01	− 0.083 (0.020)	0.013	0.002	4.57	Intron	GANC	None
	AX-116127463	129.08	0.01	− 0.100 (0.021)	0.017	0.003	5.53	Intron	GANC	
	AX-116127466	129.09	0.01	− 0.097 (0.020)	0.018	0.003	6.02	Intron	GANC	
	AX-116702728	129.17	0.01	− 0.083 (0.018)	0.015	0.003	5.30	Intron	VPS39	
	AX-116127269	127.69	0.03	− 0.047 (0.012)	0.012	0.002	4.23	Non-coding	None <sup>a</sup>	
	AX-116317099	36.59	0.47	0.020 (0.005)	0.016	0.003	4.35	Non-coding	NCR2	Birthweight variability [66]; number mummified piglets [67]; age at puberty [68]
7	AX-116317159	36.80	0.49	0.019 (0.004)	0.015	0.003	4.78	Non-coding	NRN1	
	AX-116317552	38.26	0.36	0.018 (0.005)	0.013	0.002	4.03	Intron	CUL9	
	AX-116317553	38.26	0.36	0.018 (0.005)	0.013	0.002	4.07	Intron	CUL9	
	AX-116317740	38.97	0.36	− 0.019 (0.005)	0.015	0.003	4.43	Non-coding	None <sup>a</sup>	
	AX-116689108	41.80	0.01	0.106 (0.018)	0.020	0.003	8.50	Non-coding	ADGRF1; ADGRF5	
	AX-116596013	18.07	0.05	− 0.039 (0.009)	0.012	0.002	4.60	Intron	LMBR1	Litter weight [69]; litter size [69]; litter weight of piglets born alive [69]; birthweight [70]
18	AX-116596024	18.19	0.09	− 0.030 (0.007)	0.013	0.002	4.83	Non-coding	KLF14; TSGA13; COPG2	
	AX-116596140	18.88	0.04	0.045 (0.011)	0.01	0.002	4.08	Intron	UBE2H	
	AX-116596204	19.30	0.07	− 0.034 (0.008)	0.012	0.002	4.53	Intron	AHCYL2	
	AX-116596216	19.37	0.09	− 0.029 (0.007)	0.011	0.002	5.01	Intron	AHCYL2	
	AX-116596225	19.40	0.06	− 0.041 (0.009)	0.015	0.003	5.56	Intron	AHCYL2	
	AX-116596239	19.45	0.05	− 0.039 (0.009)	0.013	0.002	4.82	Intron	SMO	

SSC *Sus scrofa* chromosomes; SNPs on each SSC present confirmed candidate QTL region

MAF minor allele frequency

<sup>a</sup> Obtained with a log-transformed variance of residuals

<sup>b</sup> Within ± 50 kb

**Table 6** Most significant (*in italics*) and suggestive SNPs in identified vQTL for litter size variability<sup>a</sup>

SSC Name	Position (Mb)	MAF	SNP effect (with SE)	Genetic variance explained	Phenotypic variance explained	- log(p-value)	SNP location	Candidate genes	Previously detected QTL
7	AX-116317026	0.016	0.055 (0.013)	0.007	0.002	4.59	Non-coding	None <sup>a</sup>	Birthweight variability [66]; number mummified piglets [67]; age at puberty [68]
	AX-116317099	0.470	0.022 (0.004)	0.016	0.004	5.93	Non-coding	NCR2	
	AX-116317136	0.480	- 0.020 (0.004)	0.014	0.004	5.36	Non-coding	FOXP4	
	AX-116317145	0.310	- 0.025 (0.005)	0.019	0.005	5.60	Non-coding	FOXP4	
	AX-116317156	0.310	- 0.026 (0.005)	0.019	0.005	5.76	Exon	FOXP4	
	AX-116317159	0.490	0.017 (0.004)	0.011	0.003	4.88	Non-coding	MDFI; FOXP4	
	AX-116317341	0.332	0.017 (0.004)	0.008	0.002	4.06	Intron	TRERF1	
	AX-116317343	0.332	0.017 (0.004)	0.008	0.002	4.06	Intron	TRERF1	
	AX-116317552	0.361	0.017 (0.004)	0.009	0.002	4.04	Intron	CUL9	
	AX-116317553	0.361	0.017 (0.004)	0.009	0.002	4.05	Intron	CUL9	
	AX-116689108	0.010	0.110 (0.016)	0.018	0.005	11.57	Non-coding	ADGRF1; ADGRF5	
7	AX-116320104	0.312	0.020 (0.004)	0.012	0.003	5.15	Non-coding	None <sup>a</sup>	Teat number [71]; number mummified piglets [67]
	AX-116320268	0.248	0.019 (0.005)	0.009	0.002	4.41	Intron	SH3GL3	
	AX-116320489	0.345	0.017 (0.004)	0.009	0.002	4.25	Non-coding	FSD2; AP3B2	
	AX-116320491	0.332	0.02 (0.004)	0.012	0.003	5.10	Non-coding	FSD2; AP3B2	
	AX-116320493	0.330	0.021 (0.004)	0.013	0.003	5.43	Intron	AP3B2	
	AX-116689308	0.332	0.019 (0.004)	0.011	0.003	4.71	Intron	AP3B2	
	AX-116320498	0.332	0.019 (0.004)	0.011	0.003	4.53	Intron	AP3B2	
	AX-116320499	0.339	0.020 (0.004)	0.012	0.003	5.30	Intron	AP3B2	
	AX-116320501	0.331	0.019 (0.004)	0.011	0.003	4.53	Intron	AP3B2	
	AX-116320502	0.242	0.019 (0.005)	0.009	0.002	4.52	Intron	AP3B2	
	AX-116321160	0.069	- 0.028 (0.007)	0.007	0.002	4.00	Non-coding	PSTPIP1; RCN2	
	AX-116321525	0.057	- 0.035 (0.008)	0.009	0.002	4.35	Intron	ARID3B	
11	AX-116416565	0.018	- 0.048 (0.012)	0.006	0.002	4.30	Intron	FLT1	Teat number [72]; number mummified piglets [67]; uterine horn length [73]
	AX-116416787	0.036	0.041 (0.010)	0.008	0.002	4.43	Non-coding	None <sup>a</sup>	
	AX-116416839	0.05	0.038 (0.008)	0.009	0.002	5.50	Intron	KATNAL1	
	AX-116416845	0.036	0.041 (0.010)	0.008	0.002	4.49	Intron	KATNAL1	
	AX-116779844	0.036	0.041 (0.010)	0.008	0.002	4.53	Non-coding	USPL1	
	AX-116416878	0.038	0.038 (0.010)	0.007	0.002	4.14	Non-coding	USPL1	
	AX-116779846	0.039	0.037 (0.009)	0.007	0.002	4.28	Non-coding	USPL1; ALOX5AP	
	AX-116851433	0.039	0.037 (0.009)	0.007	0.002	4.28	Non-coding	USPL1; ALOX5AP	
	AX-116650934	0.231	0.020 (0.005)	0.01	0.003	4.16	Non-coding	FRY; ZARTL	

SSC Sus Scrofa chromosomes; two clusters of SNPs on SSC7 and one on SSC11 present confirmed candidate QTL regions

MAF: minor allele frequency

<sup>a</sup> Obtained with Double Hierarchical GLM

<sup>b</sup> Within ± 50 kb

genetic variance obtained in the GWAS were 0.012 for LnVarTNB and 0.014 for varTNB. The inflation factors ( $\lambda$ ) for both variability phenotypes are in Additional file 1: Fig. S1. Only one identified vQTL region overlapped between LnVarTNB and varTNB, i.e. a region on SSC7, which contained the most significant SNP detected for both variability traits, AX-116689108. This SNP had a low minor allele frequency ( $MAF = 0.01$ ) with only three animals being homozygous for the least frequent genotype and 253 animals being heterozygous. The least frequent allele was associated with greater litter size variability. This low  $MAF$  could be explained by the selection history of the population, with strong selection for increased litter size over the last decades, because greater litter size variation could reduce the average litter size.

We also investigated the LD between AX-116689108 and its surrounding SNPs to check if the mapping of the most significant SNP was correct. This showed that the highest LD was between AX-116689108 and the SNPs in the vQTL region on SSC7, which confirms that the significant SNP is properly mapped.

Additional vQTL regions were identified on SSC1 and SSC18 for LnVarTNB and on SSC7 and SSC11 for varTNB. These GWAS results provide support for considering the variability phenotypes obtained with the two methods as two different traits, since different genomic regions were identified for each of them.

Based on the positions of the significant and suggestive SNPs for the two variability traits, we identified several candidate genes (see Tables 5 and 6) within  $\pm 50$  kb from each SNP. However, not all the significant SNPs were located within a known gene region. This was the case for the most significant SNP detected on SSC7 for both traits and for most of the suggestive SNPs for varTNB. Several previously reported QTL related to reproduction traits in pigs were located within the identified vQTL regions (Tables 5 and 6). Interestingly, SNP AX-116317698 at 38.81 Mb on SSC7 in the vQTL region for variability of litter size was previously found to be associated with TNB (see Additional file 1: Fig. S2 and Additional file 2 Table S1).

## Discussion

The main aim of our study was to detect novel genomic regions that are involved in the genetic control of the phenotypic variability of litter size in Large White pigs. In addition to this objective, we also compared two methods to estimate variability phenotypes, i.e. LnVar and DHGLM. This is the second study that reports vQTL for the variability of litter size and the first one to study the genomic differences between variability phenotypes obtained with two methods.

## Comparison of methods

We compared the following two methods for estimating litter size variability: the simple approach of estimating log-transformed variance residuals (LnVar) and a DHGLM. The latter was previously used for analysis of part of the same data as used here by Sell-Kubiak et al. [25]. The LnVar is a simpler method than DHGLM to compute variability phenotypes and requires less computation time, making it more suitable for application in breeding programs. The DHGLM method was described in detail by Berghof et al. [15] and Dobrzański et al. [28]. Dobrzański et al. [28] also tested whether accounting for the “parity curve” in litter size, i.e. a change in average litter size in subsequent parities [40], affected the variability phenotype. This was proposed since extending the model to better fit the phenotypic data should, in theory, yield more accurate estimates of the residual variance. However, no differences in residual variance were found when applying these more complex animal models [28].

The DHGLM is an interesting approach to study the variability of traits because it enables analysis of variation even when only one observation per animal is available [14]. This of course comes with a cost, since more observations per animal increases the heritability of trait variability [41]. It is important since trait variability tends to have a very low heritability in the classical sense [6], whereas in exponential models it should be considered as a measure of the reliability of EBV for variability and as such does not reflect the magnitude of the genetic variance (for this, the genetic coefficient of variation on the SD level is used). Due to its complexity, DHGLM was compared with simpler methods: log-transformed variance of birth weight per litter in pigs [14], log-transformed squared estimated residuals of yearling weight in Nellore beef cattle [19], and log-transformed variance of residual of body weight in layer hens (LnVar) [15]. Sell-Kubiak et al. [14] and Berghof et al. [15] reported high similarity between the variability traits obtained with the compared methods, whereas the Iung et al. [19] indicated that DHGLM provided more accurate estimates. However, only Sell-Kubiak et al. [14] and Iung et al. [19] reported comprehensive comparative studies, whereas Berghof et al. [15] merely presented the genetic correlation between traits. Interestingly, in the study of Sell-Kubiak et al. [14] the correlation between EBV for the studied methods (0.88) was slightly higher than that found between LnVar and DHGLM (0.86) in the current study, whereas in the study of Iung et al. [19] the correlation was only 0.45. In those studies, the accuracy of EBV was always higher for DHGLM than for the simpler method.

Overall, our results indicate that the two methods used to estimate variability phenotypes of litter size (LnVar

and DHGLM) do not produce identical traits, based on comparing estimates of variance components, estimated breeding values, and theoretical accuracies of EBV and by a three-fold cross-validation. Therefore, our results do not confirm the study of Berghof et al. [15] that concluded that these two methods yielded the same trait when analyzing body weight variability in layer hens. In addition, the GWAS results of our study reveal that these two traits are controlled by different genomic regions, given that only the region on SSC7 was significant for both traits. This is in line with the presence of some reranking of EBV when comparing LnVar and DHGLM, which after deregression were used as response variable in GWAS. Thus, our results indicate that LnVar and DHGLM yield, to a large extent, genetically different phenotypes of variability. Moreover, we also showed in the three-fold cross-validation that DHGLM provides more accurate EBV. This means that the two methods should not be treated as interchangeable and although DHGLM has a longer computation time and is more difficult to implement in a real-life breeding program, we recommend it over the simpler methods. This is in line with the findings of Sell-Kubiak et al. [14] and Lung et al. [19].

### Significant SNPs and candidate genes

Within each vQTL region for variability of litter size, several candidate genes were identified. We focused only on genes that are located within  $\pm 50$  kb from the most significant SNP in each vQTL region. The most significant SNP (AX-116689108) for both variability phenotypes was located on SSC7 within a non-coding region but notably in the middle of a regulatory element (genomic evolutionary rate profiling, GERP) [39]. In our study of 2015 on variability of litter size, which included a much smaller number of animals from the same population ( $N=2067$ ), we also identified a SNP on SSC7 that was significantly associated with varTNB\_2015 [25]. This SNP was located at 43.76 Mb based on the *Sus scrofa* build 10.2 [25] but at 38.26 Mb in the *Sus scrofa* build 11.1 used in the current study. However, in the current study, this SNP was not significant [ $-\log_{10}(P\text{-value})=4.05$ ]. The LD between this SNP and the most significant SNP from the current study (at 41.8 Mb on SSC7) was only 0.11. Although this is not a strong LD, it indicates that this region on SSC7 plays an important role in litter size variability.

The two genes that are located within  $\pm 50$  kb from the most significant SNP on SSC7 for both traits are *ADGRF1*—*adhesion G protein-coupled receptor F1* (SSC7:41,806,393–41,853,657 bp) and *ADGRF5*—*adhesion G protein-coupled receptor F5* (SSC7:41,640,258–41,769,286 bp). The *ADGRF1* GO term annotation relates it to transmembrane signaling receptor activity, for

which *ADGRF5* is an important paralog [27]. *ADGRF1* is involved in neurodevelopment of the brain and its functions are related to the effect of docosahexaenoic acid on the brain [42, 43]. *ADGRF5* was recently reported to be involved in the prevention of paraquat-related lung injuries and to be important for the function of the respiratory system [44] and in the process of vascularization [45, 46]. Moreover, knockout mice that lack *ADGRF5* (in combination with a knockout of another gene *adhesion G protein-coupled receptor ADGRL4*) show perinatal lethality in 50% of the animals [46]. These functions can have an important effect on the blood supply to developing fetuses. Both *ADGRF1* and *ADGRF5* are expressed in human ovary, uterus, prostate, and testis tissues [27].

Another identified vQTL region for varTNB was also located on SSC7, with the most significant SNP being AX-116320493 and the candidate gene being *AP3B2*—*adaptor related protein complex 3 subunit beta 2*, which is expressed in the testis and prostate [27]. Mutations of this gene have been reported to result in neurologic disorders in humans [47].

The two other candidate genes for LnVarTNB were *GANC*—*glucosidase alpha, neutral C* for the identified vQTL on SSC1 and *AHCYL2* *adenosylhomocysteinase like 2* on SSC18. In humans, *GANC* is expressed in several organs related to reproduction in females (breast, uterus, cervix, and ovary) and in males (testis, prostate, and seminal vesicle) [27]. *GANC* encodes a member of the glycosyl family, which is a key enzyme in the metabolism of glycogen and is associated with susceptibility to diabetes [48, 49]. *AHCYL2* is involved in the development of congenital heart defects in humans, and a GWAS in Chinese cattle also described it as a candidate gene associated with displacement of the abomasum [50]. It is also expressed in the ovary and uterus and in the testis and prostate [27].

The last of the most significant SNPs for varTNB was AX-116416839 located on SSC11 within the candidate gene *KATNAL1* *katanin catalytic subunit A1 like 1*. This gene is associated mostly with an expression in neurons [51], but it has been shown to have a role in the regulation of the Sertoli cell microtubules, which if disturbed can cause male infertility [52]. Since the functions of the candidate genes are not directly linked with female reproduction, it would be useful to further study the existence of causative mutations underlying variability of litter size.

### Conclusions

We identified six novel genomic regions that are associated with the variability of litter size in pigs. Only one vQTL region, on SSC7, partially overlapped between the variability traits obtained with the two methods (LnVar and

DHGLM) used here and, in both cases, it contained the same most significant SNP. Both our current findings and those of a previous study of the same population, provide strong evidence for a causative mutation on SCC7 for litter size variability. However, future studies based on sequence data are needed to confirm the genomic regions involved in the control of variability of litter size, since the most significant SNP on SCC7 was detected within a non-coding region. In addition, our results indicate that the LnVar and DHGLM methods that were used to estimate the variability of litter size produced phenotypically and genetically different traits, and that DHGLM yields much more accurate results. Thus, we recommend the use of DHGLM over the simpler method to study and implement selection on the variability of litter size in breeding programs.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12711-021-00692-5>.

**Additional file 1: Figure S1.** Expected and observed  $-\log_{10}$  P-values of SNPs associated with litter size variability defined as LnVarTNB and varTNB with inflation factor ( $\lambda$ ) that were estimated using the R package QQperm [74]. **Figure S2.** Manhattan plot of the genome-wide association study for total number born in the Large White pig population.

**Additional file 2: Table S1.** SNPs significantly associated with the total number born with chromosome number (SSC) and minor allele frequency (MAF).

### Acknowledgements

The authors would like to thank Barbara Harlizius, Arne Gjuvland and Esinam Nancy Amuzu-Aweh for their valuable comments during the development of this study.

### Authors' contributions

ESK initiated and designed the study, prepared the phenotypic data, performed the analysis on phenotypes and post-GWAS evaluation as well as wrote the paper. ML prepared the genotypic data, performed the GWAS, and contributed to the writing of the manuscript. EFK was a discussion partner and contributed to the writing of the manuscript. All authors read and approved the final manuscript.

### Funding

This project was financed by the National Science Centre, Poland (NCN SONATA Grant No. 2016/23/D/NZ9/00029; ESK is the grant holder) whereas publication of the manuscript was possible by the statutory funding of the Department of Genetics and Animal Breeding at Poznan University of Life Sciences (grant no. 506.534.04.00). ESK also acknowledges the financial support of the Polish Ministry of Science and Higher Education (Grant No. 1021/STYP/12/2017).

### Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available because they are part of the commercial breeding program of Topigs Norsvin. They are however available upon reasonable request (contact: [egbert.knol@topignorsvin.com](mailto:egbert.knol@topignorsvin.com)).

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Genetics and Animal Breeding, Poznan University of Life Sciences, Poznań, Poland. <sup>2</sup>Topigs Norsvin Research Centre, Beuningen, The Netherlands. <sup>3</sup>Topigs Norsvin, Curitiba, Brazil.

Received: 15 July 2021 Accepted: 15 December 2021

Published online: 03 January 2022

### References

1. Bruijning M, Metcalf CJE, Jongejans E, Ayroles JF. The evolution of variance control. *Trends Ecol Evol.* 2020;35:22–33.
2. Shen X, Pettersson M, Rönnegård L, Carlborg Ö. Inheritance beyond plain heritability: variance-controlling genes in *Arabidopsis thaliana*. *PLoS Genet.* 2012;8:e1002839.
3. Queitsch C, Sangster TA, Lindquist S. Hsp90 as a capacitor of phenotypic variation. *Nature.* 2002;417:618–24.
4. Rutherford S, Hirate Y, Swalla BJ. The Hsp90 capacitor, developmental remodeling, and evolution: the robustness of gene networks and the curious evolvability of metamorphosis. *Crit Rev Biochem Mol.* 2007;42:355–72.
5. Sangster TA, Bahrami A, Wilczek A, Watanabe E, Schellenberg K, McLellan C, et al. Phenotypic diversity and altered environmental plasticity in *Arabidopsis thaliana* with reduced Hsp90 levels. *PLoS One.* 2007;2:e648.
6. Hill WG, Mulder HA. Genetic analysis of environmental variation. *Genet Res (Camb).* 2010;92:381–95.
7. Ibáñez-Escriche N, Varona L, Sorensen D, Noguera JL. A study of heterogeneity of environmental variance for slaughter weight in pigs. *Animal.* 2008;2:19–26.
8. Ros M, Sorensen D, Waagepetersen R, Dupont-Nivet M, SanCristobal M, Bonnet JC, et al. Evidence for genetic control of adult weight plasticity in the snail *Helix aspersa*. *Genetics.* 2004;168:2089–97.
9. Sonesson AK, Ødegård J, Rönnegård L. Genetic heterogeneity of within-family variance of body weight in Atlantic salmon (*Salmo salar*). *Genet Sel Evol.* 2013;45:41.
10. Sae-Lim P, Khaw HL, Nielsen HM, Puvanendran V, Hansen Ø, Mortensen A. Genetic variance for uniformity of body weight in lumpfish (*Cyclopterus lumpus*) used a double hierarchical generalized linear model. *Aquaculture.* 2020;514:734515.
11. Mulder HA, Gienapp P, Visser ME. Genetic variation in variability: Phenotypic variability of fledging weight and its evolution in a songbird population. *Evolution.* 2016;70:2004–16.
12. Yousefi Zonuz A, Alijani S, Rafat SA. Genetic heterogeneity of residual variance of hatch weight in Mazandaran native chicken. *Br Poult Sci.* 2019;60:366–72.
13. Sell-Kubiak E, Wang S, Knol EF, Mulder HA. Genetic analysis of within-litter variation in piglets' birth weight using genomic or pedigree relationship matrices. *J Anim Sci.* 2015;93:1471–80.
14. Sell-Kubiak E, Bijma P, Knol EF, Mulder HA. Comparison of methods to study uniformity of traits: application to birth weight in pigs. *J Anim Sci.* 2015;93:900–11.
15. Berghof TVL, Bovenhuis H, Mulder HA. Body weight deviations as indicator for resilience in layer chickens. *Front Genet.* 2019;10:1216.
16. Yang J, Loos RJF, Powell JE, Medland SE, Speliotes EK, Chasman DI, et al. FTO genotype is associated with phenotypic variability of body mass index. *Nature.* 2012;490:267–72.
17. Tropf FC, Lee SH, Verweij RM, Stulp G, van der Most PJ, De Vlaming R, et al. Hidden heritability due to heterogeneity across seven populations. *Nat Hum Behav.* 2017;1:757–65.
18. Rönnegård L, Felleki M, Fikse F, Mulder HA, Strandberg E. Genetic heterogeneity of residual variance-estimation of variance components using double hierarchical generalized linear models. *Genet Sel Evol.* 2010;42:8.

19. lung LHS, Neves HHR, Mulder HA, Carvalheiro R. Genetic control of residual variance of yearling weight in Nellore beef cattle. *J Anim Sci*. 2017;95:1425–33.
20. Rönnegård L, Valdar W. Detecting major genetic loci controlling phenotypic variability in experimental crosses. *Genetics*. 2011;188:435–47.
21. Rönnegård L, Valdar W. Recent developments in statistical methods for detecting genetic loci affecting phenotypic variability. *BMC Genet*. 2012;13:63.
22. Nelson RM, Pettersson ME, Li X, Carlborg Ö. Variance heterogeneity in *Saccharomyces cerevisiae* expression data: trans-regulation and epistasis. *PLoS One*. 2013;8:e79507.
23. Geiler-Samerotte K, Bauer C, Li S, Ziv N, Gresham D, Siegal M. The details in the distributions: why and how to study phenotypic variability. *Curr Opin Biotechnol*. 2013;24:752–9.
24. SEGES. Annual Report—Results. Copenhagen. 2020. <https://pigresearchentre.dk/About-us/Annual-reports>. Accessed 30 Jun 2021.
25. Sell-Kubiak E, Duijvesteyn N, Lopes MS, Janss LLG, Knol EF, Bijma P, et al. Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. *BMC Genomics*. 2015;16:1049.
26. Animal Genome Project. Pig QTL data base. <http://www.animalgenome.org/QTLdb/pig.html>. Accessed 30 Jun 2021.
27. Stelzer G, Rosen R, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: from gene data mining to disease genome sequence analysis. *Curr Protoc Bioinformatics*. 2016. <https://doi.org/10.1002/cpbi.5>.
28. Dobrzański J, Mulder HA, Knol EF, Szwaczkowski T, Sell-Kubiak E. Estimation of litter size variability phenotypes in Large White sows. *J Anim Breed Genet*. 2020;137:559–70.
29. Gilmour AR, Gogel BJ, Cullis BR, Thompson R. ASReml user guide release 3.0. Hemel Hempstead: VSN Int Ltd. 2009. <https://www.vsn.co.uk/>. Accessed 30 June 2021.
30. SanCristobal-Gaudy M, Elsen J-M, Bodin L, Chevalet C. Prediction of the response to a selection for canalisation of a continuous trait in animal breeding. *Genet Sel Evol*. 1998;30:423–51.
31. SanCristobal-Gaudy M, Bodin L, Elsen JM, Chevalet C. Genetic components of litter size variability in sheep. *Genet Sel Evol*. 2001;33:249–71.
32. Sorensen D, Waagepetersen R. Normal linear models with genetically structured residual variance heterogeneity: A case study. *Genet Res*. 2003;82:207–22.
33. Felleki M, Lee D, Lee Y, Gilmour AR, Rönnegård L. Estimation of breeding values for mean and dispersion, their variance and correlation using double hierarchical generalized linear models. *Genet Res*. 2012;94:307–17.
34. Garrick DJ, Taylor JF, Fernando RL. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol*. 2009;41:55.
35. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*. 2014;15:478.
36. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88:76–82.
37. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet*. 2014;46:100–6.
38. 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. *Am J Hum Genet*. 2007;81:559–75.
39. Ensembl. Sscrofa 11.1. <http://www.ensembl.org>. Accessed 30 Jun 2021.
40. Sell-Kubiak E, Knol EF, Mulder HA. Selecting for changes in average "parity curve" pattern of litter size in Large White pigs. *J Anim Breed Genet*. 2019;136:134–48.
41. Berghof TVL, Poppe M, Mulder HA. Opportunities to improve resilience in animal breeding programs. *Front Genet*. 2019;9:692.
42. Kim HY, Spector AA. N-Docosahexaenoylethanolamine: A neurotrophic and neuroprotective metabolite of docosahexaenoic acid. *Mol Aspects Med*. 2018;64:34–44.
43. Lee JW, Huang BX, Kwon HS, Rashid MA, Kharebava G, Desai A, et al. Orphan GPR110 (ADGRF1) targeted by N-docosahexaenoylethanolamine in development of neurons and cognitive function. *Nat Commun*. 2016;7:13123.
44. Kubo F, Ariestanti DM, Oki S, Fukuzawa T, Demizu R, Sato T, et al. Loss of the adhesion G-protein coupled receptor ADGRF5 in mice induces airway inflammation and the expression of CCL2 in lung endothelial cells. *Respir Res*. 2019;20:11.
45. Niaudet C, Petkova M, Jung B, Lu S, Laviña B, Offermanns S, et al. Adgrf5 contributes to patterning of the endothelial deep layer in retina. *Angiogenesis*. 2019;22:491–505.
46. Lu S, Liu S, Wietelmann A, Kojonazarov B, Atzberger A, Tang C, et al. Developmental vascular remodeling defects and postnatal kidney failure in mice lacking Gpr116 (Adgrf5) and Etd1 (Adgrl4). *PLoS One*. 2017;12:e0183166.
47. Assoum M, Philippe C, Isidor B, Perrin L, Makrythanasis P, Sondheimer N, et al. Autosomal-recessive mutations in AP3B2, adaptor-related protein complex 3 beta 2 subunit, cause an early-onset epileptic encephalopathy with optic atrophy. *Am J Hum Genet*. 2016;99:1368–76.
48. Hirschhorn R, Huie ML, Kasper JS. Computer assisted cloning of human neutral  $\alpha$ -glucosidase C (GANC): a new paralog in the glycosyl hydrolase gene family 31. *Proc Natl Acad Sci USA*. 2002;99:13642–6.
49. Dewey S, Lai X, Witzmann FA, Sohal M, Gomes AV. Proteomic analysis of hearts from Akita mice suggests that increases in soluble epoxide hydrolase and antioxidative programming are key changes in early stages of diabetic cardiomyopathy. *J Proteome Res*. 2013;12:3920–33.
50. Huang H, Cao J, Guo G, Li X, Wang Y, Yu Y, et al. Genome-wide association study identifies QTLs for displacement of abomasum in Chinese Holstein cattle. *J Anim Sci*. 2019;97:1133–42.
51. Hatakeyama E, Hayashi K. KATNAL1 is a more active and stable isoform of katanin, and is expressed dominantly in neurons. *Biochem Biophys Res Commun*. 2018;507:389–94.
52. Smith LB, Milne L, Nelson N, Eddie S, Brown P, Atanassova N, et al. KATNAL1 regulation of sertoli cell microtubule dynamics is essential for spermiogenesis and male fertility. *PLoS Genet*. 2012;8:e1002697.
53. Mackay TFC, Lyman RF. *Drosophila* bristles and the nature of quantitative genetic variation. *Philos T R Soc Lond B Biol Sci*. 2005;360:1513–27.
54. Ordas B, Malvar RA, Hill WG. Genetic variation and quantitative trait loci associated with developmental stability and the environmental correlation between traits in maize. *Genet Res (Camb)*. 2008;90:385–95.
55. Paré G, Cook NR, Ridker PM, Chasman DI. On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the Women's Genome Health Study. *PLoS Genet*. 2010;6:e1000981.
56. Ayroles JF, Buchanan SM, O'Leary C, Skutt-Kakarika K, Grenier JK, Clark AG, et al. Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proc Natl Acad Sci USA*. 2015;112:6706–11.
57. Perry GML, Nehrke KW, Bushinsky DA, Reid R, Lewandowski KL, Hueber P, et al. Sex modifies genetic effects on residual variance in urinary calcium excretion in rat (*Rattus norvegicus*). *Genetics*. 2012;191:1003–13.
58. Wolc A, Arango J, Settar P, Fulton JE, O'Sullivan NP, Preisinger R, et al. Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. *Anim Genet*. 2012;43:87–96.
59. Jimenez-Gomez JM, Corwin JA, Joseph B, Maloof JN, Kliebenstein DJ. Genomic analysis of QTLs and genes altering natural variation in stochastic noise. *PLoS Genet*. 2011;7:e1002295.
60. Mulder HA, Crump R, Calus MPL, Veerkamp R. Unraveling the genetic architecture of environmental variance of somatic cell score using high-density single nucleotide polymorphism and cow data from experimental farms. *J Dairy Sci*. 2013;96:7306–17.
61. Ek WE, Rask-Andersen M, Karlsson T, Enroth S, Gyllensten U, Johansson Å. Genetic variants influencing phenotypic variance heterogeneity. *Hum Mol Genet*. 2018;27:799–810.
62. Wang X, Liu X, Deng D, Yu M, Li X. Genetic determinants of pig birth weight variability. *BMC Genet*. 2016;17:41.
63. Wang Y, Ding X, Tan Z, Ning C, Xing K, Yang T, et al. Genome-wide association study of piglet uniformity and farrowing interval. *Front Genet*. 2017;8:194.
64. lung LHS, Mulder HA, Neves HHR, Carvalheiro R. Genomic regions underlying uniformity of yearling weight in Nellore cattle evaluated under different response variables. *BMC Genomics*. 2018;19:619.
65. Hussain W, Campbell M, Jarquin D, Walia H, Morota G. Variance heterogeneity genome-wide mapping for cadmium in bread wheat reveals novel genomic loci and epistatic interactions. *Plant Genome*. 2020;13:e20011.
66. Wang X, Liu X, Deng D, Yu M, Li X. Genetic determinants of pig birth weight variability. *BMC Genet*. 2016;17:15.

67. Onteru SK, Fan B, Du ZQ, Garrick DJ, Stalder KJ, Rothschild MF. A whole-genome association study for pig reproductive traits. *Anim Genet.* 2011;43:18–26.
68. Tart JK, Johnson RK, Bundy JW, Ferdinand NN, McKnite AM, Wood JR, et al. Genome-wide prediction of age at puberty and reproductive longevity in sows. *Anim Genet.* 2013;44:387–97.
69. Coster A, Madsen O, Heuven HCM, Dibbits B, Groenen MAM, van Arendonk JAM, et al. The imprinted gene DIO3 is a candidate gene for litter size in pigs. *PLoS One.* 2012;7:e31825.
70. Jiao S, Maltecca C, Gray KA, Cassady JP. Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: II. genome-wide association. *J Anim Sci.* 2014;92:2846–60.
71. 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.
72. Zhuang Z, Ding R, Peng L, Wu J, Ye Y, Zhou S, et al. Genome-wide association analyses identify known and novel loci for teat number in Duroc pigs using single-locus and multi-locus models. *BMC Genomics.* 2020;21:344.
73. Rosendo A, Iannuccelli N, Gilbert H, Riquet J, Billon Y, Amigues Y, et al. Microsatellite mapping of quantitative trait loci affecting female reproductive tract characteristics in Meishan × Large White F2 pigs. *J Anim Sci.* 2012;90:37–44.
74. Petrovski S, Wang QL. Permutation based QQ plot and inflation factor estimation. Version 1.0.1. 2016. <https://github.com/cran/QQperm/blob/master/DESCRIPTION/>. Accessed 15 Dec 2021.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

