

SCIENTIFIC REPORTS



OPEN

Searching for new loci and candidate genes for economically important traits through gene-based association analysis of Simmental cattle

Jiangwei Xia*, Huizhong Fan*, Tianpeng Chang, Lingyang Xu, Wengang Zhang, Yuxin Song, Bo Zhu, Lupei Zhang, Xue Gao, Yan Chen, Junya Li & Huijiang Gao

Single-marker genome-wide association study (GWAS) is a convenient strategy of genetic analysis that has been successful in detecting the association of a number of single-nucleotide polymorphisms (SNPs) with quantitative traits. However, analysis of individual SNPs can only account for a small proportion of genetic variation and offers only limited knowledge of complex traits. This inadequacy may be overcome by employing a gene-based GWAS analytic approach, which can be considered complementary to the single-SNP association analysis. Here we performed an initial single-SNP GWAS for bone weight (BW) and meat pH value with a total of 770,000 SNPs in 1141 Simmental cattle. Additionally, 21836 cattle genes collected from the Ensembl Genes 83 database were analyzed to find supplementary evidence to support the importance of gene-based association study. Results of the single SNP-based association study showed that there were 11 SNPs significantly associated with bone weight (BW) and two SNPs associated with meat pH value. Interestingly, all of these SNPs were located in genes detected by the gene-based association study.

Carcass and meat-quality traits have attracted much attention from breeders in the beef cattle industry. Meat-quality traits are mainly measured by composition, quality, and palatability factors such as visual appearance, smell, firmness, juiciness, tenderness, and flavor. Some researchers find that meat pH value is highly correlated with other meat-quality measurements (e.g. drip loss and texture score) and carcass yield (e.g. carcass weight, loin depth, loin length)¹. Improving the meat-pH value has become a high priority for the beef industry to satisfy consumer preferences. On the other hand, the proportion of genetic variation explained by single-nucleotide polymorphism (SNP) based genome-wide association study (GWAS) of bone weight (BW) and pH value are often significantly lower than the heritability estimates for the traits². For example, the heritability for BW is as high as 41% in our analysis. However, the 12 genetic loci identified for BW to account for only ~4.2% of the phenotypic variance in BW, which means that many genetic variants with smaller effects failed to be detected by GWAS. Therefore, association analysis of complex traits for BW and pH value in Simmental cattle is not sufficient, and further study is required to detect more loci.

It is well known that traditional GWAS is an individual-marker-based analysis that has been very successful in identifying disease loci in humans and economically important traits in domestic animals³⁻⁵. However, single-SNP analysis often focuses on only a few of the most significant SNPs in the genome, and these loci only explain a small proportion of the genetic risk for diseases or complex traits^{6,7}. This limitation may be improved by employing a gene-based GWAS analytic approach. A gene-based association analysis can combine genetic information for all SNPs in a gene, increase the capability to find novel genes, and generate more informative results.

Different approaches have been used to identify genes that are associated with traits of interest⁸⁻¹⁰. One of the best known gene-based algorithms is the Gene-based Association Test using Extended Simes (GATES) method,

Institute of Animal Science, Chinese Academy of Agricultural Science, Beijing, China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.L. (email: jl1@iascaas.net.cn) or H.G. (email: gaohj111@sina.com)

Trait	Mean	Standard deviation	Maximum	Minimum
BW/kg	40.00	6.58	80.00	19.00
pH	5.63	0.38	7.16	4.00

Table 1. Descriptive statistics of two cattle economical traits.

which combines the p values of the SNPs within a gene to obtain an overall p value for the association of the entire gene⁹. This method does not consider other factors, such as gene size and linkage disequilibrium (LD) between markers. As a result, it often produces more false discoveries. Another well-known gene-based GWAS algorithm was proposed by Capomaccio *et al.* in 2015, and this method uses the Multiple Species Gene-Based Association Suite (MUGBAS) for discrete traits and a set-based test for discrete and continuous traits¹⁰. Nonetheless, the set-based test requires heavy computation, therefore, limits its application at a genome-wide level. An efficient genome-wide gene-based association method was developed, we performed a modified gene-based analysis for GWAS studies¹¹. For a given gene contains several SNPs, we first used principal component analysis (PCA) to extract PCs, and then ranked all of these PCs based on the significance of their statistical association with a trait of interest. Finally, we calculated the gene's statistical value using Fisher's combination test for gene association¹². This procedure was used to test whether the set of genes was significantly associated with the traits of interest. In this study, we focused on genes associated with the traits of BW and meat pH value.

Materials and Methods

Ethics statement. The study was approved by the Science Research Department of the Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS) (Beijing, China). All procedures were in strict accordance with the guidelines proposed by the China Council on Animal Care. Using animals and private land in this study were approved by the respective owners.

Animal resources and phenotypes. We established the Simmental cattle population in the Ulgai, Xilingol league, Inner Mongolia, China. The population consisted of 1141 young Simmental cattle born between 2009 and 2014. After weaning, the cattle were transferred to the Jinweifuren cattle farm (Beijing) for fattening in a uniform feeding and management environment. The cattle were observed for growth and developmental traits until slaughter at 16–18 months of age. Our study focused primarily on phenotypic traits associated with carcass quality and meat quality; therefore, during the period of slaughter, we measured the traits in strict accordance with the guidelines proposed by the Institutional Meat Purchase Specifications for fresh beef. First, this study was performed on the trait bone weight (BW), which was measured in half of the cattle carcass. After removing the exposed meat from the bone, the weight of the remaining bone was defined as BW. For the pH value, we used a steak from the twelfth rib at slaughter and measured pH at three locations using a Mettler Toledo pH meter (Mettler Toledo, Greifensee, Switzerland). Summary statistics of the two traits were given in Table 1.

Genotyping and quality control. Blood samples were collected along with the regular quarantine inspection of the farms. Genomic DNA was extracted from blood samples using a TIAN amp Blood DNA Kit (Tiangen Biotech Company Limited, Beijing, China), and DNA with an A260/280 ratio ranging between 1.8 and 2.0 was subjected to further analysis. The Illumina Bovine HD Bead Chip contained 774,660 SNPs was used for individual genotyping. The SNPs were uniformly distributed across the whole bovine genome. The average distance between consecutive markers is 3.43 Kb with a standard deviation of 4.38 Mb. The genotyping platform adopted in this study was Illumina (San Diego, CA, USA) Infinium II Assay. Samples were genotyped using Illumina BEADSTUDIO (Inc.9885 Towne Centre Drive, San Diego, CA 92121 USA) and SNP chips were scanned using Infinium Genome Studio.

Regarding quality control, we used the PLINK software (v1.9, <http://pngu.mgh.harvard.edu/~purcell/plink/>) to remove individuals and SNPs based on the following criteria. All markers with call rates <90%, minor allele frequencies (MAFs) <5%, occurrence of the genotype in <5 individuals, or severe departure from Hardy-Weinberg equilibrium (with $p < 10^{-6}$) were excluded from the analysis. All animals with missing genotypes >10% or SNP Mendel error rate >2% were removed from the data. In addition, all SNPs with uncertain map positions were excluded from the analysis. Ultimately, a total of 1141 individuals and 677,855 SNPs remained in the final data for the subsequent analysis.

Gene-based association analysis

Phenotypic preparation. Before the association study, phenotypic values of the two traits were adjusted by effects of year, farm, gender, fattening days (the date since entering fattening farm to slaughtering), entering weight (live weight when entering fattening farm). In addition, we also used the corresponding population structure matrix constructed by the first five PCs from a portion of SNPs over the bovine genome to correct the population structure (similar with the PCs calculation in single marker GWAS), using a general linear model. In other words, the residuals after fitting the above fixed effects were treated as the phenotypic values of traits for the association studies.

Gene collection and definition of gene locus. One of the major steps of a gene-based analysis is the assignment of SNPs to genes. All genes including coding sequences, non-coding sequences and pseudogenes, were defined based on the Bos taurus UMD 3.1 Sequence. Sequences were all downloaded from the Ensembl Genes 83 database in the BioMart (<http://asia.ensembl.org/biomart/martview>). SNPs were mapped to genes

according to their physical distance: a SNP was mapped to the gene whose coding sequence had an overlap with a 20 kb range around the gene¹³. Only genes with at least five mapped SNPs were included in the analysis. Ultimately, a total of 21836 genes was mapped onto the downloadable genes in this analysis.

Two-stage-gene-based association study. *Stage 1: PC construction within a gene.* In the first stage, we constructed principal components (PCs) within each gene. We treated each SNP within a gene as a variable and calculated the variance-covariance matrix for all SNPs. We then calculated the eigenvalues and eigenvectors of the covariance matrix. The PCs were obtained by multiplying the eigenvectors by the SNP genotype matrix. Finally, we selected the top PCs that contributed 85% of the total variation of SNP data. These PCs were treated as the independent variables for subsequent regression analyses.

Stage 2: Fisher's combination test for each gene. In the second stage, the PCs within each gene were treated as independent variables for regression analysis and significance tests. The response variables were the two traits after correction for the fixed effects. Because the PCs were independent, the test for association of a trait and a PC was done independently. The simple correlation coefficient for the corrected phenotypes and PCs could be used to calculate the p values of PCs.

In the Fisher's combination test for each gene, χ^2 was constructed by combining K independent p values for PCs, as follows:

$$Z_F = -2 \sum_{i=1}^K \log P_i \quad (1)$$

Under the null model, this test statistic followed a χ^2 distribution with degree of freedom 2K. A new p-value for the gene was calculated from this Chi-square distribution.

Bonferroni correction was adopted for multiple tests of all genes in the genome. A gene was considered significant at the genome-wide significance level if the nominal p value was less than 0.05/N. For the gene-based association analysis, a total of 21836 genes was mapped onto the available genes in this analysis. We therefore chose $N = 21836$ and defined the gene-based genome-wide significance level as $0.05/21836 = 2.29 \times 10^{-6}$.

Single-marker association analysis. The compressed mixed linear model (CMLM) was used cluster analysis to assign similar individuals to the same groups and fit the "Q + K" matrix into the model to improve statistical power¹⁴. We used CMLM to detect individual markers using the default parameters provided by the Genome Association and Prediction Integrated Tool (GAPIT)¹⁵. First, a principle components analysis (PCA) was performed and a kinship matrix was calculated using the GAPIT package in R. The "Q" matrix was determined by the PCA to account for effects due to population structure, and the kinship matrix (K) was calculated using the VanRaden algorithm¹⁶ to replace the incomplete pedigrees. To avoid the bias caused by potential LD and reduce computing time, a set of SNPs were randomly selected with 'SNP.fraction' as input parameter for the PCs calculations¹⁵. The fraction of SNP number can be controlled by "Ratio" parameter in GAPIT software. We set SNP.fraction to 10%, and finally correct the population structure using the first five PCs estimated from a portion of SNPs. For each trait, the model was

$$Y = Wv + \beta X + \mu Z + e \quad (2)$$

where Y is the vector of phenotypic value; v is the vector of unknown fixed effect of the current marker; β is the vector of fixed effects, including years, farms, gender, fattening days, entering weight, and PCs; μ is a vector of random additive genetic effects corresponding to the clustered groups with an assumed $N(0, \sigma^2 K)$ distribution, where σ^2 is the additive genetic variance and K is the compressed kinship matrix; W is a vector of the SNP genotype indicators and has a value of 0, 1, or 2, corresponding to AA, AB, BB (B being the minor allele); X and Z are the incidence matrices for β and μ , respectively; and e is a vector of random residual effects with an assumed $N(0, I\sigma_e^2)$, where σ_e^2 is the residual error variance. For each SNP, a t-test was used to examine the association between the SNP and the trait.

Quantile-Quantile (Q-Q) plots were used to assess how well the model captured population structure and familial relatedness. In the Q-Q plot, the y-axis represented the negative logarithms of the p values from the models, and the x-axis represented the expected value under the null hypothesis (SNPs didn't associate with the traits). Because a total of 677,855 SNPs was used for this analysis, the number of potential associated SNPs after Bonferroni correction ($0.05/677855 = 7.38 \times 10^{-8}$) was too small, resulting in very low statistical power¹⁷. Therefore, a suggestive significance threshold for p value ($1/677855 = 1.47 \times 10^{-6}$) was used in this analysis¹⁸⁻²⁰.

Results

Population stratification assessment. Figure 1 shows the population stratification of the Simmental population. The PCA result shows five separate clusters. Population stratification due to different genetic backgrounds and farms was considered as an underlying confounder in the association analysis.

Significant SNPs. Significant SNPs were shown in Table 2, which also included Bos Taurus Autosome, SNP position in the genome, nearest known gene, MAFs, and p value ($< 1.47 \times 10^{-6}$). Figure 2A and B shows the Manhattan plots for all SNPs. To summarize, the number of significant SNPs identified by CMLM was 11 for BW and 2 for pH value. For BW, most of the significant SNPs were located in or near *LAP3*, *LCORL*, *FAM184B*, and *NCAPG* on BTA6. The two SNPs significant for pH value were located near *S100A10* on BTA3. The Q-Q plots for the two traits (Fig. 3A and B) suggested that there was no inflation or systematic bias in this study. Most of the

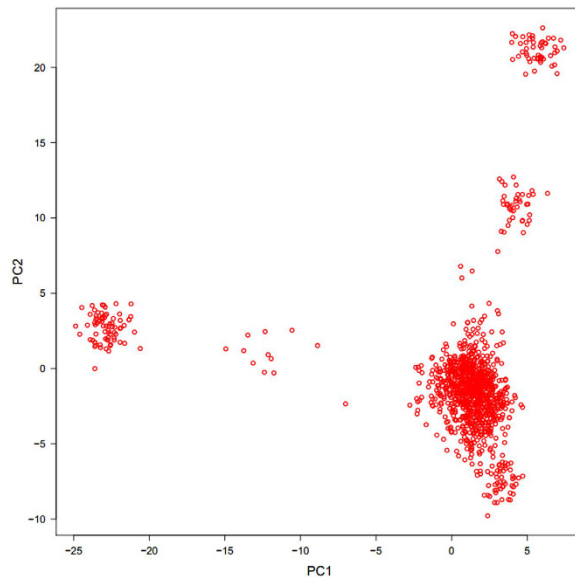


Figure 1. Principal components (PC) plot. The second principal component (PC2) plotted against the first principal component (PC1).

SNP	BTA	Position	P value	MAF	Nearest Gene Name	Distance
BovineHD0600010713 ^a	6	38686919	2.36E-07	0.3623	FAM184B	14613
BovineHD0600010716 ^a	6	38704872	8.77E-07	0.3660	FAM184B	32566
Hapmap26308-BTC-057761 ^a	6	38576012	7.84E-08	0.3836	LAP3	in
Hapmap30134-BTC-034283 ^a	6	38464203	2.48E-07	0.2661	LAP3	110456
BovineHD0600010646 ^a	6	38468901	6.22E-07	0.1321	LAP3	105758
BovineHD4100004580 ^a	6	38852093	1.24E-07	0.3952	LCORL	in
BovineHD0600010755 ^a	6	38866381	7.34E-07	0.3672	LCORL	in
BovineHD4100004575 ^a	6	38830725	7.34E-07	0.3672	LCORL	13269
BovineHD4100004577 ^a	6	38837159	7.34E-07	0.3672	LCORL	6835
BovineHD4100004578 ^a	6	38840174	7.34E-07	0.3672	LCORL	3820
BovineHD0600010736 ^a	6	38762470	1.42E-06	0.3172	NCAPG	3544
BovineHD030000648 ^b	3	18808560	4.96E-07	0.3856	S100A10	in
BovineHD0300006049 ^b	3	18812141	7.81E-07	0.2458	S100A10	1843

Table 2. Significant SNPs identified for BW and pH value traits by single-marker GWAS method ($p < 1.47 \times 10^{-6}$). SNP = SNP name in BovineHD panel. BTA = Bos Taurus Autosome. Position = position (bp) on UMD3.1 P value = p-values calculated by CMLM. MAF = Minor Allele Frequency. Nearest Gene Name = name of the nearest gene. Distance = distance between SNP and nearest gene. Note: SNP with a superscript “a” was identified by the trait of BW only, SNP with a superscript “b” was identified by the trait of pH only. And similar with Tables 3 and 4.

points were concentrated along a diagonal line because the GWAS model sufficiently accounted for the population structure, and only a small number of SNPs were associated with the traits.

Significant Genes. For the gene-based association analysis, the Manhattan plots were shown in Fig. 2C for BW and Fig. 2D for pH value. Table 3 displays the significant genes detected by the gene-based method, including their starting and ending positions in the genome, the ensemble IDs, Bos Taurus Autosome, and the p values. The total number of significant genes identified by the gene-based GWAS was 12 for BW and 3 for pH value. For trait BW, a total of 10 detected genes were located on BTA6, 1 was located on BTA1, and 1 was located on BTA16, and four already detected genes (*LAP3*, *LCORL*, *FAM184B*, and *NCAPG*) on BTA6 in the gene-based association analysis were all found in the single-marker analysis. For pH value, we found 3 genes, all located on BTA3. Similar to the single-marker analysis, we found one already detected gene (*S100A10*) on BTA3 in the gene-based association. In summary, most of the significant SNPs detected by the individual-SNP analysis were also detected in the gene-based association analysis. The Q-Q plots for the test statistics of the gene-based analysis are shown in Fig. 3C and D for BW and pH value, respectively.

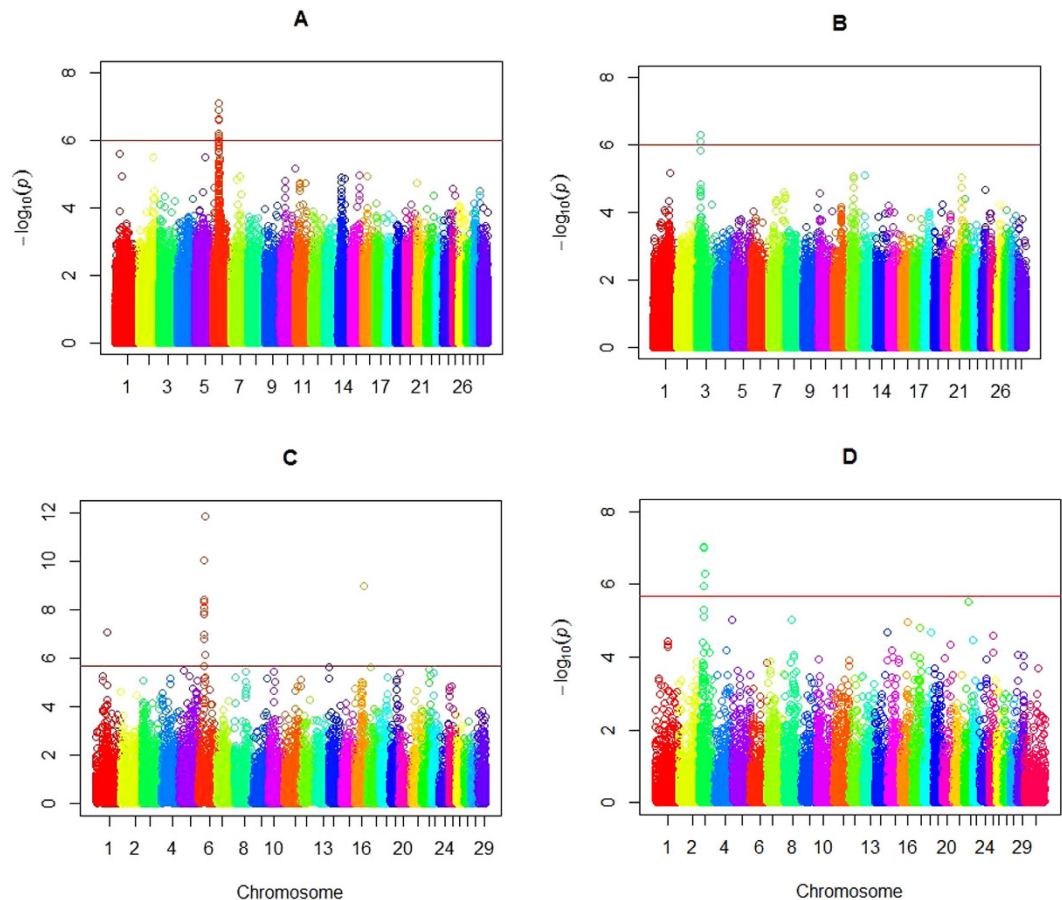


Figure 2. Manhattan plots of $-\log_{10}(p)$ values for two traits from the single-SNP method and the gene-based method. Panels A and B are the plots for bone weight and pH values, respectively, in the single-marker analysis. Panels C and D are the plots for bone weight and pH values, respectively, in the gene-based association analysis. The 29 chromosomes are color coded. The red horizontal line indicates genome-wide significance level from Bonferroni correction ($-\log_{10}(1/677855)$) for the single-marker analysis. The genome-wide significance level for the gene-based method is ($-\log_{10}(0.05/21836)$).

Discussion

In this study, a gene-based association strategy was used to identify new associations of genetic variants with BW and meat pH value in Simmental cattle. The gene-based association analysis combined genetic information given by all the single nucleotide polymorphisms (SNPs) in a gene. Our method orthogonalized the SNPs within each gene using PCA so that the Fisher's method can be used to calculate a new p-value for a gene from p-values of multiple independent PCs. Of course, choosing PCs to formulate the statistic for a gene takes advantage of information on those SNPs. Therefore, the gene-based method can increase the probability to find novel genes, and generate more informative results than the single-marker method. While using the single marker GWAS, none of the SNPs reached the genome-wide significance threshold due to the stringent criterion, which forced us to lower the threshold to pick some of the significant SNPs.

In addition to identify additional loci using the gene-based association analysis approach, we confirmed previously reported associations with genes related to BW. There were a large number of significant SNPs associated with BW located on BTA6^{21,22}. Similar to the single-marker results, four already detected genes (*LAP3*, *LCORL*, *FAM184B*, and *NCAPG*) on BTA6 were also found in the gene-based association analysis for the trait of BW. The genes *LAP3*, *LCORL*, and *FAM184B* have been shown to be candidate genes for carcass or growth traits in cattle^{23–25}, the most significant SNP (Hapmap26308-BTC-057761) located within *LAP3* was also found in black cattle as the second strongest association²⁶. In addition, some significant SNPs were in the genes *NCAPG*, *LAP3* and *LCORL*. For *NCAPG*, it encodes a non-structural maintenance of chromosomes (-SMC) condensin I complex and leads to the change of amino acid Ile442 to Met442, which is associated with foetal growth and carcass weight in cattle^{25,27,28}. In addition, the polymorphism in *NCAPG* shows significant associations with carcass weight, carcass yield estimate, and lipid deposition and has been identified as a candidate causative variant for bovine carcass weight quantitative trait locus (QTL)^{23,27,29–31}. For *LAP3*, it participates in oxytocin hydrolysis and acts as promoter polymorphism and milk production, and is highly expressed in skin and mammary and adipose tissue^{23,32,33}. For *LOCAL*, it encodes a ligand-dependent nuclear receptor co-repressor, may affect skeletal frame size and adult height in human, horse, and cattle^{34–36}, and has a strong correlation with the nearest gene *NCAPG*. The

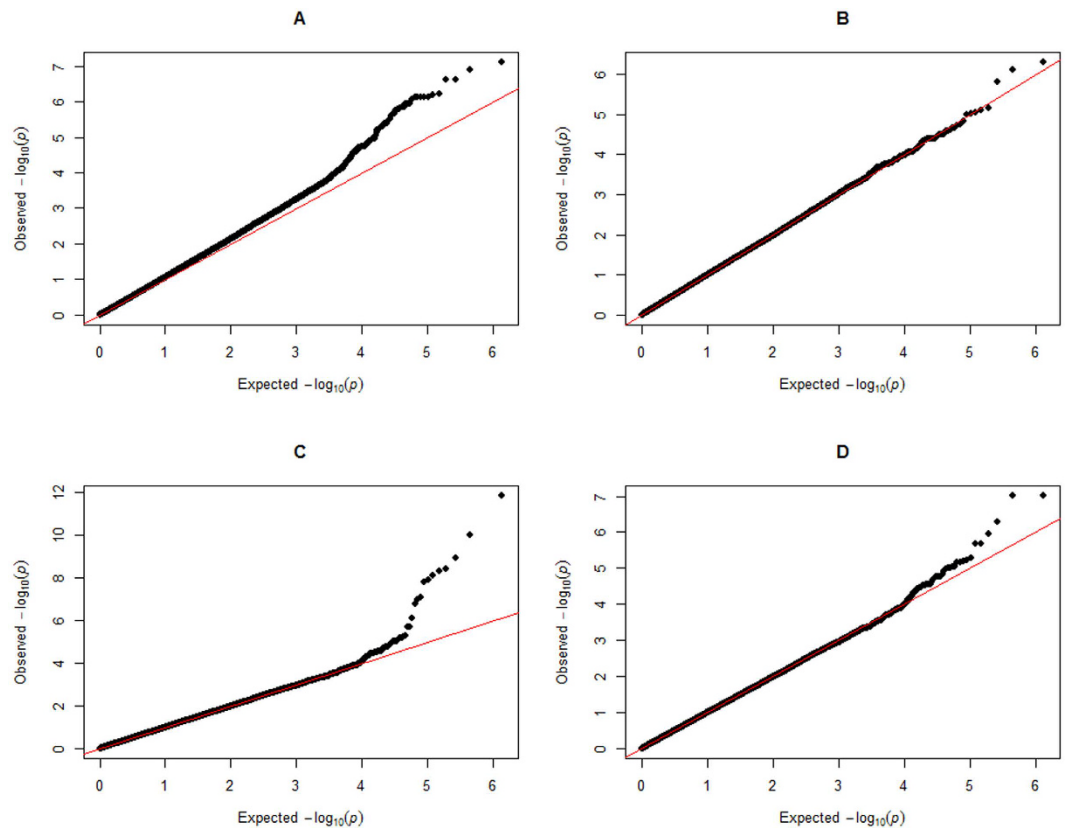


Figure 3. The Quantile-Quantile plot of p-values. The observed negative logarithms of the p values in GWAS using the gene-based method and the single-marker method for two traits are plotted against their expected values under the null hypothesis of no association with the trait. Panels A and B show the plots for bone weight and pH value, respectively, in the single-marker analysis. Panels C and D show the plots for bone weight and pH value, respectively, in the gene-based association analysis.

Ensemble ID	BTA	Start(bp)	End(bp)	P value	Gene symbol
ENSBTAG00000005108 ^a	6	41236589	41640789	1.41E-12	SLIT2
ENSBTAG000000021582 ^a	6	38765969	38812051	9.60E-11	NCAPG
ENSBTAG000000018303 ^a	16	59132781	59456250	1.12E-09	PAPPA2
ENSBTAG000000046561 ^a	6	38840894	38992112	3.93E-09	LCORL
ENSBTAG000000045804 ^a	6	38541579	38541674	4.71E-09	novel gene
ENSBTAG00000005932 ^a	6	38614370	38672306	8.01E-09	FAM184B
ENSBTAG000000011187 ^a	6	37355568	37457493	1.29E-08	FAM13A
ENSBTAG000000019441 ^a	6	38603034	38608841	1.53E-08	MED28
ENSBTAG000000032819 ^a	1	71072051	71085633	8.44E-08	MUC20
ENSBTAG00000005989 ^a	6	38574590	38600027	1.07E-07	LAP3
ENSBTAG000000011973 ^a	6	38754415	38754897	1.60E-07	DCAF16
ENSBTAG000000047743 ^a	6	41709362	43021626	7.47E-07	KCNIP4
ENSBTAG000000018203 ^b	3	26319356	26347752	9.65E-08	CD101
ENSBTAG000000033217 ^b	3	16376436	16403812	9.78E-08	TPM1
ENSBTAG000000015147 ^b	3	18799612	18810545	5.19E-07	S100A10

Table 3. Significant genes identified for BW and pH value traits by gene-based GWAS method. Ensemble ID = gene ID name in Ensemble database. BTA = Bos Taurus Autosome. Start(bp) = the gene start position (bp) on Ensemble database. End(bp) = the gene end position (bp) on Ensemble database. P value = p-values calculated by Fisher's combination test. Gene symbol: the corresponding gene name.

LCORL/NCAPG locus has been associated with mammalian stature³⁶, direct calving ease, feed intake, gain, meat and carcass traits across multiple breeds of cattle^{23,24,37}. Snelling and collaborators detected a 0.57-Mb (37.96 Mb to 38.53 Mb) segment in *LCORL/NCAPG* locus on BTA6 associated with feed intake, gain, meat and growth traits

Gene(s) Abbreviation	Comments	References
<i>SLIT2</i> ^a	A genome-wide significant association with cattle carcass weight, broiler chicken cross has been reported. Cellular Function and Maintenance, Skeletal and Muscular System Development and Function.	46–48
<i>PAPPA2</i> ^a	Pregnancy associated plasma protein A2 (PAPP-A2) affects bone size and shape and contributes to natural variation in postnatal growth in mice. Researchers found SNP near this PAPP2 to be associated with postpartum anestrus interval in Tropical Composite cattle.	49–51
<i>FAM13A</i> ^a	FAM13A (family with sequence similarity 13, member A) gene, which has been shown to be associated with mastitis in Jersey cows. FAM13A is associated with skeletal muscle mass. The principal determinants of skeletal muscle mass are the muscle fiber number and muscle fiber size.	52,53
<i>MED28</i> ^a	MED28 is significant association with bovine carcass weight. It is most highly expressed in the bovine pineal gland, lymphoreticular tissue, uterus, abomasum, and expressed in lower quantities in the liver, reticulum, intestine and kidney	54,55
<i>MUC20</i> ^a	This gene encodes a member of the mucin protein family. Mucins are high molecular weight glycoproteins secreted by many epithelial tissues to form an insoluble mucous barrier.	56,57
<i>KCNIP4</i> ^a	The KCNIP4 gene (Kv channel-interacting protein 4) has a role in the calcium ion binding, and in potassium and voltage-gated ion channel activity. It has a significant association with yearling weight Canchim Beef Cattle.	58,59
<i>DCAF16</i> ^a	DCAF16 takes part in the process of protein ubiquitination and acts as Cul4-RING E3 ubiquitin ligase complex in cellular component. Some researchers indicate a significant difference in the expression of the DCAF16 genes in fetal and adult bovine longissimus muscle.	35,60
<i>CD101</i> ^b	A genome-wide significant association with cattle Milking Speed, pig feed efficiency has been reported. Plays a role as inhibitor of T-cells proliferation induced by CD3. Inhibits expression of IL2RA on activated T-cells and secretion of IL2. Inhibits tyrosine kinases that are required for IL2 production and cellular proliferation, blocking Ca ²⁺ + Flux.	61,62
<i>TPM1</i> ^b	This gene associated with Muscle and Fat Tissues of Native Korean Cattle. Tropomyosin 1 (Alpha) is a member of the tropomyosin family of highly conserved, widely distributed actin-binding proteins involved in the contractile system of striated and smooth muscles and the cytoskeleton of non-muscle cells.	63,64

Table 4. Selected non-overlapping expressed genes for two traits.

in a crossbred population of beef cattle. In addition, *LAP3*, *NCAPG*, and *LCORL* have been considered potential positional and functional candidate genes for direct calving ease, lean growth, and fat deposition^{24,27}. Overall, we suggest a 2.6-Mb (37.3 Mb–39.9 Mb) segment on BTA6 as a candidate region for BW. Similar to the results of single-marker analysis, we found a common gene (*S100A10*) for the trait pH value with the gene-based association analysis. *S100A10* encodes S100 calcium-binding protein A10, which plays a role in calcium ion binding and ion channel binding³⁸. *S100A10* induces the dimerization of ANXA2/p36, and it may function as a regulator of protein phosphorylation in that the ANXA2 monomer is the preferred target (*in vitro*) of tyrosine-specific kinase³⁹. Some researchers found that *S100A10* is associated with residual feed intake in Angus cattle⁴⁰. Relevant to our research, we found that it is also associated with the meat trait marbling score in Simmental beef cattle⁴¹. This gene is thought to be involved in regulation of cell cycle progression and differentiation. It has been implicated in major depression (downregulated in depressed humans and in animal models), suicide (downregulated in peripheral blood of attempters and in prefrontal cortex of suicide completers), and bipolar disorder (upregulated in peripheral blood)^{42–44}. In Table 4, for the remaining associated genes, we provided the gene symbols, a brief description of the potential relevance of the gene/gene product to BW and pH based on its function and relevant references.

The gene-based association study may have its own limitations. For example, if a gene contains only one causal variant, inclusion of a large number of null SNPs in the gene-based method may dilute this gene's significance. In addition, the ± 20 kb boundaries defined for a gene are based on prior reports⁴⁵. This way of defining the gene boundaries is quite subjective. Inadaptable boundaries may occur so that some SNPs may be included in multiple genes. It may also be difficult to definitively identify the causal gene when multiple adjacent genes are statistically significant. For these reasons, the gene-based method may not be seen as a replacement for the traditional single-marker analysis, but rather as complementary to it.

In summary, we applied a gene-based association analysis that is complementary to GWAS and identified important genes associated with traits BW and pH value in beef cattle. A series of bioinformatics analyses provide supportive evidence that the gene-based association analysis is useful. We believe that subsequent studies, including further exploration and analysis of these genes, may reveal more causal networks underlying carcass traits in cattle.

References

- Zhang, C. *et al.* Genome-wide association studies (GWAS) identify a QTL close to PRKAG3 affecting meat pH and colour in crossbred commercial pigs. *BMC Genetics* **16**, 1–12, doi: 10.1186/s12863-015-0192-1 (2015).
- Matika, O. *et al.* Genome-wide association reveals QTL for growth, bone and *in vivo* carcass traits as assessed by computed tomography in Scottish Blackface lambs. *Genetics, Selection, Evolution: GSE* **48**, 11, doi: 10.1186/s12711-016-0191-3 (2016).
- Donnelly, P. Progress and challenges in genome-wide association studies in humans. *Nature* **456**, 728–731 (2008).
- Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* **6**, 95–108 (2005).
- Bolormaa, S. *et al.* A genome-wide association study of meat and carcass traits in Australian cattle. *Journal of animal science* **89**, 2297–2309, doi: 10.2527/jas.2010-3138 (2011).
- Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nature genetics* **42**, doi: 10.1038/ng.546 (2010).
- Bush, W. S. & Moore, J. H. Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology* **8**, e1002822, doi: 10.1371/journal.pcbi.1002822 (2012).

8. Gudmundsson, J. *et al.* Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nature genetics* **41**, 1122–1126, doi: 10.1038/ng.448 (2009).
9. Li, M. X., Gui, H. S., Kwan, J. S. & Sham, P. C. GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am J Hum Genet* **88**, 283–293, doi: 10.1016/j.ajhg.2011.01.019 (2011).
10. Capomaccio, S., Milanese, M., Bomba, L., Vajana, E. & Ajmone-Marsan, P. MUGBAS: a species free gene-based programme suite for post-GWAS analysis. *Bioinformatics (Oxford, England)* **31**, 2380–2381, doi: 10.1093/bioinformatics/btv144 (2015).
11. Mukherjee, S. *et al.* Gene-based GWAS and -biological pathway analysis of the resilience of executive functioning. *Brain imaging and behavior* **8**, 110–118, doi: 10.1007/s11682-013-9259-7 (2014).
12. Fan, H. *et al.* Pathway-Based Genome-Wide Association Studies for Two Meat Production Traits in Simmental Cattle. *Scientific reports* **5**, 18389, doi: 10.1038/srep18389 (2015).
13. Capomaccio, S. *et al.* Searching new signals for production traits through gene-based association analysis in three Italian cattle breeds. *Animal genetics* **46**, 361–370, doi: 10.1111/age.12303 (2015).
14. Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* **42**, 355–360, doi: http://www.nature.com/ng/journal/v42/n4/supinfo/ng.546_S1.html (2010).
15. Lipka, A. E. *et al.* GAPIT: genome association and prediction integrated tool. *Bioinformatics (Oxford, England)* **28**, 2397–2399, doi: 10.1093/bioinformatics/bts444 (2012).
16. VanRaden, P. M. Efficient Methods to Compute Genomic Predictions. *J Dairy Sci* **91**, 4414–4423, doi: http://dx.doi.org/10.3168/jds.2007-0980 (2008).
17. Nakagawa, S. A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral Ecology* **15**, 1044–1045, doi: 10.1093/beheco/arl107 (2004).
18. Sun, Y. *et al.* Genome-wide linkage analysis and association study identifies loci for polydactyly in chickens. *G3 (Bethesda)* **4**, 1167–1172, doi: 10.1534/g3.114.011338 (2014).
19. Wu, Y. *et al.* Genome-Wide Association Studies Using Haplotypes and Individual SNPs in Simmental Cattle. *PLoS One* **9**, e109330, doi: 10.1371/journal.pone.0109330 (2014).
20. Xiong, X. *et al.* Genome-wide association analysis reveals genetic loci and candidate genes for meat quality traits in Chinese Laiwu pigs. *Mamm Genome* **26**, 181–190, doi: 10.1007/s00335-015-9558-y (2015).
21. Snelling, W. M. *et al.* Genome-wide association study of growth in crossbred beef cattle. *Journal of animal science* **88**, 837–848, doi: 10.2527/jas.2009-2257 (2010).
22. Snelling, W. M. *et al.* Partial-genome evaluation of postweaning feed intake and efficiency of crossbred beef cattle. *Journal of animal science* **89**, 1731–1741, doi: 10.2527/jas.2010-3526 (2011).
23. Lindholm-Perry, A. K. *et al.* Association, effects and validation of polymorphisms within the NCAPG - LCORL locus located on BTA6 with feed intake, gain, meat and carcass traits in beef cattle. *BMC genetics* **12**, 103, doi: 10.1186/1471-2156-12-103 (2011).
24. Bongiorno, S., Mancini, G., Chillemi, G., Pariset, L. & Valentini, A. Identification of a Short Region on Chromosome 6 Affecting Direct Calving Ease in Piedmontese Cattle Breed. *PLoS One* **7**, e50137, doi: 10.1371/journal.pone.0050137 (2012).
25. Eberlein, A. *et al.* Dissection of genetic factors modulating fetal growth in cattle indicates a substantial role of the non-SMC condensin I complex, subunit G (NCAPG) gene. *Genetics* **183**, 951–964, doi: 10.1534/genetics.109.106476 (2009).
26. Nishimura, S. *et al.* Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle. *BMC genetics* **13**, 1–11, doi: 10.1186/1471-2156-13-40 (2012).
27. Setoguchi, K. *et al.* Cross-breed comparisons identified a critical 591-kb region for bovine carcass weight QTL (CW-2) on chromosome 6 and the Ile-442-Met substitution in NCAPG as a positional candidate. *BMC genetics* **10**, 43, doi: 10.1186/1471-2156-10-43 (2009).
28. Olsen, H. G. *et al.* Mapping of a milk production quantitative trait locus to a 420-kb region on bovine chromosome 6. *Genetics* **169**, 275–283, doi: 10.1534/genetics.104.031559 (2005).
29. Setoguchi, K. *et al.* The SNP c.1326T>G in the non-SMC condensin I complex, subunit G (NCAPG) gene encoding a p.Ile442Met variant is associated with an increase in body frame size at puberty in cattle. *Animal genetics* **42**, 650–655, doi: 10.1111/j.1365-2052.2011.02196.x (2011).
30. Hoshiba, H. *et al.* Comparison of the effects explained by variations in the bovine PLAG1 and NCAPG genes on daily body weight gain, linear skeletal measurements and carcass traits in Japanese Black steers from a progeny testing program. *Animal science journal = Nihon chikusan Gakkaiho* **84**, 529–534, doi: 10.1111/asj.12033 (2013).
31. Weikard, R. *et al.* Metabolomic profiles indicate distinct physiological pathways affected by two loci with major divergent effect on Bos taurus growth and lipid deposition. *Physiological genomics* **42A**, 79–88, doi: 10.1152/physiolgenomics.00120.2010 (2010).
32. Ju, Z. *et al.* Functional characterization of genetic polymorphisms identified in the promoter region of the bovine PEPS gene. *DNA and cell biology* **31**, 1038–1045, doi: 10.1089/dna.2011.1555 (2012).
33. Zheng, X. *et al.* Single nucleotide polymorphisms, haplotypes and combined genotypes of LAP3 gene in bovine and their association with milk production traits. *Molecular biology reports* **38**, 4053–4061, doi: 10.1007/s11033-010-0524-1 (2011).
34. Horikoshi, M. *et al.* New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nat Genet* **45**, 76–82, doi: 10.1038/ng.2477 (2013).
35. Metzger, J., Schrimpf, R., Philipp, U. & Distl, O. Expression levels of LCORL are associated with body size in horses. *PLoS One* **8**, e56497, doi: 10.1371/journal.pone.0056497 (2013).
36. Pryce, J. E., Hayes, B. J., Bolormaa, S. & Goddard, M. E. Polymorphic regions affecting human height also control stature in cattle. *Genetics* **187**, 981–984, doi: 10.1534/genetics.110.123943 (2011).
37. Nishimura, S. *et al.* Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle. *BMC genetics* **13**, 40, doi: 10.1186/1471-2156-13-40 (2012).
38. Tzang, R.-F. *et al.* Association study of p11 gene with major depressive disorder, suicidal behaviors and treatment response. *Neuroscience Letters* **447**, 92–95, doi: http://dx.doi.org/10.1016/j.neulet.2008.09.063 (2008).
39. Gerke, V., Koch, W. & Thiel, C. Primary structure and expression of the Xenopus laevis gene encoding annexin II. *Gene* **104**, 259–264, doi: http://dx.doi.org/10.1016/0378-1119(91)90259-E (1991).
40. Al-Husseini, W. *et al.* Expression of candidate genes for residual feed intake in Angus cattle. *Animal genetics* **45**, 12–19, doi: 10.1111/age.12092 (2014).
41. Xia, J. *et al.* Genome-wide association study identifies loci and candidate genes for meat quality traits in Simmental beef cattle. *Mammalian Genome* **27**, 246–255 (2016).
42. Svenningsson, P. & Greengard, P. p11 (S100A10) — an inducible adaptor protein that modulates neuronal functions. *Current Opinion in Pharmacology* **7**, 27–32, doi: http://dx.doi.org/10.1016/j.coph.2006.10.001 (2007).
43. Zhang, L. *et al.* P11 (S100A10) as a potential biomarker of psychiatric patients at risk of suicide. *Journal of Psychiatric Research* **45**, 435–441, doi: http://dx.doi.org/10.1016/j.jpsychires.2010.08.012 (2011).
44. Zhang, L. *et al.* P11 expression and PET in bipolar disorders. *Journal of Psychiatric Research* **45**, 1426–1431, doi: http://dx.doi.org/10.1016/j.jpsychires.2011.06.006 (2011).
45. Weng, L. *et al.* SNP-based pathway enrichment analysis for genome-wide association studies. *BMC Bioinformatics* **12**, 99, doi: 10.1186/1471-2105-12-99 (2011).

46. Lia, Y. I., Lee, J. H., Lee, Y. M. & Kim, J. J. Application of Linkage Disequilibrium Mapping Methods to Detect QTL for Carcass Quality on Chromosome 6 Using a High Density SNP Map in Hanwoo. *Asian-Australasian Journal of Animal Sciences* **24**, pp 457–462 (2011).
47. Liu, R. *et al.* Identification of loci and genes for growth related traits from a genome-wide association study in a slow- × fast-growing broiler chicken cross. *Genes & Genomics* **37**, 829–836, doi: 10.1007/s13258-015-0314-1 (2015).
48. Zu, S., Bartik, Z., Zhao, S., Sillen, U. & Nordenskjöld, A. Mutations in the ROBO2 and SLIT2 genes are rare causes of familial vesico-ureteral reflux. *Pediatric Nephrology* **24**, 1501–1508, doi: 10.1007/s00467-009-1179-9 (2009).
49. Fortes, M. R. S. *et al.* Genomic regions associated with fertility traits in male and female cattle: Advances from microsatellites to high-density chips and beyond. *Animal Reproduction Science* **141**, 1–19, doi: <http://dx.doi.org/10.1016/j.anireprosci.2013.07.002> (2013).
50. Christians, J. K., de Zwaan, D. R. & Fung, S. H. Y. Pregnancy Associated Plasma Protein A2 (PAPP-A2) Affects Bone Size and Shape and Contributes to Natural Variation in Postnatal Growth in Mice. *PLoS One* **8**, e56260, doi: 10.1371/journal.pone.0056260 (2013).
51. Costa, R. B. *et al.* Genome-wide association study of reproductive traits in Nellore heifers using Bayesian inference. *Genetics, Selection, Evolution: GSE* **47**, 67, doi: 10.1186/s12711-015-0146-0 (2015).
52. Gutiérrez-Gil, B. *et al.* Application of Selection Mapping to Identify Genomic Regions Associated with Dairy Production in Sheep. *PLoS One* **9**, e94623, doi: 10.1371/journal.pone.0094623 (2014).
53. Lee, Y. *et al.* Identification of candidate SNP (single nucleotide polymorphism) for growth and carcass traits related to QTL on chromosome 6 in Hanwoo (Korean cattle). *Asian-Aust. J. Anim. Sci* **21**, 1703–1709 (2008).
54. Malik, S. & Roeder, R. G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature reviews. Genetics* **11**, 761–772, doi: 10.1038/nrg2901 (2010).
55. Lindholm-Perry, A. K. *et al.* Association, effects and validation of polymorphisms within the NCAPG - LCORL locus located on BTA6 with feed intake, gain, meat and carcass traits in beef cattle. *BMC genetics* **12**, 1–13, doi: 10.1186/1471-2156-12-103 (2011).
56. Li, G., Zhang, H., Lv, J., Hou, P. & Wang, H. Tandem Repeats Polymorphism of MUC20 Is an Independent Factor for the Progression of Immunoglobulin A Nephropathy. *American Journal of Nephrology* **26**, 43–49 (2006).
57. Higuchi, T. *et al.* MUC20 Suppresses the Hepatocyte Growth Factor-Induced Grb2-Ras Pathway by Binding to a Multifunctional Docking Site of Met. *Molecular and Cellular Biology* **24**, 7456–7468, doi: 10.1128/mcb.24.17.7456-7468.2004 (2004).
58. Zhang, Z., Ding, X., Liu, J., Zhang, Q. & de Koning, D. J. Accuracy of genomic prediction using low-density marker panels. *Journal of Dairy Science* **94**, 3642–3650, doi: <http://dx.doi.org/10.3168/jds.2010-3917> (2011).
59. Buzanskas, M. E. *et al.* Genome-Wide Association for Growth Traits in Canchim Beef Cattle. *PLoS One* **9**, e94802, doi: 10.1371/journal.pone.0094802 (2014).
60. Liu, Y., Duan, X., Chen, S., He, H. & Liu, X. NCAPG is differentially expressed during longissimus muscle development and is associated with growth traits in Chinese Qinchuan beef cattle. *Genetics and Molecular Biology* **38**, 450–456 (2015).
61. Bouloc, B., Geissmann, Fraitag, Andry, Teillac, Bensussan, Revuz, Boumsell, Wechsler & Bagot (2000) CD101 expression by Langerhans cell histiocytosis cells. *Histopathology* **36**, 229–232, doi: 10.1046/j (2000).
62. Abo-Ismael, M. In 10th World Congress on Genetics Applied to Livestock Production. (Asas) (2014).
63. Hershberger, R. E. *et al.* Coding sequence rare variants identified in MYBPC3, MYH6, TPM1, TNNC1 and TNNI3 from 312 patients with familial or idiopathic dilated cardiomyopathy. *Circulation. Cardiovascular genetics* **3**, 155–161, doi: 10.1161/circgenetics.109.912345 (2010).
64. Sahana, G., Kadlecová, V., Hornshøj, H., Nielsen, B. & Christensen, O. F. A genome-wide association scan in pig identifies novel regions associated with feed efficiency trait. *Journal of animal science* **91**, 1041–1050 (2013).

Acknowledgements

This work was supported by the National Natural Science Foundations of China (31472079, 31372294, 31402039 and 31201774), the Cattle Breeding Innovative Research Team (cxgc-ias-03), Chinese Academy of Agricultural Sciences Fundamental Research Budget Increment Projects (2013ZL031 and 2014ZL006), Chinese Academy of Agricultural Sciences Foundation (2014ywf-yb-4), and Beijing Natural Science Foundation (6154032).

Author Contributions

H.J.G. and J.Y.L. conceived and designed the experiments. J.W.X. and Z.H.F. conducted the experiments. P.T.C., Y.L.X., and G.W.Z. analysed the data. X.Y.S., X.G., Y.C., and L.P.Z. assisted in conducting the experiments. H.J.G. and J.Y.L. supervised the study. J.W.X. wrote the manuscript. All authors have read and approved the final manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Xia, J. *et al.* Searching for new loci and candidate genes for economically important traits through gene-based association analysis of Simmental cattle. *Sci. Rep.* **7**, 42048; doi: 10.1038/srep42048 (2017).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017