

# SCIENTIFIC REPORTS



OPEN

## Multi-strategy genome-wide association studies identify the DCAF16-NCAPG region as a susceptibility locus for average daily gain in cattle

Received: 22 April 2016  
Accepted: 04 November 2016  
Published: 28 November 2016

Wengang Zhang<sup>1</sup>, Junya Li<sup>1</sup>, Yong Guo<sup>2</sup>, Lupei Zhang<sup>1</sup>, Lingyang Xu<sup>1</sup>, Xue Gao<sup>1</sup>, Bo Zhu<sup>1</sup>, Huijiang Gao<sup>1</sup>, Hemin Ni<sup>2</sup> & Yan Chen<sup>1</sup>

Average daily gain (ADG) is the most economically important trait in beef cattle industry. Using genome-wide association study (GWAS) approaches, previous studies have identified several causal variants within the *PLAG1*, *NCAPG* and *LCORL* genes for ADG in cattle. Multi-strategy GWASs were implemented in this study to improve detection and to explore the causal genes and regions. In this study, we conducted GWASs based on the genotypes of 1,173 Simmental cattle. In the SNP-based GWAS, the most significant SNPs (rs109303784 and rs110058857,  $P = 1.78 \times 10^{-7}$ ) were identified in the *NCAPG* intron on BTA6 and explained 4.01% of the phenotypic variance, and the independent and significant SNP (rs110406669,  $P = 5.18 \times 10^{-6}$ ) explained 3.32% of the phenotypic variance. Similarly, in the haplotype-based GWAS, the most significant haplotype block, Hap-6-N1416 ( $P = 2.56 \times 10^{-8}$ ), spanned 12.7 kb on BTA6 and explained 4.85% of the phenotypic variance. Also, in the gene-based GWAS, seven significant genes were obtained which included *DCAF16* and *NCAPG*. Moreover, analysis of the transcript levels confirmed that transcripts abundance of *NCAPG* ( $P = 0.046$ ) and *DCAF16* ( $P = 0.046$ ) were significantly correlated with the ADG trait. Overall, our results from the multi-strategy GWASs revealed the *DCAF16-NCAPG* region to be a susceptibility locus for ADG in cattle.

With the recent emergence of genome-wide association studies (GWASs)<sup>1</sup>, major advances have been made in the understanding and practice of functional gene discovery and quantitative trait locus (QTL) mapping<sup>2-4</sup>. Although the Single Nucleotide Polymorphism (SNP)-based GWAS has been useful for identifying causal variants<sup>5</sup>, this strategy has its limitations. This approach overlooks the interaction between SNPs within a gene, misses weak signals that aggregate within related SNP sets, and incurs a severe penalty for multiple testing<sup>6</sup>.

To increase the statistical power and limit the false discovery rate (FDR) associated with GWAS analyses, GWASs have been improved using haplotype-based<sup>7-10</sup> and gene-based<sup>11-14</sup> strategies to assess complex and quantitative traits in human and domestic animals. The haplotype-based GWAS has high statistical power<sup>15,16</sup> and aims to identify causal haplotypes with specific combinations<sup>7</sup>. Haplotype-based GWASs have recently identified susceptibility haplotypes or blocks for coronary artery disease<sup>7</sup>, low-density lipoprotein cholesterol<sup>8</sup>, triglyceride levels<sup>9</sup>, and boar taint<sup>10</sup>.

Because the gene-based GWAS analysis involves all variants within a gene, it has reduced the number of required tests and is more powerful than the simple SNP-based GWAS<sup>17-19</sup>. Several gene-based GWAS methods have been developed, including the genetic similarity gene-based GWAS<sup>20</sup>, entropy-based joint analysis<sup>21</sup>, and extended Simes procedure association analysis<sup>17</sup>. Risk genes have been successfully identified for several human diseases, including multiple sclerosis<sup>11</sup>, hypertension<sup>13</sup>, and Alzheimer's disease<sup>14</sup>. However, there has been little gene-based GWAS research on quantitative traits in domestic animals.

<sup>1</sup>Cattle Genetics and Breeding Group, Institute of Animal Science (IAS), Chinese Academy of Agricultural Sciences (CAAS), Beijing 100193, China. <sup>2</sup>Animal Science and Technology College, Beijing University of Agriculture (BUA), Beijing 102206, China. Correspondence and requests for materials should be addressed to H.N. (email: nihemin@aliyun.com) or Y.C. (email: chenyan0204@163.com)

| Years | Breeds                                   | Method <sup>1</sup> | Position (Mbp) <sup>2</sup>  | Candidate genes <sup>3</sup>     | Reference |
|-------|--|---------------------|--|----------------------------------|-----------|
| 2004  | M1 line <sup>4</sup>                     | QTL                 | 2: 6.5–14.4<br>6: 6.6–8.5  |                                  | 54        |
| 2007  | Angus, Charolais, Alberta Hybrid crosses | QTL                 | 5: 71.4–71.66:<br>10.9–11.1<br><b>6: 29.0–29.6</b><br>7: 44.2–46.9 |                                  | 55        |
| 2008  | Continental × British <sup>5</sup>       | Association         | 29: 38.1–39.9  | <i>IGF2</i>                      | 56        |
| 2011  | Crossbred                                | Association         | <b>6: 21.0–22.0</b>  | <i>NCAPG</i>                     | 40        |
| 2011  | Wagyu                                    | Association         | 1: 60.9–61.1<br>6: 4.9–9.3   | <i>GHSR</i>                      | 57        |
| 2012  | Angus                                    | GWAS                | <b>6: 27.0–29.1</b>  |                                  | 58        |
| 2012  | M1 line                                  | QTL                 | 5: 32.0–34.2   |                                  | 59        |
| 2012  | Brangus heifers                          | GWAS                | <b>6: 31.7–32.4</b><br>10: 33.1–33.5                               |                                  | 60        |
| 2013  | Angus, Charolais, Piedmontese            | GWAS                | 1: 51.3–51.4   |                                  | 61        |
| 2013  | Angus × Simmental crossbred              | GWAS                | 2: 11.3–11.6<br>15: 11.3–11.6                                      | <i>NCKAP5</i> ,<br><i>PHOX2A</i> | 62        |
| 2013  | Nanyang <sup>6</sup>                     | Association         | 7: 24.3–24.5   | <i>PROPI</i>                     | 63        |
| 2015  | Nanyang                                  | Association         | 22: 16.9–16.9  | <i>CIDEC</i>                     | 64        |

**Table 1. Average daily gain (ADG)-associated quantitative trait loci (QTL) in cattle.** Note: <sup>1</sup>“GWAS” refers to the genome-wide association study, “Association” refers to the candidate gene association analysis, and “QTL” refers to the QTL mapping linkage association analysis. <sup>2</sup>QTL positions that were reported as associated with the ADG trait. <sup>3</sup>Candidate genes that were reported as associated with the ADG trait. <sup>4</sup>The M1 line of Beefbooster Inc. was developed from an Angus base and has been under selection for over 30 years. <sup>5</sup>Samples were hybrid beef steers sired by Angus, Charolais, or University of Alberta Hybrid bulls. Dams were from three composite lines. <sup>6</sup>Nanyang is a breed of Chinese native cattle. The QTL regions that are reported to surround the *DCAF16-NCAPG* locus are indicated in Boldface.

In the beef cattle industry, average daily gain (ADG) is an economically important growth trait that contributes to the production efficiency and economic benefits of graziery. Table 1 lists the ADG-associated QTL positions and candidate genes that have been reported in cattle. Notably, *PLAG1* and *NCAPG-LCORL*, known loci that are linked to adult human height<sup>22–24</sup>, have been associated with growth traits and body size in cattle<sup>25–29</sup>. The dissection of a QTL and the fine mapping of QTNs involved in bovine stature have been reported for the *PLAG1* gene<sup>28</sup>. The mechanism of the effect of *PLAG1* on growth and fertility has been clearly illustrated, and *PLAG1* knockout mice have highlighted the importance of *PLAG1* in postnatal growth and reproduction<sup>30</sup>.

Using 1,173 samples genotyped by Illumina BovineHD Beadchip, multi-strategy GWASs were performed to explore candidate genes or QTL regions for the ADG trait in Simmental cattle. Transcripts abundance of candidate genes were also examined and validated to be associated with ADG trait in this study. Identification of the promising candidate genes for further studies will greatly dissect the molecular mechanisms underlying ADG trait in cattle and has the practical in breeding program for the improvement of carcass weight in breeding program.

## Materials and Methods

**Ethics statement.** All animal procedures were conducted in strict accordance with the guidelines proposed by the Chinese Council on Animal Care, and all protocols were approved by the Science Research Department of the Institute of Animal Science, Chinese Academic of Agriculture Sciences (Beijing, China). The use of animals and private land in this study was approved by their respective owners.

**Phenotype Data.** The resource population consisted of 1,173 Simmental cattle that were born between 2008 and 2013 in Ulagai, Inner Mongolia. After weaning, all calves were transferred to a fattening farm in Beijing and fattened in the same pens for 8–12 months. All cattle were fed with identical feed, which consisted of silage, brewer’s grain, bean dregs, breadcrumbs, and maize. We measured each bull’s body weight at the following five time points: birth, upon entering the fattening farm, 12 months of age, 18 months of age, and before slaughter. The growth curve analyses closely followed the linear regression during the fattening period (see Supplementary Fig. S1), and the slope of the regression line therefore represented the average daily gain (ADG) during the fattening period.

**Genotype Data.** The genotypes of the 1,173 beef cattle were obtained by Illumina BovineHD BeadChip, which included 774,660 SNPs. Quality control procedures were carried out using PLINK 1.7 software<sup>31</sup> to remove SNPs with a call rate less than 95%, a minor allele frequency (MAF) less than 0.05 and a significant deviation from the Hardy-Weinberg equilibrium ( $P < 10^{-5}$ ); moreover, animals with more than 10% missing genotypes were removed from the dataset. Missing alleles were imputed using Beagle 4.1 software<sup>32</sup> to guarantee the accuracy and effectiveness of the statistics<sup>33,34</sup>.

**Gene Annotation.** A total of 24,596 genes were downloaded from the Ensembl Genes database (<http://www.ensembl.org/index.html>, UMD3.1), including the coding and non-coding RNA. To address the regulatory regions and linkage disequilibrium in SNPs<sup>18,35</sup>, we defined the gene boundary as  $\pm 50$  kb upstream and downstream of the gene. Each gene was covered by three or more SNPs in the genotyping BeadChip, and 23,856 genes remained to be analyzed.

**SNP-based GWAS.** A standard MLM for GWAS was performed by extending the Henderson notation as follows:

$$y = \mu + Wv + X\beta_i + Zu + e \quad (1)$$

where  $y$  represented a vector of ADG,  $\mu$  represented the population mean,  $v$  represented a vector of fixed effects,  $\beta_i$  denoted the effect of the  $i$ th SNP,  $u$  represented a vector of the polygenic effects and  $e$  represented the residual.  $W$ ,  $X$  and  $Z$  represented the incidence matrices for  $v$ ,  $\beta_i$  and  $u$ .  $Z$  was the genetic additive matrix constructed by SNPs, termed as *kinship*. As described by Lopes<sup>36</sup>, we built kinship using 50,000 random SNPs across autosomes. In this model, we considered sex, birth year, calving season and population stratification as fixed effects.

The percent phenotypic variance that was explained by a single significant SNP was calculated as follows:

$$V_{pi} = \frac{2p_i q_i \beta_i^2}{\sigma_p^2} \times 100\% \quad (2)$$

where  $p_i$  and  $q_i$  represented the allele frequencies for the  $i$ th SNP,  $\beta_i$  denoted the effect of the  $i$ th SNP, and  $\sigma_p^2$  represented the phenotypic variance. The R package *heritability* (<https://cran.r-project.org/web/packages/heritability/index.html>) was used to estimate the ADG-associated heritability and genetic variance.

**Haplotype-based GWAS.** Haplotype-based GWAS was performed using the method proposed by Gregersen VR *et al.*<sup>10</sup>. Haplotype blocks were established based on pairwise measures of the linkage disequilibrium (LD)<sup>37</sup> and implemented using the PLINK 1.7 software with a block window that was less than 100 kb. The haplotype block estimation option was `--blocks --ld-window-kb 100`. After the haplotype block partitioning, haplotypes for each sample were calculated using a standard expectation-maximum (EM) algorithm, and the program was conducted using the R package *haplo.stats* (URL: <https://cran.r-project.org/web/packages/haplo.stats/index.html>). Haplotype association analyses were implemented in the R package *lme4* (URL: <https://cran.r-project.org/web/packages/lme4/index.html>) using the MLM equation as follows:

$$y = \mu + Wv + \sum_{i=1}^t H_j \beta_{ij} + Zu + e \quad (3)$$

where  $y$  represented a vector of ADG,  $\mu$  represented the population mean,  $v$  represented a vector of fixed effects,  $\beta_{ij}$  represented the effect of the  $i$ th haplotype in the  $j$ th block (which contained  $t$  haplotypes),  $u$  denoted polygenic effects for each individual, and  $e$  represented the residual.  $W$ ,  $H_j$  and  $Z$  were the incidence matrices for  $v$ ,  $\beta_{ij}$  and  $u$ . A *Chi-square* hypothesis test with  $df=1$  was used to calculate the significance level of the haplotype block as follows:

$$Chi_j = \frac{\max(\hat{\beta}_{ij})^2}{\text{Var}(\hat{\beta}_{ij})} \quad (4)$$

where  $\max(\hat{\beta}_{ij})$  denoted the maximum effective haplotype at the  $j$ th block, and  $\text{Var}(\hat{\beta}_{ij})$  represented the variance of  $\hat{\beta}_{ij}$  obtained via mixed model equations.

The percent phenotypic variance ( $V_{pj}$ ) explained by the  $j$ th block was calculated using a two-step approach. Firstly, the effect of haplotypes at the  $j$ th block was estimated using least square (LS) method and all  $j$ th block haplotypes were clustered into two groups (G1 and G2) based on the estimated effects. Each sample was defined as 0, 1, and 2 (G1/G1, G1/G2 and G2/G2) according to the EM results. We then calculated  $V_{pj}$  as follows:

$$V_{pj} = \frac{2p q \beta^2}{\sigma_p^2} \quad (5)$$

where  $\beta$  represented the regression coefficient of the phenotype on the indicator (0, 1 and 2), and  $p$  and  $q$  indicated the frequencies of G1 and G2, respectively.

**Gene-based GWAS.** We conducted a gene-based GWAS method using a principal component analysis (PCA) according to the method proposed by Kai Wang *et al.*<sup>38</sup>. First, principle components (PCs) were constructed based on an intragenic SNP indicator, and we selected the PCs based on a cumulative contributed proportion  $>85\%$ . Second, the estimate breeding value (EBV) was calculated based on genomic best linear prediction (GBLUP) with fixed effects (sex, birth year, calving season and population stratification) and random effects (polygenic effects). Third, the effectiveness of each PC and statistical hypothesis test was calculated. The general linear model was:

$$EBV = Xb_i + e \quad (6)$$

where  $b_i$  represented the regression coefficient of the phenotype on the PC,  $X$  represented the vector of the PC, and  $e$  represented the residual. The following *Chi-square* hypothesis testing ( $df=1$ ) formula was used:

$$Chi = \frac{b_i^2}{Var(b_i)} \quad (7)$$

For each gene, we selected the minimum P-value for the PCs when the PC number exceeded two. The significant threshold was set based on the permutation testing to overcome false positive discovery<sup>39</sup>. Thus, 1,000 permutation cycles were performed (23,856,000 multiple tests), and the 240,000<sup>th</sup> highest value represented the cut-off point for the 1% level of significance.

**Gene expression level.** To validate whether the explored gene resulting from the three GWAS methods was associated with ADG trait, transcript abundance in longissimus dorsi muscle tissue was measured. We selected 28 steers randomly in 2014. Longissimus dorsi muscle samples were collected from steers at slaughter and stored in liquid nitrogen. Total RNA was isolated using the TRIzol Reagent total RNA extraction kit (Invitrogen, Carlsbad, CA, USA) and precipitated with ethanol.

Primers were designed using the Primer 5 software and were approximately 200 bp in length (Supplementary Table S1). Real-time PCR was performed to examine the expression level of selected genes using the SYBR<sup>®</sup> Fast qPCR Mix (Takara Bio, Otsu, Japan) with the Applied Biosystems<sup>®</sup> 7500 Real-Time PCR Systems (Applied Biosystems, Foster City, CA, USA). Expression values were normalized to *GAPDH* as the internal control. The mean fold change in expression of the target genes was calculated using the  $2^{-\Delta\Delta Ct}$  method.

Correlation analyses were conducted using R version 3.2.2 (<https://cran.r-project.org>, 18/3/2016). Correlations were derived for all candidate genes expression and phenotypic data with 28 random steers from the same year. General linear model (GLM) was used and the fixed effects included calving season and population stratification effects.

## Results and Discussion

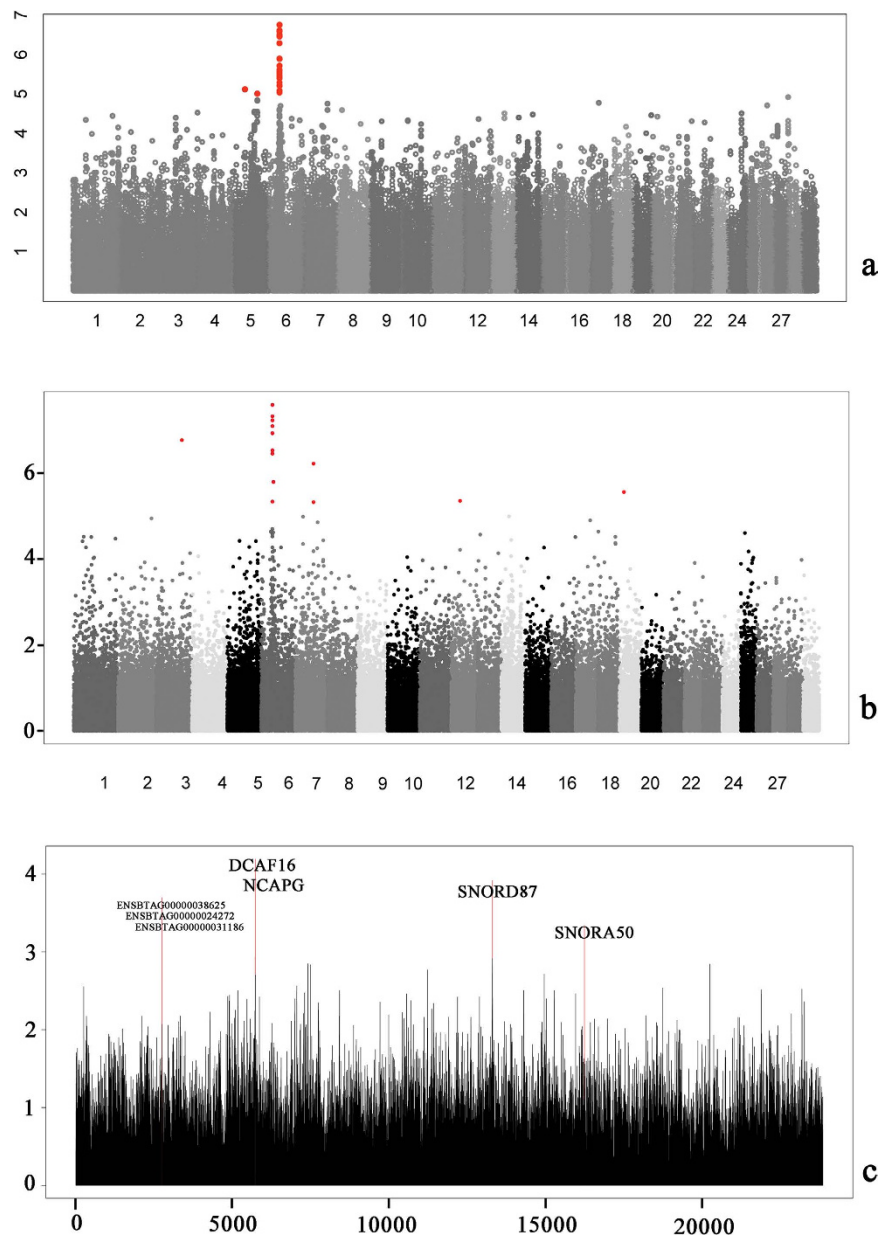
**Phenotype description and genetic parameters.** The phenotypic distribution followed a Gaussian distribution with a mean of 0.98 kg/day, a maximum of 1.87 kg/day, a minimum of 0.54 kg/day, and a standard deviation (SD) of 0.16 kg/day. The heritability ( $h^2$ ) of the average daily gain (ADG) was 0.48, with an additive genetic variance ( $V_a$ ) equal to 0.012.

Following the quality control and imputation, 1,141 samples with 669,742 SNPs remained. Cleaned SNPs were uniformly distributed over the whole bovine genome with a mean inter-marker space of 4.52 kb.

**SNP-based GWAS results.** In this study, we used three strategies to perform a genome-wide association study (GWAS) for the ADG trait in beef cattle (Fig. 1). In the SNP-based association, we identified 40 distinct SNPs (Supplementary Table S2) that exceeded the suggested significance thresholds ( $P < 10^{-6}$ ), 38 of which were located within BTA6 (Fig. 1a). Here, we identified the most significant SNPs, rs109303784 and rs110058857, on BTA6 with identical P-values of  $1.78 \times 10^{-7}$ . The distance between the two significant SNPs was 680 bp, which were in complete linkage disequilibrium ( $r^2 = 1$ ) and explained 4.01% of the phenotypic variance. Rs109303784 and rs110058857 were both located upstream of *NCAPG* and downstream of the *DCAF16* gene according to the Ensembl genome database (<http://www.ensembl.org/index.html>). Figure 2 showed the regional  $-\log_{10}$  (P-value) of the significant SNPs that surround the *DCAF16-NCAPG* locus on BTA6. We also calculated the LD levels, with the two peak SNPs denoted by different colors. Notably, we found that rs110406669 ( $P = 5.18 \times 10^{-6}$ ) had a low LD with the two peak SNPs and independently explained 3.32% of the phenotypic variance. Moreover, two other prominent SNPs, rs109028700 (BTA5:43111315) and rs137683327 (BTA5:84944556), were located on BTA5 and explained 2.59% and 2.87% of the phenotypic variance, respectively.

**Haplotype-based GWAS results.** A total of 93,732 blocks were identified, and these blocks comprised 615,355 SNPs. The maximum length was 99.9 Kb, and the minimum length was 0.4 Kb. Fourteen significant haplotype blocks (shown in Table 2) were obtained at the suggested threshold ( $P < 10^{-5}$ ) across 5 chromosomes (BTA3, BTA6, BTA7, BTA12, and BTA19). Similar to the SNP-based GWAS, 7 associated haplotype blocks that surrounded rs109303784 and rs110058857 were found on BTA6 (Fig. 2). The most significant block, Hap-6-1416 ( $P = 2.56 \times 10^{-8}$ ), spanned 22.8 Kb and was located in upstream of *NCAPG* at a distance of 12.7 Kb with rs109303784. The Hap-6-N1416 block explained 4.85% of the phenotypic variance and had 7 distinct haplotypes (GTGGATA, GTGAATA, GTAAATA, ACAGGCG, ACAAGCG, ACAAATA and ATAAATA, referred to as Haplo1, Haplo2, Haplo3, Haplo4, Haplo5, Haplo6 and Haplo7) with frequencies of 0.13%, 2.67%, 4.94%, 19.36%, 5.74%, 1.34% and 65.82%, respectively. The average effect was 0.24 kg/day, with the minimum in Haplo3 of 0.08 kg/day and the maximum in Haplo5 of 0.45 kg/day.

In contrast to the SNP-based GWAS results, no prominent block was found on BTA5, but 5 blocks were identified on BTA3, 7, 12 and 19. However, no gene regions or coding domains coincided with these blocks. Notably, Hap-3-N3218 ( $P = 1.7 \times 10^{-7}$ ) on BTA3 contained 3 extragenic SNPs (rs109934393, rs43349539 and rs43348574) that explained 6.22% of phenotypic variance. These results indicated that unknown functional regions or regulatory elements may exist around this identified block.

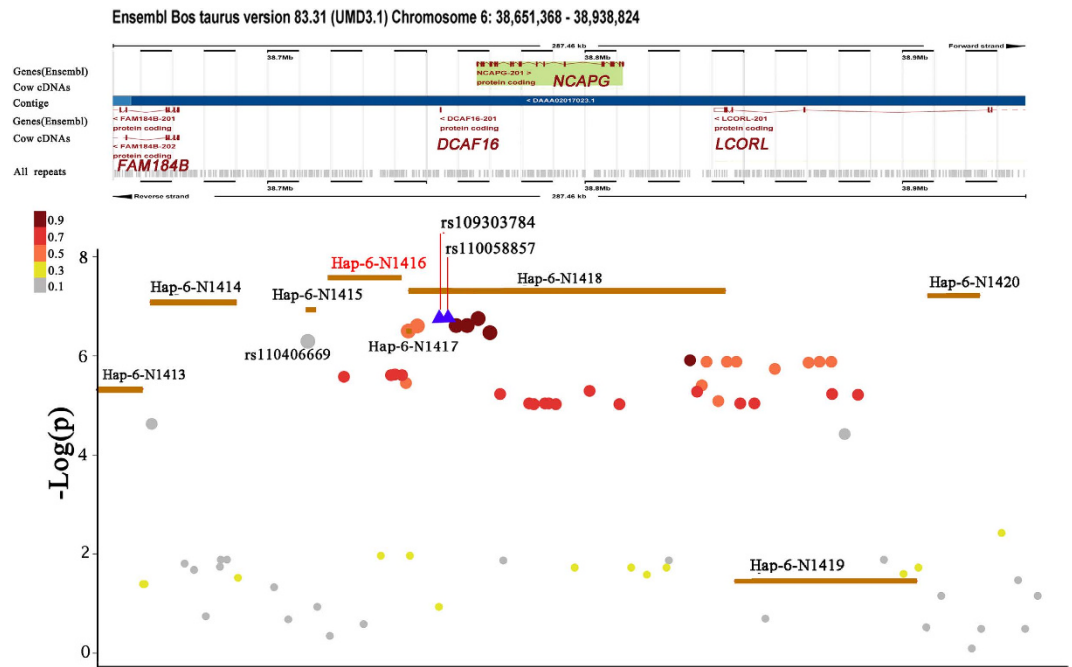


**Figure 1. Results of the multi-strategy GWAS for average daily gain.** (a) Manhattan plots for the SNP-based GWAS. (b) Manhattan plots for the haplotype-based GWAS. (c) Log<sub>10</sub> (P-value) values of 23,856 genes in the gene-based GWAS.

**Gene-based GWAS results.** A total of 24,616 genes were annotated in ENSEMBLE database. For the gene-based association, 23,856 genes with an average 34.7 SNPs per gene were analyzed. And other 760 genes were excluded, since they included less than three SNPs or not were located in autosomes (sex chromosome or mitochondria DNA). The 1,000 permutation-cycle results suggested a set P-value of  $10^{-3}$  with a FDR < 1%. Seven genes were identified for ADG in this study (Table 3). Specifically, *DCAF16* and *NCAPG* were implicated by the SNP- and Haplotype-based association results. We also found two small nucleolar RNAs, SNORD50 and SNORD87, with identical functions in the modification process of other small nuclear RNAs (snRNAs). Additionally, two uncharacterized proteins—ENSBTAG00000038625 and ENSBTAG00000024272—were obtained. These results indicated that the gene-based method can identify functional genes or loci which are previously unverified and provide a possible structural basis for further gene functional validation studies.

***DCAF16-NCAPG* locus associated with ADG.** Taken together, 163 significant SNPs were identified by three GWAS strategies (The SNPs in the gene-based set were SNPs within significant genes). Venn diagram summarizing the three strategies results was shown in Fig. 3. Here, the SNP- and haplotype-based GWAS approaches returned a distinct set of 8 and 44 prominent SNPs, respectively. Five genes—*PTPRR*, *LMNTD1*, *FAM114A2*, *C8A* and *STARD13*—were proximal to these 52 significant SNPs, suggesting associations for some of these genes with





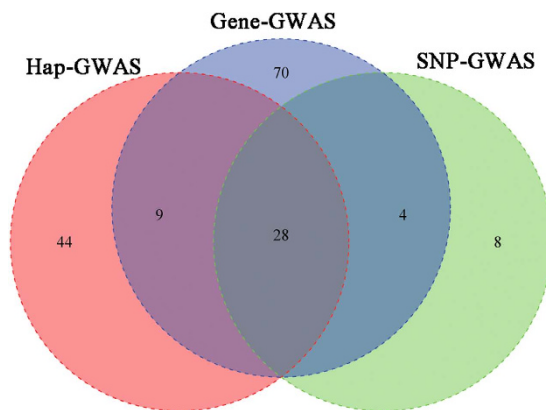
**Figure 2.** Regional  $-\log_{10}(P\text{-value})$  plot of the SNP-based and haplotype-based association around the *DCAF16-NCAPG* locus on BTA6: 38.6–39.0(Mb). The yellow bar represents the block position. The purple triangle represents the two most significant SNPs (rs109303784 and rs110058857,  $r^2 = 1$ ). SNPs were colored based on their LDs with two most significant SNPs as follows: red SNPs with LDs at  $r^2 > 0.9$ , pink SNPs with LDs at  $r^2 > 0.7$ , orange SNPs with LDs at  $r^2 > 0.5$ , yellow SNPs with LDs at  $r^2 > 0.3$  and grey SNPs with LDs at  $r^2 < 0.3$ . The size of the plots indicates the significance level of SNPs in the SNP-based GWAS. The positions of all RefSeq genes were downloaded from the ENSEMBL database.

| Hap. <sup>1</sup> | SNP  | Start    | End      | Len. <sup>2</sup> | Chr <sup>3</sup> | P-value               | Effect <sup>4</sup> |
|-------------------|--|----------|----------|-------------------|------------------|-----------------------|---------------------|
| Hap-6-N1416       | rs137684828 rs133288426 rs136487046 rs110302064<br>rs109776587 rs110281825             | 38711873 | 38734752 | 22,879            | 6                | $2.56 \times 10^{-8}$ | 4.85%               |
| Hap-6-N1418       | 27 <sup>5</sup>  | 38737206 | 38837159 | 99,953            | 6                | $4.75 \times 10^{-8}$ | 4.60%               |
| Hap-6-N1420       | rs132741545 rs137111419  | 38900275 | 38917456 | 17,181            | 6                | $5.88 \times 10^{-8}$ | 4.64%               |
| Hap-6-N1414       | rs136542559 rs110770764 rs110867784 rs110780166<br>rs109035277 rs134219500 rs109920396 | 38655605 | 38682695 | 27,090            | 6                | $8.02 \times 10^{-8}$ | 4.38%               |
| Hap-6-N1415       | rs110406669 rs134225464  | 38704872 | 38707716 | 2,844             | 6                | $1.18 \times 10^{-7}$ | 4.98%               |
| Hap-3-N3218       | rs109934393 rs43349539 rs43348574  | 89890121 | 89893239 | 3,118             | 3                | $1.70 \times 10^{-7}$ | 6.22%               |
| Hap-6-N1417       | rs109448564 rs109436638  | 38735901 | 38736441 | 540               | 6                | $2.95 \times 10^{-7}$ | 3.59%               |
| Hap-6-N1409       | rs110983998 rs109705804 rs110638909 rs110338374<br>rs110878984 rs109627413 rs136745840 | 38477781 | 38495361 | 15,411            | 6                | $3.50 \times 10^{-7}$ | 3.87%               |
| Hap-7-N2309       | rs137459122 rs136096935 rs136896599  | 67220287 | 67227674 | 7,387             | 7                | $6.02 \times 10^{-7}$ | 3.73%               |
| Hap-6-N1536       | rs43457333 rs43457339 rs43457349 rs43457352 rs43457353<br>rs43455999                   | 42130983 | 42147873 | 16890             | 6                | $1.60 \times 10^{-6}$ | 2.00%               |
| Hap-19-N540       | rs110686148 rs135073323  | 13534451 | 13535938 | 1,487             | 19               | $2.68 \times 10^{-6}$ | 3.26%               |
| Hap-12-1130       | rs137645749 rs41600431 rs137228197 rs135990217<br>rs29021098                           | 28042816 | 28052265 | 9,449             | 12               | $4.42 \times 10^{-6}$ | 2.79%               |
| Hap-6-N1413       | rs109849093 rs110793327 rs109063701 rs109622396  | 38644886 | 38653201 | 8,315             | 6                | $4.59 \times 10^{-6}$ | 3.16%               |
| Hap-7-N2310       | rs133024924 rs134650380 rs135673337 rs132656697<br>rs134402892                         | 67237316 | 67243709 | 6,393             | 7                | $4.74 \times 10^{-6}$ | 2.52%               |

**Table 2.** Significant haplotypes from the haplotype-based GWAS. Note: <sup>1</sup>Haplotype name. <sup>2</sup>Length (bp). <sup>3</sup>Chromosome. <sup>4</sup>Phenotypic variance explained by the haplotype. <sup>5</sup>Twenty-seven SNPs were listed as follows: rs135282158 rs137844992 rs38746212 rs109303784 rs110058857 rs110412527 rs109414269 rs110841247 rs109885206 rs109554838 rs110443327 rs109322908 rs109861428 rs110062110 rs109002201 rs110024295 rs133222819 rs137268410 rs109801011 rs110426625 rs135260513 rs109795992 rs110030099 rs110908263 rs110386170 rs110419157 rs109924798.

| ENSEMBL ID         | Description  | P-value               | FDR    | Number of SNPs | Chr |
|--------------------|--|-----------------------|--------|----------------|-----|
| ENSBTAG00000011973 | DDB1 and CUL4 associated factor 16 ( <i>DCAF16</i> )       | $6.45 \times 10^{-5}$ | 0.0016 | 30             | 6   |
| ENSBTAG00000021582 | non-SMC condensin I complex, subunit G ( <i>NCAPG</i> )    | $1.12 \times 10^{-4}$ | 0.0043 | 42             | 6   |
| ENSBTAG00000043234 | Small nucleolar RNA SNORD87                                | $1.22 \times 10^{-4}$ | 0.0045 | 17             | 14  |
| ENSBTAG00000038625 | Uncharacterized protein                                    | $2.01 \times 10^{-4}$ | 0.0045 | 3              | 3   |
| ENSBTAG00000024272 | Uncharacterized protein                                    | $2.01 \times 10^{-4}$ | 0.0049 | 4              | 3   |
| ENSBTAG00000031186 | Guanylate binding protein family, member 6 ( <i>GBP6</i> ) | $2.11 \times 10^{-4}$ | 0.0057 | 4              | 3   |
| ENSBTAG00000043192 | Small nucleolar RNA SNORA50                                | $4.57 \times 10^{-4}$ | 0.0084 | 44             | 18  |

**Table 3. Seven significant ADG-associated genes based on the gene-based GWAS.**



**Figure 3. Venn diagram summarizing the association analyses results of the three strategies.** The number represents the interaction and the remaining significant SNPs identified in three GWAS methods.

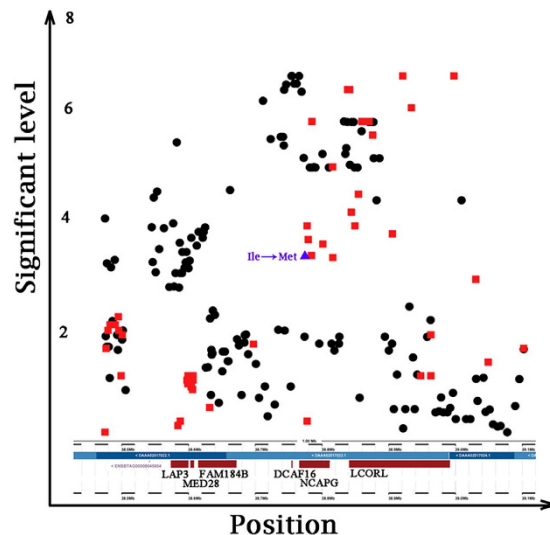
the ADG trait. We focused on the intersection of candidate SNPs identified by the three GWASs methods with the highest ADG trait-associated accuracy, which included 28 significant SNPs located at 38.6–39.0 Mb on BTA6. Figure 2 showed a schematic diagram of the region, which contains four annotated genes—*FAM184B*, *DCAF16*, *NCAPG*, and *LCORL*—from the Ensembl genome database.

*DCAF16*, which was near to the peak SNPs for SNP-based GWAS approach, was the most significant gene ( $P = 6.45 \times 10^{-5}$ ) for gene-based GWAS analysis. Similarly, the most significant block, Hap-6-N1416 ( $P = 2.56 \times 10^{-8}$ ), was also located downstream of *DCAF16* (physical distance = 19,663 bp) according to the Ensembl database. *DCAF16* may function as a substrate receptor for the *CUL4-DDB1 E3* ubiquitin-protein ligase complex, which is involved in two pathways that promote protein modifications and ubiquitination. *NCAPG*, which was also identified by three GWAS methods simultaneously, encodes a subunit of the condensin complex, which is responsible for the condensation and stabilization of chromosomes during mitosis and meiosis. The associated pathways involved the cell cycle, mitosis and the mitotic prometaphase. Numerous studies<sup>25,26,40–52</sup> have confirmed that *NCAPG* has strong effects on the body sizes and growth traits of human and domestic animals. According to the association analyses from Lindholm-Perry's results<sup>40</sup>, 47 SNPs within or near the gene boundaries of the three candidate genes (*NCAPG*, *LCORL* and *LAP3*) were genotyped. Figure 4 showed a comparison of these association study results with our SNP-based GWAS results. In contrast to our results, the most significant SNPs were located in the *LCORL* gene. However, most of the significant SNPs from these two analyses were located around the BTA6: 38.78 (Mb) region near the downstream region of *DCAF16*, suggesting that this region might be a more effective QTL for ADG trait in cattle.

Additionally, a missense mutation (c.1326 T > G, indicated in Fig. 4 by a purple triangle) was identified in exon 9 of *NCAPG* by several association<sup>26,45</sup> and linkage analyses<sup>25,41</sup>. The resulting amino acid change of Ile442 to Met442 in the encoded protein has been shown to be a candidate causative variation of the growth trait in cattle. Significant selection regions that affect the statures of European and African cattle cohorts were identified in *NCAPG* by multiple signal selection analyses<sup>49</sup>. GWAS analyses in horses<sup>42,43,46,47</sup> and cattle<sup>48,51</sup> indicated that the *NCAPG-LCORL* locus or closed regions were significantly associated with body size and growth traits.

Based on our results and previous reports<sup>44,52</sup>, we tested *DCAF16*, *NCAPG*, and *LCORL* expression in muscle tissues. Longissimus dorsi muscle samples from 28 steers with ADG phenotypes were collected. General linear model (GLM) results showed *DCAF16* and *NCAPG* expression is significantly associated with ADG trait (Table 4) and correlations between ADG and genes expression were presented in (Supplementary Fig. S3). No significant difference was detected for the *LCORL* gene. Our results were concordant with the results presented by Perry *et al.*<sup>44</sup> that abundance of *NCAPG* was associated with ADG in the muscle tissue muscle from cows.

In the NCBI database, the *NCAPG* gene has one reference transcript (Genebank accession number: NM\_001102376) and two predicted transcripts (XM\_005207785 and XM\_015471561), which were derived by a computational analysis using transcriptome data from 11 Hereford cattle. The differences between the three



**Figure 4. Regional plot of our GWAS results versus association analysis results by Lindholm-Perry<sup>40</sup>.** The black circles represent the  $-\text{Log}_{10}$  (P-value) of our SNP-based GWAS, and the red squares represent the  $-\text{Log}_{10}$  (P-value) of the previous association analysis. The purple triangle represents SNP c.1326 T > G, which is the Ile442 to Met442 amino acid change, in exon 9 of NCAPG.

| Gene         | Effect (SE) <sup>1</sup> | P-value |
|--------------|--------------------------|---------|
| NCAPG        | 0.018 ( $\pm 0.009$ )    | 0.046   |
| LCORL        | 0.050 ( $\pm 0.032$ )    | 0.138   |
| DCAF16       | 0.030 ( $\pm 0.012$ )    | 0.046   |
| NM_001102376 | 0.003 ( $\pm 0.006$ )    | 0.597   |
| XM_005207785 | 0.012 ( $\pm 0.006$ )    | 0.050   |
| XM_015471561 | 0.007 ( $\pm 0.004$ )    | 0.074   |

**Table 4. Target gene expression in muscle tissue and estimated effects for ADG.** Note: <sup>1</sup>Estimate effect in kg/day and standard error of unit of transcript abundance in fold change of genes on ADG.

transcripts occur in exon 1 (Supplementary Fig. S4). Three transcripts primers were designed using the Primer 5 software (Supplementary Table S3). We demonstrated the existence of three transcripts in Simmental cattle using reverse transcription polymerase chain reaction (RT-PCR) (Supplementary Fig. S2), and the PCR production sequences were consistent with those reported in the NCBI database. To address the significant association between each transcripts abundance and the ADG trait, we also tested the expression levels of three transcripts. GLM results showed XM\_005207785 ( $P = 0.050$ ) expression was significantly associated with ADG, while no significant correlation were found in NM\_001102376 ( $P = 0.597$ ) and XM\_015471561 ( $P = 0.074$ ) transcripts (Table 4).

Overall, *DCAF16* and *NCAPG* have been simultaneously explored by the three GWAS methods, and statistical analysis have proven that *DCAF16* and one of *NCAPG* transcripts (XM\_005207785) abundance were associated with ADG trait, indicating that the *DCAF16-NCAPG* region is a susceptibility locus for the ADG trait in cattle.

Furthermore, we noticed that the independent and significant SNP (rs110406669) from the SNP-based GWAS was located 5' upstream with a distance of 30,695 bp to XM\_005207785. Two peak SNPs were located in intron 1 of XM\_005207785 and upstream with a distance of 6,970/7,650 bp to *DCAF16*. We then searched the transcription factor-binding (TF) site around candidate regions using the Tfsitescan software on the MIRAGE WWW server (<http://www.ifti.org/cgi-bin/ifti/Tfsitescan.pl>). The regions, which contained  $\pm 5$  Kb flanking sequences of the obtained significant SNPs (rs109303784, rs110058857 and rs110406669), were analyzed, and Table 5 showed the Tfsitescan results. The distances between the two most significant TF sites identified here—Nmp4-COL1A1-sit and AT2-VIRE—and the significant SNPs were 130 bp and 178 bp, respectively. It has been shown that Nmp4-COL1A1-sit influences cell structure and function during extracellular matrix remodeling in osteoblasts<sup>53</sup>. The protein product of AT2-VIRE, the AT2 receptor, is widely and abundantly expressed in fetal tissues and plays a pivotal role in cell differentiation and growth<sup>31</sup>. Moreover, similar TF site sequences were found upstream of the *NCAPG* gene in various species (Supplementary Table S4). Taken together, we proposed that Nmp4-COL1A1-sit, AT2-VIRE or other TF sites are probably involved in the regulation of *DCAF16* or *NCAPG* transcript expression in association with the ADG trait.



| Significant SNP             | Site (Length)                             | Position <sup>1</sup> (bp) | Occurrence            | Exp. Value <sup>2</sup>                 | Significance Level <sup>3</sup> | Distance <sup>4</sup> (bp) |
|-----------------------------|---|----------------------------|-----------------------|---|---------------------------------|----------------------------|
| rs110406669                 | Zp2-Ebox (14)                             | -34345                     | 1                     | $3.81 \times 10^{-4}$                   | *                               | 3627                       |
|                             | Fkh/HNF-3-subcl' (15)                     | -31870                     | 1                     | $2.82 \times 10^{-4}$                   | *                               | 1152                       |
|                             | <b>Nmp4-COL1A1-sit' (20)</b>              | <b>-30848</b>              | <b>1</b>              | <b><math>1.31 \times 10^{-7}</math></b> | <b>***</b>                      | <b>130</b>                 |
|                             | flk1-HBE' (14)                            | -29997                     | 1                     | $3.81 \times 10^{-4}$                   | *                               | 721                        |
|                             | CTF/NF1-CF1-2 (11)                        | -27444                     | 1                     | $5.67 \times 10^{-4}$                   | *                               | 3274                       |
|                             | delta-globin-PY (10)/pyr factor site (10) | -25949                     | 1                     | $7.50 \times 10^{-4}$                   | *                               | 4769                       |
| rs110058857 and rs110058857 | gamma-globin-he' (13)                     | -25836                     | 1                     | $4.65 \times 10^{-4}$                   | *                               | 4882                       |
|                             | Sp1-U2snR.2 (11)                          | +6963                      | 1                     | $5.67 \times 10^{-4}$                   | *                               | 4892                       |
|                             | PRDI_1)' (14)                             | +7799                      | 1                     | $3.81 \times 10^{-4}$                   | *                               | 4056                       |
|                             | Elf-1-FCER1A' (11)                        | +7828                      | 1                     | $5.67 \times 10^{-4}$                   | *                               | 4027                       |
|                             | mTH-Pitx3-site- (14)                      | +9508                      | 1                     | $3.81 \times 10^{-4}$                   | *                               | 2347                       |
|                             | CBF-CYP2E1 (14)                           | +10839                     | 1                     | $3.81 \times 10^{-4}$                   | *                               | 1016                       |
|                             | <b>AT2-VIRE (16)</b>                      | <b>+11677</b>              | <b>1</b>              | <b><math>2.71 \times 10^{-5}</math></b> | <b>***</b>                      | <b>178</b>                 |
|                             | C/EBP-IL-8-1 (14)                         | +11954                     | 1                     | $3.81 \times 10^{-4}$                   | *                               | 99                         |
| IgHC.11' (12)               | +14748                                    | 1                          | $1.42 \times 10^{-4}$ | *                                       | 2893                            |                            |
| SV40.10(11)/SV40.4(11)      | +15055                                    | 2                          | $5.67 \times 10^{-4}$ | *                                       | 3200                            |                            |

**Table 5. List of transcription factor-binding (TF) sites around the NCAPG-LCORN locus.** Note: <sup>1</sup>“Position” indicates distance (in nucleotides) relative to the beginning of XM\_005207785; negative indicates upstream of XM\_015471561. <sup>2</sup>Experience Value. <sup>3</sup>Significance level is represented by an asterisk (\*), where \* indicates  $10^{-4} < P\text{-Value} < 10^{-3}$ , and \*\*\* indicates  $P\text{-Value} < 10^{-4}$ . <sup>4</sup>Distance between TF and significant SNPs.

## Conclusion

In this study, we performed multi-strategy GWASs to investigate average daily gain (ADG) in the Simmental beef cattle. Forty significant SNPs in the SNP-based GWAS, 14 significant haplotype blocks in the haplotype-based GWAS, and 7 prominent genes in the gene-based GWAS were identified. Two genes, *DCAF16* and *NCAPG*, were demonstrated to be associated with ADG by all three GWAS methods. Most importantly, the significant SNPs within the *NCAPG-DCAF16* region were strongly associated with the ADG trait, with phenotypic variance of approximately 4%, suggesting the existence of causal variants in this region. Moreover, we have also shown that *DCAF16* and *NCAPG* expression were significantly associated with ADG. Our findings provide insights into the understanding of the genetic mechanisms underlying ADG trait in cattle, and these results inform future NGS-GWAS analyses of causal variants for the ADG trait. Moreover, multi-strategy GWASs represents a powerful approach to the search and analysis of susceptibility loci-related traits.

## References

- Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385–389 (2005).
- Lavinsky, J. *et al.* Genome-Wide Association Study Identifies Nox3 as a Critical Gene for Susceptibility to Noise-Induced Hearing Loss. *PLoS Genet* **11**, doi: 10.1371/journal.pgen.1005094 (2015).
- Wolf, Z. T. *et al.* Genome-Wide Association Studies in Dogs and Humans Identify ADAMTS20 as a Risk Variant for Cleft Lip and Palate. *PLoS Genet* **11**, doi: 10.1371/journal.pgen.1005059 (2015).
- Fontanesi, L. *et al.* A genome wide association study for backfat thickness in Italian Large White pigs highlights new regions affecting fat deposition including neuronal genes. *BMC Genomics* **13**, doi: 10.1186/1471-2164-13-583 (2012).
- Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am J Hum Genet* **90**, 7–24 (2012).
- Listgarten, J. *et al.* A powerful and efficient set test for genetic markers that handles confounders. *Bioinformatics* **29**, 1526–1533 (2013).
- Tregouet, D. A. *et al.* Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. *Nat Genet* **41**, 283–285 (2009).
- Zhang, Q. S., Browning, B. L. & Browning, S. R. Genome-wide haplotypic testing in a Finnish cohort identifies a novel association with low-density lipoprotein cholesterol. *Eur J Hum Genet* **23**, 672–677 (2015).
- Wu, Y. *et al.* Genome-Wide Association Studies Using Haplotypes and Individual SNPs in Simmental Cattle. *PLoS One* **9**, doi: 10.1371/journal.pone.0109330 (2014).
- Gregersen, V. R. *et al.* Genome-wide association scan and phased haplotype construction for quantitative trait loci affecting boar taint in three pig breeds. *BMC Genomics* **13**, doi: 10.1186/1471-2164-13-22 (2012).
- Lin, X., Deng, F. Y., Lu, X. & Lei, S. F. Susceptibility Genes for Multiple Sclerosis Identified in a Gene-Based Genome-Wide Association Study. *J Clin Neurol* **11**, 311–318 (2015).
- Hibar, D. P. *et al.* Voxelwise gene-wide association study (vGeneWAS): Multivariate gene-based association testing in 731 elderly subjects. *Neuroimage* **56**, 1875–1891 (2011).
- Yang, H. C. *et al.* Identification of IGF1, SLC4A4, WWOX, and SFMBT1 as Hypertension Susceptibility Genes in Han Chinese with a Genome-Wide Gene-Based Association Study. *PLoS One* **7**, doi: 10.1371/journal.pone.0032907 (2012).
- Hong, M. G. *et al.* Genome-wide and gene-based association implicates FRMD6 in Alzheimer disease. *Hum Mutat* **33**, 521–529 (2012).
- Clark, A. G. The role of haplotypes in candidate gene studies. *Genet Epidemiol* **27**, 321–333 (2004).
- Schaid, D. J. Evaluating associations of haplotypes with traits. *Genet Epidemiol* **27**, 348–364 (2004).
- Li, M. X., Gui, H. S., Kwan, J. S. H. & Sham, P. C. GATES: A Rapid and Powerful Gene-Based Association Test Using Extended Simes Procedure. *Am J Hum Genet* **88**, 283–293 (2011).
- Liu, J. Z. *et al.* A Versatile Gene-Based Test for Genome-wide Association Studies. *Am J Hum Genet* **87**, 139–145 (2010).
- Kang, G. L., Jiang, B. & Cui, Y. H. Gene-based Genomewide Association Analysis: A Comparison Study. *Curr Genomics* **14**, 250–255 (2013).

20. Buil, A. *et al.* A new gene-based association test for genome-wide association studies. *BMC Proc* **3** Suppl 7, S130, doi: 10.1186/1753-6561-3-S7-S130 (2009).
21. Kang, G. & Zuo, Y. Entropy-based joint analysis for two-stage genome-wide association studies. *J Hum Genet* **52**, 747–756 (2007).
22. Lindholm-Perry, A. K. *et al.* Association, effects and validation of polymorphisms within the NCAPG-LCORN locus located on BTA6 with feed intake, gain, meat and carcass traits in beef cattle. *Bmc Genet* **12**, doi: 10.1186/1471-2156-12-103 (2011).
23. Kneeland, J. *et al.* Identification and fine mapping of quantitative trait loci for growth traits on bovine chromosomes 2, 6, 14, 19, 21, and 23 within one commercial line of *Bos taurus*. *J Anim Sci* **82**, 3405–3414 (2004).
24. Nkrumah, J. D. *et al.* Primary genome scan to identify putative quantitative trait loci for feedlot growth rate, feed intake, and feed efficiency of beef cattle. *Journal of Animal Science* **85**, 3170–3181 (2007).
25. Sherman, E. L. *et al.* Polymorphisms and haplotypes in the bovine neuro peptide Y, growth hormone receptor, ghrelin, insulin-like growth factor 2, and uncoupling proteins 2 and 3 genes and their associations with measures of growth, performance, feed efficiency, and carcass merit in beef cattle. *Journal of Animal Science* **86**, 1–16 (2008).
26. Komatsu, M. *et al.* Genetic association between GHSR1a 5'UTR-microsatellite and nt-7(C > A) loci and growth and carcass traits in Japanese Black cattle. *Animal Science Journal* **82**, 396–405 (2011).
27. Rolf, M. M. *et al.* Genome-wide association analysis for feed efficiency in Angus cattle. *Anim Genet* **43**, 367–374 (2012).
28. Li, C. *et al.* The identification of common haplotypes on bovine chromosome 5 within commercial lines of *Bos taurus* and their associations with growth traits. *Journal of Animal Science* **80**, 1187–1194 (2002).
29. Peters, S. O. Bayesian genome wide association analyses of growth and yearling ultrasound measures of carcass traits in Brangus heifers (vol 90, pg 3398, 2012). *Journal of Animal Science* **91**, 1522–1522 (2013).
30. Lu, D. *et al.* Genome-wide association analyses for growth and feed efficiency traits in beef cattle. *Journal of Animal Science* **91**, 3612–3633 (2013).
31. Serao, N. V. L. *et al.* Bivariate Genome-Wide Association Analysis of the Growth and Intake Components of Feed Efficiency. *PLoS One* **8**, doi: 10.1371/journal.pone.0078530.t001 (2013).
32. Pan, C. Y. *et al.* A critical functional missense mutation (H173R) in the bovine PROP1 gene significantly affects growth traits in cattle. *Gene* **531**, 398–402 (2013).
33. Wang, J. *et al.* Haplotypes in the promoter region of the CIDEC gene associated with growth traits in Nanyang cattle. *Scientific Reports* **5**, doi: 10.1038/srep12075 (2015).
34. Kim, J. J. *et al.* Identification of 15 loci influencing height in a Korean population. *J Hum Genet* **55**, 27–31 (2010).
35. Gudbjartsson, D. F. *et al.* Many sequence variants affecting diversity of adult human height. *Nature Genetics* **40**, 609–615 (2008).
36. Weedon, M. N. *et al.* Genome-wide association analysis identifies 20 loci that influence adult height. *Nature Genetics* **40**, 575–583 (2008).
37. 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 (2009).
38. 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 (2011).
39. Pausch, H. *et al.* Genome-Wide Association Study Identifies Two Major Loci Affecting Calving Ease and Growth-Related Traits in Cattle. *Genetics* **187**, 289–297 (2011).
40. Karim, L. *et al.* Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nature Genetics* **43**, 405–413 (2011).
41. Cho, Y. S. *et al.* A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nature Genetics* **41**, 527–534 (2009).
42. Juma, A. R., Damdimopoulou, P. E., Grommen, S. V. H., Van de Ven, W. J. M. & De Groef, B. Emerging role of PLAG1 as a regulator of growth and reproduction. *Journal of Endocrinology* **228**, R45–R56 (2016).
43. Horiuchi, M. *et al.* Stimulation of different subtypes of angiotensin II receptors, AT1 and AT2 receptors, regulates STAT activation by negative crosstalk. *Circ Res* **84**, 876–882 (1999).
44. Browning, B. L. *et al.* Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–71 (2013).
45. Nielsen, R. *et al.* Genotype and SNP calling from next-generation sequencing data. *Nature reviews Genetics* **12**, 443–451 (2011).
46. Browning, S. R. *et al.* Haplotype phasing: existing methods and new developments. *Nature reviews Genetics* **12**, 703–714 (2011).
47. Qanbari, S. *et al.* Classic Selective Sweeps Revealed by Massive Sequencing in Cattle. *PLoS Genet* **10**, doi: 10.1371/journal.pgen.1004148 (2014).
48. Lopes, M. S. *et al.* Improved estimation of inbreeding and kinship in pigs using optimized SNP panels. *Bmc Genet* **14**, doi: 10.1186/1471-2156-14-92 (2013).
49. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–5 (2005).
50. Wang, K. & Abbott, D. A principal components regression approach to multilocus genetic association studies. *Genetic Epidemiology* **32**, 108–118 (2008).
51. Churchill, G. A. & Doerge, R. W. Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–71 (1994).
52. 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 Genet* **10**, doi: 10.1186/1471-2156-10-43 (2009).
53. Makvandi-Nejad, S. *et al.* Four loci explain 83% of size variation in the horse. *PLoS One* **7**, e39929 (2012).
54. Signer-Hasler, H. *et al.* A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PLoS One* **7**, e37282 (2012).
55. Lindholm-Perry, A. K. *et al.* Adipose and muscle tissue gene expression of two genes (NCAPG and LCORN) located in a chromosomal region associated with cattle feed intake and gain. *PLoS One* **8**, e80882 (2013).
56. 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. *Anim Sci J* **84**, 529–34 (2013).
57. Tetens, J., Widmann, P., Kuhn, C. & Thaller, G. A genome-wide association study indicates LCORN/NCAPG as a candidate locus for withers height in German Warmblood horses. *Anim Genet* **44**, 467–71 (2013).
58. Boyko, A. R. *et al.* Genomic analysis establishes correlation between growth and laryngeal neuropathy in Thoroughbreds. *BMC Genomics* **15**, doi: 10.1186/1471-2164-15-259 (2014).
59. Saatchi, M., Schnabel, R. D., Taylor, J. F. & Garrick, D. J. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC Genomics* **15**, doi: 10.1186/1471-2164-15-442 (2014).
60. Randhawa, I. A., Khatkar, M. S., Thomson, P. C. & Raadsma, H. W. Composite Selection Signals for Complex Traits Exemplified Through Bovine Stature Using Multibreed Cohorts of European and African *Bos taurus*. *G3 (Bethesda)* **5**, 1391–401 (2015).
61. Widmann, P. *et al.* Systems biology analysis merging phenotype, metabolomic and genomic data identifies Non-SMC Condensin I Complex, Subunit G (NCAPG) and cellular maintenance processes as major contributors to genetic variability in bovine feed efficiency. *PLoS One* **10**, e0124574 (2015).

62. Sahana, G., Hoglund, J. K., Guldbrandtsen, B. & Lund, M. S. Loci associated with adult stature also affect calf birth survival in cattle. *BMC Genet* **16**, 47 (2015).
63. 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. *Genet Mol Biol* **38**, 450–6 (2015).
64. Thunyakitpisal, P. *et al.* Cloning and functional analysis of a family of nuclear matrix transcription factors (NP/NMP4) that regulate type I collagen expression in osteoblasts. *J Bone Miner Res* **16**, 10–23 (2001).

### Acknowledgements

This work was funded in part by National Natural Science Foundation of China (31402039), Beijing Natural Science Foundation (6154032), Chinese Academy of Agricultural Sciences Foundation (2014ywf-yb-4), National Natural Science Foundation of China (31372294), National Natural Science Foundation of China (31472079), Cattle Breeding Innovative Research Team (cxgc-ias-03), National Beef Cattle Industrial Technology System (CARS-38), and Project of College Innovation Improvement under Beijing Municipality (PXM2016\_014207\_000012). The authors would like to thank the staff at the cattle experimental unit in Beijing and Ulagai for caring of animals and collection biological samples.

### Author Contributions

J.Y.L. and Y.C. conceived and designed the experiments. W.G.Z., H.M.N., and Y.G. performed the experiments and wrote the manuscript. L.P.Z. and X.G. performed GWAS analysis. L.Y.X., H.J.G., and B.Z. collected phenotype records and fixed effects data. All the authors have read and approved the final manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

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

**How to cite this article:** Zhang, W. *et al.* Multi-strategy genome-wide association studies identify the DCAF16-NCAPG region as a susceptibility locus for average daily gain in cattle. *Sci. Rep.* **6**, 38073; doi: 10.1038/srep38073 (2016).

**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) 2016