

Genome-wide association study of the backfat thickness trait in two pig populations

Dandan ZHU¹, Xiaolei LIU¹, Rothschild MAX², Zhiwu ZHANG³, Shuhong ZHAO¹, Bin FAN (✉)¹

¹ Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

² Department of Animal Science, Iowa State University, Ames, IA 50011, USA

³ Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853, USA

Abstract Backfat thickness is a good predictor of carcass lean content, an economically important trait, and a main breeding target in pig improvement. In this study, the candidate genes and genomic regions associated with the tenth rib backfat thickness trait were identified in two independent pig populations, using a genome-wide association study of porcine 60K SNP genotype data applying the compressed mixed linear model (CMLM) statistical method. For each population, 30 most significant single-nucleotide polymorphisms (SNPs) were selected and SNP annotation implemented using *Sus scrofa* Build 10.2. In the first population, 25 significant SNPs were distributed on seven chromosomes, and SNPs on SSC1 and SSC7 showed great significance for fat deposition. The most significant SNP (ALGA0006623) was located on SSC1, upstream of the *MC4R* gene. In the second population, 27 significant SNPs were recognized by annotation, and 12 SNPs on SSC12 were related to fat deposition. Two haplotype blocks, M1GA0016251-MARC0075799 and ALGA0065251-MARC0014203-M1GA0016298-ALGA0065308, were detected in significant regions where the *PIPNC1* and *GHI* genes were identified as contributing to fat metabolism. The results indicated that genetic mechanism regulating backfat thickness is complex, and that genome-wide associations can be affected by populations with different genetic backgrounds.

Keywords backfat thickness, SNP chip, genome-wide association study, compressed mixed linear model, pig

1 Introduction

In the pork industry, fat deposition traits such as backfat thickness and fat percentage are considered of major economic importance. Research on how genetic variants affect fat deposition could be useful both theoretically and practically. By using candidate gene and quantitative trait loci (QTL) mapping strategies, genes, such as *MC4R* (melanocortin-4 receptor) [1] and *FTO* (fat mass and obesity-associated) [2], were successfully identified as contributing to fat deposition. As a result of low mapping density of microsatellite markers, only a few quantitative trait loci have been recognized as associated with fat deposition [3,4]. Due to the rapid development of technology for high throughput sequencing, the current porcine 60K SNP chip provides a higher density than previous chips and can improve precision in locating QTL regions [5].

A number of economically important traits in pigs have been investigated in genome-wide association (GWA) studies [6,7]. However, in most of the GWA studies only one experimental population was used, and the results have not been verified in independent populations. The aims of this study were to detect potential genetic variants in the backfat thickness trait in two independent populations of pigs using GWA study and to investigate how this is affected by genetic background.

2 Materials and methods

2.1 Animals and traits

Two pig populations were used; a population composed of 820 sows from two genetic lines including Large White pigs and Large White × Landrace cross (Population 1) and

an F_2 population built by intercrossing Berkshire boars and Yorkshire sows from which 208 F_2 pigs were randomly selected (Population 2). Phenotype assessments of the tenth rib backfat thickness were recorded as previously described [4,7], when individual pigs reached a body-weight of about 130 kg.

2.2 SNP chip genotyping and quality control

All pigs were genotyped by Illumina PorcineSNP60 BeadChip, containing 62163 SNPs across the entire genome. The gPLINK software [8] was used for quality control of genotyped data according to the following criteria: (1) call rate > 0.9 , (2) minor allele frequency (MAF) > 0.01 , and (3) significant divergence from Hardy-Weinberg Equilibrium with P -values $> 10^{-6}$. For Population 1, a total of 56440 SNPs passed the quality control and were retained in the dataset. For Population 2, 13678 SNPs were removed and 49485 SNPs were kept for further analyzes. In addition, a total of 41635 SNPs were selected for cluster analysis of the two populations.

2.3 Statistical analysis

GWA analysis was performed with the Genome Association and Prediction Integrated Tool (GAPIT) using the compressed mixed linear model (CMLM) method [9]. The CMLM is described as follows:

$$Y = X_{\alpha} + P_{\beta} + K_{\mu} + e$$

where Y is the vector of observed phenotypes, X is a matrix representing genotypes, P is the principal component matrix for population structure, K is the kinship matrix. X_{α} and P_{β} are regarded as fixed effects and K_{μ} and e are regarded as random effects.

Population structure represents non-random distribution of genotypes among individuals within a population, and population structure and unequal relatedness among individuals in a population are the two major causes of false positive results in association studies [9,10]. In the CMLM method, population structure is fitted as a fixed

effect, whereas kinship among individuals is incorporated as the variance-covariance structure of the random effect for individuals. In addition, all individuals will be clustered into groups to give better control of population structure and kinship so as to reduce the false positive association rate [9]. SNPs are sorted by respective P -values and SNPs showing high significance for the trait will have the smallest P -values. Thirty SNPs with the lowest P -values were chosen for annotation by Ensembl *Sus scrofa* Build 10.2 (http://asia.ensembl.org/Sus_scrofa/Info/Index?db=core). A cluster analysis of populations was performed by principle component analysis (PCA) using GAPIT.

Haplotype block analysis was performed on the genome region containing the significantly associated SNPs in Population 2 and run with the default parameters using HAPLOVIEW program [11].

3 Results and discussion

Population stratification is one of the major factors influencing the validity of a GWA study [10]. The PCA clustering of two populations is shown in Fig. 1. For Population 1, the individuals from two genetic lines were classified into one cluster, suggesting there was no significant genetic difference between these lines. In addition, sows from Population 1 were obviously separated from those from Population 2.

GWA results for the tenth rib backfat thickness trait are shown in Fig. 2. Different associations between two pig populations were evident. The tenth rib backfat thickness trait is not only controlled by multiple genes but can also be affected to varying degrees by other environmental factors. For quantitative traits, it is difficult to confirm QTL or genes that have common effects and the effects, such as fatness and pork quality, could be verified in different pig populations [12]. Furthermore, the separation of two populations in clustering analysis (Fig. 1) suggested the existence of genetic differences between these populations, which might partly explain the different GWA results.

For Population 1 (Fig. 2a), the 30 most significant SNPs

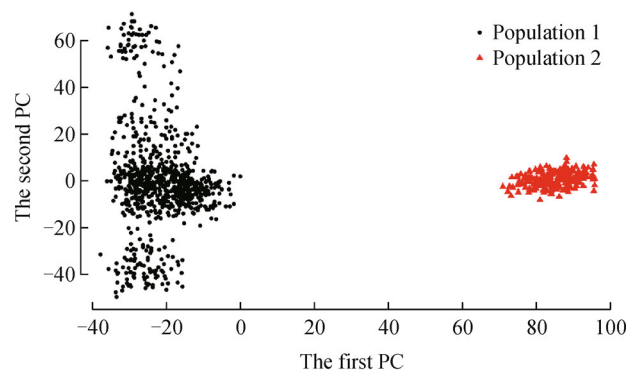


Fig. 1 Cluster analysis plotted by first versus second principal components (PC) for the two pig populations

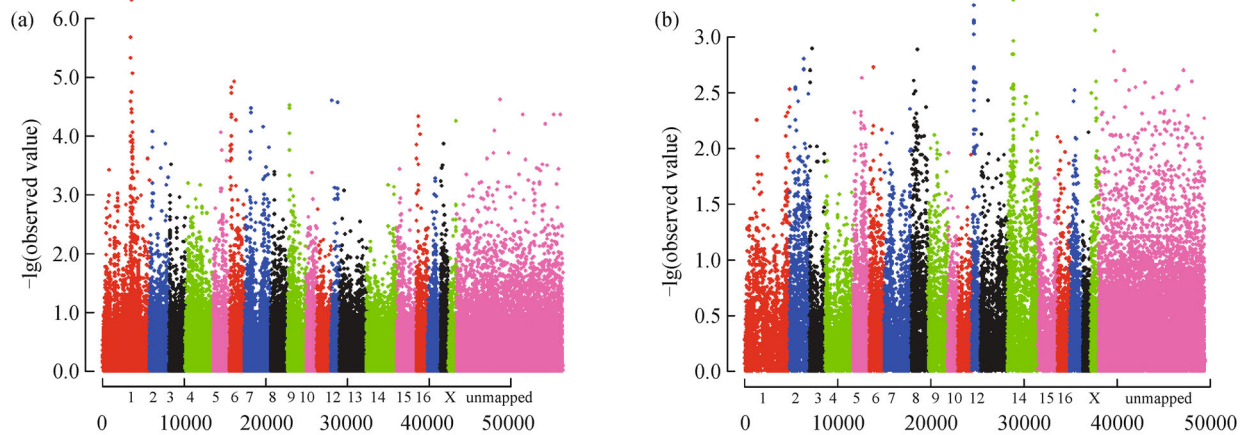


Fig. 2 Manhattan plots of genome-wide association studies with tenth rib backfat thickness in two pig populations. (a) Plot for population 1; (b) plot for population 2. SNPs on each chromosome are labeled with a different color on the X-axis as indicated, and the Y-axis is the negative log base 10 of the P -values. SNPs with tighter associations with the trait will have a larger Y-coordinate value.

were chosen for annotation and five SNPs were in disassembled scaffolds of *Sus scrofa* 10.2. Among the 25 mapped SNPs, the largest number were on SSC1 (9), and the rest were distributed on SSC6 (6), SSC7 (3), SSC9 (2), SSC12 (2), SSC16 (2) and SSCX (1). The location of the SNPs on SSC1 and SSC7 were consistent previous reports [7,13], and the SNPs on SSC1 showed the greatest association with the trait. The top four significant SNPs were located on SSC1 (ALGA0006623, ALGA0006599, INRA0004898 and DRGA0001605). Additionally, the most significant SNP (ALGA0006623) was located at around 528 kb upstream of *MC4R* gene. *MC4R*, a G-protein coupled receptor, has been implicated in mediating the effect of leptin on food intake and energy balance, and has previously been found to be associated with fat deposition in pig [1]. Also, *CCBE1* (collagen and calcium binding EGF domains 1) gene (within gene, ASGA0005017) on SSC1, has been related to fat metabolism through DAVID (<http://david.abcc.ncifcrf.gov/>) functional annotation [7]. Two SNPs (H3GA0020700 and MARC0069646) corresponded to a fat QTL region between TNFB (58 cM) and S0102 (70.1 cM) on SSC7 [13].

In Population 2 (Fig. 2b), the 27 most significant SNPs were mapped in the *Sus scrofa* 10.2, and 12 SNPs (M1GA0016251, MARC0075799, ASGA0053328, ALGA0065206, ALGA0065212, ALGA0065251, MARC0014203, M1GA0016298, ALGA0065308, ASGA0053453, MARC0040976 and ASGA0053456) were located within a 1.78 Mb segment (between 14.14 and 15.92 Mb) on SSC12, which corresponded to a QTL region associated with fat metabolism [14]. There were eight SNPs located within five annotated genes: *PITPNC1* (phosphatidylinositol transfer protein, cytoplasmic 1), *TEX2* (testis expressed 2), *FTSJ3* (FtsJ homolog 3),

MAP3K3 (mitogen-activated protein kinasekinase 3) and *MARCH10* (membrane-associated ring finger (C3HC4) 10, E3 ubiquitin protein ligase). Two haplotype blocks were identified on SSC12 for these 12 SNPs (Fig. 3). The first haplotype block, M1GA0016251-MARC0075799, was within an 18 kb fragment, located in the *PITPNC1* gene, which has been associated with fatty acid transport or metabolism. *PITPNC1*, also known as M-rdgB beta, encodes a protein that belongs to the Nir/rdgB family and has been implicated in a broad spectrum of cellular functions, such as regulation of lipid trafficking and metabolism [14]. The second haplotype block, ALGA0065251-MARC0014203-M1GA0016298-ALGA0065308, was within a 401 kb fragment, identified around 16 genes. Of them, *GHI* (growth hormone 1) has been reported to be associated with childhood obesity [15]. *GHI* is an endocrine factor secreted from the anterior pituitary gland that serves to synchronize growth and metabolism. *GHI* deficiency in mammals has been found to be associated with stunted somatic growth and increased adipose tissue [15]. Also, *GHI* gene was found to be related to fat deposition in pigs [12].

4 Conclusions

GWA studies for the tenth rib backfat thickness trait were undertaken on two pig populations of distinct genetic backgrounds, and the associated genes and genome regions have been identified. This not only confirmed the reported candidate genes but also revealed novel genes. The results differed between the two populations indicating that the genetic mechanisms regulating backfat thickness are complex and GWA results can be affected by the genetic background of the population studied.

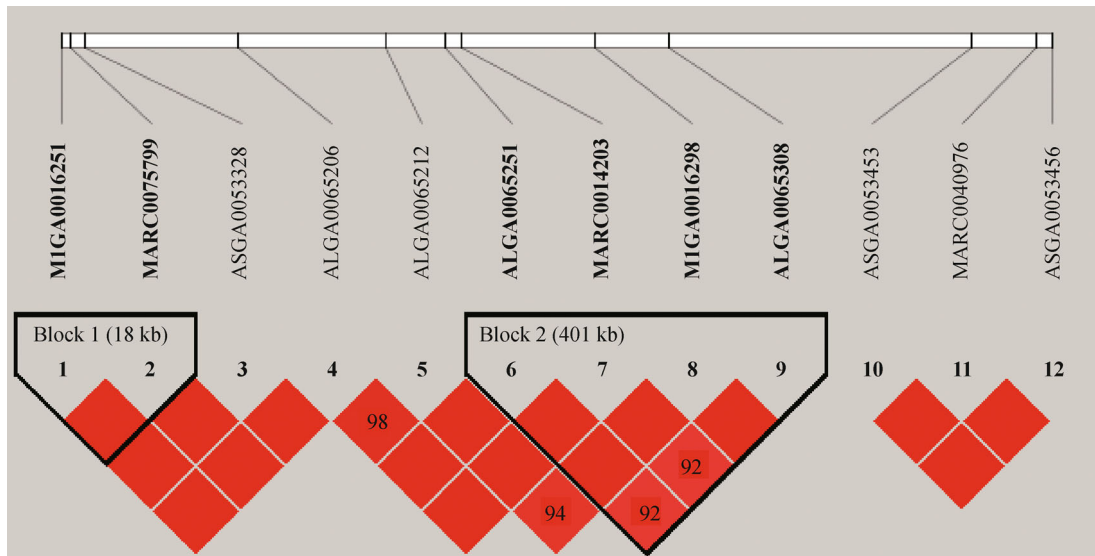


Fig. 3 Haplotypes on a 1.78 Mb region on SSC12 containing all the significant SNPs associated with the tenth rib backfat thickness for pigs in Population 2 obtained with the HAPLOVIEW program. The solid lines indicate the two blocks identified.

Acknowledgements This study was supported by the National Science Foundation of China (31172192), New Century Excellent Talents (NCET-11-0646) and Fundamental Research Funds for the Central Universities (2011JQ009, 2012PY009).

Compliance with ethics guidelines Dandan Zhu, Xiaolei Liu, Rothschild Max, Zhiwu Zhang, Shuhong Zhao and Bin Fan declare that they have no conflict of interest or financial conflicts to disclose.

All applicable institutional and national guidelines for the care and use of animals were followed.

References

- Kim K S, Larsen N, Short T, Plastow G, Rothschild M F. A missense variant of the porcine melanocortin-4 receptor (*MC4R*) gene is associated with fatness, growth, and feed intake traits. *Mammalian Genome*, 2000, 11(2): 131–135
- Fan B, Du Z Q, Rothschild M F. The fat mass and obesity-associated (*FTO*) gene is associated with intramuscular fat content and growth rate in the pig. *Animal Biotechnology*, 2009, 20(2): 58–70
- Bidanel J P, Milan D, Iannuccelli N, Amigues Y, Boscher M Y, Bourgeois F, Caritez J C, Gruand J, Le Roy P, Lagant H, Quintanilla R, Renard C, Gellin J, Ollivier L, Chevalet C. Detection of quantitative trait loci for growth and fatness in pigs. *Genetics Selection Evolution*, 2001, 33(3): 289–309
- Malek M, Dekkers J C, Lee H K, Baas T J, Rothschild M F. A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. I. Growth and body composition. *Mammalian Genome*, 2001, 12(8): 630–636
- Ramos A M, Crooijmans R P, Affara N A, Amaral A J, Archibald A L, Beever J E, Bendixen C, Churcher C, Clark R, Dehais P, Hansen M S, Hedegaard J, Hu Z L, Kerstens H H, Law A S, Megens H J, Milan D, Nonneman D J, Rohrer G A, Rothschild M F, Smith T P, Schnabel R D, Van Tassell C P, Taylor J F, Wiedmann R T, Schook L B, Groenen M A. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS ONE*, 2009, 4(8): e6524
- Do D N, Ostersen T, Strathe A B, Mark T, Jensen J, Kadarmideen H N. Genome-wide association and systems genetic analyses of residual feed intake, daily feed consumption, backfat and weight gain in pigs. *BMC Genetics*, 2014, 15(1): 27
- Fan B, Onteru S K, Du Z Q, Garrick D J, Stalder K J, Rothschild M F. Genome-wide association study identifies Loci for body composition and structural soundness traits in pigs. *PLoS ONE*, 2011, 6(2): e14726
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M A, Bender D, Maller J, Sklar P, de Bakker P I, Daly M J, Sham P C. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 2007, 81(3): 559–575
- Zhang Z, Ersoz E, Lai C Q, Todhunter R J, Tiwari H K, Gore M A, Bradbury P J, Yu J, Arnett D K, Ordovas J M, Buckler E S. Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 2010, 42(4): 355–360
- Pearson T A, Manolio T A. How to interpret a genome-wide association study. *The Journal of the American Medical Association*, 2008, 299(11): 1335–1344
- Barrett J C, Fry B, Maller J, Daly M J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 2005, 21(2): 263–265
- Li M, Wu H, Luo Z, Xia Y, Guan J, Wang T, Gu Y, Chen L, Zhang K, Ma J, Liu Y, Zhong Z, Nie J, Zhou S, Mu Z, Wang X, Qu J, Jing L, Wang H, Huang S, Yi N, Wang Z, Xi D, Wang J, Yin G, Wang L, Li N, Jiang Z, Lang Q, Xiao H, Jiang A, Zhu L, Jiang Y, Tang G, Mai M, Shuai S, Li N, Li K, Wang J, Zhang X, Li Y, Chen H, Gao X, Plastow G S, Beck S, Yang H, Wang J, Wang J, Li X, Li R. An atlas of DNA methylomes in porcine adipose and muscle tissues. *Nature Communications*, 2012, 3: 850

13. Yue G, Stratil A, Cepica S, Schröffel JJr, Schröfelova D, Fontanesi L, Cagnazzo M, Moser G, Bartenschlager H, Reiner G, Geldermann H. Linkage and QTL mapping for *Sus scrofa* chromosome 7. *Journal of Animal Breeding and Genetics*, 2003, 120(s1): 56–65
14. Muñoz M, Alves E, Corominas J, Folch J M, Casellas J, Noguera J L, Silió L, Fernández A I. Survey of SSC12 regions affecting fatty acid composition of intramuscular fat using high-density SNP data. *Frontiers in Genetics*, 2011, 2: 101
15. Wabitsch M. Overweight and obesity in European children: definition and diagnostic procedures, risk factors and consequences for later health outcome. *European Journal of Pediatrics*, 2000, 159 (S1): S8–S13