



Københavns Universitet

**Additive and epistatic genome-wide association for growth and ultrasound scan measures of carcass-related traits in Brahman cattle**

Ali, A. A.; Khatkar, M. S.; Kadarmideen, Haja; Thomson, P. C.

*Published in:*  
Journal of Animal Breeding and Genetics

*DOI:*  
[10.1111/jbg.12147](https://doi.org/10.1111/jbg.12147)

*Publication date:*  
2015

*Document version*  
Publisher's PDF, also known as Version of record

*Citation for published version (APA):*  
Ali, A. A., Khatkar, M. S., Kadarmideen, H., & Thomson, P. C. (2015). Additive and epistatic genome-wide association for growth and ultrasound scan measures of carcass-related traits in Brahman cattle. *Journal of Animal Breeding and Genetics*, 132(2), 187-197. <https://doi.org/10.1111/jbg.12147>



INVITED REVIEW

# Additive and epistatic genome-wide association for growth and ultrasound scan measures of carcass-related traits in Brahman cattle

A.A. Ali<sup>1</sup>, M.S. Khatkar<sup>1</sup>, H.N. Kadarmideen<sup>2</sup> & P.C. Thomson<sup>1</sup>

<sup>1</sup> Faculty of Veterinary Science, University of Sydney, Camden, NSW, Australia

<sup>2</sup> Faculty of Health and Medical Sciences, University of Copenhagen, Frederiksberg C, Denmark

## Keywords

Beef cattle; bovine; epistasis; genome-wide association studies; insulin-like growth factor-1; single nucleotide polymorphism.

## Correspondence

P.C. Thomson, 425 Werombi Road, Camden, NSW 2570, Australia.

Tel: +61 2 9351 1654;

Fax: +61 2 9351 1693;

E-mail: peter.thomson@sydney.edu.au

Received: 14 November 2014;

accepted: 2 February 2015

## Summary

Genome-wide association studies are routinely used to identify genomic regions associated with traits of interest. However, this ignores an important class of genomic associations, that of epistatic interactions. A genome-wide interaction analysis between single nucleotide polymorphisms (SNPs) using highly dense markers can detect epistatic interactions, but is a difficult task due to multiple testing and computational demand. However, it is important for revealing complex trait heredity. This study considers analytical methods that detect statistical interactions between pairs of loci. We investigated a three-stage modelling procedure: (i) a model without the SNP to estimate the variance components; (ii) a model with the SNP using variance component estimates from (i), thus avoiding iteration; and (iii) using the significant SNPs from (ii) for genome-wide epistasis analysis. We fitted these three-stage models to field data for growth and ultrasound measures for subcutaneous fat thickness in Brahman cattle. The study demonstrated the usefulness of modelling epistasis in the analysis of complex traits as it revealed extra sources of genetic variation and identified potential candidate genes affecting the concentration of insulin-like growth factor-1 and ultrasound scan measure of fat depth traits. Information about epistasis can add to our understanding of the complex genetic networks that form the fundamental basis of biological systems.

## Introduction

Growth and body composition traits in northern Australian cattle can be impacted by factors such as external and internal parasites, other diseases, heat, humidity and seasonal variation in the quantity of natural pasture availability and its generally low quality. Genetic improvement has been achieved, as a means of improving production efficiently in tropical cattle, by selection for appropriate traits related to growth, body composition and tolerance to environmental stressors (Burrow 2012).

Phenotypic selection to improve growth rates of beef cattle in the tropics has been successful using traditional selection practices over multiple generations based on cattle performance records and pedigree information (Koots *et al.* 1994). However, the genetic gain might be accelerated if polymorphisms responsible for growth differences can be determined at the genome level. Chromosomal regions, genes and specific polymorphisms associated with growth have been investigated previously (Casas *et al.* 2003; Allan *et al.* 2007). A number of studies identified chromosomal regions harbouring variation affecting growth

and reproduction in cattle (Snelling *et al.* 2010; Bolormaa *et al.* 2011; Hawken *et al.* 2012). Recently, systems biology approaches, including gene networks, have been applied to the genetic dissection of complex phenotypes (Fortes *et al.* 2013; Kadarmideen 2014). However, most GWAS have focused on identifying single-locus effects, without considering interaction between loci, that is 'epistasis'; when epistasis is ignored, some important loci may remain undetected. In its simplest definition, epistasis refers to an interaction between a pair of loci (or gene–gene interactions), in which the phenotypic effect of genotype at one locus depends on the genotype at the second locus.

Testing the interactions between all pairs of single nucleotide polymorphisms (SNPs) using the currently available high-density bovine SNP chip, which can comprise more than 700K SNPs (Rincon *et al.* 2011), is a difficult task due to statistical complexity (e.g. multiple testing) and computational burden (Marchini *et al.* 2005). For these reasons, epistasis is not yet routinely analysed in complex trait studies.

Rather than testing all possible pair-wise comparisons, a more practical strategy might be to examine a subset of SNPs which could have influence on a trait of interest. The objective of this study was to demonstrate a three-stage approach: (i) fit a model without the SNP to estimate the variance components using an iterative REML procedure; (ii) a SNP analysis using these variance components, thus avoiding REML iteration; and (iii) use the significant SNPs for genome-wide epistasis analysis (GWEA) using field data for growth and real-time ultrasound scan measures for subcutaneous fat thickness in Brahman cattle.

## Materials and methods

All animal procedures were approved by the J.M. Rendel Laboratory Animal Ethics Committee (CSIRO, Queensland) as approvals TBC107 (1999–2009) and RH225-06 (2006–2010).

### Animals, traits and genotypes

The Brahman breed heifers used in this study were born from 1999 to 2003, using artificial insemination and natural service, on four cooperative properties and the former 'Belmont' research station of CSIRO at Rockhampton, Queensland, Australia. Cattle were part of the Northern Breeding Project of the Cooperative Research Centre for Beef Genetic Technologies (Beef CRC) in the tropical regions of northern Australia (Upton *et al.* 2001).

Animals were progeny of 54 sires; the sire parentage was determined by DNA fingerprinting (Vankan & Burns 1997). A total of 1027 females were recorded for growth and real-time ultrasound scan measures for fat thickness. Details of the population and traits measured have been described previously (Barwick *et al.* 2009). Briefly, traits were recorded on animals at two occasions: (i) when the mean age of heifers was 18 months, near the end of the first postweaning wet season (W1) and (ii) when the mean age was 24 months, near the end of the dry season (D2). Growth traits recorded were as follows: live weight (WT, kg), average daily gain (ADG, g/day), body condition score (BCS, recoded scores to 1–15 scale), hip height (HH, cm) and serum concentration of the insulin-like growth factor-1 (IGF-1, ng/ml). Traits measured with ultrasound scan were rump fat thickness at the P8 position (SP8, mm), fat depth measured between the 12th and 13th ribs (SRIB, mm) and area of the *logissimus thoracis et lumborum* muscle area at the 12th/13th rib (SEMA, cm<sup>2</sup>). Of the 1027 animals with the growth and ultrasound scan records, 584 individuals were sampled randomly for genotyping. Blood samples were collected by venipuncture of a tail vein, and genomic DNA was extracted. DNA samples were genotyped at 9075 SNP markers using the Affymetrix bovine 10K SNP array (Khatkar *et al.* 2007). The SNPs in this panel are distributed across the genome with an average spacing of approximately 325 kb between SNPs.

To ensure reliable quality data for the GWAS analysis, the following exclusion measures were applied: a SNP or animal with call rate <90%, SNP with minor allele frequency (MAF) of <0.01, SNP not in Hardy–Weinberg equilibrium ( $p \leq 0.0001$ ), SNP not on an autosome or with location unknown. After quality control, data on 565 animals and 6714 SNP markers were available for further analysis.

The genotypes for each SNP were coded as 0, 1 or 2 copies of the allele B for genotypes AA, AB and BB, respectively. Missing genotypes were imputed using BEAGLE (Browning & Browning 2011). All SNPs were mapped to the current UMD 3.1 build of the bovine genome sequence assembled by the University of Maryland.

### Statistical analyses

#### Quantitative analysis

Descriptive statistics and genetic parameters (e.g. variance components and heritability) were estimated for all traits considered. Univariate animal models were used to estimate fixed effects and variance

components using REML methods in the absence of a SNP. The heritability estimates were calculated as the proportion of additive genetic variance to the phenotypic variance.

The following linear mixed model was fitted to the phenotypic data:

$$Y_i = \mu + \text{CalveM} + \text{DAge} + \text{Herd} + \text{Cohort} + \text{Animal} + \varepsilon_i, \quad (1)$$

where  $Y_i$  = observation of trait on animal  $i$ ,  $\mu$  = overall mean, CalveM = calving month (levels, 1–8), DAge = dam age (six levels, 3–8 years), Herd = property of origin (five levels, 1–5 properties), Cohort = combination of herd of origin and year of birth (11 levels, 1–11) are fixed effects. Animal and  $\varepsilon_i$  (residual error) are random effects. The vector of random animal effects,  $\mathbf{a} = \{\text{Animal}\}$  in eqn (1), was assumed to be normally distributed with zero mean and variance  $\text{var}(\mathbf{a}) = \mathbf{A}\sigma_{\text{Animal}}^2$ , where  $\mathbf{A}$  is the numerator relationship matrix across all animals derived from the available pedigree information, and  $\sigma_{\text{Animal}}^2$  is the additive polygenic variance component. Data from all the 1027 animals with growth and ultrasound records were used for model fitted using eqn (1).

Similarly, the vector of random errors  $\boldsymbol{\varepsilon} = \{\varepsilon\}$  is assumed to have mean zero and variance  $\text{var}(\boldsymbol{\varepsilon}) = \mathbf{I}\sigma_{\varepsilon}^2$ , where  $\mathbf{I}$  is an identity matrix;  $\sigma_{\varepsilon}^2$  is the residual variance component. Fixed effects for the traits considered have been described previously (Barwick *et al.* 2009). All analyses, including subsequent models, were performed using ASREML-R software (Butler *et al.* 2009).

#### Additive genome-wide association model

The additive effect of a SNP was modelled as a covariate, being the number of copies of the allele (0, 1 or 2). The following linear mixed model was fitted to the data:

$$Y_i = \mu + \text{CalveM} + \text{DAge} + \text{Herd} + \text{Cohort} + \beta\text{SNP} + \text{Animal} + \varepsilon_i, \quad (2)$$

where all terms are as defined in eqn (1), and SNP represents the additive effect of the SNP (0, 1 or 2 copies of the allele at the SNP, as a covariate). The additive polygenic variance and residual variances estimated using eqn (1), that is  $\hat{\sigma}_{\text{Animal}}^2$  and  $\hat{\sigma}_{\varepsilon}^2$ , are assumed for this model. This allows the mixed model equations to be solved noniteratively at each SNP, thus reducing the computational burden. The additive effect of each SNP was tested using a Wald test against a null hypothesis of a SNP effect  $\beta = 0$ . Corrections for

multiple testing were performed using the  $q$ -value false discovery rate (FDR) package in R software (Storey & Tibshirani 2003) which uses characteristics of the  $p$ -value distribution to compute  $q$ -value thresholds. This method maps  $p$ -value into corresponding  $q$ -values by first estimating the proportion of loci not associated with the trait of interest ('null'  $p$ -values,  $\pi_0$ ). A SNP was consequently declared significant at a threshold of  $q$ -value  $\leq 0.01$  which, for example, may correspond to  $p$ -value  $\leq 10^{-6}$  in particular set of tests. We also considered a more moderate significance threshold where a SNP was considered to have a suggestive association with  $q$ -value  $\leq 0.05$ . In total, 565 animals with both phenotype and genotype data were utilized in eqn (2) and subsequently in eqn (3).

#### Genome-wide epistatic association model

Following the procedure proposed by Marchini *et al.* (2005), a subset of the most significant SNPs identified by the additive association models were selected for epistatic testing. The threshold for selecting candidate SNPs for testing was a  $p$ -value of 0.01 from the association model in eqn (2). This relaxed threshold was chosen to capture SNP loci with small effects on the trait of interest. However, not testing all pair-wise SNP combinations substantially reduces the computational and multiple testing burdens.

The effect of two loci (e.g. SNP<sub>1</sub> and SNP<sub>2</sub>) on a trait in question was estimated using a mixed model as follows:

$$Y_i = \mu + \text{CalveM} + \text{DAge} + \text{Herd} + \text{Cohort} + \text{SNP}_1 + \text{SNP}_2 + \text{SNP}_1.\text{SNP}_2 + \text{Animal} + \varepsilon_i, \quad (3)$$

where SNP<sub>1</sub> and SNP<sub>2</sub> are three-level factors for genotypes (e.g. AA, AB and BB) at SNP<sub>1</sub> and SNP<sub>2</sub>, and SNP<sub>1</sub>.SNP<sub>2</sub> is the interaction between SNP<sub>1</sub> and SNP<sub>2</sub> genotypes as an indicator for an epistasis effect. All other terms in the model were the same as in eqn (2). Note that while the SNP effect for the additive association model was treated as a covariate (0, 1 or 2 copies of the variant allele) to maximize the power of association detection, for estimating interactions (e.g. epistasis) it is necessary to treat the effect of the SNPs as factors. A separate model was fitted for each pair-wise combination of selected SNPs. Only first-order interactions between pairs of SNPs were considered in the model, as inclusion of all possible higher order interactions is computationally intractable. Corrections for multiple testing was made using the  $q$ -value procedure and evaluated at several FDR thresholds (0.1, 0.01, 0.001, 0.0001 and 0.00001). A search for

significant interacting SNPs in a gene was conducted using the public domain dbSNP database in NCBI (<http://www.ncbi.nlm.nih.gov/>).

The selection and evaluating a subset of SNPs that meet a predefined threshold cut-off in a single-SNP association test were subsequently followed up in a pair-wise SNP epistatic analysis, such an approach would not only reduce significantly the computational run time and storage requirement, but also provide advantage in statistical power since the penalty due to multiple testing would not be great as in all pair-wise combination test of SNP markers.

## Results

Trait means and heritability for the growth and ultrasound scan measures of carcass-related traits are presented in Table 1. The number of records differs slightly across the two seasons, and fewer animals were recorded for hip height in the wet season and IGF-1 in the dry season. The trait means varied between wet and dry season which may be mainly due to the differences in the age of animals. Mean W1ADG was higher in the wet season than the dry season. The heritability for growth and carcass-related traits was moderate to high and ranged from

0.14 ± 0.07 to 0.60 ± 0.13, with the large standard errors reflecting a small sample size. Heritability estimates were similar between traits measured in wet and dry seasons.

### Genome-wide associations

Results of the GWAS analyses for growth and carcass-related traits are presented in Tables 2 and 3 and Figure 1. Manhattan plots for other ultrasound-scanned carcass traits are given in Figures S1 and S2 in the supporting information. Tables and plots included all the significant SNP results. A  $p$ -value of  $1.0 \times 10^{-6}$  which corresponds to a  $q$ -value of 0.01 was considered to identify SNPs with a significant association with a trait being tested. With this threshold, six SNPs were identified as being significantly associated with at least one trait considered: these SNPs were detected on BTA 14 at genome positions from 21 to 35 Mb, and most of these SNPs were associated with more than one trait. The most significant SNP association ( $p$ -value =  $4.9 \times 10^{-11}$ ;  $q$ -value =  $3.3 \times 10^{-7}$ ; MAF = 0.498) was rs29020688 on BTA 14 at 24 Mb for D2SP8 of the rump fat depth measured in the dry season.

As also shown in Tables 2 and 3, the other significant SNPs were rs29010515, rs29010516, rs29010471, rs29011751 and rs29025824 for different traits considered. The rs29020688, rs29010515 and rs29010516 markers were consistently associated with IGF-1, rump fat depth at P8 and fat depth between the 12th and 13th ribs measured in the wet and dry seasons. Q-Q plots of these traits displayed an excess of small observed  $p$ -values compared with the expected theoretical distribution; this excess indicates a likely true associations of SNPs with traits, rather than an associations by chance (see supporting information, Figures S3 to S5).

### Genome-wide epistatic associations

The numbers of SNPs with nominal  $p$ -values  $\leq 0.01$  that were selected from the single-SNP GWAS study are given in Table S1 in the supporting information. The mean number of SNPs available per trait for SNP-SNP interaction analyses was fairly similar ( $89.4 \pm 6.5$  and  $88.1 \pm 10.8$ ) for traits measured in wet and dry seasons, respectively.

The number of significant epistatic associations at five different false discover rate (FDR) thresholds (FDR at 0.1, 0.01, 0.001, 0.0001 and 0.00001) is presented in Table 4. Significant epistasis at the lowest FDR (0.1) was observed for the same three traits in

**Table 1** Descriptive statistics and heritability for growth and ultrasound scan measures of carcass-related traits in Brahman cattle ( $n = 1027$ )

Time	Trait	Number of animals	Mean ± SD	Heritability ± SE
W1	W1WT	1027	228 ± 43.84	0.35 ± 0.11
	W1ADG	1021	607.96 ± 148.26	0.20 ± 0.09
	W1BCS	954	5.30 ± 1.35	0.41 ± 0.11
	W1HH	616	127.5 ± 4.90	0.42 ± 0.13
	W1IGF-1	916	182.65 ± 84.27	0.32 ± 0.11
	W1SP8	1027	3.73 ± 1.93	0.51 ± 0.12
	W1SRIB	1027	1.98 ± 1.02	0.60 ± 0.13
	W1SEMA	1024	44.13 ± 6.64	0.24 ± 0.09
	D2	D2WT	1026	320.05 ± 58.71
D2ADG		1022	138.25 ± 233.88	0.14 ± 0.07
D2BCS		1026	5.44 ± 1.40	0.42 ± 0.12
D2HH		941	132.4 ± 4.90	0.40 ± 0.12
D2IGF-1		759	215.39 ± 92.26	0.41 ± 0.13
D2SP8		1025	3.21 ± 1.76	0.44 ± 0.11
D2SRIB		1025	1.92 ± 1.00	0.47 ± 0.11
D2SEMA		1023	44.08 ± 8.85	0.33 ± 0.11

W1, traits measured in the wet season; D2, traits measured in the dry season; SD, standard deviation; SE, standard error; WT, live weight; ADG, average daily gain; BCS, body condition score; HH, hip heights; IGF-1, insulin-like growth factor-1; SP8, ultrasound scan of rump fat depth at P8 site; SRIB, ultrasound scan of fat depth measure between 12th and 13th ribs; SEMA, ultrasound scan of eye muscle area.

**Table 2** Significant SNP associations for insulin-like growth factor-1 and ultrasound scan of fat depth measured in wet season for female Brahman cattle

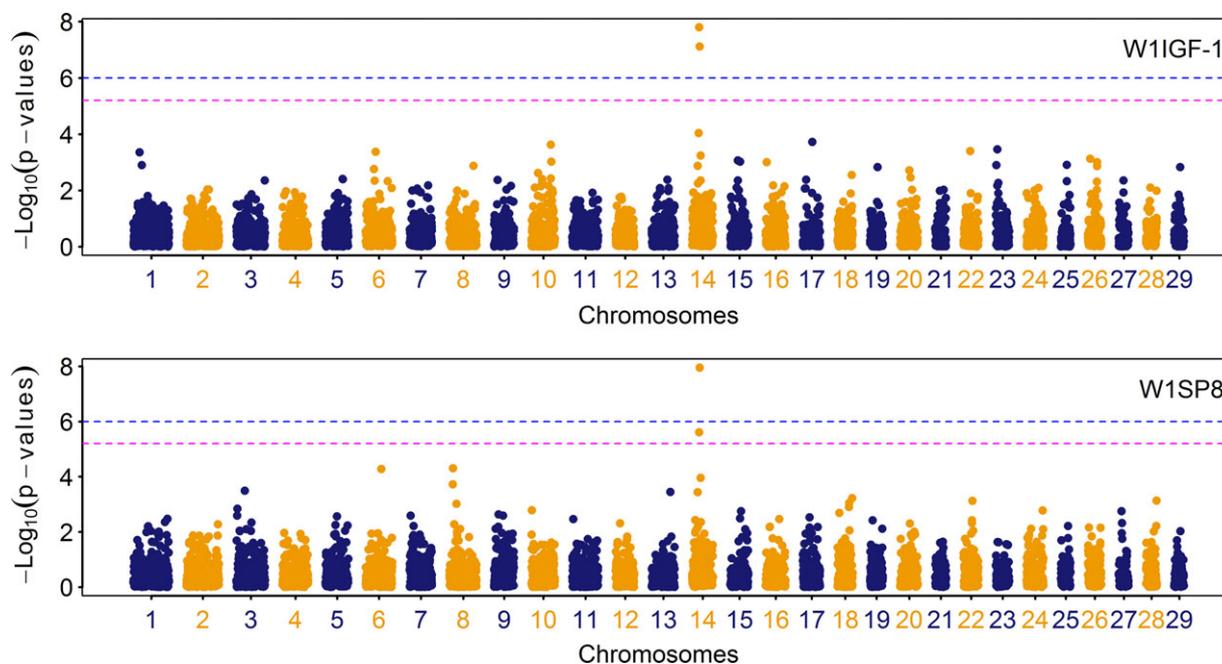
Trait	SNP	BTA	Position	Allele	MAF	p-Value	q-Value
W1IGF-1	rs29010516	14	22 382 053	A/T	0.492	$1.6 \times 10^{-8}$	$5.2 \times 10^{-5}$
	rs29010515	14	22 382 090	G/T	0.492	$1.6 \times 10^{-8}$	$5.2 \times 10^{-5}$
	rs29020688	14	24 365 206	A/G	0.498	$7.8 \times 10^{-8}$	<0.001
W1SP8	rs29010471	14	21 128 800	C/T	0.138	$9.4 \times 10^{-7}$	0.001
	rs29020688	14	24 365 206	A/G	0.498	$1.1 \times 10^{-8}$	$7.4 \times 10^{-5}$
	rs29010516	14	22 382 053	A/T	0.492	$2.5 \times 10^{-6}$	0.005
W1SRIB	rs29010515	14	22 382 090	G/T	0.492	$2.5 \times 10^{-6}$	0.005
	rs29010516	14	22 382 053	A/T	0.492	$3.7 \times 10^{-7}$	0.008
	rs29010515	14	22 382 090	G/T	0.492	$3.7 \times 10^{-7}$	0.008
	rs29020688	14	24 365 206	A/G	0.498	$6.6 \times 10^{-7}$	0.004

W1IGF-1, insulin-like growth factor-1; W1SP8, ultrasound scan of rump fat depth at P8 site; W1SRIB, ultrasound scan of fat depth between the 12th and 13th ribs measured in the wet season; MAF, minor allele frequency.

both wet and dry seasons, namely IGF-1, rump fat depth at P8 site and fat depth measured between the 12th and 13th ribs. The highest number of significant epistasis effects across each FDR threshold was detected for IGF-1 trait measured in the wet season. Indeed, this is the only trait for which any significant epistasis was detected at FDR thresholds greater than the lowest threshold. Only two significant epistatic associations were identified for rump fat depth at P8

site. No significant epistasis was observed for WT, HH, BCS and EMA traits in either wet or dry seasons.

The graphical representations of the genome-wide association results for IGF-1 measured in the wet season are presented as an image plot in Figure 2. The additional image plots for other ultrasound-scanned carcass and serum concentration of IGF-1 traits are given in Figures S6 to S10 in the supporting information. These heatmap plots show the strongest epistatic



**Figure 1** Manhattan plots of genome-wide association. Top: for W1IGF-1, bottom: for ultrasound scan for W1SP8. The significance threshold (horizontal blue dashed line) is  $p\text{-value} \leq 10^{-6}$  ( $q\text{-value} < 0.01$ ) and suggestive significance (horizontal red dashed line) is  $p\text{-value} 5.2 \times 10^{-5}$  ( $q\text{-value} < 0.05$ ). Note the significant association on chromosome 14.

**Table 3** Significant SNP associations for insulin-like growth factor-1 and ultrasound scan of fat depth measured in dry season for female Brahman cattle

Trait	SNP	BTA	Position	Allele	MAF	p-Value	q-Value
D2IGF-1	rs29010516	14	22382053	A/G	0.492	$1.3 \times 10^{-7}$	0.002
	rs29020688	14	24365206	A/G	0.498	$4.9 \times 10^{-6}$	0.026
D2SP8	rs29020688	14	24365206	A/G	0.498	$4.9 \times 10^{-11}$	$3.3 \times 10^{-7}$
	rs29010516	14	22382053	A/T	0.492	$4.0 \times 10^{-7}$	0.001
	rs29010515	14	22382090	G/T	0.492	$4.0 \times 10^{-7}$	0.001
	rs29011751	14	30096108	C/T	0.438	$6.7 \times 10^{-7}$	0.001
	rs29025824	14	35755282	A/T	0.354	$8.2 \times 10^{-6}$	0.011
D2SRIB	rs29020688	14	24365206	A/G	0.498	$1.0 \times 10^{-8}$	$6.7 \times 10^{-5}$
	rs29010516	14	22382053	A/T	0.492	$7.7 \times 10^{-7}$	0.002
	rs29010515	14	22382090	G/T	0.492	$7.7 \times 10^{-7}$	0.002
	rs29011751	14	30096108	C/T	0.438	$1.4 \times 10^{-6}$	0.023

D2IGF-1, insulin-like growth factor-1; SP8, ultrasound scan of rump fat depth at P8 site; D2SRIB, ultrasound scan of fat depth between the 12th and 13th ribs measured in the dry season.

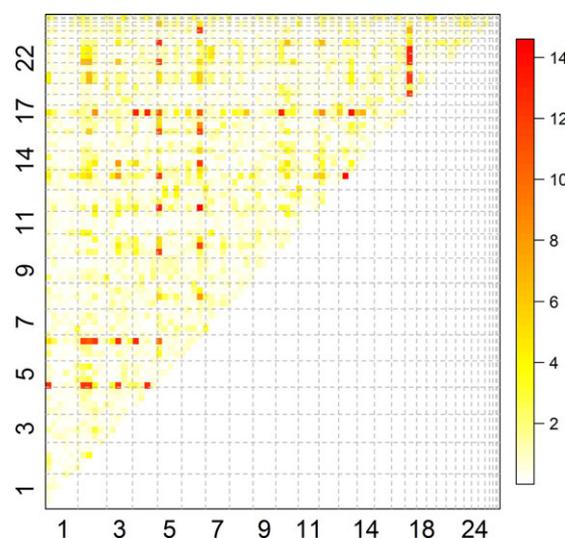
**Table 4** Number of significant epistatic associations at five false discovery rates (FDR) threshold for growth and carcass traits in tropical Brahman heifers

Trait		FDR threshold				
		0.1	0.01	0.001	0.0001	0.00001
Wet season						
W1WT	NS					
W1BCS	NS					
W1HH	NS					
W1IGF-1	229	136	89	43	24	
W1EMA	NS					
W1P8	2					
W1RIB	7					
Dry season						
D2WT	NS					
D2BCS	NS					
D2HH	NS					
D2IGF-1	8					
D2EMA	NS					
D2P8	2					
D2RIB	4					

Prefix W1, traits measured in the wet season; prefix D2, traits measured in the dry season; Traits: WT, live weight (kg); BCS, body condition score (score 1–12); HH, hip height (cm); IGF-1, serum level of insulin-like growth factor-1(mg/ml); EMA, ultrasound scan for eye muscle area (cm<sup>2</sup>); P8, ultrasound scan for fat depth at P8 site (mm); RIB, ultrasound scan for rib fat depth (mm).

signals or the so-called hot spots (dark red colour spots) where significant epistatic association exists at genome-wide level. The scale on the right-hand side of the plot represents  $-\log_{10}(p\text{-value})$ . The threshold  $-\log_{10}(p\text{-value})$  of each heatmap plot was indicated in the caption.

Based on the  $q$ -value threshold of 0.01, 136 significant epistatic interactions were identified for

**Figure 2** Genome-wide epistasis association for insulin-like growth factor-1 measured in the wet season. The heatmap legend scale (right side) is on  $-\log_{10}(p\text{-value})$  scale. Note values corresponding to  $-\log_{10}(p\text{-value}) \geq 5$  ( $q\text{-value} < 0.01$ ) indicate significant epistatic signals.

WIIGF-1 (see supporting information, Table S2). These involved 56 SNPs spanning 19 chromosomes, namely BTA 5–10, 13–20, 22, 23 and 25–29; the largest number of SNPs ( $n = 12$  SNPs) was on BTA 10. Of the 136 epistatic pairs detected, 22 epistatic associations involving SNPs on BTA 5, 6, 8, 9, 10, 15, 16, 17, 19, 22, 23, 25 and 26, and 29 were highly significant ( $q\text{-value} \leq 1.4 \times 10^{-10}$ ). The strongest significant epistatic effects for WIIGF-1 ( $q\text{-value} = 6.2 \times 10^{-12}$ ) were detected between rs29022513 on BTA 10 (86 513 542 bp) and rs29020759 on BTA 16 (7 817 810 bp), and between rs29016126 on BTA 17 (44 319 169 bp) and rs29013846 also on BTA 17

(18 809 070 bp). An additional 23 significant epistatic interactions involving 32 SNPs on 17 chromosomes (BTA 1, 2, 4, 5, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 22, 25 and 27) were detected using a more lenient threshold  $q$ -value of 0.1 for W1P8, W1RIB, D2IGF-1, D2P8 and D2RIB traits (see supporting information Table S3). Of these 32 SNPs detected in epistatic interactions, 11 SNPs on BTA 1, 4, 5, 9, 10, 16, 17, 25 and 26 were associated with W1RIB; 10 SNPs on BTA 2, 11, 12, 14 and 22 were associated with D2IGF-1; seven SNPs on BTA 1, 4, 5, 12, 14, 25 and 27 were associated with D2RIB; four SNPs on BTA 2, 14, 20 and 22 were associated with W1P8; and three SNPs on BTA 1, 13 and 15 were epistatically associated with D2P8. The strongest of the FDR = 0.1 epistatic associations was observed between rs29011033 on BTA 26 at 39 051 127 bp position and rs29017479 on BTA 9 at 75 753 669 bp position ( $q$ -value = 0.0018), and between rs29021134 on BTA 12 at 59 528 070 bp and rs29025462 on BTA 27 at 25 615 010 bp ( $q$ -value = 0.0009) for W1RIB and D2RIB, respectively.

Table S4 presents the genes that the SNPs showing significant epistasis are located in: we call these positional epistatic genes. Twenty epistatic interactions between SNPs on 15 chromosomes (BTA 1, 5, 6, 8, 10, 13, 14, 16, 17, 18, 19, 20, 23, 25 and 26) were observed for W1IGF1 and D2SP8 traits. These epistatic SNP pairs in Table S4 were the only SNP pairs for which a gene could be found in at least one SNP. However, there were no cases where both members of an epistatic pair of SNPs were located in the same gene. These epistatic SNPs were found to be located in the following genes: SNX30 (BTA 8; 103709538...103811065), PRKDC (BTA 14; 107610798...107695100), MTSS1 (BTA 14; 16970557...17132800), SUS4 (BTA 16; 27388681...27522205), ARHGAP35 (BTA 18; 54425390...54497692), CDH18 (BTA 20; 53409771...54015683), CPNE5 (BTA 23; 10631429...10735195), PRKCB (BTA 25; 21759551...22137203), PRKG1 (BTA 26; 6901760...8343635) and DZANK1 (BTA 13; 38782686...38842977), and SNPs: rs29016259 (BTA 1), rs29023212 (BTA 5), rs29023213 (BTA 5), rs29012836 (BTA 6), rs29016280 (BTA 6), rs29022513 (BTA 10), rs29016126 (BTA 17), rs29013846 (BTA 17), rs29018088 (BTA 19) and rs29014419 (BTA 23). Nine of these ten epistatic signals were associated with W1IGF-1.

## Discussion

Genome-wide pair-wise SNP interaction using high-density SNP chip genotypes is a difficult task due to

statistical complexity (e.g. multiple testing) and computational runtime problems. To overcome these problems, we employed a three-stage approach: (i) fit a model without SNPs to estimate the variance components; (ii) a single-SNP analysis using these variance components, thus avoiding iteration; and (iii) a subset of the significant SNPs that met a lenient threshold  $p$ -value of 0.01 was selected for GWEA using field data for growth and real-time ultrasound scan measures for subcutaneous fat thickness in Brahman cattle.

## Genome-wide associations

Using 6714 quality-controlled SNP markers of the bovine 10K SNP chip, significant genetic variants associated with IGF-1, ultrasound scan for rump fat at P8 site and rib fat depth traits were identified. Several SNPs were significantly associated with more than one trait; hence, there is some evidence of pleiotropic effects. Overall, six SNPs were significantly associated with at least one trait. All of these SNPs were located on chromosome 14, in the genomic regions between 21 and 35 Mb. The strongest signal was detected with rs29020688 on BTA 14 at the genomic location of 24 Mb for fat depth at P8 site measured in the dry season. This rs29020688 marker was located in the intron region of *Bostaurus* XK, Kell blood group complex subunit-related family, member 4 (XKR4) gene (Gene ID: 517598 on chromosome 14; 24295405...24614310), and it is an XK-related protein 4 coding gene. Recently, Bolormaa *et al.* (2011) and Porto Neto *et al.* (2012) identified XKR4 as a possible candidate gene affecting rump fat depth in Australian indicine and composite cattle.

However, our results indicate that the XKR4 gene was also associated with IGF-1 and fat depth between the 12th and 13th ribs. If this finding is reproduced by other studies using high-density SNP markers genotyped with a large number of animals, the SNP rs29020688 marker in the XKR4 gene could potentially be used as a biomarker for subcutaneous fat composition trait. The SNP rs29010471 on BTA 14 at genome position of 21 128 800 bp was associated with IGF-1 level. This SNP is located in the intron of *B. taurus* protein kinase, DNA-activated, catalytic polypeptide (PRKDC) gene (Gene ID: 512740, chromosome 14, 21037822...21164094). To date, no other study has reported this association. Other significant SNPs detected on BTA 14 were rs29010515, rs29010516, rs29011751 and rs29025824 for IGF-1 and scanned fat depth traits. These SNPs were located

in nongene or noncoding regions. However, the SNPs that fall within noncoding regions may still have consequences for gene splicing and transcript factor binding or may be in linkage disequilibrium with the causal mutations. Some of these associations found in this study are consistent with the locations of QTL identified using linkage mapping in cattle, bearing in mind that the confidence intervals on QTL locations are usually large (Visscher *et al.* 1996), ranging from 20 to 30 cM, where hundreds of genes may be located in a region of this size (Ron & Weller 2007).

The association of a number of SNPs with fat depth traits on BTA 14 was consistent with meta-QTL score computed from the combined results of QTLs reported from previous QTL mapping studies (Casas *et al.* 2003; Mizoshita *et al.* 2004; Choi *et al.* 2006). Wibowo *et al.* (2008) reviewed over 40 studies with QTL reported on BTA 14 and identified multiple QTLs that span large chromosomal regions including the same region identified in the present study.

Interestingly, Snelling *et al.* (2010) identified strong SNP associations that fall within the same region on BTA 14 for several growth traits in cross-bred beef cattle, a study conducted as part of the Cycle VII US Meat Animal Research Centre (USMARC). Recently, Bolormaa *et al.* (2011) and Hawken *et al.* (2012) detected significant SNP markers in genome region 22–25 Mb on BTA 14 associated with a wide variety of traits, including age at puberty, serum IGF-1, hip height and scanned fat depth at P8 traits. They used the 50k SNP bovine chip genotype data obtained from Australian Brahman and composite tropical beef cattle.

Similarly, Fortes *et al.* (2012) identified the same region in BTA 14 associated with IGF-1 and scrotal circumference in Australian Brahman bulls genotyped with the bovine SNP 50k chip. This region contains 23 positional genes, including pleiomorphic adenoma gene 1 (PLAG1), proenkephalin (PENK), ribosomal protein S20 (RPS20) and moloney murine sarcoma viral oncogene homologue (MOS).

Furthermore, Karim *et al.* (2011) showed that this genome region was associated with bovine stature and identified pleiomorphic adenoma gene 1 (PLAG1) as a possible causative gene for stature in dairy cattle studied in New Zealand. Thus, genomic region 21–39 Mb in BTA 14 appears to have pleiotropic effects on growth, fat depth and reproductive traits across different cattle breeds.

#### Genome-wide epistasis association

Significant epistatic signals were detected for serum concentration of IGF-1 and ultrasound scan measures

for fat depth traits using a genome-wide association analysis method. Using a FDR threshold of 0.01, 136 genome-wide epistatic signals were detected on 15 different chromosomes for WIIGF-1 (see supporting information, Table S2), with the highest number of epistatic signals located on chromosome 10. Overall, there were 56 SNPs spanning the genome involved in these 136 interactions. Twenty epistatic interactions, associated with WIIGF1 and D2SP8 traits, were between SNPs located on 15 chromosomes (BTA 1, 5, 6, 8, 10, 13, 14, 16, 17, 18, 19, 20, 23, 25 and 26) (see supporting information, Table S4). Of these epistatic interactions, 10 SNPs were located in the following genes: SNX30 (BTA 8), DZANK1 (BTA 13), PRKDC (BTA 14), MTSS1 (BTA 14), SUS4 (BTA 16), ARHGAP35 (BTA 18), CDH18 (20), CPNE5 (BTA 23), PRKCB (BTA 25) and PRKG1 (BTA 26). The large numbers of epistatic interaction were associated with WIIGF1, and some of these epistatic signals show strong significant association ( $p$ -value =  $6.1 \times 10^{-13}$ ;  $q$ -value =  $5.2 \times 10^{-11}$ ). Surprisingly, SNPs with the strongest associations detected using standard GWAS did not appear in any significant epistatic association. This in fact may suggest that many of epistatic SNPs may show small effect when tested individually and could be missed through GWAS alone.

The effects of these genes became evident only after fitting epistasis models. One of the drawbacks of the standard genome-wide association studies is that they consider only one SNP at a time, so it does not allow for additive or interaction effects of other loci. Hence, GWAS are unable to explain the full genetic variation in complex traits and diseases, more due to statistical issues than biological issues. Indeed, Khatib *et al.* (2009) showed that models that fitted gene–gene epistatic interactions in eight candidate genes in the POU1F1 signalling pathway for fertility traits in dairy cows revealed significant contribution to the variation in fertility traits compared with models without interactions between genes. Similarly, Suchocki *et al.* (2010) assessed epistatic interactions between four candidate genes, namely BNA1A1, DGAT1, LEP and LEPR, and their effects on milk traits in Jersey and Polish Holstein-Friesian cows and found evidence of epistasis between DGAT1 and LEPR, as well as between LEPR and BTN1A1 for milk composition traits.

The model for epistasis in eqn (3) contains a term with 4 degrees of freedom (SNP<sub>1</sub>–SNP<sub>2</sub>) and contains an extensive range of epistatic expression. If this model was re-parameterised to code the 3-level factors by additive ( $a_1$ ,  $a_2$ ) and dominance ( $d_1$ ,  $d_2$ ) effects,

then the epistatic interactions could be dissected into additive  $\times$  additive, dominance  $\times$  dominance and additive  $\times$  dominance components. This could be further explored in future; however, considerations of power to detect each component may need further consideration.

The present study has some limitations, namely the small sample size and low-density SNP panel, even though this was considered high density at the time the experiment was conducted. Often in the single-SNP GWAS models, the small-effect SNPs will not be detected because of the stringent threshold (e.g.  $p$ -value  $\leq 10^{-5}$ ); hence by lowering the cut-off, we will be able to capture those epistatic SNPs (Kadarmideen 2014). In our epistasis association model, a subset of significant SNPs at a lenient threshold of  $p$ -value  $\leq 0.01$  was selected from the traditional genome-wide association model for pair-wise SNP interaction tests; FDR using the  $q$ -value method was computed to control false positives. However, it is possible that this method, in the interests of tractability, will miss detecting some epistatic interactions. Furthermore, only first-order interactions were evaluated in the epistatic model, to reduce computational burden; hence, only a small subset of all possible epistatic effects was investigated. Despite these shortfalls, the overall evidence points towards the detection of a large number of epistatic loci located in 19 positional genes affecting serum levels of IGF-1. The results of the present study suggest that the expression of IGF-1 may be under the influence of complex epistatic pathways; hence, genetic studies on this trait should take into account these epistatic effects.

Higher resolution genome scans with a larger number of animals genotyped and across cattle breeds will allow a more accurate assessment of the genetic basis of regulation of growth and carcass traits.

It should also be noted that estimating pair-wise epistatic (gene–gene) interactions is becoming increasingly important in modelling the genetic basis of complex traits via ‘systems genetics’ methods such as Weighted Interaction SNP Hub (WISH) networks shown by Kogelman & Kadarmideen (2014) and applied to obesity and metabolic traits in the porcine model for obesity (Kogelman *et al.* 2014).

In summary, we demonstrated the usefulness of modelling epistasis in the analysis of complex traits and successfully identified potential genes that affecting serum concentration of IGF-1 and ultrasound-scanned fat rump depth traits using field data from an Australian Brahman cattle population. Information about epistasis can add to our understanding of the

complex genetic networks that form the fundamental basis of biological systems.

## Acknowledgements

We acknowledge the CRC for Beef Genetic Technologies for providing the data used in this study. Haja N. Kadarmideen’s time on this project was partially supported by the EU-FP7 Marie Curie Actions – Career Integration Grant (CIG-293511), awarded to him in 2011. Abdirahman A. Ali was the recipient of a CSIRO Animal, Food and Health Sciences PhD scholarship. We would like to dedicate this study to John James, in acknowledgement of his tremendous contribution to animal breeding and genetics, and in particular, for the support he has given to countless scientists working in this field.

## References

- Allan M.F., Thallman R.M., Cushman R.A., Echtenkamp S.E., White S.N., Kuehn L.A., Casas E., Smith T.P.L. (2007) Association of a single nucleotide polymorphism in SPP1 with growth traits and twinning in a cattle population selected for twinning rate. *J. Anim. Sci.*, **85**, 341–347.
- Barwick S.A., Johnston D.J., Burrow H.M., Holroyd R.G., Fordyce G., Wolcott M.L., Sim W.D., Sullivan M.T. (2009) Genetics of heifer performance in ‘wet’ and ‘dry’ seasons and their relationships with steer performance in two tropical beef genotypes. *Anim. Prod. Sci.*, **49**, 367–382.
- Bolormaa S., Porto Neto L.R., Zhang Y.D., Bunch R.J., Harrison B.E., Goddard M.E., Barendse W. (2011) A genome-wide association study of meat and carcass traits in Australian cattle. *J. Anim. Sci.*, **89**, 2297–2309.
- Browning B.L., Browning S.R. (2011) A fast, powerful method for detecting identity by descent. *Am. J. Hum. Genet.*, **88**, 173–182.
- Burrow H.M. (2012) Importance of adaptation and genotype  $\times$  environment interactions in tropical beef breeding systems. *Animal*, **6**, 729–740.
- Butler D.G., Cullis B.R., Gilmour A.R., Gogel B.J. (2009) Mixed Models for S Language Environments. ASReml-R Reference Manual, Version 3. Queensland Department of Primary Industries and Fisheries, Brisbane, Australia.
- Casas E., Shackelford S.D., Keele J.W., Koohmaraie M., Smith T.P.L., Stone R.T. (2003) Detection of quantitative trait loci for growth and carcass composition in cattle. *J. Anim. Sci.*, **81**, 2976–2983.
- Choi I.S., Kong H.S., Oh J.D., Yoon D.H., Cho B.W., Choi Y.H., Kim K.S., Choi K.D., Lee H.K., Jeon G.J. (2006) Analysis of microsatellite markers on bovine chromosomes 1 and 14 for potential allelic association with car-

- cass traits in Hanwoo (Korean cattle). *Asian-Aust. J. Anim. Sci.*, **19**, 927–930.
- Fortes M.R.S., Lehnert S.A., Bolormaa S., Reich C., Fordyce G., Corbet N.J., Whan V., Hawken R.J., Reverter A. (2012) Finding genes for economically important traits: Brahman cattle puberty. *Anim. Prod. Sci.*, **52**, 143–150.
- Fortes M.R.S., Li Y.T., Collis E., Zhang Y.D., Hawken R.J. (2013) The IGF1 pathway genes and their association with age of puberty in cattle. *Anim. Genet.*, **44**, 91–95.
- Hawken R.J., Zhang Y.D., Fortes M.R.S., Collis E., Barris W.C., Corbet N.J., Williams P.J., Fordyce G., Holroyd R.G., Walkley J.R.W., Barendse W., Johnston D.J., Prayaga K.C., Tier B., Reverter A., Lehnert S.A. (2012) Genome-wide association studies of female reproduction in tropically adapted beef cattle. *J. Anim. Sci.*, **90**, 1398–1410.
- Kadarmideen H.N. (2014) Genomics to systems biology in animal and veterinary sciences: progress, lessons and opportunities. *Livestock Sci.*, **166**, 232–248.
- Karim L., Takeda H., Lin L., Druet T., Arias J.A.C., Baurain D., Cambisano N., Davis S.R., Farnir F., Grisart B., Harris B.L., Keehan M.D., Littlejohn M.D., Spelman R.J., Georges M., Coppieters W. (2011) Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nat. Genet.*, **43**, 405–413.
- Khatib H., Huang W., Wang X., Tran A.H., Bindrim A.B., Schutzkus V., Monson R.L., Yandell B.S. (2009) Single gene and gene interaction effects on fertilization and embryonic survival rates in cattle. *J. Dairy Sci.*, **92**, 2238–2247.
- Khatkar M.S., Zenger K.R., Hobbs M., Hawken R.J., Cavanagh J.A.L., Barris W., McClintock A.E., McClintock S., Thomson P.C., Tier B., Nicholas F.W., Raadsma H.W. (2007) A primary assembly of a bovine haplotype block map based on a 15,036-single-nucleotide polymorphism panel genotyped in Holstein-Friesian cattle. *Genetics*, **176**, 763–772.
- Kogelman L.J., Kadarmideen H.N. (2014) Weighted Interaction SNP Hub (WISH) network method for building genetic networks for complex diseases using whole genome genotype data. *BMC Syst. Biol.*, **8**(Suppl 2), S5.
- Kogelman L.J.A., Pant S.D., Fredholm M., Kadarmideen H.N. (2014) Systems genetics of obesity in an F2 pig model by genome-wide association, genetic network and pathway analyses. *Front. Genet.*, **5**, 214.
- Koots K.R., Gibson J.P., Smith C., Wilton J.W. (1994) Analyses of published genetic parameter estimates for beef production traits. 1. Heritability. *Anim. Breed. Abstracts*, **62**, 309–338.
- Marchini J., Donnelly P., Cardon L.R. (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat. Genet.*, **37**, 413–417.
- Mizoshita K., Watanabe T., Hayashi H., Kubota C., Yamakuchi H., Todoroki J., Sugimoto Y. (2004) Quantitative trait loci analysis for growth and carcass traits in a half-sib family of purebred Japanese Black (Wagyu) cattle. *J. Anim. Sci.*, **82**, 3415–3420.
- Porto Neto L.R., Bunch R.J., Harrison B.E., Barendse W. (2012) Variation in the XKR4 gene was significantly associated with subcutaneous rump fat thickness in indicine and composite cattle. *Anim. Genet.*, **43**, 785–789.
- Rincon G., Webber K.L., Van Eenennaam A.L., Golden B.L., Medrano J.F. (2011) Hot topic: performance of bovine high-dense genotyping platforms in Holsteins and Jerseys. *J. Dairy Sci.*, **94**, 6116–6121.
- Ron M., Weller J.I. (2007) From QTL identification in livestock – winning by points rather than knock-out: a review. *Anim. Genet.*, **38**, 429–439.
- Snelling W.M., Allan M.F., Keele J.W., Kuehn L.A., McDanel T., Smith T.P.L., Sonstegard T.S., Thallman R.M., Bennett G.L. (2010) Genome-wide association study of growth in crossbred beef cattle. *J. Anim. Sci.*, **88**, 837–848.
- Storey J.D., Tibshirani R. (2003) Statistical significance for genome-wide studies. *Proc. Natl Acad. Sci. USA*, **100**, 9440–9445.
- Suchocki T., Komisarek J., Szyda J. (2010) Testing candidate gene effects on milk production traits in dairy cattle under various parameterizations and modes of inheritance. *J. Dairy Sci.*, **93**, 2703–2717.
- Upton W., Burrow H.M., Dundon A., Robinson D.L., Farrell E.B. (2001) CRC breeding program design, measurements and database: methods that underpin CRC research results. *Aust. J. Exper. Agri.*, **41**, 943–952.
- Vankan D.M., Burns B.M. (1997) DNA fingerprinting – how it works and applications for the beef industry. *Proc. Assoc. Adv. Anim. Breed. Genet.*, **12**, 433–437.
- Visscher P.M., Thompson R., Haley C.S. (1996) Confidence intervals in QTL mapping by bootstrapping. *Genetics*, **143**, 1013–1020.
- Wibowo T.A., Gaskins C.T., Newberry R.C., Thorgaard G.H., Michal J.J., Jiang Z. (2008) Genome assembly anchored QTL map of bovine chromosome 14. *Int. J. Biol. Sci.*, **4**, 406–414.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** Manhattan plots of genome-wide association. Top: for W1SRIB, bottom: for D2IGF-1.

**Figure S2.** Manhattan plots of genome-wide association. Top: for D2SP8, bottom: for D2SRIB.

**Figure S3.** The quantile–quantile plot of the observed distribution of the  $-\log_{10}(p\text{-values})$  for W1IGF-1.

**Figure S4.** The quantile–quantile plot of the observed distribution of the  $-\log_{10}(\text{p-values})$  for W1SP8.

**Figure S5.** The quantile–quantile plot of the observed distribution of the  $-\log_{10}(\text{p-values})$  for W1SRIB.

**Figure S6.** Genome-wide epistasis association for insulin-like growth factor-1 measured in the dry season.

**Figure S7.** Genome-wide epistasis association for ultrasound scan rump fat depth at P8 site measured in the wet season.

**Figure S8.** Genome-wide epistasis association for ultrasound scan for rump fat depth at P8 site measured in the dry season.

**Figure S9.** Genome-wide epistasis association for ultrasound scan rib fat depth measured in the wet season.

**Figure S10.** Genome-wide epistasis association for ultrasound scan rib fat depth measured in the dry season.

**Table S1.** Number of SNPs with nominal p-value  $\leq 0.01$  selected from single-SNP GWAS scan.

**Table S2.** Genome-wide significant epistasis at threshold  $q$ -value of 0.01 for serum levels of insulin-like growth factor-1 measured in wet season (WIIGF-1).

**Table S3.** Genome-wide significant epistasis at threshold  $q$ -value of 0.1 for ultrasound-scan for carcass traits measured in the wet and dry seasons and insulin-like growth factor-1 measured in the dry season.

**Table S4.** Positional epistatic genes associated with insulin-like growth factor-1 and ultrasound scan for rump fat depth at P8 site in Brahman cattle.