# Genome-wide association studies for feedlot and growth traits in cattle

S. Bolormaa, B. J. Hayes, K. Savin, R. Hawken, W. Barendse, P. F. Arthur, R. M. Herd and M. E. Goddard

*J ANIM SCI* published online January 14, 2011

# Genome-wide association studies for feedlot and growth traits in cattle

S. Bolormaa[*,†,1], B. J. Hayes[†], K. Savin[*,†], R. Hawken[*,&], W. Barendse[*,&], P.F. Arthur[*,§], R.M. Herd [*,‡] and M. E. Goddard[*,†,**]

[*]*Cooperative Research Centre for Beef Genetic Technologies, Armidale, NSW, Australia*

[†]*Biosciences Research Division, Department of Primary Industries, Bundoora, VIC, Australia*

[&]*CSIRO Livestock Industries, Bioscience Precinct, St. Lucia, QLD, Australia*

[§]*Industry and Investment NSW, Elizabeth Macarthur Agricultural Institute, Menangle, NSW, Australia*

[‡]*Industry and Investment NSW, Beef Industry Centre of Excellence, Armidale, NSW, Australia*

[**]*Department of Agriculture and Food Systems, University of Melbourne, Parkville, VIC, Australia*

[1]Corresponding author:

Bolormaa Sunduimijid

Mailing address:

Biosciences Research Division,

Department of Primary Industries,

1 Park Drive, Bundoora,

Victoria 3083, Australia

Phone: +61 3 90327010

Email: bolormaa.sunduimijid@dpi.vic.gov.au





**Published Online First on January 14, 2011 as doi:10.2527/jas.2010-3079**

**ABSTRACT:** A genome wide-association study (GWAS) for production traits in cattle was carried out using genotype data from the 10K Affymetrix and the 50K Illumina SNP chips. The results for residual feed intake (RFI), body weight and hip height in 3 beef breed types (*Bos indicus*, *Bos taurus* and *B.indicus* × *B.taurus*), and for stature in dairy cattle, are presented. The aims were to discover SNP associated with all traits studied, but especially RFI, and further to test the consistency of SNP effects across different cattle populations and breed types. The data were analysed within datasets and within breed types using a mixed model and fitting 1 SNP at a time. In each case the number of significant SNP was more than expected by chance alone. A total of 75 SNP from the reference population with 50K chip data were significant ($P < 0.001$) for RFI, with a false discovery rate of 68%. These 75 SNP were mapped on 24 different bovine chromosomes (BTA). Of the 75 SNP, the 9 most significant SNP were detected on BTA 3, 5, 7 and 8 with $P \leq 6.0 \times 10^{-5}$. In a population of Angus cattle divergently selected for high and low RFI and 10K chip data, 111 SNP were significantly ($P < 0.001$) associated with RFI, with a false discovery rate of 7%. Approximately 103 of these SNP were therefore likely to represent true positives. Due to the small number of SNP common to both the 10K and 50K SNP chips, only 27 SNP were significantly ($P < 0.05$) associated with RFI in the 2 populations. However, other chromosome regions were found that contained SNP significantly associated with RFI in both datasets although no SNP within the region showed a consistent effect on RFI. SNP effects were consistent between datasets only when estimated within the same breed type.

**Key Words**: Beef and Dairy Cattle, feed intake, height, RFI, SNP, weight

## INTRODUCTION

In genomic selection, the estimation of breeding values is based on genetic markers. This is particularly useful for traits that are very expensive to measure such as residual feed intake (**RFI**). Genomic selection relies on linkage disequilibrium (**LD**) between genetic markers, such as SNP, and quantitative trait loci (**QTL**) that affect the trait. This LD generates an association between some markers and the trait. In beef cattle, some studies (Barendse et al., 2007; Nkrumah et al., 2007; Sherman et al., 2009) have reported associations between markers and RFI. For instance, Barendse et al. (2007), using a commercial SNP chip containing approximately 10,000 (10K) SNP, analysed 8,786 polymorphic SNP in 189 Australian beef cattle, chosen on the basis of being phenotypically high and low for RFI, and detected 161 SNP associated with RFI at $P < 0.01$. Development of a commercial 50,000 (50K) SNP chip provided the opportunity to conduct a more powerful genome-wide association study (**GWAS**) for RFI.

Whether markers associated with a trait are to be used for genomic selection or for mapping the QTL to a chromosomal region, it is necessary to confirm in independent populations the associations that have been discovered in 1 population. Often such attempts at confirmation have been unsuccessful (e.g., Pryce et al., 2010a). Failure to confirm associations could be due to 3 reasons: (1) the original discovery was a false positive, (2) the association is specific to that breed either because the QTL does not segregate in another breed or because the phase or strength of LD differs between breeds, or (3) lack of statistical power in either the discovery or validation population or both populations. In this paper, the importance of these 3 reasons for failure to validate associations is examined.

Ideally, a SNP allele that is associated with an increase in a trait, such as RFI, in 1 breed will be also associated with an increase in other breeds. However, cattle breeds differ in LD phase between markers in the 50K SNP chip (de Roos et al., 2008, 2009) so it is expected that

they will differ in LD phase between SNP and QTL. This might mean that the association between SNP and QTL is still significant in the second breed but reversed in sign. More likely, the association is simply weak and not significant. If the QTL is segregating in the second breed, it is likely that different SNP, close to the QTL, will now show a significant association with the QTL. Therefore, to confirm the detection of a QTL, 3 types of evidence are investigated: Is the SNP, discovered to be associated with the trait in the discovery population, significantly associated with the trait in the second (validation) population; is the direction of the association the same; and is there another SNP in the same vicinity that shows a significant association with the trait under investigation?

As well as RFI, we also present data on GWAS for weight and hip height, which are model quantitative traits and for which data is more widely available than for RFI. To do this, GWAS using data from 3 breed types of beef cattle (*Bos taurus*, *Bos indicus* and their crosses) and 2 SNP chips (10K and 50K) were conducted. A GWAS for stature or height in dairy cattle using the same 50K SNP chip was also carried out to determine if associations could be confirmed across dairy and beef breeds.

Genetic correlations between traits could be due to QTL that have pleiotropic effects on multiple traits or could be due to closely linked QTL each affecting different traits. Nkrumah et al. (2007) found QTL affecting dry matter intake, feed conversion ratio and average daily gain together in similar locations on the bovine genome map. Indeed, Moore et al. (2008) pointed out that it is important to investigate the effects of QTL on other traits when studying the molecular basis of RFI to avoid unfavourably correlated responses when selecting for RFI. Therefore, we investigated if SNP associated with RFI were also associated with feed intake or growth rate. If a trait has no phenotypic correlation with RFI then SNP should rarely be associated with both traits

entirely due to false discovery. Therefore if SNP have a significant effect on 2 uncorrelated traits more often than expected by chance, this is evidence that the association is real.

The objectives of this study were to detect SNP associated with RFI, growth and height in 3 breed types of beef cattle (*Bos taurus*, *Bos indicus* and their crosses) and in dairy cattle, and to validate SNP effects across different datasets and breed types.

## MATERIALS AND METHODS

### *SNP data*

The SNP marker data used in this study was obtained from 2 different sources: one used the BovineSNP50K BeadChip (Illumina, San Diego) and the other one used the Parallele SNP10K chip (Affymetrix, Santa Clara, CA). The 50K SNP were at random positions with approximately equal spacing (median interval of 37kbp) along the bovine genome (Matukumalli et al., 2009). The 10K SNP were with mean inter-marker spacing of 258 kbp (Fidanza et al., 2001). The SNP were ordered by chromosome position using Bovine Genome Build 4.0 (http://www.ncbi.nlm.nih.gov/projects/genome/guide/cow/).

**Beef data.** Using the 50K chip, 53,798 SNP were genotyped. Preliminary investigations on genotype dataset showed that all genotypes had more than 95% quality scores and the proportion of missing genotypes was less than 2.1%. A minor allele frequency (**MAF**) < 0.05 was found for 16,008 SNP and 8,469 SNP deviated from Hardy-Weinberg equilibrium (**HWE**: $P < 0.0001$). However, these were not removed from the further analyses. Out of the initial 53,798 SNP, 50,650 were polymorphic and included in the GWAS. Additionally, 8,201 SNP, which were genotyped using the 10K chip, were evaluated for their effects on RFI. There were 2,390 SNP on both the 10K and 50K SNP chips.

**Dairy data.** A pre-edited genotype dataset consisting of 39,048 SNP loci (Hayes et al., 2009) was used for the association analyses.

### *Animals and populations*

**Beef cattle.** Phenotype and genotype data held in 3 cattle databases was used (Table 1). From the Beef Co-operative Research Centre Phase I (**CRCI**) records, phenotypic records on RFI and growth traits on 852 steers with 50K SNP genotype data were obtained (Table 2). These steers were from 7 different pure breeds of 3 breed types. The 4 breeds (Angus, Murray Grey, Shorthorn and Hereford) were *Bos taurus* (**Bt**), 1 breed (Brahman) was *Bos indicus* (**Bi**) and 2 breeds (Santa Gertrudis and Belmont Red) were **Bt×Bi** synthetic breeds (Johnston et al., 2003). From the Beef CRC Phase II (**CRCII**) dataset, records were obtained for 1,456 cows with 50K SNP chip data plus weight and height data. These cows were from 2 breed types: Bi (Brahman) and Bt×Bi crosses (Tropical Composites) (Barwick et al., 2009). Thirdly, records for 379 Angus (Bt) cattle that had been genotyped using the 10K SNP chip were obtained. These cattle were from the divergent RFI selection lines based at the Trangie Agricultural Research Centre, New South Wales, Australia (Arthur et al., 2001a). Although Trangie selection line, CRCI and CRCII animals are different, they could be related due to common ancestors.

**Dairy cattle.** Data for bulls with 50K SNP chip data were extracted from the Australian Dairy Herd Improvement Scheme (**ADHIS**) database. There were 588 Holstein bulls that received estimated breeding values based on the phenotypic records of their daughters before 2005 (reference dataset) and 117 Holstein bulls proven between 2005 and 2007 (validation dataset; Table 1).

### *Traits studied*

**Beef cattle.** The CRCI steers were approximately 1-year-old before being recorded in a research feedlot for 4 traits: residual feed intake (RFI), average daily gain (**ADG**), daily feed intake (**DFI**), and metabolic mid-weight (**mMWT**), and before the feedlot period, for post weaning hip height (**pwHH**), following standard procedures described by Johnston et al. (2003) and Robinson and Oddy (2004). RFI is a measure of feed efficiency and is calculated as the difference in feed intake, above or below that expected or predicted on the basis of metabolic weight and growth rate (Arthur et al., 2001b). The CRCII heifers were recorded for first post-weaning wet season weight (**w1LWT**) and hip height (**w1HH**) (Barwick et al., 2009). The Angus cattle with 10K SNP data were bulls and heifers and were measured for the feedlot traits at a younger age than the CRCI steers (Arthur et al., 2001b).

**Dairy cattle**. Deregressed estimated breeding values (**EBV**) for stature in the Holstein reference and validation datasets were used for GWAS. Further details of how deregressed EBV were calculated are given in Pryce et al. 2010b. Stature EBV for dairy bulls are calculated from 2-year-old daughters' phenotypes measured at the highest point of the sacrum at the hip bones and converted to a scale of 1 to 9, where 1 is approximately 130 cm and 9 is 150 cm.

### *Estimate of LD in the populations studied*

The linkage disequilibrium (LD) between pairs of SNP markers ($r^2$) was used to estimate the extent of LD in the populations studied. The average pair-wise $r^2$ for each population was calculated using the LDMAX procedure in GOLD (Abecasis and Cookson, 2000) using the conventional measure of $r^2$ (Hill and Robertson, 1968; Devlin and Risch, 1995). Then the average of LD estimates ($\bar{r}^2$) for each population was calculated in every 10kbp-interval and it was corrected for sample size ( = $\bar{r}^2$ - 1/N, where N is sample size).

***Statistical analyses***

**Phenotype and model used.** The association between each SNP and each of the traits was assessed by a regression analysis using the ASReml software (Gilmour et al., 2002). The mixed model applied was: trait ~ mean + fixed effects + $SNP_i$ + animal + error; with animal and error fitted as random effects. The *i*th SNP ($SNP_i$) was fitted as a covariate effect. Fixed effects were different for the CRCI and CRCII datasets. For CRCI dataset, breed, herd of origin, sex, year of measurement, season, market-weight destination and nutritional treatment were fitted as class variables and age deviation from group mean was fitted as a covariate. Whereas for CRCII data the effects of breed, herd of origin, sire group, cohort, calving month and their first degree interactions were fitted as fixed effects (Barwick et al., 2009). The fixed effects used for the Angus Trangie selection line dataset were contemporary group and linear covariate for age (Arthur, 2001b). Using the same model without fitting $SNP_i$, estimates of heritability in beef cattle were calculated based on the genotyped animals and their 5-generation-ancestors (Table 2).

In the dairy cattle, the estimates of heritabilities of deregressed EBVs for stature were calculated. SNP were evaluated for their effects on stature using a mixed model fitting the mean, SNP as fixed effect and random animal effect.

**Significance of SNP**. The SNP were tested for a significant association with particular traits at different probability thresholds (Table 3). In a GWAS there are many thousands of significance tests performed. Therefore, the number of SNP that were significant to the number expected by chance was compared using a false discovery rate (**FDR**) as: $FDR = \frac{P(1-s)}{s(1-P)}$, where $P$ is a defined probability threshold and $s$ is a proportion of SNP that are nominally

significant at the defined threshold ( = number of significant SNP divided by number of total SNP). This is equivalent to the FDR formula of Storey (2002). Then the number of true positive SNP equals to (1 - FDR) multiplied by the number of significant SNP at a particular probability threshold.

The correlations of SNP effects between RFI, ADG and DFI were estimated. SNP effects with high standard errors are sometimes large but the effects are poorly estimated. Therefore, the SNP effects were divided by their standard error before correlations of the SNP effects were calculated.

**Validation of SNP.** The SNP that were significantly associated with RFI in the 50K SNP data were tested for an association in the 10K SNP data. There were only 2,390 SNP in common between the 2 datasets. Therefore, we also tested whether significant associations with RFI were found within the same 1-Mbp-intervals in both datasets using chi-square tests. If a 1-Mbp-interval did not contain any SNP in one of 50K (reference) or 10K (validation) datasets or in both datasets, then the particular 1-Mbp- intervals were removed. After removing those 1-Mbp-intervals, each 1-Mbp-interval was scored as containing or not containing 1 or more significant SNP from the 10K and from the 50K datasets. The data is then in the form of a 2×2 table in which 1-Mbp-regions are classified as significant or not significant in 2 different experiments. We tested the significance of the agreement between experiments using a chi-square test. However, some 1-Mbp-regions contain more SNP than others and so may be more likely to contain a significant SNP. This would bias the chi-square test. Therefore, we carried out the permutation tests to establish an appropriate significance threshold for the chi-square statistic. We did permutation tests with 10,000 repetitions to derive the distribution of the test statistic under the null hypothesis to calculate the significance of the association in 1-Mbp-interval

between the 2 datasets. The permutation test was performed using the reference dataset (e.g., 50K data) with real effects and the validation dataset (e.g., 10K data) but with the significance status of SNP permutated across the genome. The number of SNP considered being significant in the validation data was the same as the number of significant SNP ($P < 0.05$) with real effects but for the permutation test, significant SNP were chosen at random. A chi-square test for each of 10,000 permutations was calculated and this empirical distribution of chi-square statistics under the null hypothesis was used to test the significance of the association in the same 1 Mbp.

For weight and height traits, a validation test of SNP associations was carried out using the results from the analyses of the same 50K SNP chip data in 2 different populations (CRCI and CRCII cattle). Additionally, the 3 breed types within the CRCI data (Bt, Bi and Bt×Bi) were also analysed separately as well as in a joint analysis. Similarly, the 2 breed types (Bi and Bt×Bi) represented in the CRCII dataset were analysed separately as well as jointly. The number of records in each breed line is given in Table 2. The number of SNP that were significant in both CRC datasets was counted and for these SNP, 2 parameters to assess the agreement between the results were calculated: the correlation between SNP effects in the 2 datasets, and the proportion of SNP in which the effects were in the same direction; that is, the proportion in which the same SNP allele increased the trait.

Similarly, a validation of SNP for stature in the dairy reference and validation populations was carried out by examining the proportion of SNP effects with the same direction in the 2 datasets. Finally, the SNP significantly associated with height in the beef and dairy cattle were compared. This was done at the same SNP position and within 1Mbp regions.

Information about particular genes, located near SNP significantly associated with RFI, was extracted from on-line sources: http://www.ensembl.org/index.html , http://www.genecards.org/cgi-bin/cardsearch.pl#top and http://www.uniprot.org.

## RESULTS

### *Summary statistics*

Raw means, standard deviations and heritability estimates are given in Table 2. Heritability estimates are based on small sample sizes and so subject to large standard errors. In the Trangie animals, the estimate of heritability ($h^2$) for RFI is biased upwards because animals with extreme phenotypes for RFI were chosen for the experiment. The estimates of heritability using full RFI dataset (n = 1,177) was 0.39 (Artur et al., 2001a). The proportion of genetic variance relative to the total variance was used to calculate 'heritability' for deregressed EBVs of stature in both Holstein reference and validation populations (Table 2). This measures the reliability of the progeny test rather than the heritability of the raw trait.

### *LD in the populations studied*

The average $r^2$ ($\bar{r}^2$) declined as a function of distance between markers. LD in all populations decreased rapidly over short distances (Figure 1), but remained slightly above zero over long distances. The dairy cattle population had the highest levels of LD. LD was highest in Bt, followed by Bt×Bi, and followed by Bi.

### *RFI*

A total of 75 SNP from the 50K-chip data for the CRCI steers were significant ($P < 0.001$) for RFI, with a FDR of 67% (Table 3). These 75 SNP were mapped on 24 different bovine chromosomes (BTA). Out of 75 SNP, the 9 most significant SNP were detected on BTA 3, 5, 7 and 8 with $P \leq 6.0 \times 10^{-5}$. A broad peak including the 3 most significant SNP (of these 9 SNP) was detected between 86-94Mbp of BTA 8 (Figure 2). Of these 3 SNP on BTA 8, 2 were in high LD ($r^2 = 0.58$) with each other, the other SNP was in lower LD ($r^2 < 0.16$). In Trangie population (10K chip data), 111 SNP were significantly associated with RFI, with a FDR of 7% (Table 3). Approximately, 103 of these SNP are therefore likely to represent true positives.

There was no separate dataset for RFI based on the 50K chip that could be used for validation. Therefore, the results from the 10K SNP chip were used to validate those from the 50K SNP chip. Of the 2,390 SNP in common between the 2 datasets (50K and 10K SNP data), 27 of them were significant at $P < 0.05$ in both datasets (Table 4). This is not convincingly more than expected by chance. Every Mbp interval was also classified as containing or not containing one or more significant SNP in both experiments. Out of 2,131 intervals, 406 intervals included at least one (sometimes up to 11) significant SNP ($P < 0.05$) in both 50K and 10K datasets ($\chi^2 = 9.25$; Table 5) and this $\chi^2$ is significant based on the 10,000 permutation tests. As an example of this tendency to find QTL for RFI in the same region in both datasets, the 10K data also show a high and broad peak for RFI on BTA 8 at a position of 82 to 94Mbp (Figure 2).

***Pleiotropy of SNP affecting RFI***

RFI was moderately correlated with DFI ($r_P = 0.56$), and weakly with ADG ($r_P = 0.12$), although this latter correlation was expected to be zero. ADG and DFI had a high, positive correlation ($r_P = 0.63$). A similar number of significant SNP ($P < 0.001$) were detected for ADG

and DFI (Table 3). The correlation between SNP effects estimated for RFI and DFI was moderately positive (0.58), whereas the correlation for ADG and DFI was high (0.71; Table 6). RFI is a measurement that is corrected phenotypically for ADG and the correlations of SNP effects reflected this. The number of SNP that were significant ($P < 0.05$) for both RFI and DFI or for both ADG and DFI was 651 and 973, respectively. The proportion of these significant SNP effects in the same direction was 100% (Table 6).

Because RFI and ADG are almost uncorrelated, SNP are not expected to be associated with both traits unless there are QTL with a pleiotropic effect on both traits. In total, 162 SNP ($P < 0.05$) had significant effects for both RFI and ADG, which is no more than expected by chance, and the proportion of effects of these SNP in the same direction were 40% (Table 5 & 6). When 1-Mbp-intervals were considered instead of individual SNP, the number of intervals which contain one or more significant SNP ($P < 0.05$) for both RFI and ADG was 845 out of 2,530 one-Mbp-intervals (Table 5). This was not more than expected by chance ($\chi^2 = 2.22$). Also, there were 576 intervals (out of 2,530) containing SNP that were significant ($P < 0.05$) for all of RFI, ADG and mMWT (e.g., Table 7).

### *Growth traits using 50K chip*

**Beef cattle.** The number of significant SNP at a threshold of $P < 0.001$ (Table 3) was 78 for metabolic mid-weight in the feedlot and 75 for post-weaning height in the CRCI population, but 156 for end of wet season weight and 134 for height in the larger CRCII data. Consequently, the FDRs were lower in the CRCII data (32% and 38%) compared with the CRCI data (65% and 67%).

Table 8 shows the number of SNP that were significant for weight or for height in one of the CRCI populations and one of the CRCII populations. Table 8 also gives the correlations of SNP effects for weight and height between the CRCI and CRCII data sets across breed types, as well as the proportion of SNP whose effects were in the same direction. The number of SNP significant in both CRCI and CRCII datasets is no more than expected by chance and, in most cases the proportion of effects in the same direction does not depart significantly from the 50% expected by chance. The exception is for the Bt×Bi datasets where 72% of SNP effects are in the same direction for weight and 66% for height (Table 8). When one-Mbp-intervals were considered, the number of regions contains a significant SNP ($P < 0.05$) in both datasets was 959, which is more than the number expected by chance ($\chi^2 = 6.27$; Table 5), but not significant by the permutation test.

**Dairy cattle**. The proportion of SNP that were significant in the Holstein reference dataset was higher than in the smaller validation dataset and as high as in the larger CRCII dataset (Table 3). This resulted in an FDR for the dairy reference dataset of 22% at $P < 0.001$, but higher (56%) for the dairy validation set. When the same threshold of $P < 0.05$ was set in both reference and validation datasets, the number of SNP that were significant in both datasets was 215 which is not greater than expected by chance (Table 5) but the proportion of significant SNP effects in same direction was 68%. By considering 1Mbp-intervals, the number of intervals containing at least one significant SNP ($P < 0.05$) in both reference and validation datasets was 640 ($\chi^2 = 2.12$, which is not significant).

**Validation of SNP across beef vs dairy cattle breeds**. The number of SNP that were significant ($P < 0.05$) for both dairy stature and beef hip height was 202 which is somewhat more than expected by chance (Table 5). When the presence of significant SNP within 1-Mbp-

intervals along each chromosome was examined, the chi-square test was also not significant, $\chi^2 = 2.91$ (Table 5). As shown in Table 5, there were a number of 1-Mbp-intervals containing SNP significant ($P < 0.05$) for both the dairy and beef height trait. For dairy stature and beef w1HH there were 857 1-Mbp-intervals containing SNP significant for these 2 traits. For dairy stature and beef pwHH there were 821 1-Mbp-intervals containing SNP significant for these 2 traits. This suggested that some significant SNP in the dairy reference dataset are near significant SNP in the beef datasets (in this case within the same 1-Mbp-intervals) but this trend was not significant by permutation test. In general, there were regions of the genome that contain many significant SNP for height and weight in different populations such as the broad peak detected on BTA 3 from 102.159Mbp to 109,411Mbp (Figure 3) which contained significant SNP in both beef and dairy datasets.

## DISCUSSION

This paper discussed the results of a GWAS for traits related to weight, height and feed intake in beef and dairy cattle genotyped using 50K and 10K SNP chips. Although more significant associations between SNP genotype and trait than expected by chance were found, many of the FDR were disappointingly high (Table 3). Some of the factors affecting FDR are apparent from the results. The larger sample size (in the CRCII than in the CRCI and in dairy reference than in dairy validation) was associated with lower FDR. The low FDR in the 10K dataset was most likely due to the use of lines of Angus cattle selected for high and low RFI. This increased the range of breeding values for RFI and so increased the power of the analysis. Part of the variation in the Trangie selection line data could be due to genetic drift but we have attempted to correct for this by fitting an animal model in the analysis. The higher 'heritability'

of progeny means compared with single animal phenotypes was associated with lower FDR in the dairy cattle than in the beef cattle GWAS using the same number of cattle. Also, the use of a single breed in the dairy cattle experiment compared with 7 beef breeds across *B. indicus* and *B. taurus* would also have contributed to the greater FDR in the beef cattle experiment as explained below.

At the density of SNP used here, the phase of LD would not be expected to be consistent across cattle breeds (deRoos et al., 2008). Consequently when multiple breeds are combined the associations between a SNP and a QTL are likely to be in different directions in different breeds and hence partially cancel out. Fitting a model with effect of SNP nested within breed (results not included) lacked the power to accurately determine the phase of LD due to the small number of animals within each breed.

The FDR was reduced slightly by using a more stringent *P*-value in the significance test but at the cost of reducing the number of true associations detected. Therefore, rather than rely on a very stringent significance test, confirmation of associations discovered between SNP and traits was sought by confirming them in an independent population of cattle. The most obvious confirmation would be to find the same SNP significant in both datasets. However, this was rarely the case, due to the lack of power in both discovery and validation datasets and to the use of different breeds for validation and discovery. Variation in LD phase between breeds means that a SNP that is significant in one breed may not be significant in another breed even if the same QTL is segregating in both breeds.

The most powerful confirmation test appears to be finding that SNP that are significant in both populations have effects in the same direction more often than expected by chance. For instance, the number of SNP that were significant in Bt×Bi populations from both CRCI and

CRCII, was no more than expected by chance. However, among those SNP significant in both populations, 72% had an effect in the same direction for height and 66% for weight. Similar results are shown for the reference and validation Holstein populations. The negative correlations between CRCI Bi and CRCII Bi populations (Table 8) could be due to the very small number of animals (78) in CRCI Bi dataset. This test is powerful because it tests a very specific null hypothesis, that is, that 50% of SNP will have effects in opposite directions. Unfortunately, this null hypothesis appears to be true unless the 2 populations are from the same or closely related breeds.

Further evidence that the associations found were real is provided by finding SNP in the same 1 Mbp region significantly associated with RFI in the 2 independent datasets (50K and 10K). Even though we found associations in the 1-Mbp-interval for live weight and hip height datasets within beef breeds ($\chi^2$ = 6.27 and 5.68, respectively), the permutation tests with 10,000 repetitions showed that they are not significant. This might suggest that those significant associations in 1-Mbp-interval for live weight and hip height were due to the unequal distribution of SNP across the genome. If there is an unequal number of SNP in each 1-Mbp-interval, the chi-square test is inappropriate. Although the permutation test is better, it is less powerful. Therefore, some important findings could be among the results that were not significant by the permutation tests.

The low power of GWAS with < 1000 animals is indirect evidence that there are few QTL affecting these traits with large effects and most QTL have small effects. However, the evidence showed that some of the associations are real and in particular those found in more than one dataset are unlikely to be false discoveries.

We have identified several chromosome regions that appear to contain polymorphisms or QTL affecting RFI. For example a region on BTA 8 from 86Mbp to 94 Mbp contains several SNP that were significant for RFI in the 50K or 10K experiments as well as one SNP that was significant in both. There were also SNP significantly associated with ADG and mMWT in this region (Figure 2). There were also several significant associations with RFI in both the 10K and 50K datasets at 51.05-51.77 Mbp on BTA 5. These results could be due to a single QTL that is in LD with SNP some Mbp away or they could reflect more than one QTL in these regions. Within this region is a gene encoding hydroxysteroid (17-beta) dehydrogenase 3 (HSD17B3), which is important for steroid metabolism and another gene encoding SRC homology 2 domain-containing-transforming protein C3 (**SHC3**). SRC is a signal transduction protein, involved in recognition of phosphorylated tyrosine. In humans, SHC3 play a role as a signalling adaptor that couples activated growth factor receptors to signalling pathway in neurons and is also involved in the signal transduction pathways of neurotrophin-activated Trk receptors in cortical neurons.

Genetic correlations between traits imply that QTL have pleiotropic effects on multiple traits. High correlations of SNP effects were found between RFI and DFI as well as between ADG and DFI. However, when correlated traits are analysed, the sampling errors tend to be correlated so they do not represent independent evidence for the existence of a QTL. Therefore, uncorrelated traits such as RFI and ADG are useful to investigate the pleiotropic effects of QTL (Nkrumah (2007) found similar results). As example, 3 SNP on BTA 2 situated near 109,093,402 bp (within 42 kbp in both sides) had significant effects (with probability thresholds between $8.0\times10^{-4}$ and $2.0\times10^{-2}$) for the feedlot RFI, ADG and DFI traits. The gene for insulin like growth factor binding protein 2 (IGFBP2) is located on BTA2 near 109Mbp.

On the other hand, it is also possible that the association between a SNP and more than one trait reflects the effect of multiple QTL each affecting a single trait rather than one QTL affecting multiple traits (pleiotropy). With the current density of SNP marker panels it is difficult to distinguish multiple QTL in close proximity. It may be possible to distinguish between these QTL using denser SNP panels.

There are other reports of QTL for RFI in cattle (Barendse et al., 2007; Nkrumah et al., 2007; Sherman et al., 2009). Perhaps due to lack of power and high FDR, there is not a close agreement between the studies despite some overlap between the cattle used by Barendse et al. (2007) and the 50K SNP database used in this report. However, there are some chromosomal regions where significant associations with RFI are in common with the other studies (Table 9). Barendse et al. (2007) also detected RFI QTL ($P$ = 0.006) on BTA 8 situated at 21.2cM. This SNP was significant with the Angus (10K SNP) and CRCI (50K SNP) data ($P$ = 0.044 & $P$ = 0.003, respectively). Also, Barendse et al. (2007) identified 2 SNP at stringent thresholds ($P <$ 0.0009) on BTA 1 and 20, but these 2 SNP (at same position) had no effect on RFI in the Angus and CRCI datasets used here. However there were significant neighbouring SNP on BTA 1 and on BTA 20 within a 1-Mbp-intervals: 3 RFI 50K SNP at ($P < 0.032$) and one 10K SNP ($P$ = 0.035), respectively (Table 9). Sherman *et al.* (2009) in Canadian cattle (Angus, Charolais & composites) mapped 2 QTL for RFI ($P = 1.2\times10^{-5}$ & $7.6\times10^{-5}$) on BTA 1 and 3, respectively. The putative positions of their RFI QTL on these 2 chromosomes were in approximately the same positions as 2 significant SNP in this study ($P < 2.7\times10^{-3}$ at 6 Mbp on BTA 1 and $P < 3.5\times10^{-4}$ at 82 Mbp on BTA 3; Table 9). Other significant SNP from the present GWASs found on BTA 8, 11, 17, 18, 21, 22, 24, 25 and 26 were also near to those reported by Sherman et al. (2009) and Nkrumah et al. (2007) (Table 9). In Sherman et al. (2009), the most significant QTL

for DFI ($P = 1.38\times10^{-10}$) was found on BTA 7 at 54cM. The closest significant SNP to this DFI QTL in the present study was observed at 55.4Mbp with threshold of $P = 0.004$. Nkrumah et al. (2007) also reported the association of SNP with ADG. The 7 QTL for ADG from their study were near SNP ($P < 0.008$) affecting ADG on BTA 7, 11, 14, 17, 18, 20 and 28 in this study.

A group of SNP for stature in dairy population, height and weight for beef cattle were found to be significant $P < 0.001$ on BTA 5 situated at region of 120.9-121.5Mbp. Schrooten et al. (2000) testing German Holstein Friesian cattle using 277 microsatellite markers (MS) found an indication of suggestive QTL for stature, chest width and birth weight on BTA5 (at 122cM). Furthermore, Hiendleder et al. (2003) reported significant QTL for stature on BTA 6 (at 66cM) in the German Holstein breed using MS markers. In the present study, 3 SNP positioned at 66Mbp on BTA 6 were found to be significant for hip height in CRC beef cattle datasets.

The most notable high and narrow peak for the size of effect was observed on BTA 2 near 20Mbp in the Holstein validation set. Dozens of SNP on BTA 3, which were located from 102.159 Mbp to 109,411 Mbp (Figure 3) were associated with different growth traits in the beef datasets as well as stature in the dairy dataset.

## CONCLUSION

The GWAS reported here found more significant associations between SNP and traits than expected by chance. FDR was lower in the analyses with a larger number of animals, a more stringent significance test, using one breed only, a more highly heritable measurement, and in a population with large genetic variance for RFI due to divergent selection. The direction of the effect of a SNP on traits such as weight and height was only consistent within a breed probably due to the inconsistency of LD phase between breeds. This implies that the power to detect SNP

when all breed types are analysed together is reduced because the association between the SNP and the trait is not consistent across breeds. The ability to confirm an association in an independent dataset is greatest if the confirmation is carried out in the same breed as used for the discovery of the association. In this case, the most powerful confirmation test is that the direction of the association between a SNP and a trait is the same in the discovery and validation datasets. Although most effects of QTL on RFI appear to be small, associations have been found in more than one dataset between RFI and SNP located on chromosomes 5 and 8.

## ACKNOWLEDGEMENTS

We acknowledge the funding from Cooperative Research Centre for Beef Genetic Technologies, Armidale, New South Wales in Australia for this study. We are thankful to Dr. Toni Reverter (CSIRO Livestock Industries, QLD, Australia) for advice regarding which individuals to genotype and for providing a pedigree file. Also, we would like to thank Dr. David Johnston (Animal Genetics and Breeding Unit (AGBU), NSW, Australia) for his advice and comments and Dr. Yuandan Zhang (AGBU, NSW, Australia) and Dr. Bruce Tier (AGBU, NSW, Australia) for cleaning the genotype files.

## LITERATURE CITED

Abecasis, G. R., and W. O. Cookson. 2000. Gold--graphical overview of linkage disequilibrium. Bioinformatics. 16:182-183.

Arthur, P. F., J. A. Archer, R. M. Herd, and G. J. Melville. 2001a. Response to selection for net feed intake in beef cattle. Proc. Assoc. Advmt. Anim. Breed. Genet. Queenstown. New Zealand. 30 Jul - 2 Aug. 14:135-138.

Arthur, P. F., J. A. Archer, D. J. Johnston, R. M. Herd, E. C. Richardson, and P. F. Parnell. 2001b. Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. J. Anim. Sci. 79:2805–2811.

Barendse, W., A. Reverter, R. J. Bunch, B. E. Harrison, W. Barris, and M. B. Thomas. 2007. A validated whole-genome association study of efficient food conversion in cattle. Genetics 176:1893-1905.

Barwick, S. A., M. L. Wolcott, D. J. Johnston, H. M. Burrow, and M. T. Sullivan. 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.

De Roos, A. P. W., B. J. Hayes, R. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein Friesian, Jersey and Angus cattle. Genetics. 179:1503–1512

De Roos, A. P. W., B. J. Hayes, R. Spelman, and M. E. Goddard. 2009. Reliability of genomic predictions across multiple populations. Genetics. 183:1545-1553.

Devlin, B., and N. Risch. 1995. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics. 29:311-322.

Fidanza, J., M. Glazer, D. Mutnick, G. McGall and C. Frank. 2001. High capacity substrates as a platform for a DNA probe array genotyping assay. Nucleosides Nucleotides Nucleic Acids. 20:533-538.

Gilmour, A. R., B. J. Gogel, B. R. Gullis, S. J. Welham, and R. Thompson. 2002. Asreml user guide release 1.0. VSN International Ltd. Hemel Hempstead, U.K.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, K. Savin, C. P. van Tassell, T. S. Sonstegard, and M. E. Goddard. 2009. A validated genome wide association study to breed cattle adapted

to an environment altered by climate change. PLoS ONE 4(8): e6676. doi:10.1371/journal.pone.000667.

Hill, W.G. and A. Robertson. 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38:226-231

Johnston, D. J., A. Reverter, H. M. Burrow, V. H. Oddy, and D. L. Robinson. 2003. Genetic and phenotypic characterisation of animal, carcass, and meat quality traits from temperate and tropically adapted beef breeds. 1. Animal measures. Austr. J. Agric. Res. 54:107-1183.

Lettre, G., A. U, Jackson, C. Gieger , F. R. Schumacher, S. I. Berndt, S. Sanna, S. Eyheramendy, B. F. Voight, J. L. Butler, C. Guiducci, and et al. 2008. Identification of ten loci associated with height highlights new biological pathways in human growth. Nat. Genet. 40:584-591.

Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan M. P. Heaton, J. O'Connell, S. S. Moore, T. P. Smith, T. S. Sonstegard, and C. P. Van Tassell. 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4:e5350.

Moore, S. S., F. D. Mujibi, and E. L. Sherman. 2008. Molecular basis for residual feed intake in beef cattle. J. Anim. Sci. 1910. doi: 10.2527/jas.2008-1418.

Nkrumah, J. D. et al. 2007. Primary genome scan to identify putative quantitative trait loci for feedlot growth rate, feed intake, and feed efficiency of beef cattle. J. Anim. Sci. 85:3170-3181.

Pryce, J. E., M. Haile-Mariam, K. Verbyla, P. J. Bowman, M. E. Goddard, and B. J. Hayes. 2010a. Genetic markers for lactation persistency in primiparous Australian dairy cows. J. Dairy Sci. 93:2202-2214.

Pryce, JE., S. Bolormaa, A.J. Chamberlain, P.J. Bowman, K. Savin, M.E. Goddard and B.J. Hayes. 2010b. A validated genome wide association study in two dairy cattle breeds for milk production and fertility traits using variable length haplotypes. J. Dairy Sci. 93: 2827-3404.

Robinson, D.L. and V.H. Oddy (2004) Genetic parameters for feed efficiency, fatness, muscle area and feeding behaviour of feedlot finished beef cattle. Livest. Prod. Sci. 90:255–270. doi: 10.1016/j. livprodsci.2004.06.011.

Schrooten, C., H. Bovenhuis, W. Coppieters, and J. A. M. Van Arendonk. 2000. Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle J. Dairy Sci. 83:795-806.

Sherman, E. L., J. D. Nkrumah, C. L. R. Bartusiak, B. Murdoch, and S. S. Moore. 2009. Fine mapping quantitative trait loci (QTL) for feed intake and efficiency in beef cattle. J. Anim. Sci. 87: 37-45.

Storey, J. D. 2002. A direct approach to false discovery rates. J. R. Stat. Soc. Series B 64:479-498.

**Table 1.** Breed type and trait description

| | | Beef dataset | | | Dairy dataset | |
|---|---|---|---|---|---|---|
| | | CRCI | CRCII | Trangie | HF-ref | HF-val |
| Breed type | Bt | ✓ | | ✓ | ✓ | ✓ |
| | Bi | ✓ | ✓ | | | |
| | Bt×Bi | ✓ | ✓ | | | |
| SNP chip | 50K | ✓ | ✓ | | ✓ | ✓ |
| | 10K | | | ✓ | | |
| Traits | RFI | ✓ | | | | |
| | DFI | ✓ | | | | |
| | ADG | ✓ | | | | |
| | mMWT | ✓ | | | | |
| | w1LWT | | ✓ | | | |
| | pwHH | ✓ | | | | |
| | w1HH | | ✓ | | | |
| | Stature | | | | ✓ | ✓ |

* Bt = *Bos taurus*; Bi = *Bos indicus*; Bt×Bi = crosses between *Bos taurus* and *Bos indicus*; Trangie = Angus Trangie selection line; RFI = residual feed intake, kg/day; DFI = daily feed intake, kg/day; ADG = average daily gain, kg/day; mMWT = metabolic mid-weight, $kg^{0.75}$; w1LWT = "end of wet season 1" weight, kg; pwHH = post-weaning hip height, cm; w1HH = "end of wet season 1" hip height, cm; stature = height, cm; CRC = Cooperative Research Centre

**Table 2.** Number of records (N), mean, standard deviation (SD) and estimates of heritability ($h^2$) and associated standard errors (SE) for all traits studied

| Trait | CRC | N | Bt | Bi | Bt×Bi | ALL-mean | ALL-SD | ALL-$h^2$ | ALL-$h^2$_SE |
|---|---|---|---|---|---|---|---|---|---|
| *Beef Cattle* | | | | | | | | | |
| RFI-10K | - | 379 | 379 | - | - | -0.20 | 1.3 | 0.89 | 0.09 |
| RFI | I | 852 | 486 | 78 | 288 | -0.04 | 1.2 | 0.18 | 0.13 |
| DFI | I | 852 | 486 | 78 | 288 | 12.3 | 2.1 | 0.16 | 0.13 |
| ADG | I | 852 | 486 | 78 | 288 | 1.40 | 0.4 | 0.24 | 0.14 |
| mMWT | I | 852 | 486 | 78 | 288 | 93.8 | 11.4 | 0.31 | 0.15 |
| w1LWT | II | 1456 | - | 590 | 866 | 301.3 | 44.3 | 0.61 | 0.11 |
| pwHH | I | 812 | 466 | 65 | 281 | 116.4 | 6.5 | 0.25 | 0.18 |
| w1HH | II | 1224 | - | 360 | 864 | 126.0 | 5.8 | 0.60 | 0.12 |
| *Dairy Cattle* | | | | | | | | | |
| Stature-ref | - | 588 | - | - | - | 0.54 | 0.72 | 0.78 | 0.12 |
| Stature-val | - | 117 | - | - | - | 0.55 | 0.80 | 0.71 | 0.13 |

- = data not available; RFI = residual feed intake, kg/day; DFI = daily feed intake, kg/day; ADG = average daily gain, kg/day; mMWT = metabolic mid-weight, $kg^{0.75}$; w1LWT = ”end of wet season 1” weight, kg; pwHH = post-weaning hip height, cm; w1HH = ”end of wet season 1” hip height, cm;, stature = height, cm.

**Table 3**. Number of significant SNP and false discovery rate (FDR) at different thresholds ($P < 0.01$) for all traits studied in beef and dairy cattle

| | No. of SNP at *P* | | | FDR (%) at *P* | | |
|---|---|---|---|---|---|---|
| Trait | <0.0001 | <0.001 | <0.01 | <0.0001 | <0.001 | <0.01 |
| *Beef Cattle* | | | | | | |
| RFI-10K | 36 | 111 | 468 | 2 | 7 | 17 |
| RFI | 11 | 75 | 615 | 46 | 67 | 82 |
| DFI | 8 | 76 | 624 | 63 | 67 | 81 |
| ADG | 11 | 83 | 698 | 46 | 61 | 72 |
| mMWT | 6 | 78 | 694 | 84 | 65 | 73 |
| w1LWT | 29 | 156 | 935 | 17 | 32 | 54 |
| pwHH | 13 | 75 | 632 | 39 | 67 | 80 |
| w1HH | 26 | 134 | 833 | 19 | 38 | 60 |
| *Dairy Cattle* | | | | | | |
| Stature-ref | 26 | 173 | 912 | 15 | 22 | 42 |
| Stature-val | 9 | 70 | 589 | 43 | 56 | 66 |

* RFI = residual feed intake; DFI = daily feed intake; ADG = average daily gain; mMWT = metabolic mid-weight; w1LWT = ”end of wet season 1” weight; pwHH = post-weaning hip height; w1HH = ”end of wet season 1” hip height; stature = height; 10K = using 10K chip

**Table 4**. Significant SNP ($P < 0.05$) for RFI at particular positions and within one-Mbp-intervals in both 10K and 50K datasets

| SNP name | BTA | Position (Mbp) | 50K *P* | 10K *P* | BTA | Location (Mbp)[&] | 50K No.SNP[†] | 10K No.SNP[†] | 50K $P_{min}$* | 10K $P_{min}$* |
|---|---|---|---|---|---|---|---|---|---|---|
| 352323 | 1 | 103459113 | 0.0265 | 0.0015 | 2 | 22-23 | 1 | 2 | 0.0053 | 0.0009 |
| 347872 | 1 | 140599889 | 0.0313 | 0.0081 | 2 | 24-25 | 2 | 3 | 0.0003 | 0.0101 |
| 352046 | 1 | 35152843 | 0.0415 | 0.0326 | 2 | 63-64 | 2 | 1 | 0.0002 | 0.0338 |
| 345175 | 2 | 113984723 | 0.0143 | 0.0186 | 3 | 105-106 | 7 | 2 | 0.0000 | 0.0032 |
| 348132 | 2 | 133058384 | 0.0064 | 0.0095 | 4 | 41-42 | 3 | 2 | 0.0001 | 0.0211 |
| 353948 | 2 | 83913947 | 0.0336 | 0.0154 | 4 | 91-92 | 2 | 6 | 0.0002 | 0.0010 |
| 350236 | 3 | 24374862 | 0.0147 | 0.0028 | 5 | 51-52 | 2 | 6 | 0.0021 | 0.0001 |
| 342691 | 5 | 116152845 | 0.0310 | 0.0324 | 5 | 75-76 | 3 | 4 | 0.0004 | 0.0073 |
| 349813 | 5 | 90670437 | 0.0034 | 0.0468 | 5 | 85-86 | 2 | 5 | 0.0010 | 0.0246 |
| 347570 | 8 | 21215935 | 0.0033 | 0.0438 | 5 | 110-111 | 2 | 2 | 0.0000 | 0.0084 |
| 347480 | 8 | 93873871 | 0.0001 | 0.0264 | 7 | 102-103 | 3 | 4 | 0.0000 | 0.0011 |
| 349887 | 9 | 36408577 | 0.0255 | 0.0216 | 8 | 2-3 | 2 | 3 | 0.0025 | 0.0004 |
| 352056 | 11 | 16684267 | 0.0489 | 0.0210 | 8 | 86-87 | 6 | 3 | 0.0000 | 0.0031 |
| 344077 | 11 | 51903129 | 0.0344 | 0.0264 | 8 | 90-91 | 7 | 2 | 0.0078 | 0.0001 |
| 352038 | 12 | 81260515 | 0.0306 | 0.0104 | 8 | 93-94 | 6 | 3 | 0.0001 | 0.0199 |
| 342868 | 14 | 59149744 | 0.0101 | 0.0399 | 8 | 104-105 | 4 | 1 | 0.0009 | 0.0392 |
| 343856 | 16 | 16731684 | 0.0170 | 0.0381 | 9 | 14-15 | 1 | 1 | 0.0086 | 0.0009 |
| 346839 | 16 | 33946102 | 0.0234 | 0.0270 | 9 | 60-61 | 3 | 1 | 0.0044 | 0.0001 |
| 345143 | 16 | 46045401 | 0.0167 | 0.0200 | 10 | 18-19 | 2 | 1 | 0.0006 | 0.0242 |
| 349182 | 18 | 24355937 | 0.0407 | 0.0009 | 11 | 1-2 | 2 | 3 | 0.0040 | 0.0008 |
| 352299 | 18 | 43829131 | 0.0161 | 0.0036 | 12 | 55-56 | 2 | 4 | 0.0064 | 0.0000 |
| 345848 | 18 | 45787269 | 0.0214 | 0.0010 | 17 | 10-11 | 3 | 2 | 0.0008 | 0.0317 |
| 353716 | 19 | 17750262 | 0.0239 | 0.0017 | 17 | 43-44 | 4 | 1 | 0.0045 | 0.0001 |
| 353167 | 19 | 19698761 | 0.0317 | 0.0487 | 17 | 57-58 | 5 | 4 | 0.0024 | 0.0002 |
| 353494 | 20 | 51402608 | 0.0315 | 0.0294 | 18 | 3-4 | 4 | 2 | 0.0041 | 0.0008 |
| 354432 | 26 | 2527236 | 0.0188 | 0.0061 | 20 | 33-34 | 5 | 1 | 0.0006 | 0.0135 |
| 348792 | 26 | 32256982 | 0.0396 | 0.0064 | 24 | 10-11 | 2 | 1 | 0.0003 | 0.0336 |
| | | | | | 25 | 12-13 | 1 | 4 | 0.0039 | 0.0005 |
| | | | | | 27 | 21-22 | 2 | 1 | 0.0084 | 0.0004 |
| | | | | | 28 | 36-37 | 4 | 2 | 0.0009 | 0.0115 |

BTA = No. of chromosome; [&] - indicates one-Mbp range; No.SNP = number of significant SNP ($P < 0.05$) for both 50K & 10K datasets within one-Mbp-intervals; $P$ = $F$-probabilities; *$P_{min}$ = $P$ value of SNP were found to be the lowest amongst significant SNP ($P < 0.05$) across 2 datasets at particular locations of one-Mbp-interval

**Table 5**. Validation of individual SNP or 1-Mbp- chromosome regions affecting traits in 2 datasets.

| Trait & dataset | SNP/interval | Total No | No. Sig.* data1 | No. Sig.† data2 | No. Sig. both data1 &data2 | $\chi^2$ | *P* value |
|---|---|---|---|---|---|---|---|
| *RFI based 50K & 10K chips in beef cattle* | | | | | | | |
| 50K* & 10K† | Individual SNP | 2390 | 130 | 369 | 27 | 2.99 | 0.084 |
| 50K* & 10K† | 1-Mbp-interval | 2131 | 1207 | 660 | 406 | 9.25 | 0.012 |
| *RFI & ADG based 50K chip in beef cattle* | | | | | | | |
| RFI-all* & ADG-all† | Individual SNP | 50633 | 2826 | 2995 | 162 | 0.18 | 0.672 |
| RFI-all* & ADG-all† | 1-Mbp-interval | 2530 | 1417 | 1476 | 845 | 2.22 | NS |
| *LWT§ based 50K chip in beef cattle* | | | | | | | |
| CRCI*1 & CRCII† | Individual SNP | 49957 | 2699 | 3050 | 176 | 0.58 | 0.446 |
| CRCI* & CRCII† | 1-Mbp-interval | 2532 | 1585 | 1484 | 959 | 6.27 | NS |
| *HH* based 50K chip in beef cattle* | | | | | | | |
| CRCI* & CRCII† | Individual SNP | 49926 | 2846 | 3037 | 198 | 4.04 | 0.044 |
| CRCI* & CRCII† | 1-Mbp-interval | 2532 | 1543 | 1492 | 938 | 5.68 | NS |
| *Stature based 50K chip in dairy cattle* | | | | | | | |
| HF-ref* & HF-val† | Individual SNP | 39040 | 3215 | 2462 | 215 | 0.86 | 0.354 |
| HF-ref* & HF-val† | 1-Mbp-interval | 2527 | 1370 | 1147 | 640 | 2.12 | NS |
| *Growth based 50K chip in beef & dairy cattle* | | | | | | | |
| pwHH-all* & HF-ref† | Individual SNP | 38367 | 2191 | 3157 | 202 | 3.02 | 0.074 |
| pwHH-all* & HF-ref† | 1-Mbp-interval | 2530 | 1492 | 1371 | 821 | 1.03 | NS |
| w1HH-all* & HF-ref† | Individual SNP | 38521 | 2614 | 3176 | 227 | 0.71 | 0.399 |
| w1HH-all* & HF-ref† | 1-Mbp-interval | 2530 | 1543 | 1371 | 857 | 2.91 | NS |

* = dataset 1; † = dataset 2; § $Bt \times Bi_{CRCI}$ & $Bt \times Bi_{CRCII}$ , No = number; Sig. = significant; - = not available; $\chi^2$ = chi-square value; *P* value = probability of $\chi^2$ using permutation tests with 10,000 repetitions; NS = not significant

**Table 6**. Pleiotropic effects of SNP across feedlot traits in the 50K CRCI data*

| | RFI | DFI | ADG | mMWT |
|---|---|---|---|---|
| RFI | **2826** | 0.58 | -0.05 | -0.06 |
| DFI | 651 (100%) | **2733** | 0.71 | 0.68 |
| ADG | 162 (40%) | 973 (100%) | **2995** | 0.60 |
| mMWT | 182 (31%) | 961 (100%) | 786 (100%) | **3141** |

*on diagonal = number of significant SNP ($P < 0.05$) for each trait, above diagonal = correlations of SNP effects between traits, below diagonal = number of significant SNP in both traits (proportion of significant SNP effects to be in same direction for both traits )

**Table 7**. Significant SNP ($P < 0.05$) within 1-Mbp-interval for all of RFI, ADG and mMWT in CRCI animals

| BTA | Position | RFI No.SNP | ADG No.SNP | mMWT No.SNP | RFI $P_{min}$* | ADG $P_{min}$* | mMWT $P_{min}$* |
|---|---|---|---|---|---|---|---|
| 2 | 106-107 | 2 | 3 | 1 | 0.0001 | 0.0019 | 0.0009 |
| 3 | 51-52 | 6 | 1 | 4 | 0.0004 | 0.0199 | 0.0124 |
| 3 | 84-85 | 4 | 2 | 3 | 0.0002 | 0.0075 | 0.0356 |
| 3 | 105-106 | 7 | 1 | 1 | 0.0000 | 0.0445 | 0.0131 |
| 4 | 46-47 | 1 | 2 | 2 | 0.0258 | 0.0000 | 0.0004 |
| 4 | 91-92 | 2 | 4 | 1 | 0.0002 | 0.0114 | 0.0239 |
| 6 | 41-42 | 3 | 2 | 5 | 0.0002 | 0.0132 | 0.0067 |
| 6 | 111-112 | 3 | 3 | 3 | 0.0007 | 0.0095 | 0.0066 |
| 8 | 86-87 | 6 | 2 | 2 | 0.0000 | 0.0047 | 0.0155 |
| 8 | 87-88 | 5 | 2 | 3 | 0.0001 | 0.0140 | 0.0121 |
| 8 | 88-89 | 7 | 1 | 2 | 0.0009 | 0.0382 | 0.0301 |
| 8 | 89-90 | 7 | 3 | 2 | 0.0006 | 0.0104 | 0.0053 |
| 8 | 104-105 | 4 | 2 | 2 | 0.0009 | 0.0253 | 0.0187 |
| 9 | 78-79 | 4 | 3 | 1 | 0.0003 | 0.0037 | 0.0450 |
| 10 | 18-19 | 2 | 1 | 2 | 0.0006 | 0.0432 | 0.0103 |
| 11 | 46-47 | 1 | 2 | 10 | 0.0483 | 0.0002 | 0.0002 |
| 14 | 17-18 | 6 | 5 | 5 | 0.0005 | 0.0008 | 0.0081 |
| 16 | 25-26 | 1 | 4 | 1 | 0.0417 | 0.0006 | 0.0005 |
| 17 | 10-11 | 3 | 3 | 2 | 0.0008 | 0.0358 | 0.0194 |
| 17 | 37-38 | 2 | 1 | 2 | 0.0005 | 0.0492 | 0.0051 |
| 19 | 38-39 | 2 | 1 | 2 | 0.0003 | 0.0417 | 0.0346 |
| 20 | 30-31 | 1 | 2 | 2 | 0.0001 | 0.0085 | 0.0047 |
| 22 | 45-46 | 1 | 1 | 2 | 0.0003 | 0.0356 | 0.0243 |
| 23 | 18-19 | 1 | 1 | 1 | 0.0432 | 0.0001 | 0.0009 |
| 23 | 49-50 | 5 | 3 | 1 | 0.0002 | 0.0274 | 0.0448 |

No.SNP = number of significant SNP ($P < 0.05$) for all 3 traits found within one-Mbp-intervals; *$P_{min}$ = $P$ value of SNP were found to be the lowest amongst significant SNP ($P < 0.05$) across all 3 traits at particular locations of one-Mbp-interval

**Table 8**. Validation of SNP for body weight and hip height ($P < 0.05$) between CRCI and CRCII datasets across breed types

| $P < 0.05$ | $ALL_{CRCI}$ : $ALL_{CRCII}$ | $Bt_{CRCI}$ : $Bi_{CRCII}$ | $Bt_{CRCI}$ : $Bt{\times}Bi_{CRCII}$ | $Bi_{CRCI}$ : $Bi_{CRCII}$ | $Bi_{CRCI}$ : $Bt{\times}Bi_{CRCII}$ | $Bt{\times}Bi_{CRCI}$ : $Bi_{CRCII}$ | $Bt{\times}Bi_{CRCI}$ : $Bt{\times}Bi_{CRCII}$ |
|---|---|---|---|---|---|---|---|
| ***Body weight*** | | | | | | | |
| No. of SNP | 242 | 206 | 209 | 245 | 203 | 159 | 176 |
| correlation | 0.26 | -0.10 | -0.04 | -0.21 | 0.03 | 0.07 | 0.43 |
| % of same dir | 61 | 46 | 48 | 39 | 50 | 53 | 72 |
| ***Hip height*** | | | | | | | |
| No. of SNP | 208 | 161 | 149 | 213 | 226 | 197 | 198 |
| correlation | 0.12 | 0.00 | 0.02 | 0.09 | -0.17 | 0.10 | 0.36 |
| % of same dir | 53 | 47 | 51 | 55 | 43 | 54 | 66 |

No. = number of SNP; correlation = correlation of corrected SNP; and % of same dir = proportion of the direction of the effect, which was the same; ALL = all CRCI or CRCII animals; Bt = *Bos taurus*; Bi = *Bos indicus*; Bt×Bi = crosses of *Bos taurus* and *Bos indicus*

**Table 9.** Significant SNP for RFI and ADG across 3 datasets (significant SNP locations close to published locations[&])

| BTA | Position[†] (Mbp) | 50K-RFI | 50K-ADG | 10K-RFI | Publication* |
|---|---|---|---|---|---|
| *Significant locations close to published locations*[&] | | | | | |
| 1 | 6 | $P = 0.0027$ | | | 3 |
| 1 | 10 | $P = 0.0007$ | | | 3 |
| 1 | 26 | $P = 0.0253$ | $P = 0.0071$ | | 1 |
| 3 | 82 | $P = 0.0003$ | | | 3 |
| 7 | 83 | | $P = 0.0007$ | | 2 |
| 4 | 60 | $P = 0.0092$ | $P = 0.0313$ | $P = 0.0306$ | 1 |
| 5 | 82 | $P = 0.0003$ | | | 1 |
| 6 | 50 | $P = 0.0184$ | $P = 0.0211$ | | 1 |
| 8 | 21 | $P = 0.0438$ | | $P = 0.0033$ | 1 |
| 8 | 21 | $P = 0.0012$ | $P = 0.0073$ | $P = 0.0438$ | 1 |
| 8 | 80 | $P = 0.0060$ | | | 2 |
| 11 | 19 | | $P = 0.0079$ | | 2 |
| 11 | 30 | $P = 0.0072$ | | | 3 |
| 12 | 55 | $P = 0.0064$ | $P = 0.0193$ | $P = 0.0000$ | 1 |
| 13 | 14 | | $P = 0.0154$ | $P = 0.0121$ | 1 |
| 14 | 74 | | $P = 0.0016$ | | 2 |
| 17 | 9 | | $P = 0.0042$ | | 2 |
| 17 | 18 | $P = 0.0045$ | | | 2 |
| 17 | 56 | $P = 0.0026$ | | | 3 |
| 18 | 28 | $P = 0.0020$ | | $P = 0.0185$ | 1; 3 |
| 18 | 47 | | $P = 0.0049$ | | 2 |
| 18 | 64 | $P = 0.0048$ | | | 2 |
| 20 | 2 | | | $P = 0.0347$ | 1 |
| 20 | 65 | | $P = 0.0029$ | | 2 |
| 21 | 4 | $P = 0.0079$ | | | 3 |
| 22 | 26 | $P = 0.0044$ | | | 3 |
| 22 | 51 | $P = 0.0079$ | $P = 0.0247$ | $P = 0.0484$ | 1 |
| 24 | 4 | $P = 0.0076$ | | | 3 |
| 25 | 14 | $P = 0.0039$ | | | 3 |
| 28 | 23 | | $P = 0.0003$ | | 2 |

* - 1 = Barendse et al. (2007); 2 = Nkrumah et al. (2007) (total number of marker associated with RFI was 8 at chromosome-wise threshold of $P < 0.05$); 3 = Sherman et al. (2009) total number of marker associated with RFI was 19 at chromosome-wise threshold of $P < 0.05$; BTA = No. of chromosome; [&] = http://www.ncbi.nlm.nih.gov/projects/genome/guide/cow/ ;

[†] - All publications referenced the estimated QTL position in cM. We have assumed 1cM approximately equals 1 Mbp to compare the results

**Figure 1**. Relationship between genetic distances and values of linkage disequilibrium (mean $r^2$ corrected for sample size) between SNP markers in different breed types (beef *Bos taurus* (Bt), *Bos indicus* (Bi), and crosses of Bt×Bi, and dairy Bt (all Holstein bulls)).

**Figure 2**. Significant SNP ($P < 0.05$) for residual feed intake (RFI), average daily gain (ADG) and daily feed intake (DFI) on BTA 8.

* -10K indicates using 10K SNP chip dataset whereas 50K indicates using 50K SNP chip dataset .

**Figure 3**. Significant SNP ($P < 0.05$) for growth (mMWT, w1LWT, pwHH, w1HH and stature) in beef CRC cattle and Holstein (HF) bulls on BTA 3.

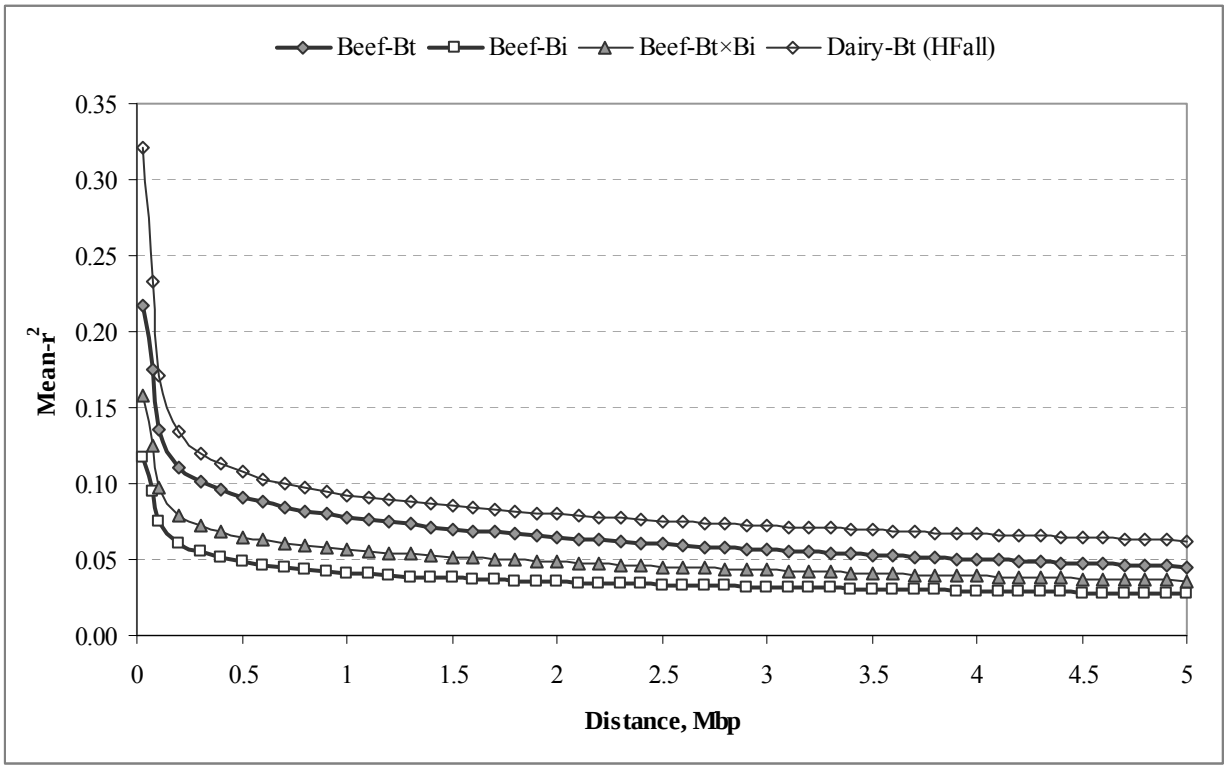


**Figure 1**.

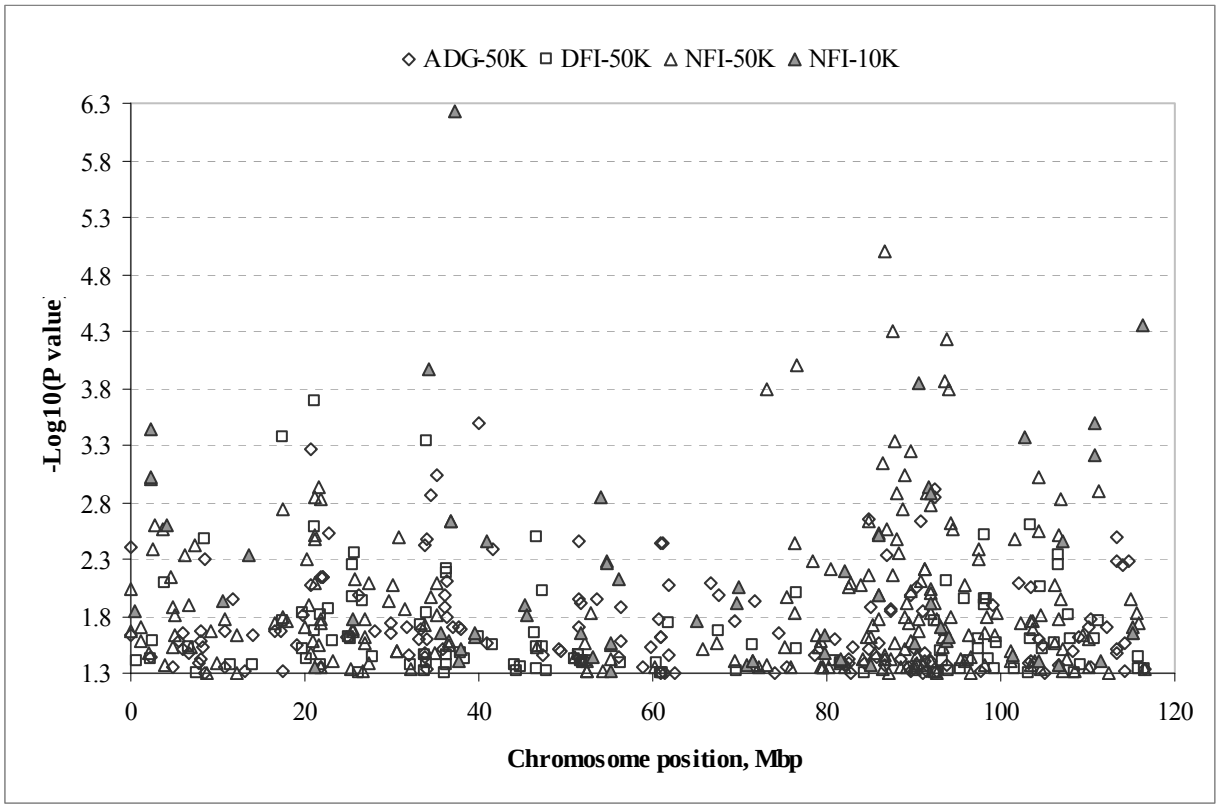


**Figure 2**.

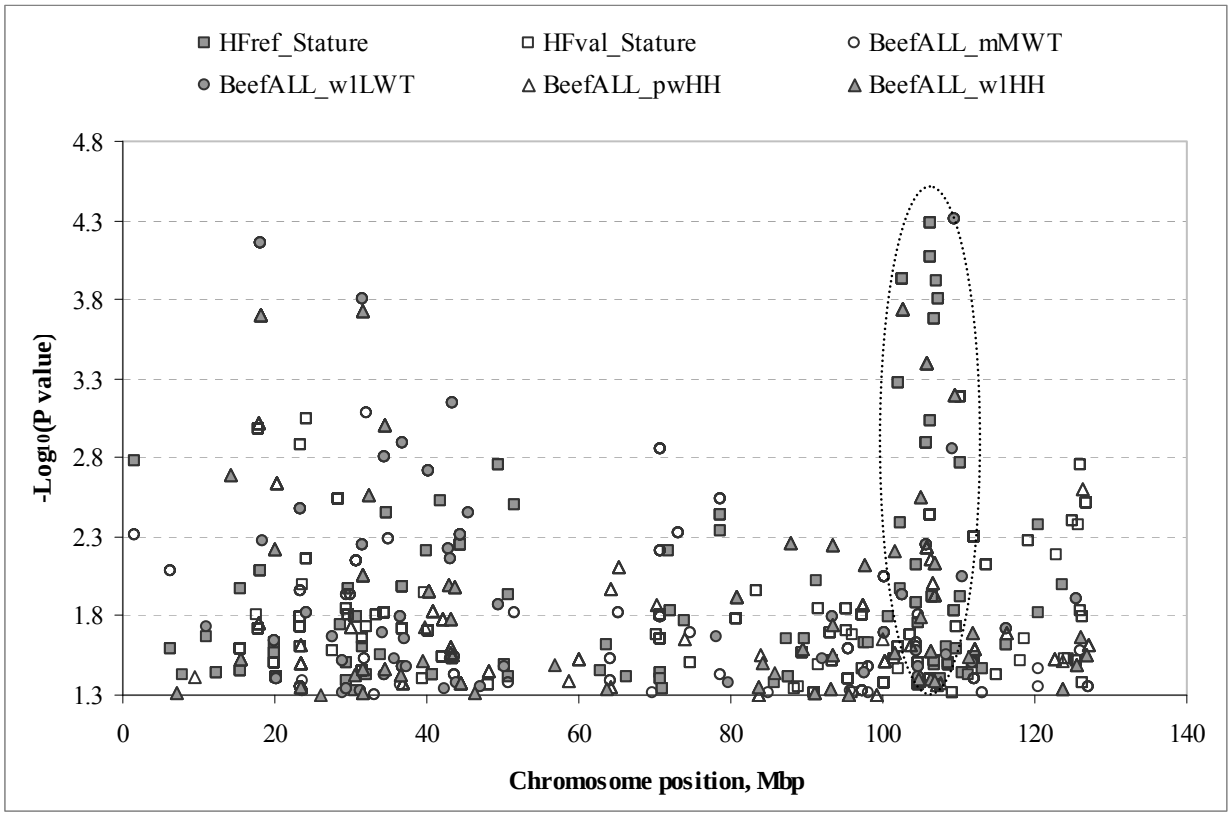


**Figure 3**.