final report

Project code: B.BSC.0049

Prepared by: William Barendse

CSIRO Livestock Industries

Date published: 16 October 2009

ISBN: 9781741919134

PUBLISHED BY
Meat & Livestock Australia Limited
Locked Bag 991
NORTH SYDNEY NSW 2059

Genes for meat and carcass quality

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication.

Abstract

Meat and carcass quality are commercially important traits of beef cattle but meat quality in particular has proved to be hard to improve using genetic methods. One option is to identify DNA markers that could be used to predict breeding values or phenotypes. In this study, we used the genome wide association methodology to identify likely genes affecting meat and carcass quality. The traits of most interest were marbling or intramuscular fat percent (IMF), meat tenderness (LLPF) and meat yield (RBY) and rump fat tickness (P8FAT). We have performed two genome wide association studies, the most recent using 54,000 DNA markers on 1,035 cattle with meat and carcass quality phenotypes associated with the Beef CRC DNA bank and database. So far, 279 highly significant ($P < 0.001$) associations for the four traits were identified, more than would be expected by chance. Several of these associations have now been confirmed in large independent studies. The largest confirmed effect accounts for 2.2% of the residual phenotypic variance in rump fat thickness, this was larger than the combined effect of the Calpain and Calpastatin genes for meat tenderness in the same type of cattle, and it was the largest gene effect reported so far for meat or carcass quality, after the myostatin double muscling mutation. Four genes have been confirmed that each explain as much variability for IMF as Calpain or Calpastatin explain for meat tenderness and should be made available for commercial release as genetic tests.

Executive summary

Why was the Work Done:
This work was performed to provide tools (DNA markers) to select cattle for meat and carcass quality traits. A small number of DNA markers for meat quality, especially for marbling and meat tenderness, were in use in the Beef Industry before this project began. However, except for the two markers for meat tenderness in the Calpain 1 and Calpastatin genes and the double muscling mutations for meat yield in the Myostatin gene, there were no markers for other meat or carcass quality that were reported to have large, reliable effects on either marbling, intramuscular fat, rump fat, meat tenderness or meat yield. There was therefore a clear need for a large number of additional DNA markers, particularly for marbling or intramuscular fat, which would allow for improved genetic testing. The aim of the project, therefore, was to identify as many DNA markers for meat and carcass quality as possible. Markers that were identified would then be tested in additional, independent samples to get unbiased estimates of their effects. Finally, those markers that gave consistent effects on their traits would be made available for genetic testing. This last activity would include DNA markers that had been identified in previous studies, particularly in other countries to determine whether those markers had any efficacy in Australian cattle and conditions. Our aim was not to identify the best methods for using the marker information or the best methods to introduce the DNA markers to the Australian industry – in the absence of DNA markers such activities would be theoretical.

What was Achieved:
Up to date, we have identified 279 DNA markers that were highly significantly associated ($P < 0.001$) to one of the four traits intramuscular fat (IMF), meat tenderness (LLPF), retail beef yield (RBY) and rump fat thickness (P8FAT), and for which there was substantial information based on hundreds of animals from a total of seven cattle breeds. The trait IMF had at least a third more significant DNA markers than any of the other three traits, with 87. There were 63, 64 and 65 DNA markers each, for P8FAT, LLPF and RBY, respectively.

We identified at least five moderate to large genetic effects. For quantitative traits most of the genes are now known to account for small amounts of the variance, although there are a few known exceptions. We made every effort to determine whether there might be some DNA markers that account for moderate to large genetic effects, because selection decisions are easier and can be done with more confidence when large genes are present in a selection index. We identified one such large effect on P8FAT. Although rump fat thickness had the fewest markers on the fewest chromosomes evidently that did not prevent it from also having the largest single genetic effect. This large genetic effect was confirmed in a completely separate sample of cattle. This gene accounts for 2.2% of the residual variance of P8FAT in the discovery population, which is larger than the combined effects of the Calpain 1 and Calpastatin genes on meat tenderness. It appears to have positive effects on fertility traits in tropical cattle as well, but discussion of that is outside the scope of this report. We also identified four genes associated with IMF each accounting for more than 1% of the residual variance in a confirmation sample. Our work on meat tenderness showed that there were no genes larger than Calpain 1 or Calpastatin in our population sample, although we identified additional DNA markers that showed effects on this trait. These additional DNA markers have been evaluated further to determine which of these would be useable in commercial testing.

When and How can the Industry Benefit from the Work:
For this technology, the industry benefit is proportional to the amount of variability that is explained by a set of markers, because at present, the cost of markers is not sensitive to the number of tests in a DNA marker panel – that is, individuals are charged essentially as much for 10 tests as they are charged for 50 tests done on an animal. Therefore, the industry benefit increases as more markers with confirmed effects are provided for genetic testing. The industry will benefit by genetic testing of the markers and where these DNA markers are included in indices that either predict breeding values of individuals or phenotypic performance. There are

companies offering services for genetic testing in Australia, such as Pfizer Animal Genetics, Merial Australia, and Prescribe Genomics Japan through the University of Queensland. The release of the DNA markers in this study to these companies may see them implemented through the procedures of those companies. In addition, there are plans to implement Genomic Selection where DNA markers such as these will be integrated into a BREEDPLAN like tool and then with consistent year to year evaluation the DNA markers and their effects will be refined to give better prediction of breeding values and phenotypes. That system is, however, still in the planning phase and it is not clear at present who will actually implement such Genomic Selection.

Who can Benefit from the Results:

Benefits will accrue on farm, to the consumer as well as to the processing and feedlotting sectors. Most individuals who will receive a direct financial benefit are those who apply the DNA marker tests to improve their breeding stock. This includes stud and commercial breeders who use the tests, as well as those who buy stock from those who do genetic testing, because genetic improvement is cumulative and transferable. The financial benefit will be in the form of higher prices for stud animals with favourable genotypes, as well as higher prices for animals that meet certain performance standards. The consumer should benefit from better quality beef, and this may have an impact on the market share of beef compared to other meats. The feedlotting sector should benefit financially because any animal that reaches its carcass or meat quality target at an earlier age will cost less to feed and its lifetime production of methane may be lower. Finally, the meat processing sector may benefit through having carcasses and meat quality that are more uniform to specification because animals that perform to a certain standard should be easier to identify. The benefits to the feedlotters and processors should be in the form of a reduction of wasted effort.

Contents

1 Background

1.1 Background – Introduction

Marker assisted breeding is a reality although the number of DNA markers available is small. In cattle, there are two major genes affecting commercially important traits, double muscling mutations in the Myostatin (MSTN) gene and the mutation causing increased milk fat percentage in the Diacylglycerol o-acyltransferase homolog 1 (DGAT1) gene [1,2]. Three other genes with more moderate but still appreciable genetic effects on dairy traits have been identified, based on the Growth Hormone Receptor, Prolactin Receptor and the ATP-binding cassette, subfamily G, member 2 genes [3-5]. Two genes with similar moderate effects have been identified for meat quality based on the Calpain 1 (CAPN1) and Calpastatin (CAST) genes [6-9]. When this project began, there were other genes with much smaller effects than these on meat and carcass quality. Over time many more of these smaller effects have been identified, a few of these smaller genes have since been confirmed, although not in all studies, many of these genes could not be confirmed in independent studies, some of the genes have never been tested in large samples, and some of them have never been tested in follow up studies. The overall situation has been reviewed [10]. Any survey of the literature or of data shows that there are far fewer large effects than moderate to small effects. It is easy to find small genetic effects but harder to find them a second time [11], while the use of small effects has proved to be controversial in some implementations. Most of the current tests that are available are based on small effects. Nevertheless, there are now some published methods of how to integrate the information from the entire genome using DNA markers [12] irrespective of the size of the effects of the genes or of the amount of variation that they account for.

Given this state of affairs, our aim in the project was to identify as many DNA markers for meat quality as possible, search in particular for those with larger effects, and confirm as many as we could in independent large samples. Generally it has been found that larger effects have the same effect in most or all breeds. For smaller effects to be useful, in addition to be confirmed in other breeds, they have to be screened to find those that have the same favourable genotype in most or all breeds.

2 Project objectives

2.1 Project objectives

2.1.1 Project objectives - Background

The main aim of the project was to identify new DNA markers for meat and carcass quality of cattle and to confirm the effects of many of them so that they may be used in commercial testing. When the project was proposed, there was a notion that 3 markers per year would be feasible to be identified, one each for yield, intramuscular fat and meat tenderness. Over the life of the Beef CRC, to which this project was tied, it was assumed that over a five-year period that 15 DNA markers would be identified. It was also thought that these would be released as soon as they became available. It was thought that Quantitative Trait Loci (QTL) mapped in the CBX and Limousin-Jersey experiments [13,14] and other QTL studies summarised at http://www.animalgenome.org/QTLdb/ [15] would form the basis for such DNA markers. Based on the reported sizes of these QTL it was thought that between 5-10 QTL per trait would explain approximately half of the genetic variance for the trait.

2.1.2 Project objectives

The aims therefore were:

1. To identify 3 DNA markers for a meat or carcass quality trait per year using the Beef CRC Progeny test population [16] from QTL previously identified through linkage mapping
2. To confirm the association in a separate sample of cattle from the Beef CRC cattle
3. Commercialise the DNA markers releasing 3 markers per year.

2.1.3 Project objectives – Changes and project expectations

Three things changed these objectives radically soon after the project began. Firstly, the genome wide association study (GWAS) method became available for cattle in 2005. This opened the possibility of more efficient de novo discovery of QTL – it was estimated that identifying a QTL from a linkage analysis would take approximately 3 years per QTL, so identifying 15 QTL would be a large and expensive undertaking. Secondly, it was shown that most of the QTL in linkage mapping studies were grossly overestimated in size mainly because they were being discovered in sibships of 1-200 individuals. Once these QTL were tested in larger samples of 1,000 animals their effects dwindled in size so that even large QTL explained only 1-2% of the residual variance. So far, only the major genes MSTN and DGAT1 (cf. above) have effects larger than 10% of the residual variance when tested in samples of a thousand animals. Most of the QTL explained less than 0.5% of the residual variance. To explain half of the genetic variance would require hundreds of DNA markers. Thirdly, and related to this, QTL with small effect sizes require large samples of animals to confirm their effects. To ensure that sufficient power was used, the Beef CRC management, after consultation with scientists and stake holders, decided that no markers were to be commercialised until 1) they had been confirmed in large samples approaching 1,000 animals per breed for 3-5 breeds, and 2) they were included in panels of DNA markers that explained a large proportion (eg. at least 15%) of the genetic variance.

Given these changes, it is easy to see that the project would identify very many more than 15 DNA markers, that these would need to be tested more extensively than before, and that it was unlikely that any of these DNA markers would be commercialised on an annual basis as had been thought at the beginning of the project.

3 Methodology

3.1 Methodology

3.1.1 Methodology – Cattle samples

DNA was already available in a DNA bank from cattle samples of the Beef CRC Progeny Test (BCPT) project [16] and the CSIRO SBEF.018 Marbling Feedlot samples [17]. The BCPT sample consisted of 9150 cattle with a DNA sample and meat and carcass quality phenotypes, 5-generation pedigree information and consisting of seven pure breeds and a series of cross-breed cattle. The seven breeds were the taurine Angus, Hereford, Murray Grey and Shorthorn, the indicine Brahman and the tropical composite Belmont Red and Santa Gertrudis. The cross-bred animals were first generation crosses of Brahman dams and either taurine sires or tropical composite sires. This sample represented the offspring of 428 sires. The range of offspring per sire was 1-95 with an average of 21 offspring per sire. The SBEF.018 sample consisted of 1732 animals with DNA samples, breed identity, marbling score, P8FAT and eye muscle areas (EMA). Animals were mainly of the Angus (N=853) and Shorthorn (N=773) breeds with a small number (N=106) of a wide range of taurine breeds. Although these animals were not as well described as the BCPT sample, they had a much higher range of marbling scores, which is why they were included in the study for some applications. DNA for all animals were extracted using published procedures, quantified and archived [9,17].

3.1.2 Methodology – genotyping

Genotyping was performed using several previously published methods. GWAS were performed using two sets of single nucleotide polymorphisms. The first GWAS was performed using the MegAllele™ Genotyping Bovine 10K SNP Panel [18], a fully described set of SNP, by ParAllele Inc. on an Affymetrix GeneChip Scanner 3000, yielding an average spacing of 325 kb between SNP. Further details of the SNP, which are now no longer available as a genotyping panel, can be found at the link ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Btaurus/snp/Btau20050310/. The bulk of the SNP on the SNP array were obtained by comparing the genome sequence of the Hereford animal to the partial sequence of the Holstein (72.4%) and the Angus (15%) animal, with 7.5% cSNP (coding SNP) obtained from the Interactive Bovine in silico SNP database [19], and the rest from the partial sequence of the Limousin (3.1%) and the Brahman (2%) animal. Quality control of those data was performed by Bill Barendse at CSIRO in Brisbane using published methods [20]. The second GWAS was performed using the Illumina Bovine SNP50 chip [21] by Illumina Inc, yielding an average spacing of 51.9 kb between SNP. These SNP are a combination of those identified during the Bovine Genome Sequencing Project and the Bovine HapMap project [22] and those identified by reduced representation sequencing on an Illumina Genome Analyser [23]. Quality control of those data was performed by Yuandan Zhang and Bruce Tier at the Animal Genetics and Breeding Unit in Armidale.

DNA markers that were genotyped at CSIRO, either for SNP from these panels or from de novo sequencing, were followed up using three technologies. The Taqman™ assays were performed using Minor Groove Binding (MGB) assays [24] obtained from the Assays by Design pipeline using the manufacturer's instructions (Applied Biosystems Inc., Foster City, CA). The genotyping was performed on an ABI Sequence Detection System 7900HT and genotypes were scored by two individuals and sent through a quality control pipeline before they were attached to phenotypes for analysis. The SNPlex™ assays were performed usually in groups of 48 SNP at a time following the manufacturer's instructions (Applied Biosystems Inc., Foster City, CA). The SNPlex™ method did not give uniformly high returns of genotypes so although more than 1,000 animals were usually genotyped, most SNP assays gave fewer than 1,000 scorable genotypes in return. The genotypes were scored by two individuals and sent through the same quality control pipeline before they were attached to phenotypes for analysis. The Illumina Golden Gate Assays (GGT) [25,26] were genotyped in groups of 384 SNP at a time following the manufacturer's instructions (Illumina Inc., CA). The GGT assays replaced the SNPlex™ assays due to improved cost per genotype, increased number of genotypes per SNP, and improved reproducibility of genotypes. Genotypes were treated as noted above.

3.1.3 Methodology – Analysis

The standard analysis in the Beef CRC is to calculate the regression of phenotype, adjusted for fixed effects and ancestry, on the number of copies of a specified allele at each SNP. The advantage of this approach is that the resulting values can be used directly in estimating Breeding Values (BV). We fitted a full animal model in ASReml [27] using a five-generation pedigree to account for the polygenic component of inheritance as reported before [28]. We included the contemporary group fixed effects of herd, sex and kill group [29], as well as age in days at slaughter and genotype coded as number of copies of the designated allele. For IMF this resulted in the model imf ~ $N(\mu$ + herd + sex + kill group + age + genotype, $\sigma^2_e)$. The genotype was analysed as the regression of phenotype on the number of copies of an allele, which is the allele substitution effect (α). Each breed consists of a small number of herds of origin, with breed confounded with herd of origin. Replacing herd of origin with breed usually gives the same result. To estimate genotypic effects within breed, both breed and genotype nested within breed were fitted in the model. The proportion of the variance explained was obtained by comparing the residual sums of squares (RSS) of the model with SNP to the RSS of the model without SNP. Standard sample size and power calculations using the critical points of the normal distribution

were calculated as previously described [30]. Genotypes were tested for Hardy-Weinberg Equilibrium (HWE) within breeds and genotype frequencies were compared between breeds [31]. To estimate linkage disequilibrium (LD) between pairs of SNP we calculated r^2 [32] from the haplotype frequencies between the pair of SNP. Initial haplotype frequencies can be counted for any pair of SNP directly from the two-locus genotypes, except the double heterozygotes because they are uninformative for allelic phase. The haplotype frequencies used in calculating r^2 were obtained by iteration using the expectation maximization algorithm [31] starting with the direct haplotype frequency counts. Where LD patterns were calculated for large numbers of SNP in a small region, data were input into Haploview [33] and the LD between all pairs of SNP were calculated and displayed. The distribution of marbling scores between breeds and studies was performed using either a Poisson distribution in a generalized linear model in R (http://cran.r-project.org) to allow sub-tables to be analysed or using the log-likelihood test adjusted with the Williams correction [34,35]. Comparison to previously published or identified QTL was made to those mapped in the CBX and Limousin-Jersey experiments [13,14] as well as other QTL studies summarised at http://www.animalgenome.org/QTLdb/ [15]. To test if the SNP from GWAS were found in subsequent samples, where two or more samples were obtained from the BCPT, there was always no overlap between the samples in the GWAS and the samples in the confirmation study. For association studies we report the p-values as the negative base 10 logarithm of the p-value (-logP) for compact reporting. Note that a p-value = 0.01 is –logP = 2 and a p-value = 0.001 is –logP = 3, etc., so that a -logP of 3 represents a p-value ten times smaller than a -logP of 2. To combine p-values from more than one study, such as GWAS and confirmation samples, we use Fisher's formula $\chi^2_{2n} = -2\Sigma \ln P$ with df = 2n for n p-values [34].

4 Results and discussion

4.1 Results and discussion

4.1.1 Results and discussion

Due to the very large numbers of DNA markers associated with traits and the commercial sensitivity of those SNP, the identifiers for most of the SNP were not placed in this report.

Power of the study to detect QTL

How much of the bovine genome has been searched for QTL? After quality checks, there were 53,798 of the 54,001 SNP in the 54K Illumina BovineSNP50 chip with data for association analyses. There were 8,859 of the 9,276 SNP in the 10K Affymetrix Bovine SNP chip with data for association analyses. We wished to know the amount of the genome that was left unsearched by these GWAS, and the likelihood of the SNP set finding a large gene affecting the trait. Let us assume a gene with an effect of 2% of the residual variance, that is, a moderate to large QTL effect. Such a gene will show a –logP = 5.31 (ie, P = 0.0000049) in a sample of 1,035 animals, such as that used for the Illumina Bovine SNP50 GWAS. Such a sample will detect a gene affecting 2% of the residual variance at the significance threshold α = 0.001 with an approximate power of 90%. At the significance threshold used in our experiment (i.e., -logP = 3.0 or *P* = 0.001), such a gene would be found almost all of the time. The corollary is that almost all of the genes with such an effect, if they are part of the SNP chip, will be found. However, the SNP chips represent only a proportion of the polymorphisms in the bovine genome and consist of SNP that are DNA markers and not causative genes or variants. Even though these causative variants were not part of the SNP chip, they could be tracked through LD to the SNP in the SNP chip, if there was sufficient LD between the SNP in the chip and the causative variant. The drop in LD as measured by r^2 is proportional to the drop in power to detect the association. Multiplying the residual variance by r^2, to estimate the drop in effect size measured through LD, if r^2 = 0.75 then an effect of 2% of the residual variance would become 1.5%. A sample of 1,035 animals would detect an effect of that size at a significance threshold of α = 0.001 with a power of 50%. Using r^2 = 0.75 as a threshold, the average distance between SNP at which $r^2 \geq 0.75$ were intervals ≤ 10

kb between SNP in the bovine genome. Using this as a rough guide, the Illumina Bovine SNP50 chip covered 1.1 Gb of the 2.8 Gb of the bovine genome at that level. This meant that 39% of the bovine genome had been examined to detect genes with an effect of 2% of the residual variance with a power of 50% at a significance threshold of α = 0.001. Lowering these thresholds would generate more noise but would obviously allow smaller genes to be located. The Affymetrix 10K analysis, performed on 189 animals had a very much lower power to detect genes affecting traits due to smaller coverage of the genome and lower power to detect small to moderate QTL effects. There was therefore a much greater than 50% chance that there were many genes of moderate to large effect that had not been identified by these GWAS in these breeds. Other breeds should yield additional genes, and additional genes would also be discovered in GWAS that use SNP chips with a higher density of DNA markers.

Summary of the 50K Illumina GWAS, the Second GWAS

There were 87 SNP with significant (-logP > 3) associations to IMF. They were located across the bovine genome as shown in Figure 1. Due to the number of SNP (53,798) tested one would expect that 54 of the SNP would be significant by chance at –logP = 3. This would imply a false positive rate (FPR) of 54/87 or 62%. This meant that more than half of the SNP chosen for further testing at this level of significance would fail to be confirmed in a subsequent study. Of these associations there are six chromosomal regions on bovine chromosomes (BTA) 1, 3, 5, 14, 26 and X that showed greater than expected clustering of SNP with –logP > 3 and adjacent SNP with –logP > 1.3 (*P* < 0.05). Genotyping of SNP with –logP > 3 is reported below from an independent sample of 1,536 cattle of Angus, Hereford, and Brahman breed.

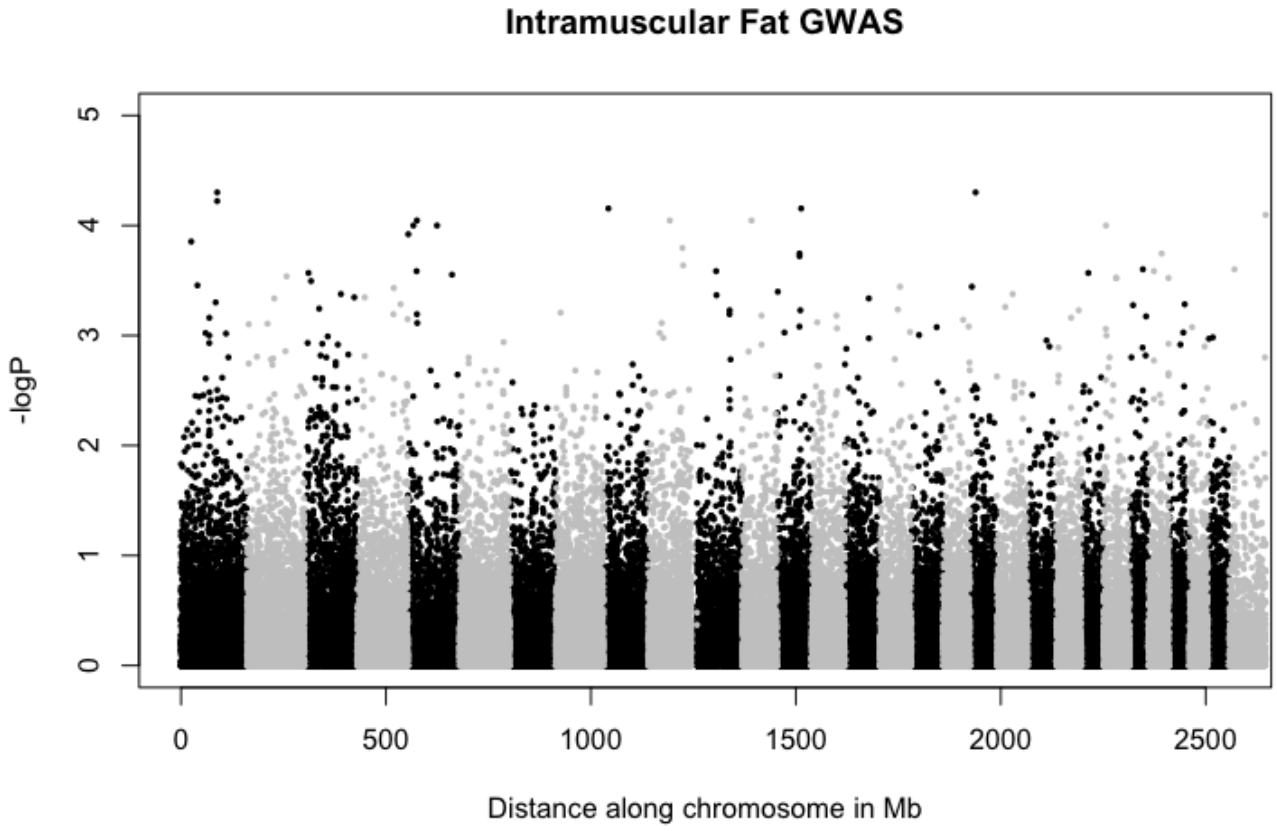

Figure 1. The location of SNP associated with IMF shown as a Manhattan plot. Odd numbered chromosomes are shown in black, even numbered chromosomes are shown in grey. Values above –logP > 3 are considered significant for further investigation.

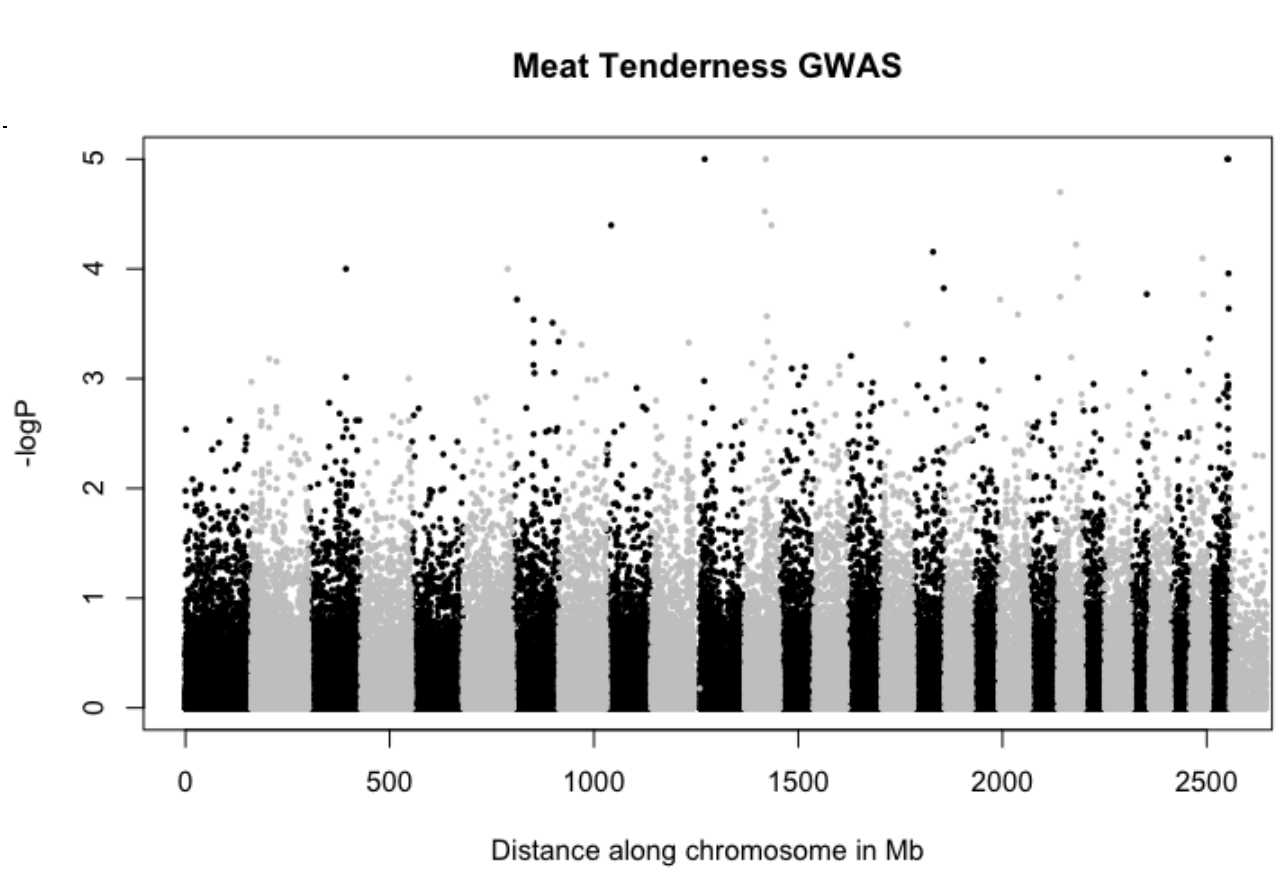

Figure 2. The location of SNP associated with LDPF shown as a Manhattan plot. Odd numbered chromosomes are shown in black, even numbered chromosomes are shown in grey. Values above –logP > 3 are considered significant for further investigation.

There were 64 SNP with significant (-logP > 3) associations to LLPF. They were located across the bovine genome as shown in Figure 2. Using the same calculation as before, this number of significant SNP implied an FPR of 84%. This meant that nearly all of the SNP chosen for further testing at this level of significance would fail to be confirmed in a subsequent study. Of these associations there are four chromosomal regions on BTA 7, 10, 22 and 29 that showed greater than expected clustering of –logP > 3 and adjacent SNP with –logP > 1.3. Of these, associations to BTA7, 10 and 29 have already been reported in other studies, only the one on BTA22 is new. Genotyping of SNP with –logP > 3 is reported below from an independent sample of 1,536 cattle of Angus, Hereford, and Brahman breed.

There were 63 SNP with significant (-logP > 3) associations to P8FAT. They were located across the bovine genome as shown in Figure 3. Using the same calculation as before, this number of significant SNP implied an FPR of 84%. This meant that nearly all of the SNP chosen for further testing at this level of significance would fail to be confirmed in a subsequent study. Of these associations BTA14 contained more than half of the associations that showed greater than expected clustering of SNP with –logP > 3 and adjacent SNP with –logP > 1.3, and BTA18, 23 and 26 also showed clusters. The clusters occurred at two locations on BTA14. We have tested SNP from one of these clusters in more detail. In the GWAS the SNP explained 2.2% of the residual variance in P8FAT. SNP from this region were highly significantly associated with P8FAT (-logP = 4.53 for the most significant) and explained 1.3% of the residual variance for P8FAT in a second independent sample of 1,533 cattle primarily of the Belmont Red and Santa Gertrudis breeds. Genotyping of other SNP with –logP > 3 is reported below from an independent sample of 1,536 cattle of Angus, Hereford, and Brahman breed.

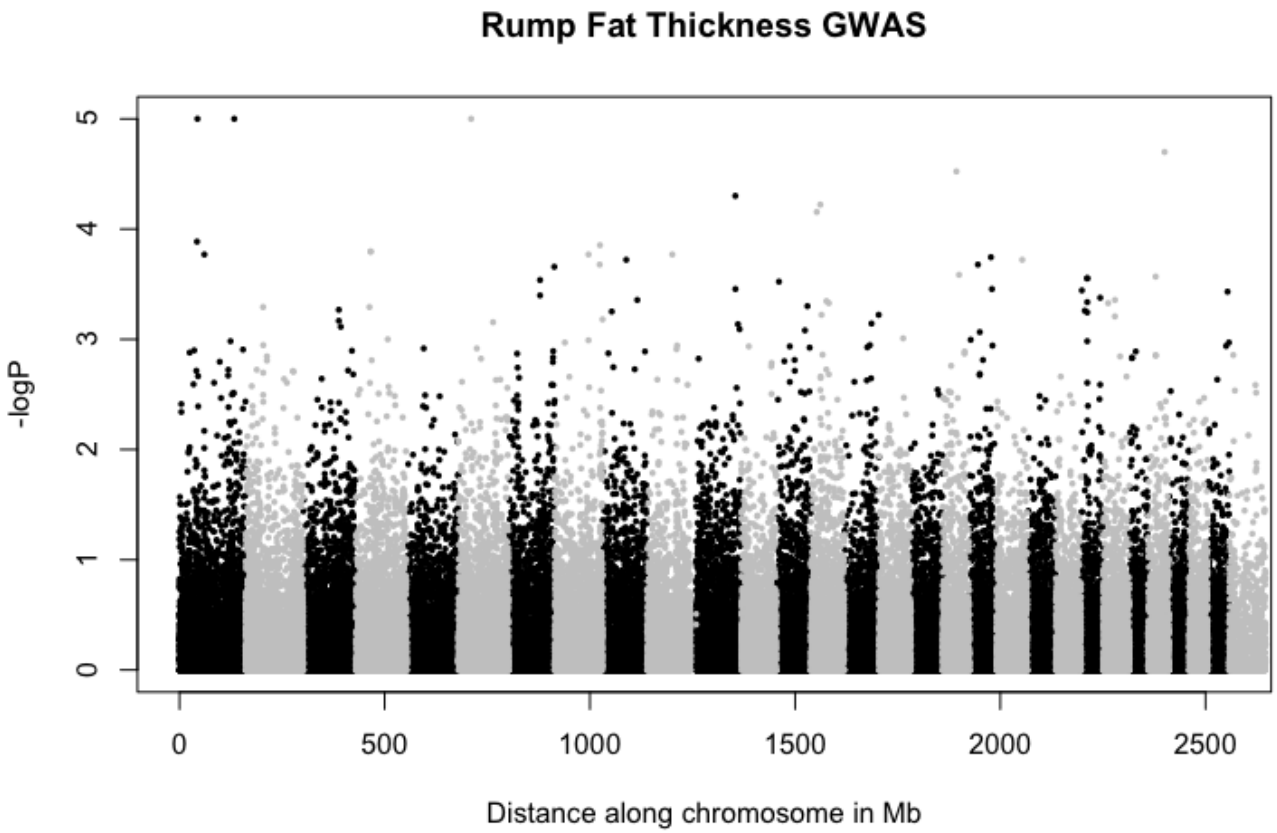

Figure 3. The location of SNP associated with P8FAT shown as a Manhattan plot. Odd numbered chromosomes are shown in black, even numbered chromosomes are shown in grey. Values above –logP > 3 are considered significant for further investigation.

There were 65 SNP with significant (-logP > 3) associations to RBY. They were located across the bovine genome as shown in Figure 4. Using the same calculation as before, this number of significant SNP implied an FPR of 84%. This meant that nearly all of the SNP chosen for further testing at this level of significance would fail to be confirmed in a subsequent study. Of these associations there are seven chromosomal regions on BTA 1, 3, 4, 10, 13, 16 and 20 that showed greater than expected clustering of SNP with –logP > 3 and adjacent SNP with –logP > 1.3.

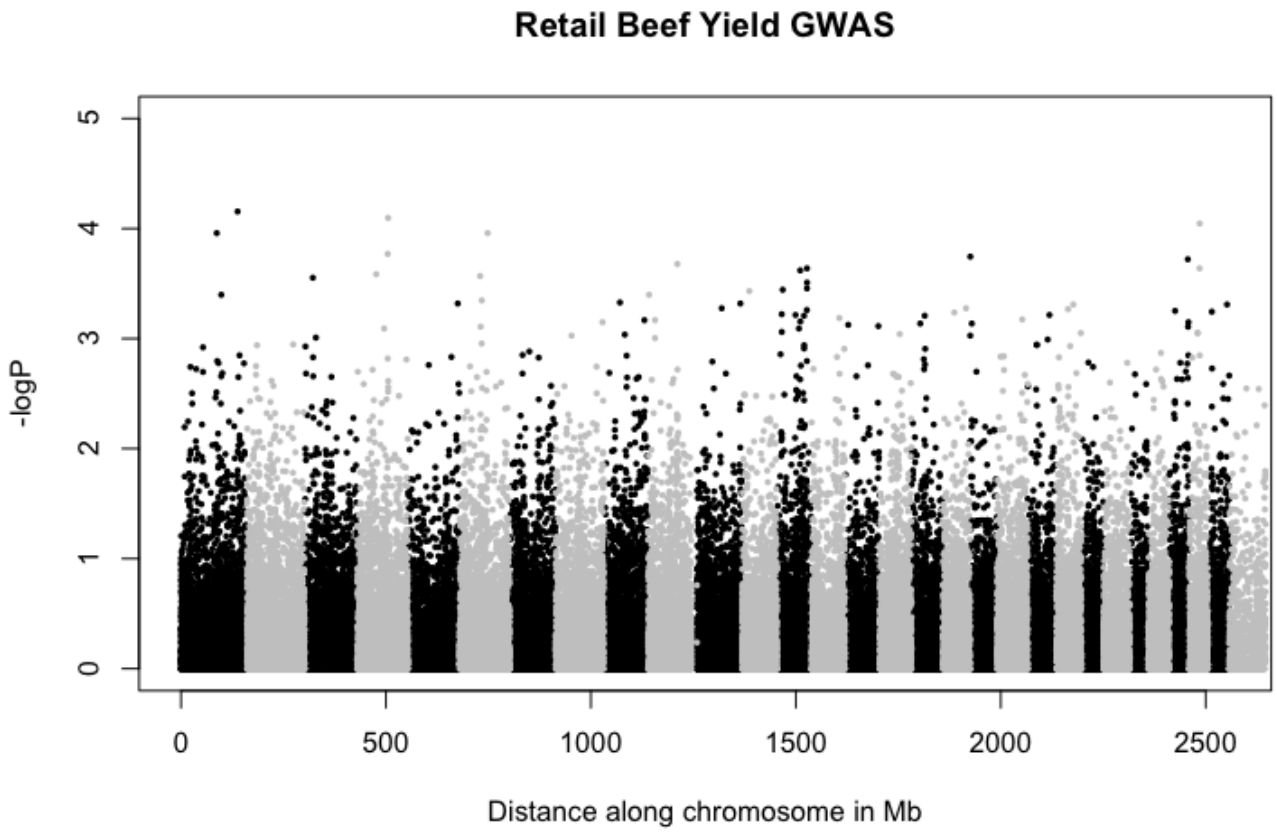

Figure 4. The location of SNP associated with RBY shown as a Manhattan plot. Odd numbered chromosomes are shown in black, even numbered chromosomes are shown in grey. Values above –logP > 3 are considered significant for further investigation.

Genetic improvement of the traits

To determine whether there was sufficient genetic variation to be used to improve the four traits, that is, to increase IMF, reduce LLPF, reduce P8FAT, and increase RBY, we plotted the allele effects for each SNP against the allele frequency of the SNP (Figure 5). If an allele that increased RBY was at a high frequency in the population then it would be of little use for future selection in this population because it would not be able to change the population in the favoured direction. There would be minimal genetic improvement. On the contrary, if the allele that increased RBY was at a low frequency then its use would be warranted because it would be useful for genetic improvement over many generations. The plots show the red or significant dots on the upper and lower edge of the allele effect by allele frequency plot (most values are in grey). Those values on the lower edge with allele frequencies < 0.5 tell the same story as those on the upper edge with allele frequencies > 0.5, and are the alleles that are useful for increasing the trait. Those on the upper edge with allele frequencies < 0.5 tell the same story as those on the lower edge with allele frequencies > 0.5, and are the alleles that are useful for decreasing the trait. Immediately it can be seen that the majority of SNP for P8FAT could be used for increasing P8FAT in the population, that is, there were few SNP with low frequencies that would decrease P8FAT. For the other traits there were approximately equal numbers of SNP for increasing or decreasing the trait. Of these traits, only P8FAT has been under strong selection using quantitative genetic methods, and is also the only one of these traits that is an obvious trait that can be observed in the live animal. This suggests that further genetic improvement in P8FAT using DNA markers will be difficult and that further Quantitative Genetic selection to decrease P8FAT in the future may be slow.

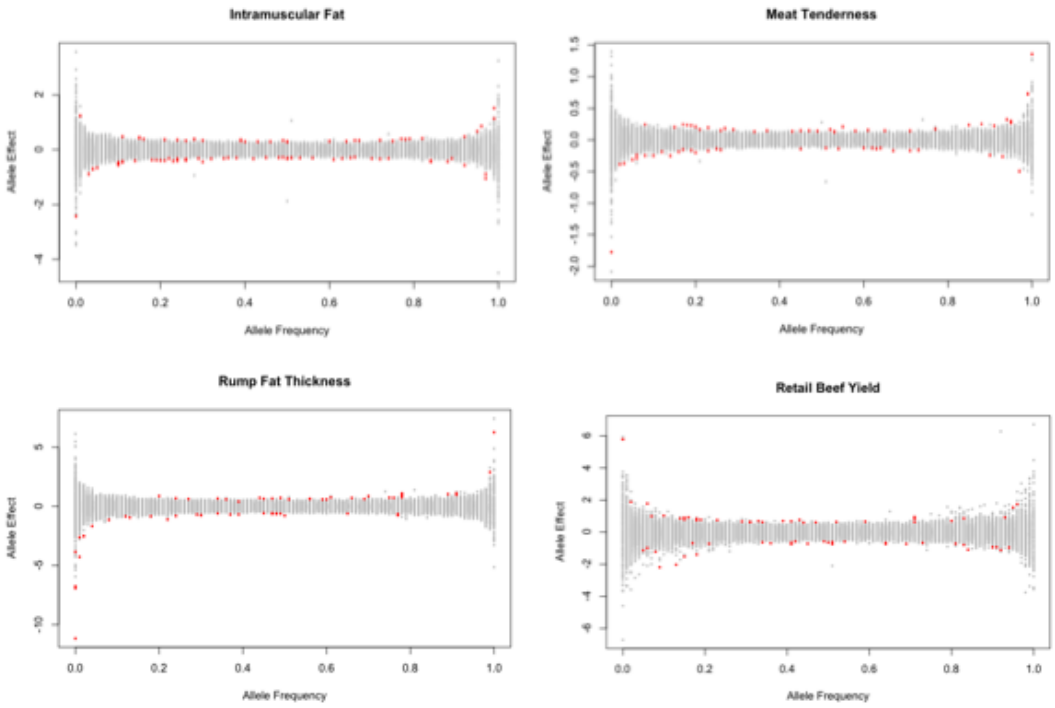

Figure 5. The relationship between allele effect and allele frequency plotted for the four traits IMF, LDPF, P8FAT, and RBY. SNP with associations of –logP > 3 are shown in red.

Correlations between the traits

The genetic correlations between these four traits in these animals have been reported [36]. IMF and P8FAT were positively correlated, they both reflect overall fatness and both were negatively correlated to RBY, as would be expected. LLPF did not show biologically significant correlations to any of the three carcass traits. Of these four traits, there were two SNP that had highly significant (-logP > 3) associations to more than one trait (Table 1). The first affected P8FAT and LLPF, where the allele that decreased the LLPF increased the P8FAT thickness. The second affected P8FAT and RBY, where the allele that decreased the P8FAT thickness increased the RBY.

Table 1. SNP with highly significant (P < 0.001) allele effects on more than one trait

P8FAT-RBY							
SNP	p_0	α_{p8fat}	SE_{p8fat}	$-logP_{p8fat}$	α_{rby}	SE_{rby}	$-logP_{rby}$
BFGL-NGS-115623	0.34	-0.596	0.177	3.09	0.626	0.178	3.32
P8FAT-LLPF							
SNP	p_0	α_{llpf}	SE_{llpf}	$-logP_{llpf}$	α_{p8fat}	SE_{p8fat}	$-logP_{p8fat}$
BTA-105204-no-rs	0.47	0.143	0.037	4.0	-0.578	0.171	3.11

p_0 is the allele frequency of allele 0, α is the allele substitution effect, SE is the standard error and –logP is the negative log of the p-value with the subscript indicating the trait of interest.

Under genomic selection, potentially all DNA markers from a SNP chip will be used even when the association does not exceed a threshold of –logP = 3. One could reconstitute the genetic correlations between traits if the DNA markers were comprehensive across the genome. Looking at all possible DNA markers it would be useful to know which SNP showed favourable results for both traits, even if these effects were small. For IMF and LLPF, the correlation of allele effects that were significant ($P < 0.05$) for both these traits was r = -0.79 (n = 165), or using all markers that have a significant effect on both traits would increase IMF and decrease LLPF. Of course, there were 2699 SNP with a significant ($P < 0.05$) effect on IMF, most of which had no effect on LLPF. These correlations overstate the relationship between SNP effects on traits because most SNP show effects only on a single trait. The other correlations (Table 2) between allele effects for SNP that were significant for two traits showed that those between LLPF and P8FAT and RBY are biologically inconsequential, because they explain little of the variance.

Table 2. Correlations between the significant (P < 0.05) allele effects for the four carcass and meat quality traits

Traits[1]				
IMF		165	169	220
LLPF	-0.79		135	166
P8FAT	0.76	0.10		233
RBY	-0.89	0.27	-0.75	

[1] Correlations are below the diagonal while sample sizes of SNP significant (P < 0.05) for both traits are above the diagonal

All other traits show high correlations between allele effects that are significant (P < 0.05) for both traits. All of the correlations are as expected. For example, the allele effects of IMF and P8FAT are positively correlated while those between RBY and the fatness traits are negatively correlated. We plotted the SNP with significant ($P < 0.05$) effects for both IMF and RBY (Figure 6).

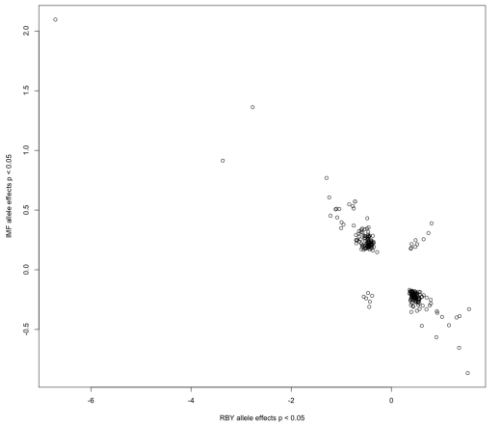

Figure 6. The plot of IMF and RBY allele effects for all SNP that have significant (P < 0.05) effects on both traits.

Confirmation study of SNP from the 50K GWAS of meat and carcass quality

After quality control of the panel of 384 SNP genotyped in the Golden Gate Assay, the confirmation study of significant (*P* < 0.001) SNP from the 50K GWAS consisted of 75 SNP for IMF, 31 SNP for P8FAT and 32 SNP for LLPF. In addition, there were 155 significant (*P* < 0.01) SNP for IMF, and a selection of candidate genes and regions of interest, totaling 335 SNP. These had been genotyped on a sample of 1,536 animals of which 1,338 animals, consisting of 655 Angus, 343 Brahman and 340 Hereford, were in the final data set after quality control. In this analysis we performed a 1-tail significance test and required that for a SNP to be considered significant that the favourable homozygote had to be the same in the 50K GWAS as in the confirmation sample (Table 3). In particular, it should be noted that a SNP was not compared to all possible traits, it was only compared to traits for which it was significant in the 50K GWAS. A small minority of loci that were significant (*P* < 0.05) in a 2-tailed test but which had the opposite favourable allele were found but these have not been reported.

Table 3. Confirmed SNP from the 50K GWAS of meat quality

SNP	Alleles	R^2	α^A	se^B	t^C
Chiller P8 fat [D]					
p001[E]					
BTA-28303-no-rs	A/G	0.0023	0.24	0.13	1.80 *
BTB-00557532	A/G	0.0055	0.54	0.20	2.73 **
BTB-00557585	A/G	0.0054	0.52	0.20	2.66 **
BTB-01530788	A/G	0.0026	0.39	0.20	1.94 *
BTB-01530836	A/G	0.0044	-0.49	0.20	2.44 **
ARS-BFGL-NGS-104268[F]	A/G	0.0049	0.39	0.15	2.64 **
IMF					
p001					
ARS-BFGL-NGS-106203	A/G	0.0064	-0.25	0.13	1.96 *
ARS-BFGL-NGS-107543	A/G	0.0040	0.12	0.07	1.74 *
ARS-BFGL-NGS-18068	A/G	0.0044	0.13	0.06	1.98 *
ARS-BFGL-NGS-34430	A/G	0.0038	-0.12	0.06	2.00 *
ARS-BFGL-NGS-75039	A/G	0.0072	-0.18	0.07	2.57 *
BTA-54207-no-rs	A/G	0.0037	0.12	0.06	1.96 *
BTA-68467-no-rs	A/G	0.0046	-0.15	0.06	2.68 **
BTB-00492076	A/G	0.0047	0.12	0.06	1.91 *
BTB-01742157	A/G	0.0055	-0.11	0.06	1.80 *
BTB-01804753	A/C	0.0065	-0.25	0.12	2.04 *
Hapmap41135-BTA-104993	A/G	0.0064	0.15	0.06	2.53 **
p01					
ARS-BFGL-NGS-103685	A/G	0.0043	-0.13	0.07	2.06 *
ARS-BFGL-NGS-106479	A/G	0.0020	0.11	0.06	1.80 *
ARS-BFGL-NGS-87787	A/C	0.0074	0.18	0.07	2.51 **
BFGL-NGS-118468	A/G	0.0119	-0.19	0.06	3.13 ***
BTA-120552-no-rs	A/T	0.0073	0.13	0.06	2.07 *
BTB-00019769	A/G	0.0056	0.24	0.08	2.88 **
BTB-00488794	A/C	0.0172	0.25	0.06	4.04 ****
BTB-00640427	A/G	0.0034	-0.16	0.06	2.41 **
BTB-00733178	A/G	0.0018	-0.10	0.06	1.66 *
BTB-01229331	A/G	0.0019	-0.12	0.06	1.93 *
Hapmap31968-BTC-056754	A/G	0.0068	0.09	0.05	1.78 *
Hapmap48049-BTA-68465	A/G	0.0016	0.10	0.06	1.71 *
Hapmap55796-rs29011172	A/T	0.0107	-0.22	0.08	2.68 **
INRA-443	A/G	0.0038	0.15	0.06	2.36 **
Chiller P8 fat tested for IMF					
p001					
ARS-BFGL-NGS-100395	A/G	0.0121	-0.23	0.08	2.99 **
LLPF					
p001					
ARS-BFGL-NGS-37441	A/G	0.0044	0.12	0.06	2.16 *
ARS-BFGL-NGS-40342	A/G	0.0014	0.08	0.05	1.71 *
ARS-BFGL-NGS-64072	A/G	0.0055	-0.16	0.07	2.32 *
BFGL-NGS-110936	A/G	0.0022	0.10	0.05	1.92 *
BTB-01033227	A/C	0.0036	0.09	0.05	2.01 *
CAPN1_1	G/C	0.0093	0.17	0.05	3.21 ***
CAST-2832	A/G	0.0150	0.22	0.05	4.05 ****
Hapmap41050-BTA-66135	A/G	0.0029	-0.09	0.05	1.79 *

* $P < 0.05$
** $P < 0.01$
*** $P < 0.001$

**** $P < 0.0001$
R^2 is obtained from the comparison of the residual sums of squares (RSS) of the model with the SNP and the RSS of the model without the SNP.
A regression estimate of the allele substitution effect, regression on number of alleles
B standard error of the regression estimate
C t-value of α/se with p-value for a 1-tail test where the favourable homozygote was the same in the 50K GWAS and the confirmation sample
D meat quality trait
E significance threshold of the SNP in the 50 K GWAS study
F SNP significant for the trait with the same favourable homozygote as the 50K GWAS but the p-value in the 50K GWAS was $P > 0.01$.

Of these SNP, there were five SNP for P8FAT thickness that were significant ($P < 0.05$) from two chromosomal regions, out of the 31 SNP significant ($P < 0.001$) of the 50K GWAS. Of these SNP, three explained $R^2 \sim 0.5\%$ of the residual variance. One these significant regions had been the most significant in the 50K GWAS in the BCPT samples. The other significant region had been the most significant region found in common between the USDA sample when compared to the BCPT sample, although it had not been the most significant region in either the USDA or the BCPT sample.

There were eight SNP for LLPF that were significant ($P < 0.05$) from three chromosomal regions, out of the 32 SNP significant ($P < 0.001$) of the 50K GWAS. Two of these SNP were already known, for CAST and CAPN1, and these explain $R^2 = 1.5\%$ and 0.9% of the residual variance respectively in this sample. One of the other SNP explains $R^2 > 0.5\%$ of the residual variance and it occurs in a third chromosomal region, which corresponds to a known QTL region affecting meat tenderness, in a gene with a plausible function, but in a region for which no diagnostic marker has yet been defined. All the other significant SNP are associated with one of these three regions.

There were 26 SNP for IMF that were significant ($P < 0.05$) from 20 chromosomal regions, 11 SNP out of the 75 significant ($P < 0.001$) SNP and 15 SNP out of the 155 significant ($P < 0.01$) SNP respectively of the 50K GWAS. Four of these chromosomal regions had more than one significant SNP. Twelve of these SNP had $R^2 > 0.5\%$ of which four SNP had $R^2 > 1\%$ of the residual variance in the confirmation sample, which is likely to be a truer indication of their real size than that found in the initial study. SNP with $R^2 > 1\%$ are of the size of CAST and CAPN1, which are the benchmark SNP that are being used in a trial EBV for meat tenderness. Based on such industry practise, these four SNP with $R^2 > 1\%$ of the residual variance would be the basis of a new set of DNA markers for IMF.

The First GWAS for meat and carcass traits

The first GWAS was performed using 189 BCPT cattle that had MQ4 measurements and they were genotyped using the 10K Affymetrix SNP chip. At the time, 10,000 SNP were thought to be of sufficient density to detect associations of SNP to traits because LD was expected to be high in cattle [37]. Indeed, initial plans for genomic selection required only a marker every centiMorgan (cM), or approximately 3,000 SNP across the whole genome [12]. The animals were unselected for the trait of MQ4. IMF, LLPF and RBY measurements were available for these individuals and so analyses for MQ4, IMF, LLPF and RBY were also performed. Such a small group of animals would have low power to detect any but the largest QTL effects, but at the time it was thought that QTL had larger effect sizes than is now known to be the case. The initial plan was that each year, as the budget allowed, an additional 200 animals would be genotyped using the same panel of SNP, until 1,000 animals were genotyped. In the interim, DNA markers taken from the GWAS would be genotyped on a separate set of animals to determine whether any genetic associations could be confirmed. After the first set of animals was analyzed it was clear in 2006 that there were as many significant results as were expected by chance for MQ4. Moreover, the estimates of LD from these and similar data showed that LD was much lower than

previously expected [38]. This implied that larger panels of SNP and larger samples of animals would be need in the GWA. We were informed that Illumina Inc. was planning to make a 6 times higher density Bovine SNP chip so use of the 10K SNP chip was discontinued. We decided to follow up a selection of the associations from the MQ4 GWAS in the hope of discovering some markers and to learn how best to analyse the data from a GWAS to identify SNP for subsequent confirmation studies.

Confirmation studies for the first GWAS of meat and carcass traits

Marbling and IMF

Initial studies of confirmation of IMF SNP was performed on the SBEF.018 cattle from an MLA funded study that had been collected at the AMH Toowoomba plant [17]. Of 48 SNP selected for confirmation, 47 proved to be polymorphic of which 3 showed significant (-logP > 1.3) regressions of marbling score on number of copies of the index allele (Table 4). Of these 3 SNP, all showed the same favourable homozygote in Angus and Shorthorn cattle. Two of the SNP had samples of n > 1,000. In these, the SNP explained approximately 0.5% of the residual phenotypic variance in marbling score. This confirmed that most QTL would be of small size.

Table 4. The significant ($P < 0.05$) associations between marbling score and SNP for IMF from the MQ4 GWAS genotyped in the combined SBEF.018 sample

SNP	n	p_0	R^2	α	SE	-logP
342743	1015	0.68	0.0061	0.118	0.047	1.90
344648	292	0.33	0.0141	0.130	0.064	1.37
354161	1049	0.79	0.0048	-0.116	0.052	1.60

p_0 is the allele frequency of the allele higher up the alphabet with a>c>g>t
R^2 is the residual phenotypic variance
α is the allele substitution effect
SE is the standard error of α
-logP is the log of the P-value from the t-test obtained from the ratio of α divided by its SE

The SNP 344648 is in the membrane bound CLEC2L gene on BTA 4 whose function is not well known. The SNP 354161 is in a non-genic region close to the CPAMD8 gene on BTA7. The SNP 342743 is in a non-genic region between the ZZEF1 and ATP2A3 genes. ATP2A3 is an intracellular pump located in the sarcoplasmic or endoplasmic reticula of cells, it catalyzes the hydrolysis of ATP and in humans is involved in the movement of Calcium ions needed for contraction of muscle. It would be relevant to energy usage and efficiency, and energy usage is related to fat stores, so this gene is a positional candidate for IMF. Knockout mutations in mice show that the ablation of the gene increased the insulin response of Pancreatic Islet cells to glucose [39], and glucose is a known factor in the development of marbling [40].

Further studies of confirmation of IMF SNP were performed on independent samples of the BCPT sample as it was thought that the SBEF.018 sample did not have enough information on pedigree and there were insufficient measurements of other traits on the animals to justify the use of that sample. In particular, the SNP were selected for effects on IMF not marbling score, so perhaps the low level of confirmation could be due to the difference between IMF and marbling score.

An additional 110 SNP, selected for effects on IMF from the MQ4 GWAS, were genotyped on approximately 1,000 BCPT animals. Of these, 4 SNP showed significant (-logP > 1.3) associations to IMF (Table 5). Of these, one was highly significantly associated and explained approximately 1.7% of the residual variance in IMF.

Table 5. The significant (*P* < 0.05) associations between IMF and SNP for IMF from the MQ4 GWAS genotyped in an independent combined BCPT sample

SNP	n	p_0	R^2	α	SE	-logP
343196	814	0.41	0.0171	0.234	0.060	3.97
343932	913	0.36	0.0089	0.177	0.062	2.32
345502	807	0.34	0.0070	0.171	0.069	1.89
349857	580	0.90	0.0083	0.363	0.155	1.71

p_0 is the allele frequency of the allele higher up the alphabet with a>c>g>t
R^2 is the residual phenotypic variance
α is the allele substitution effect
SE is the standard error of α
-logP is the log of the P-value from the t-test obtained from the ratio of α divided by its SE

The SNP 343196 is in the non-genic region between PRSS12 and NDST3 on BTA6. Neither of these is an obvious candidate. This SNP is located within 1 Megabase pair (Mb) of the FABP2 gene, a known candidate gene in fatty acid metabolism. This SNP is still significant (i.e., *P* < 0.00032) after Bonferroni correction for the number of tests performed (viz., 157 tests at α = 0.05 gives a threshold p-value = 0.00032). The amount of the residual variance explained by this SNP is moderate, approximately equal to the combined effect of the Calpain 1 and Calpastatin genes on meat tenderness. Further exploration of other SNP in this genetic region would be warranted. The SNP 343932 is in a gene poor region of BTA11 near the FANCL gene. The SNP 345502 is in the NCAM1 gene, which is the major cell adhesion molecule between neurons and between neurons and muscle. The SNP 349857 was not located to the Btau4.0 genome assembly so its annotation was not possible at present.

Meat Tenderness

Eleven SNP were chosen for associations to LLPF. Of these 11, two showed significant (-logP > 1.3) associations to LLPF in a subsequent sample (Table 6). Fewer SNP for LLPF were chosen as it was thought that initially the focus should be on IMF.

Table 6. The significant (P < 0.05) associations to LLPF in an independent combined BCPT sample

SNP	n	p_0	R^2	α	SE	-logP
343172	1,107	0.51	0.0084	-0.129	0.045	2.40
346469	1,069	0.41	0.0040	0.079	0.040	1.33

p_0 is the allele frequency of the allele higher up the alphabet with a>c>g>t
R^2 is the residual phenotypic variance
α is the allele substitution effect
SE is the standard error of α
-logP is the log of the P-value from the t-test obtained from the ratio of α divided by its SE

The SNP 343172 maps to BTA2 in the ARPC2 gene. The ARPC2 gene, actin-related protein 2/3 complex subunit 2 is an obvious positional candidate gene to effect meat tenderness. It is thought to affect the polymerization of actin in cells. The SNP 346469 is not located near a gene and is in a gene poor region of BTA11.

MQ4 scores for meat quality

MQ4 is a combined score of consumer liking including tenderness, juiciness and flavour of meat. We genotyped 42 SNP associated with MQ4 in the GWAS on an additional set of animals. There were two significant (P < 0.05) SNP associated with MQ4 (Table 7).

Table 7. The significant (P < 0.05) associations to MQ4 of SNP in an independent combined BCPT sample

SNP	n	p_0	R^2	α	SE	-logP
345261	994	0.88	0.0060	1.939	0.792	1.84
348090	521	0.47	0.0057	1.293	0.543	1.76

p_0 is the allele frequency of the allele higher up the alphabet with a>c>g>t
R^2 is the residual phenotypic variance
α is the allele substitution effect
SE is the standard error of α
-logP is the log of the P-value from the t-test obtained from the ratio of α divided by its SE

The SNP 345261 is located in the YES1 gene on BTA24. Near this gene is USP14 the ubiquitin specific peptidase 14, a cysteine protease, knock out mutations in mice show growth retardation. It is therefore a high value positional candidate gene. Further study of this gene would be warranted. The SNP 348090 is in a non-genic region next to ESRRG the estrogen-related receptor gamma.

As tenderness is a significant part of the MQ4 score, we tested these 42 SNP for effects on the objective trait of LLPF. Of these 42, two SNP were significant (-logP > 1.3) for LLPF (Table 8).

Table 8. The significant (P < 0.05) associations to LLPF of SNP to MQ4 in an independent combined BCPT sample

SNP	n	p_0	R^2	α	SE	-logP
353731	719	0.81	0.0043	-0.187	0.083	1.60
353813	1,043	0.60	0.0039	0.126	0.058	1.50

p_0 is the allele frequency of the allele higher up the alphabet with a>c>g>t
R^2 is the residual phenotypic variance
α is the allele substitution effect
SE is the standard error of α
-logP is the log of the P-value from the t-test obtained from the ratio of α divided by its SE

Neither of these SNP is in the Btau4.0 assembly so they could not be annotated.

Confirmation of markers obtained from the 10K GWAS using the 50K Illumina GWAS

To determine whether it was worth while to continue to use the 10K GWAS as a source of SNP for further analyses, we compared the results from the 10K Affymetrix GWAS to the 50K Affymetrix GWAS for the trait IMF. Of the 1,035 BCPT animals in the 50K analyses, 932 animals had not been used in the 10K GWAS. The 50K GWAS data set was re-analysed using those 932 animals using the same methods for the trait IMF as before. There were 2,326 SNP in common between the 10K and 50K GWAS. Of these, there were 156 significant at –logP > 1.3 for IMF in the 10K GWAS of which 21 were significant at –logP > 2. Of these 156 significant SNP, there were 11 that were also significant at –logP > 1.3 in the 50K GWAS. These 11 SNP (Table 9) represent 0.47% of the 2,326 SNP, of which one would expect 0.25% to be significant at –logP > 1.3 by chance. This represents a 53% FPR. There were two SNP with –logP > 2 in both GWAS, which represents an FPR of 12%. Given the number of SNP identified positive in both GWAS, further use of the 10K GWAS as a source of SNP for confirmation was discontinued, and further confirmation was then done using the 50K GWAS as a source (cf. above).

Table 9. SNP significantly (P < 0.05) associated to IMF in the 10K and 50K GWAS

SNP	Chr	Position	α_{10}	SE_{10}	-logP	α_{50}	SE_{50}	-logP	R^2_{50}
rs29021692	2	60071311	-0.341	0.123	2.20	-0.187	0.079	1.74	0.0061
ss46526426	5	101979582	0.289	0.144	1.32	0.196	0.081	1.80	0.0063
rs29013237	6	77268231	0.627	0.264	1.73	0.476	0.221	1.49	0.0049
rs29013992	6	78855304	-0.702	0.335	1.41	-0.492	0.175	2.31	0.0088
rs29015344	10	8522421	0.303	0.149	1.35	-0.224	0.086	2.04	0.0075
rs29019544	13	24711087	0.808	0.384	1.43	-0.814	0.330	1.86	0.0066
rs29026820	14	17808996	0.557	0.180	2.62	-0.223	0.084	2.09	0.0078
rs29014510	15	31343525	0.295	0.150	1.29	-0.227	0.079	2.36	0.0091
rs29020487	20	60700406	0.513	0.145	3.25	0.233	0.080	2.43	0.0094
rs29026956	21	51002612	0.518	0.211	1.81	-0.219	0.102	1.51	0.0050
rs29024708	29	44468846	-0.487	0.190	1.94	-0.334	0.104	2.86	0.0115

SNP dbSNP identifier of the single nucleotide polymorphism
Chr is the bovine chromosome
Position is the location in bp along the chromosome in the Btau4.0 assembly
α_{10} the allele effect in the 10K Affymetrix GWAS
SE_{10} the Standard error of α_{10}
-logP the negative logarthim of the P-value
α_{50} the allele effect in the 50K Illumina GWAS without the animals from the 10K GWAS
SE_{50} the Standard error of α_{50}
R^2_{50} the residual variance explained by the SNP in the 50K GWAS, the estimate at 847 degrees of freedom

Six of the 11 SNP show greater significance in the 50K GWAS than the 10K GWAS and are therefore more likely to be correct, because the 50K GWAS has more power and a greater sample size. Note that the coding of alleles in the 50K GWAS was different to that of the 10K GWAS so it is not possible to determine from Table 9 whether the same homozygote was favourable in both studies. Of the 11 SNP, rs29021692 is in a gene poor region. ss46526426 is in the MGP gene, which is involved in calcification of cartilage. rs29013237 and rs29013992 are in a megabase scale gene poor region. rs29015344 is located between TBCA and ATXN3. rs29019544 is in a large unnamed gene. rs29026820 is adjacent to the HAS2 gene, the hyaluronan synthase 2 gene. Hyaluronan is a constituent of the extracellular matrix and is involved in wound healing and blood vessel infiltration into tissues. HAS2 is a positional candidate gene for marbling and IMF. Near HAS2 is the CEBPD gene, which is part of a family of transcription factors and CEBPD is known to affect adipocyte differentiation. rs29014510 is located between SORL1 and UBASH3B. UBASH3B is a ubiquitin associated protein expressed in skeletal muscle so it is interesting but not a formal candidate gene. rs29020487 is adjacent to the FBXL7 gene. The FBXL7 gene is a member of a ubiquitin ligase complex. rs29026956 is between two unknown genes, adjacent to FBXO33, a member of a protein-ubiquitin complex. rs29024708 is located in the NRXN2 gene, which is predominantly expressed in the brain and is involved in normal neurotransmitter release.

Due to the fact that HAS2 and CEBPD are known candidate genes but the 10K GWAS showed higher significance than the 50K GWAS, we checked whether other SNP in the 50K GWAS from the same genetic region of BTA14 were more significant than rs29026820 in the 50K GWAS. There were three SNP including rs29026820 associated with IMF adjacent to the HAS2 gene with –logP > 2. The most significant was Hapmap32629-BTA-156370 located at BTA14:17898094 with –logP = 2.72.

This process could also be performed with other traits but the value of this for determining whether SNP were ready for commercialization is moot.

Confirmation of international markers

Part of the remit of the project was to take SNP that had been identified previously in other studies, preferably in international studies, and test whether the SNP were suitable for use in Australian cattle. During this project, SNP for GH1, CPE, CEBPA, FABP4, ADIPOQ and RORC were tested to confirm their effects in Australian cattle. We found that GH1 had a small effect on marbling and P8FAT and this work was published [41]. We found that the published SNP in FABP4 for marbling was not significant in our samples. We performed additional sequencing to identify a causative mutation. We found a splice site variant, which we tested and which appeared to be associated to IMF [28]. Our tests of the ADIPOQ SNP failed to find an effect on IMF and other SNP adjacent to ADIPOQ also failed to show an association. Testing of the CPE and CEBPA variants showed limited association of CPE in our cattle, but the effect showed different favourable alleles to the published research. This work has been published and we recommended that while further work on CPE may be warranted, the SNP as they stand were not suitable for commercial use in Australia [42]. Finally, we tested the RORC gene and found an effect of the same size in the BCPT sample for both marbling score and IMF as had been described previously for marbling score in the SBEF.018 sample. This work has been accepted for publication [43]. None of these SNP was of moderate to large effect and their use could only be justified as part of a panel of DNA markers.

Status of markers for commercial use

For commercial use, these results lead us to the following conclusions. Firstly, there were many of SNP that were confirmed in samples of n > 1,000 cattle for several traits. Those that have larger R^2, i.e., with values around 1% or more in samples of around 1,000 animals or more in the tables above, are likely to be found in most other studies. This comment is based on the prior experience of association mapping with genes of that size of effect. Of the SNP in the tables there were five for IMF and one for LLPF derived from the 10K Affymetrix as well as the 50K GWAS with R^2 > 1% of the residual variance in a sample of n > 1,000 animals. There were as many as that again with R^2 > 0.9% of the residual variance in a sample of n > 800 animals. In addition, there was a SNP for P8FAT from the 50K GWAS that explained 2.2% of the residual variance in the GWAS. When tested in an independent sample of 1,533 BCPT cattle it was found to have an effect of 1.3% of the residual variance, and a second independent sample of 1,338 BCPT cattle was also significant for loci from this gene. This is an unbiased estimate of the size of effect, and makes this the largest QTL affecting a meat or carcass quality trait, apart from the double muscling mutation, discovered so far.

Secondly, for IMF there are four confirmed SNP from the 50K GWAS each explaining more than 1% of the residual variance. This residual variance that is explained is the effect of the genetic variation at the SNP in terms of the variance of the phenotype after accounting for the fixed effects of management and the genetic effects of co-ancestry. It is therefore less than the full genetic variance of the SNP. This is sufficient variance with which to generate a trial EBV for IMF that includes DNA markers, because the trail EBV for LLPF uses fewer genes of the same size of effect. Additional analyses will need to be performed to allow the realization of such a trail EBV, but sufficient DNA markers have been identified to begin this process. Commercialization of this information will involve the discussion of these results with commercial genetic testers, who will then need to genotype these SNP on their resources as well as develop assays for these SNP so that they can genotype them on their machines. These are time consuming activities and can easily consume a year or more of time.

Thirdly, of the candidate gene studies that we confirmed, the SNP for GH1, RORC, and FABP4, could be added to currently available DNA marker panels for IMF.

5 Success in achieving objectives

5.1 Success in achieving objectives

5.1.1 Success in achieving objectives

Meeting the original stated objectives
The original stated objectives were to identify approximately 15 DNA markers for a total of three meat and carcass quality traits by 2011 and to have, by now commercialized three markers per year totaling nine markers. Given our results so far, we have identified 68 DNA markers that are significant ($P < 0.05$) in a second study, of which 28 are highly significant ($P < 0.01$). Three of these markers represent the same gene. So in that technical sense we have met the original stated objectives for gene discovery.

We have not commercialized any of these DNA markers but we appear to have reached the Beef CRCs revised standard, namely that the DNA markers in the first panel must exceed 4% of the genetic variance in a large confirmation sample of animals for the first set of markers (the final set will explain more than 15% of the genetic variance associated with the trait). To commercialize these results, additional analyses of these data need to be performed, in particular, analyses of the four SNP with $R^2 > 1\%$ for IMF needs to be evaluated as a panel. This is an important next step in the process. These four SNP also need to be genotyped on all animals that have IMF measurements to determine the most accurate estimates of their effects prior to being included in a trial marker assisted EBV.

6 Impact on meat and livestock industry – now and in five years time

6.1 Impact on meat and livestock industry – now and in five years time

6.1.1 Impact on meat and livestock industry – now and in five years time

The results of this study will help to realize the promised genetic improvements from molecular genetic technologies. At present, there are still few DNA markers that have solid, large effects on a trait. Markers of large and consistent effect are desperately needed for other meat and carcass quality traits, such as IMF. In this study we found new examples of such DNA markers, with several DNA markers for fatness traits that are as large or larger than the effects of Calpain and Calpastatin on meat tenderness. These DNA markers will have benefits for the breeding of cattle as soon as they are released. Furthermore, with DNA markers that are this large, we have noticed additional usages. The DNA markers in Calpain and Calpastatin have begun to reach beyond simple genetic tests and have been included in a trial Estimated Breeding Value for meat tenderness. In addition, research is on-going to include these DNA markers into the Meat Standards Australia grading, and this is significant because the MSA grading scheme is concerned with using processing and management to predict meat quality and meat tenderness irrespective of genetic differences. As research with Calpastatin in particular has shown, within 5 years of the release of markers of large effect on fatness traits, these markers will then be drawn into non-genetic approaches to managing meat and carcass quality. They may not all be used in the MSA grading scheme, but the gene identity may lead to other approaches for managing meat and carcass quality, such as nutritional supplements or special management regimes for animals with favourable genotypes at these DNA markers.

7 Conclusions and recommendations

7.1 Conclusions and recommendations

7.1.1 Conclusions and recommendations

This study discovered a set of DNA markers for meat and carcass quality and so far has confirmed the effect of some of these. This leads to a set of conclusions and recommendations.

1. Once a set of DNA markers has been confirmed, the combined effect of such a group of markers needs to be estimated as a panel, because the combined effect will not be the simple sum of the individual effects.
2. Such a panel of DNA markers needs to be tested in a third-party experiment to determine the efficacy of the panel – this would be a trial of the panel not a trial of the individual DNA markers.
3. A successful trial of the panel of DNA markers should lead to the creation of trial-EBV for these traits.
4. To search for other DNA markers affecting these traits, a GWAS study that increased the coverage of the bovine genome will be needed. We have estimated that less than half of the bovine genome has been successfully searched for genes of moderate to large effect on these traits. This is because the DNA markers we used covered the entire bovine genome with a minimum gap between DNA markers of 20 kb and an average spacing of 51.9 kb. There is space in those gaps for genes to be missed. To identify genes of smaller effect, which may explain a substantial proportion of the residual variance, higher densities of markers and larger sample sizes in the GWAS will be needed – the power of the experiment increases with the square of the sample size.
5. For the larger QTL, those with effects of approximately 1% of the residual variance or more, it should be feasible to discover the causative mutations. Discovering causative mutations leads to the fastest genetic change with lower error rates than merely using DNA markers.
6. DNA markers with confirmed effects on meat and carcass quality traits should be tested for any effects on other traits, especially feed efficiency, adaptation to the environment, and fertility. We note that the largest effect we discovered on rump fat thickness also effects age at puberty in composite cattle. Other such effects may be discovered for these DNA markers.
7. We recommend that these markers be tested in the Beef Information Nucleus sample when that become available.

8 Bibliography

1. Grobet L, Martin LJR, Poncelet D, Pirottin D, Brouwers B, et al. (1997) A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. Nature Genetics 17: 71-74.
2. Grisart B, Coppieters W, Farnir F, Karim L, Ford C, et al. (2002) Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Research 12: 222-231.
3. Blott S, Kim JJ, Moisio S, Schmidt-Kuntzel A, Cornet A, et al. (2003) Molecular dissection of a quantitative trait locus: A phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics 163: 253-266.
4. Viitala S, Szyda J, Blott S, Schulman N, Lidauer M, et al. (2006) The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle. Genetics 173: 2151-2164.
5. Cohen-Zinder M, Seroussi E, Larkin DM, Loor JJ, Everts-van der Wind A, et al. (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Research 15: 936-944.
6. Page BT, Casas E, Heaton MP, Cullen NG, Hyndman DL, et al. (2002) Evaluation of single-nucleotide polymorphisms in CAPN1 for association with meat tenderness in cattle. Journal of Animal Science 80: 3077-3085.
7. White SN, Casas E, Wheeler TL, Shackelford SD, Koohmaraie M, et al. (2005) A new single nucleotide polymorphism in CAPN1 extends the current tenderness marker test to include cattle of *Bos indicus*, *Bos taurus*, and crossbred descent. Journal of Animal Science 83: 2001-2008.
8. Barendse WJ (2002) DNA markers for meat tenderness. WO02064820. pp. 1-88.
9. Barendse W, Harrison BE, Hawken RJ, Ferguson DM, Thompson JM, et al. (2007) Epistasis between calpain 1 and its inhibitor calpastatin within breeds of cattle. Genetics 176: 2601-2610.
10. Barendse W (2008) Genetic based diagnostic tools for predicting meat quality. In: Kerry JP, Ledward D, editors. Improving the sensory and nutritional quality of fresh meat. Cambridge, England: Woodhead Publishing Limited and CRC Press LLC. pp. 292-317.
11. Barendse W (2005) The transition from quantitative trait loci to diagnostic test in cattle and other livestock. Australian Journal of Experimental Agriculture 45: 831-836.
12. Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-1829.
13. Davis GP, Hetzel DJS, Corbet NJ, Scacheri S, Lowden S, et al. (1998) The mapping of quantitative trait loci for birth weight in a tropical beef herd. 6th World Congress of Genetics Applied to Livestock Production 26: 441-444.
14. Morris CA, Cullen NG, Hickey SM, Dobbie PM, Veenvliet BA, et al. (2006) Genotypic effects of calpain 1 and calpastatin on the tenderness of cooked *m. longissimus dorsi* steaks from Jersey x Limousin, Angus and Hereford-cross cattle. Animal Genetics 37: 411-414.
15. Hu Z-L, Reecy J (2007) Animal QTLdb: beyond a repository. Mammalian Genome 18: 1-4.
16. Upton W, Burrow HM, Dundon A, Robinson DL, Farrell EB (2001) CRC breeding program design, measurements and database: methods that underpin CRC research results. Australian Journal of Experimental Agriculture 41: 943-952.
17. Barendse W, Bunch R, Thomas M, Armitage S, Baud S, et al. (2004) The TG5 thyroglobulin gene test for a marbling quantitative trait loci evaluated in feedlot cattle. Australian Journal of Experimental Agriculture 44: 669-674.
18. Hardenbol P, Yu FL, Belmont J, MacKenzie J, Bruckner C, et al. (2005) Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. Genome Research 15: 269-275.

19. Hawken RJ, Barris WC, McWilliam SM, Dalrymple BP (2004) An interactive bovine in silico SNP database (IBISS). Mammalian Genome 15: 819-827.
20. Barendse W, Harrison BE, Bunch RJ, Thomas MB, Turner LB (2009) Genome wide signatures of positive selection: The comparison of independent samples and identification of regions associated to traits. BMC Genomics 10: 178.
21. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, et al. (2009) Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLoS ONE 4: e5350.
22. Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, et al. (2009) Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. Science 324: 528-532.
23. Van Tassell CP, Smith TPL, Matukumalli LK, Taylor JF, Schnabel RD, et al. (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature Methods 5: 247-252.
24. Livak KJ (2003) SNP genotyping by the 5'-nuclease reaction. In: Kwok P-Y, editor. Methods in Molecular Biology. Totowa New Jersey: Humana Press. pp. 129-147.
25. Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, et al. (2003) Highly parallel SNP genotyping. Cold Spring Harbor Symposia on Quantitative Biology 68: 69-78.
26. Fan JB, Chee MS, Gunderson KL (2006) Highly parallel genomic assays. Nature Reviews Genetics 7: 632-644.
27. Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R (2002) ASReml User Guide Release 1.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.
28. Barendse W, Bunch RJ, Thomas MB, Harrison BE (2009) A splice site single nucleotide polymorphism of the *fatty acid binding protein 4* gene appears to be associated with intramuscular fat deposition in longissimus muscle in Australian cattle. Animal Genetics 40: 10.1111/j.1365-2052.2009.01913.x.
29. Reverter A, Johnston DJ, Perry D, Goddard ME, Burrow HM (2003) Genetic and phenotypic characterisation of animal, carcass, and meat quality traits from temperate and tropically adapted beef breeds. 2. Abattoir carcass traits. Australian Journal of Agricultural Research 54: 119-134.
30. Snedecor GW, Cochran WG (1967) Statistical Methods. Ames, Iowa, USA: The Iowa State University Press.
31. Weir BS (1996) Genetic Data Analysis II. Sunderland, MA: Sinauer Associates, Inc. 1-445 p.
32. Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theoretical and Applied Genetics 38: 226-231.
33. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263-265.
34. Sokal RR, Rohlf FJ (1981) Biometry: W.H. Freeman & Co, San Francisco.
35. Venables WN, Ripley BD (2000) Modern Applied Statistics with S-PLUS. New York: Springer Verlag New York, Inc.
36. Reverter A, Johnston DJ, Ferguson DM, Perry D, Goddard ME, et al. (2003) Genetic and phenotypic characterisation of animal, carcass, and meat quality traits from temperate and tropically adapted beef breeds. 4. Correlations among animal, carcass, and meat quality traits. Australian Journal of Agricultural Research 54: 149-158.
37. Farnir F, Coppieters W, Arranz JJ, Berzi P, Cambisano N, et al. (2000) Extensive genome-wide linkage disequilibrium in cattle. Genome Research 10: 220-227.
38. Barendse W, Reverter A, Bunch RJ, Harrison BE, Barris W, et al. (2007) A validated whole genome association study of efficient food conversion. Genetics 176: 1893-1905.
39. Arredouani A, Guiot Y, Jonas J-C, Liu LH, Nenquin M, et al. (2002) SERCA3 ablation does not impair insulin secretion but suggests distinct roles of different sarcoendoplasmic reticulum Ca2+ pumps for Ca2+ homeostasis in pancreatic β-cells. Diabetes 51: 3245-3253.
40. Pethick DW, Harper GS, Oddy VH (2004) Growth, development and nutritional manipulation of marbling in cattle: a review. Australian Journal of Experimental Agriculture 44: 705-715.

41. Barendse W, Bunch RJ, Harrison BE, Thomas MB (2006) The growth hormone 1 GH1:c.457C>G mutation is associated with relative fat distribution in intra-muscular and rump fat in a large sample of Australian feedlot cattle. Animal Genetics 37: 211-214.
42. Barendse W, Bunch RJ, Harrison BE (2009) Variation at CPE but not CEBPA appears to be associated with intramuscular fat deposition in the longissimus muscle of cattle. Animal Production Science 49: 558-562.
43. Barendse W, Bunch RJ, Harrison BE The effect of variation at the retinoic acid receptor-related orphan receptor C gene on intramuscular fat percent and marbling score in Australian cattle. Journal of Animal Science in the press.