Copyright © 2007 by the Genetics Society of America
DOI: 10.1534/genetics.107.072637

A Validated Whole-Genome Association Study of Efficient Food Conversion in Cattle

W. Barendse,[1] A. Reverter, R. J. Bunch, B. E. Harrison, W. Barris and M. B. Thomas

CSIRO Livestock Industries, Queensland Bioscience Precinct, St. Lucia 4067, Australia

Manuscript received February 26, 2007
Accepted for publication April 17, 2007

ABSTRACT

The genetic factors that contribute to efficient food conversion are largely unknown. Several physiological systems are likely to be important, including basal metabolic rate, the generation of ATP, the regulation of growth and development, and the homeostatic control of body mass. Using whole-genome association, we found that DNA variants in or near proteins contributing to the background use of energy of the cell were 10 times as common as those affecting appetite and body-mass homeostasis. In addition, there was a genic contribution from the extracellular matrix and tissue structure, suggesting a trade-off between efficiency and tissue construction. Nevertheless, the largest group consisted of those involved in gene regulation or control of the phenotype. We found that the distribution of micro-RNA motifs was significantly different for the genetic variants associated with residual feed intake than for the genetic variants in total, although the distribution of promoter sequence motifs was not different. This suggests that certain subsets of micro-RNA are more important for the regulation of this trait. Successful validation depended on the sign of the allelic association in different populations rather than on the strength of the initial association or its size of effect.

FOOD use and food conversion efficiency are important for the survival and selective advantage of an animal and underpin its ability to compete. Food is a limiting resource and as such competition for its use is a fundamental driver of selection, both natural and artificial (Darwin 1859; Byers 2000). Food conversion efficiency is a critical aspect of animal systems due to its impact on the flow of nutrients (Cahaner and Leenstra 1992; Kohn *et al.* 1997; Veerkamp 1998; Bjornsson *et al.* 2001; Robinson and Oddy 2004) and it has a major role in the overall efficiency of the enterprise. Food consumption is under physiological control (Leibowitz and Alexander 1998; Cummings *et al.* 2001) and there is genetic variability within populations for food conversion efficiency (Robinson and Oddy 2004), with a clear trade-off between such efficiency and other traits of an individual (Veerkamp 1998). In humans, food intake above metabolic needs is linked to obesity, and the homeostasis of body weight appears to be under partial genetic control (Neel *et al.* 1998).

Most of the studies performed in this area to date have addressed the homeostatic control of body mass, due to the difficulty of accurately measuring food intake in populations. Most studies are of the genes that code for hormones and their receptors that regulate these systems (Leibowitz and Alexander 1998; Sakurai *et al.* 1998; Walder *et al.* 1998; Cummings *et al.* 2001; Tanaka *et al.* 2003), such as leptin, serotonin, orexin, ghrelin, and insulin. Disruption of some of these genes leads to large changes in levels of obesity in several species. Measuring food intake with any degree of accuracy is highly complex and requires automated measurement of individual food intake, and this is difficult to do for extended periods except in some experimental animals and in livestock.

In the absence of information on overall distribution of genetic effects, one may speculate that there should be a role for many DNA variants in genes that affect processes that use energy constantly. These processes include basal metabolic rate, the control of appetite, the generation of ATP, the regulation of growth and development, and the homeostatic control of body mass. Nevertheless, the relative frequency of DNA variants in each of these systems is unknown.

Furthermore, the ideas about the kinds of genetic variation that will be responsible for quantitative trait loci (QTL) are changing rapidly so that not all the DNA variants will be of genic origin. There is debate about the proportion of the DNA variants affecting the phenotype that would come from genes (Holliday 1990; Kashi *et al.* 1997; Mattick 2001; Bejerano *et al.* 2004) compared to that from nongenic DNA. In genes, these are not only likely to be changes in the coding sequence

[1] *Corresponding author:* CSIRO Livestock Industries, Queensland Bioscience Precinct 306 Carmody Rd., St. Lucia 4067, Australia. E-mail: bill.barendse@csiro.au

but also likely to be due to differences in regulatory sequences (Xie *et al.* 2005; Clop *et al.* 2006). More than half of the ultra-conserved DNA in a mammalian species is not associated with genes, suggesting that a large proportion of the DNA elements that affect the phenotype will be nongenic; here, genic refers not only to coding sequences but also to DNA close to coding sequences that may affect the function or regulation of that coding sequence.

To identify some of the DNA variation associated with residual food intake (RFI), a measure of food conversion efficiency that includes metabolic weight, we scanned the bovine genome for associations using the whole-genome association (WGA) methodology and then confirmed a subset of the associations. Food intake can be measured accurately and conveniently in cattle for an extended time and more genetically diverse samples can be obtained for livestock than for rodent models. RFI is considered to be the preferred method of measuring food conversion efficiency since it takes into consideration the daily weight gain of the animal as well as adjusts for the metabolic weight of the individual. WGA may be applied speculatively in livestock with panels of DNA variants in the tens of thousands due to high levels of linkage disequilibrium (LD) within breeds.

MATERIALS AND METHODS

Cattle samples: The cattle DNA sample and the methods of measurement of phenotypes, including RFI, have been reported previously (Upton *et al.* 2001; Robinson and Oddy 2004). DNA and feed intake and daily gain and body-weight measurements were available for 1472 animals of seven breeds: Angus, Brahman, Belmont Red, Hereford, Murray Grey, Santa Gertrudis, and Shorthorn, representing 308 sires, 32 herds of origin across the seven breeds, and 42 feed groups. The Angus, Hereford, Murray Grey, and Shorthorn are European taurine breeds adapted to temperate environments while the Brahman, Belmont Red, and Santa Gertrudis are breeds adapted to tropical environments; the Brahman was used in the formation of the Santa Gertrudis. A subset of this sample was used for the WGA, and the subset was selected to improve the power to detect associations in a small sample. Animals were selected within the systematic or fixed effects of the experiment, so pairs of animals were matched within sex, within feed group, within herd of origin, and within market destination (domestic, Korea, and Japan); sex, feed group, herd of origin, and market destination form the contemporary group. Common sires were avoided to ensure a diversity of sires. Pairs were chosen to be discordant on phenotype, so that one member of the pair fell within the top 200 while the other member fell within the bottom 200; the pairs are not merely members of the top 200 *vs.* those of the bottom 200. This yielded a final sample for WGA of 189, consisting of 41 Angus, 21 Brahman, 24 Belmont Red, 28 Hereford, 20 Murray Grey, 28 Santa Gertrudis, and 27 Shorthorn animals: the odd numbers are due to the constraint of a 2×96-well format; three animals were unmatched, and three animals were genotyped twice, anonymously, to assess the genotyping error rate, which had a 99.92% accuracy. The sample represents 142 sires with a range of one to four offspring/sire and a median of one offspring/sire, 32 herds with a median of five steers/herd, and 37 feed groups with a median of four offspring/feed group.

The feed intake trait was RFI (Robinson and Oddy 2004) measured in kilograms per day of feed; it is the deviation between the actual feed intake and the predicted feed intake of the animal, given its average daily weight gain and metabolic weight. The heritability of RFI was estimated using the five-generation pedigree data for all the animals. The RFI values for the sample range from −3.40 to 3.81 with a mean of 0.07 and standard deviation of 1.32. The total sample from which the extremes were drawn had a mean of 0.00 and a standard deviation of 0.75.

Single nucleotide polymorphisms: Very high LD has been reported for livestock species in general (Farnir *et al.* 2000; McRae *et al.* 2002; Tenesa *et al.* 2003; Nsengimana *et al.* 2004), and high levels of LD have been reported to extend out for >1 Mbp in dairy cattle breeds (Farnir *et al.* 2000; Khatkar *et al.* 2006). As few as 3000 evenly spaced single nucleotide polymorphisms (SNPs) could therefore be sufficient for WGA to cover the genome. However, estimates of LD calculated across breeds in beef cattle show much lower LD (Barendse *et al.* 2007); to reduce LD and therefore improve the precision of gene mapping, seven breeds were used in this study.

The WGA was performed using the MegAllele Genotyping Bovine 10K SNP Panel (Hardenbol *et al.* 2005), a fully described set of SNP, by ParAllele, on an Affymetrix GeneChip Scanner 3000, yielding an average spacing of 325 kbp between SNP. The Baylor College of Medicine (BCM) SNP or the Single Nucleotide Polymorphism database (dbSNP) identifiers for the top 161 SNP are given in supplemental Table 1 at http://www.genetics.org/supplemental/. The SNP sequences can be downloaded from the Baylor College of Medicine website at ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Btaurus/snp/Btau20050310/, and the BCM identifiers can be used in dbSNP. One of the overall objectives of validation is to obtain estimates of size of effect and genetic variation due to QTL that are as accurate and as unbiased as possible (*e.g.*, Utz *et al.* 2000). In this study, the initial associations were made on animals of extreme phenotype so that a small sample will have a large power to detect associations, but using regression methods, albeit with permutation tests, will cause the allele effects to be overestimated (Lander and Botstein 1989), and so it would be important to validate the significance and size of effects in the full sample. We validated 44 of the SNPs from the WGA on the whole sample, using the SNPlex method (ABI 3130XL, ABI, Foster City, CA) in our laboratories or the Taqman assay as reported previously (Barendse *et al.* 2006, 2007) to determine whether the WGA had identified valid associations, whether the effects were similar in different breeds, and to more accurately measure the size of effect associated with each SNP. The estimates of the allele effects obtained in the WGA were compared to those in the validation sample using linear regression to determine the amount of deflation in the estimates of the additive effects. The 34 significant loci for RFI had a range of significances from $P < 0.01$ to $P < 0.0001$, partly to determine whether the level of significance was important in selecting SNPs for further validation, and so these 34 SNPs are not a panel of the most significant SNPs generated by the WGA that might be used for selecting RFI. Due to technical reasons associated with genotyping, not all the animals with genotypes in the WGA had genotypes in the validation experiment. We estimated the additive effects for these SNPs in the non-WGA residue (validation sample minus the WGA) to determine the degree of variability left over once the WGA sample was removed.

For ease of analysis of significance of SNPs and compact reporting of the data in the tables, allele 0 is the allele higher up the alphabet; this means that C is allele 1 when A is allele 0

but C is allele 0 when G is allele 1. A negative value for α means that allele 0 causes a decrease in RFI, while a positive value means that allele 1 decreases RFI. The WGA and validation genotyping use consistent allele codes.

Differences between breed genotype frequencies were tested in contingency tables (Weir 1996) using the log-likelihood *G*-test with the Williams correction (Sokal and Rohlf 1981). Due to the relatively small breed samples (*N* range of 21–41), significance was called at $P < 0.01$. The genotype frequencies were tested for Hardy–Weinberg equilibrium (HWE) using the χ^2 test. In the HWE test, observed genotype frequencies were compared only to the expected frequencies if all expected classes were >5.

Analysis: To be able to compare the WGA and validation samples, the model RFI = $N(\mu$ + contemporary group + sire, $\sigma_e)$ was fitted to the RFI values of all animals using ASReml (Gilmour *et al.* 1995, 2002). While the selecting of samples for the WGA tried to remove the effect of contemporary group and sire, the validation sample would need adjustment for these effects, and therefore comparing the two would require that both were adjusted in the same way. Contemporary group was treated as a fixed effect; it includes herd of origin, which then accounts for the breed effect. The sire was treated as a random effect. The inclusion of sire and breed controls for the average differences between breeds and sires, which controls the genetic stratification in the sample. The residual trait values were obtained once for the entire experiment. The residual trait value for each individual was then matched to its SNP genotypes for both the WGA experiment of 9260 SNPs and the validation experiment of 44 SNPs. The mean trait values for each genotype at each SNP were calculated and then these were compared one SNP at a time using *t*-tests with significance determined by 100,000 permutations of the data. In the validation experiment, statistics were calculated for the combined sample as well as for each breed separately. The *t*-value between the most significant comparison of genotype means is reported as t_{max}. The differences between all three genotype means and the allele frequencies were used to calculate α, the average effect of allele substitution of the polymorphism (Lynch and Walsh 1998). The significance threshold was determined empirically (Churchill and Doerge 1994). The false discovery rate (FDR) was determined by exploring the empirical distribution of *P*-values to estimate the proportion of observed *P*-values that follow a uniform distribution on the (0,1) interval, that is, the distribution of *P*-values corresponding to tests with true null hypotheses (Mosig *et al.* 2001; Nettleton *et al.* 2006). The proportion of the variance explained by the significant SNP was estimated using regression, where sums of squares where compared between the model without SNP to the model with SNP.

To estimate the distance over which SNPs will show significant associations in this population, the detection distance (DD) was calculated. The DD represents the distance at which a SNP can be expected to detect a DNA variant affecting the phenotype through linkage disequilibrium in the particular study. The DD was estimated from the distribution of significant associations between SNP and the trait of interest in the following way. Each SNP with $P < 0.01$ to the trait was used as an index SNP and was used once, ensuring that the observations were independent. The distance between the index SNP and the closest SNP in the same scaffold was calculated in base pairs by counting. We calculated the distances within a scaffold only, due to the uncertainty of the orientation of some scaffolds, the unknown distances between scaffolds, and in some cases the order of scaffolds in the draft bovine sequence. Therefore, some index SNPs were not included because there were no other SNPs within the scaffold that were genotyped. We did not calculate distances between SNP where there appeared to be likely mis-assembly of the scaffold, but in no case where this occurred was the adjacent SNP associated with RFI, so this exclusion did not exclude any data that might have increased the DD. The log(1/P) was calculated and the difference obtained for each pair of index and adjacent SNPs. The difference in log(1/P) or DLP and the distance between each pair of SNPs was plotted. A curve was fitted to the points and the inflection point was taken to be the maximum DD for the data.

The LD was estimated for each pair of SNPs used to estimate the DD, using r^2 calculated according to the published formula (Hill and Robertson 1968). The haplotype frequencies in each pairwise SNP comparison were estimated from two-locus genotypes of unknown phase using the EM algorithm (Weir 1996). The r^2 estimates were calculated within and across breeds. The difference in log(1/P) was plotted against the r^2 estimated across breeds for each pair of SNPs.

The promoter sequence and micro-RNA (mi-RNA) sequence motifs (Xie *et al.* 2005) were downloaded and any redundancy was fully expanded. The occurrence of each motif was counted in the 250-bp sequence flanking each SNP, and each SNP was tested to determine whether it was a polymorphism for a motif; 250 bp was chosen arbitrarily to represent the close vicinity of the SNP so it is unlikely to represent artifacts due to assembly. In addition, the number of SNPs associated with RFI with 0, 1, 2, and so on copies of a motif in the surrounding sequence were compared to the rest of the SNPs in the SNP panel in a contingency table using the log-likelihood *G*-test with the Williams adjustment (Sokal and Rohlf 1981). Furthermore, each motif was counted in the sequences surrounding the SNPs associated with RFI and compared to the number of each motif in the sequences surrounding the SNP in the rest of the panel. Since several of the classes were 0, the contingency tables were compared with a generalized linear model using the Poisson distribution implemented by S-Plus (Venables and Ripley 2000).

RESULTS

We plotted the SNP data and examined them for several characteristics (Figure 1) as a check on the performance of the panel. Of the 9260 SNPs in total, 8786 were polymorphic in this sample, showing a flat distribution of allele frequencies (Figure 1A); 76% of these SNPs showed significant differences ($P < 0.01$) in genotype frequencies when all breeds were considered, 56% of the SNPs showed differences in temperate breeds, and 31% of the SNPs showed differences in the tropically adapted breeds. The genotype frequencies were consistent with Hardy–Weinberg expectations in 94% of breed comparisons. All the X-linked loci would have been out of HWE, because the analyzed DNA samples were all from steers.

High *t*-test values for the association with RFI occurred at all frequencies, not just in the middle frequencies (Figure 1B), and the distribution of *t*-tests suggests a significance threshold of $P = 0.01$. Few of these *t*-test values were above a value of 2.5, corresponding to $P = 0.01$, and the allelic effects for the SNPs below this threshold clustered strongly around 0 (Figure 1C). The FDR, corresponding to SNPs with $P \leq 0.01$, was 17.4% and for $P \leq 0.05$ was 31.9%. Furthermore, for

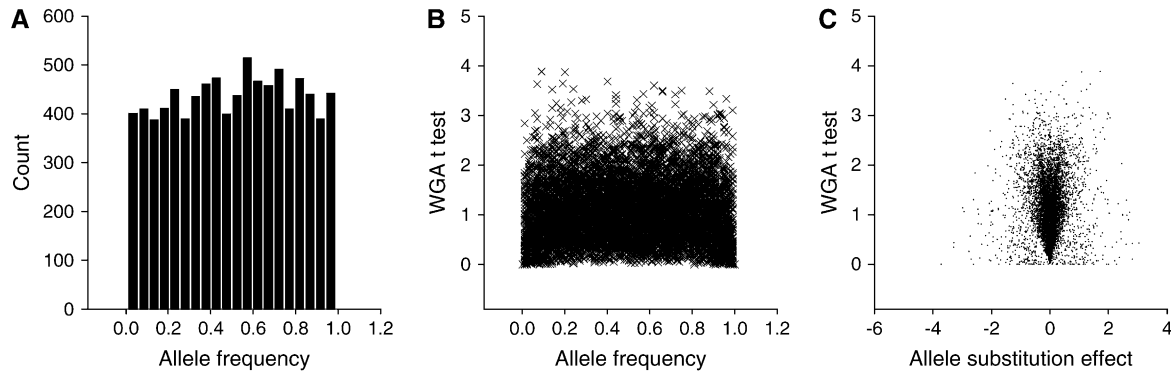

Figure 1.—(A) Histogram of allele frequencies in the whole genome association study. (B) The relationship between *t*-test and allele frequency in the whole genome association study, showing a density threshold at $\sim t = 2.5$. (C) The relationship between *t*-test and the average effect of allele substitution in the whole genome association study, showing large α-values for $t > 2.5$.

QTL with differences between means of 0.5 of a phenotypic standard deviation (σ_p) or less, *P*-values will not reach $P = 0.001$ in the sample size used for WGA in this study unless the effects are overestimated. Therefore, one should not focus only on effects with $P \leq 0.001$; otherwise, one would reject too many of the real associations to RFI. This suggests that a threshold of $P \leq 0.01$ would be useful in identifying QTL regions for further study of RFI, corresponding to a real feature in the data.

There were 161 SNPs with $P \leq 0.01$, representing 141 genetic regions, which were annotated in detail (supplemental Table S1 at http://www.genetics.org/supplemental/). Twenty of the SNPs explain 76% of the genetic variance for RFI, which has a heritability of 0.26 ± 0.07 when measured in all available animals. The amount of variance explained reached a plateau after 20 SNPs. However, when these 20 are removed, another panel of 20 SNPs explain 71% of the genetic variance, and subsequent panels of 20 explain progressively less variance. The panel of 34 SNPs used in the validation experiment (see below) explained 12% of the genetic variance in that larger sample, showing that real genetic variance is explained by some of the SNPs, but that the size of the WGA sample is not large enough to give accurate estimates of the amount of genetic variance explained by the significant SNPs.

To estimate the DD between SNPs and the QTL that they mark, we calculated the DLP between adjacent SNPs, at least one of which had $P < 0.01$ to RFI. The DLP dropped by 1 unit over a distance of much <10 kbp (Figure 2A), and the inflection point occurred at ~30 kbp. The curve fitted to this data set had an R^2 of 0.82. This is also the maximum distance found for both adjacent SNPs to be significantly associated with RFI. The shape of the curve is determined by the large number of adjacent SNPs that are <5 kbp apart where the DLP is >1. After 30 kbp, the curve is relatively flat and gradually increases from a DLP of 1.5 to that of 2.0.

The plot of the across-breed r^2 against distance shows four comparisons with $r^2 > 0.4$ but none with $r^2 \geq 0.8$ for SNPs separated by >30 kbp (Figure 2B), indicating that LD is low beyond 30 kbp for the SNPs in these breeds. The estimates of linkage disequilibrium for the pairs of

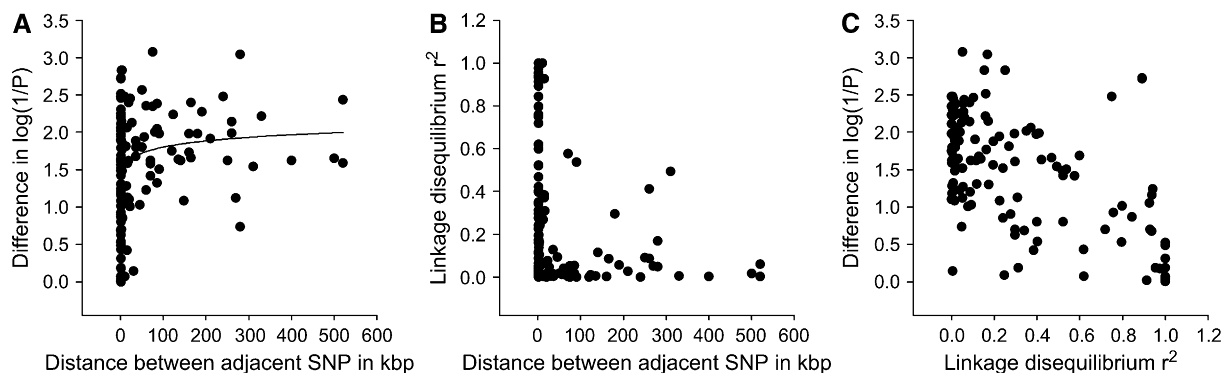

Figure 2.—(A) The reduction in significance of adjacent loci compared to the distance between adjacent loci as an estimator of the effective detection distance from the QTL. At least one member of each pair of SNP had $P < 0.01$. The most significant SNP from each region was compared to SNPs on both sides, although this was not always possible, and a maximum of three SNPs from each genetic region was used. (B) The LD (r^2) between the pairs of SNPs in a, plotted against distance between adjacent loci. (C) The reduction in significance between adjacent loci plotted against LD (r^2) for the loci in A.

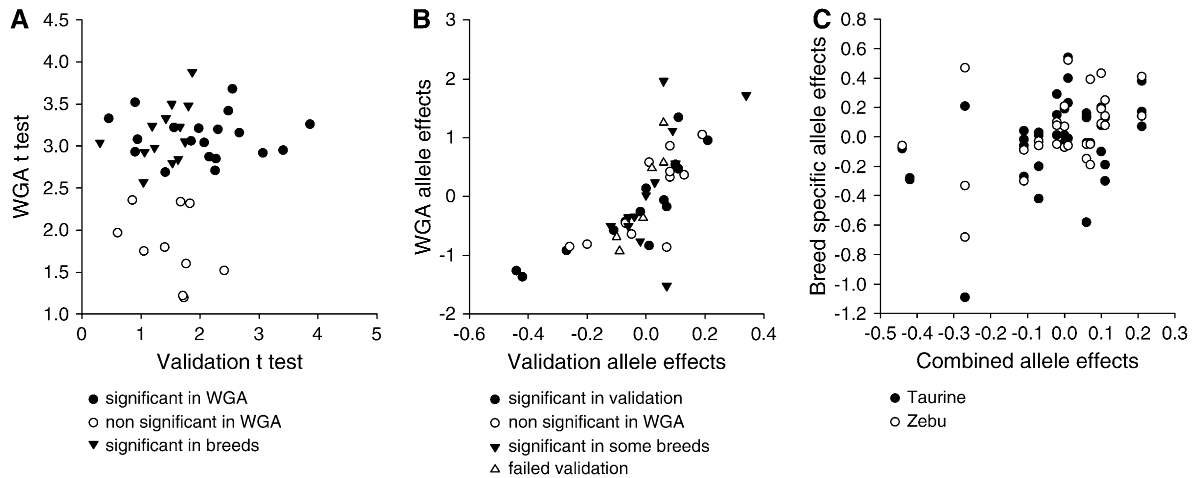

FIGURE 3.—(A) A comparison of the *t*-test value of significant SNPs of the WGA with their tests in the validation experiment. Those that were significant in only some of the breeds but not in the combined validation sample are noted. Ten of the SNPs in the validation experiment were not significant in the WGA and were included as a control. This shows that, above the significance threshold, there is no relationship between the significance of a SNP in the WGA and its significance in the validation experiment. (B) Allele substitution in the WGA compared to allele substitution in the validation sample. This shows that failure in validation is not based on the size of the allele effect in the WGA. (C) Allele substitution in different breeds plotted against the combined estimate of the allele substitution for SNPs significant overall in the validation experiment. The graph shows the α-values for all the breeds for a particular SNP in the same column, and the location of the column on the *x*-axis indicates the magnitude of the combined α-value. This shows that there are large differences between allele effects in different breeds.

index and adjacent SNPs showed a higher mean for the within-breed comparison ($r^2 = 0.35 \pm 0.01$ SE) than for the across-breed comparison ($r^2 = 0.29 \pm 0.03$ SE). Adjacent SNPs that showed an LD of $r^2 \geq 0.8$, the tag SNP threshold (CHAPMAN *et al.* 2003), had a mean DLP of 0.66 ($\pm$ 0.18 SE) while SNPs with $r^2 < 0.8$ had a mean DLP of 1.64 ($\pm$ 0.06 SE) in the difference in *P*-value for their associations to RFI; a DLP of 0.66 is equivalent to the difference between 0.01 and 0.048. There is a trend of increasing DLP with decreasing r^2 (Figure 2C).

At least 76% of the SNP with $P \leq 0.01$ in the WGA were found to have some significance in the larger sample or in at least one of the breeds of that larger sample. We tested 34 significant SNPs (Figure 3A; Table 1) in the whole sample of 1200 animals to determine how many would remain significant and to obtain more accurate estimates of the sizes of effect. The average number of WGA samples in this sample was 156 (SD = 20.41, range 90–171). Of the 189 animals in the WGA experiment, the average number of genotypes for these 34 loci was 166 (SD = 7.9, range 137–171). Animals used in the WGA represent on average 16% of the genotypes in the validation sample. Of the 34, 13 had $P < 0.05$ when all the breeds were combined, 2 of which were significant only in the total sample but not in any of the individual breeds, and a further 13 showed a significant association in at least one of the breeds; these are 76% of the total. We also tested 10 nonsignificant SNPs as a control (Figure 3A). One of those 10 (342920, Table 1) had a $P < 0.05$ in the combined sample, and $P < 0.01$ in two of the breeds. It is 20 kbp away from 342650, which had a $P = 0.023$ in the WGA.

The most significant SNPs (*e.g.*, 349908 and 351344) in the validation sample were those that had an α of the same sign in all breeds. There was no relationship between the strength of the significance or the size of the allele effect of the SNPs in the WGA on the significance of the SNPs in the validation experiment, once the significance threshold had been crossed (Figure 3, B and C). The direction of the allele substitutions were similar, in general, between the WGA and validation experiment (Figure 3B) with a significant trend line ($R^2 = 0.52$), which is consistent with most of the SNPs in the validation experiment being associated with RFI. The sizes of effects were smaller in the larger sample (Figure 3B), with a slope in the regression of 4.1. Using additive effects instead of average effects of allele substitution results in a slope of 3.9 and an $R^2 = 0.62$, indicating that the relationship is mainly due to the additive effects rather than to the dominant effects.

Removing the WGA samples from the total results in a collapse of the additive effects for most of the SNPs. The range of additive effects in the WGA sample for the 34 loci is from −0.95 to +0.91. In the whole sample the range of additive effects is from −0.29 to +0.23, but once the WGA samples are removed the range of additive effects is from −0.10 to +0.11, with 17 of the 34 additive effects ranging from −0.01 to +0.01. Removing the WGA sample results in the variance for the RFI trait dropping to 0.62 of its value in the full sample. Most of the information for the associations appears to be contained in the samples of extreme phenotype, as predicted.

The significant SNPs with α-values of the opposite sign in different breeds will lead to other SNPs that

TABLE 1

The results of validation of a set of SNPs from the WGA

Gene	SNP	N	p_0	α (kg/day)	t_{max}	P_{perm}
PLA2G5	342920	1107	0.86	−0.27[a]	2.41	0.0189
ATP1A1[b]	343136	944	0.63	−0.06	1.52	0.1224
DAG1[c]	343193	1103	0.43	−0.07[a]	2.07	0.0270
LOC506674	343500	1018	0.90	−0.42[a]	3.07	0.0029
SYT10	343617	1183	0.25	0.01[a]	2.48	0.0176
Gene unknown	343870	1044	0.49	−0.01	0.90	0.3749
YES1[b]	345261	923	0.89	0.06	1.06	0.2895
EMR3	345371	939	0.17	0.06	0.90	0.3783
XP_609078	345696	981	0.61	0.06[a]	2.26	0.0298
MED18	346024	999	0.68	0.06	1.56	0.1284
EDG1	346106	1068	0.82	−0.10	1.41	0.1663
RNF19	346634	1104	0.63	−0.07	1.85	0.0505
RPLP2[b]	346987	1094	0.14	−0.12	1.19	0.2218
GBAS	347062	1107	0.49	0.02	0.45	0.6457
Gene unknown[b]	347122	871	0.61	−0.06	1.75	0.0858
Gene unknown[b]	347283	1082	0.23	−0.02	0.30	0.7514
Gene unknown	347501	1088	0.67	−0.11	2.27	0.0284
Gene unknown	347570	1042	0.36	0.10	2.55	0.0064
XP_593840[b]	347727	1092	0.69	0.00[a]	1.80	0.0899
Gene unknown	348351	978	0.19	−0.09	0.94	0.3550
Not near gene	348902	879	0.21	0.11[a]	2.67	0.0114
NEUROD1[b]	348912	1028	0.84	0.09	1.23	0.1939
KRT1-23[b,c]	349583	1088	0.08	0.34[a]	1.87	0.0814
Not near gene	349908	1101	0.40	0.00[a]	3.41	0.0009
CROP[c,d]	350155	1108	0.66	−0.02[a]	2.30	0.0194
BSN[b,c]	350549	733	0.31	0.10[a]	1.66	0.0918
ZNF37A[d]	350777	445	0.90	−0.44	1.98	0.0474
Not near gene[b]	350995	800	0.68	0.07	1.42	0.1195
Gene unknown	351344	1174	0.25	0.21[a]	3.76	0.0001
XP_612880[b]	351382	822	0.82	0.19	1.63	0.1031
Gene unknown[b]	352079	989	0.49	0.03	1.04	0.2709
Gene unknown	353387	998	0.88	0.07	2.15	0.0160
KIAA1211[a]	353748	843	0.59	−0.04	1.53	0.1188
DMD[b]	354186	792	0.77	−0.08	1.65	0.0897
Gene unknown	354212	696	0.68	0.11	2.03	0.0455

"Not near gene" indicates that there is no gene within 150 kbp of either side; "gene unknown" means that the scaffold end is <150 kbp from one side; "SNP" is the ParAllele SNP identifier; N is the sample size; p_0 is the frequency of the 0 allele; and α (kg/day) is the average effect of allele substitution. Allele 0 is the allele higher up the alphabet. A negative value for α means that allele 0 causes a decrease in RFI, while a positive value means that allele 1 decreases RFI. P_{perm} is the P-value derived from 100,000 permutations.

[a] Dominance deviation $|k| \geq 1.0$; *i.e.*, showing dominance or overdominance, the large numbers are consistent with the low heritability of RFI.

[b] Significant in at least one of the component breeds.

[c] Within the genomic DNA of the gene.

[d] Only significant in the combined sample but not in any of the breeds.

should show consistent allele effects in all breeds. For example, the SNP 343617 (*NW_985578.1:g.47639A>G*) was very strongly associated to RFI in the WGA but was of moderate significance in the combined validation sample (Table 2), with two significant breed α-values of opposite sign tending to cancel each other out. Two of three new SNPs flanking 343617—*i.e.*, not part of the WGA but taken from genome sequencing traces in the surrounding 20 kbp and so within the DD—showed large α-values of the same sign in different breeds and one of these, *NW_985578.1:g.60065A>T*, was highly significant.

There were 5 SNPs that affected mi-RNA motifs, 16 that affected promoter sequence motifs, and 9 that were in mRNA in the 161 sequences. These represent similar proportions of the total SNPs; that is, the 5 mi-RNA SNPs represent 1.4% of the mi-RNA SNPs, the 16 promoter sequences represent 1.0% of the promoter SNPs, and the 9 mRNA SNPs represent 1.0% of the mRNA SNPs in the panel.

TABLE 2

The association between RFI and SNP near SYT10 coding sequences, with SNPs in the order in which they occur on the chromosome

SNP	N	p_0	α (kg/day)	t_{max}	P_{perm}
NW_985578.1:g.45346T> Combined	1315	0.07	−0.32	2.52	0.0134
Angus	296	0.01	−0.12	2.66	0.0082
Brahman	129	0.01	NC	0.46	0.6463
Belmont Red	291	0.06	−0.36	0.97	0.3283
Hereford	200	0.35	−0.28	2.49	0.0176
Murray Grey	75	0.00	NC	0.00	1.0000
Santa Gertrudis	225	0.01	NC	0.11	0.9136
Shorthorn	99	0.05	−0.28	0.50	0.6130
343617[a] combined	1270	0.25	0.01	2.48	0.0176
Angus	265	0.56	−0.01	0.24	0.8053
Brahman	116	0.03	NC	1.11	0.2647
Belmont Red	268	0.31	−0.06	2.27	0.0273
Hereford	188	0.21	0.54	3.08	0.0023
Murray Grey	64	0.31	0.40	0.99	0.3277
Santa Gertrudis	205	0.12	0.52	1.51	0.1270
Shorthorn	77	0.33	0.23	0.68	0.5079
NW_985578.1:g.49666G>C combined	1319	0.23	−0.09	1.66	0.0660
Angus	297	0.05	−1.03	0.18	0.8584
Brahman	128	0.05	NC	0.35	0.7347
Belmont Red	289	0.20	−0.15	1.28	0.2014
Hereford	201	0.59	−0.03	2.06	0.0404
Murray Grey	77	0.19	0.54	1.67	0.0945
Santa Gertrudis	229	0.14	−0.03	0.71	0.4774
Shorthorn	98	0.54	−0.00	0.18	0.8632
NW_985578.1:g.60065A>T combined	1308	0.07	−0.47	3.38	0.0009
Angus	296	0.01	NC[b]	2.78	0.0078
Brahman	130	0.01	NC	0.45	0.6658
Belmont Red	285	0.06	−0.35	0.93	0.3464
Hereford	199	0.34	−0.39	3.01	0.0039
Murray Grey	78	0.01	NC	0.00	1.0000
Santa Gertrudis	225	0.00	NC	0.50	0.6217
Shorthorn	95	0.06	−0.29	0.47	0.6384

NC, not calculated because it needs all genotypes and enough in each genotype so that each mean can have a standard error.

[a] 343617 is *NW_985578.1:g.47639A>G*.

[b] The heterozygotes have lower RFI values than the homozygote, consistent with the other breed comparisons.

To determine whether there were any characteristic motifs among the 161 SNPs associated with RFI, the presence and identity of each promoter or mi-RNA motif was counted within 250 bp of each SNP. Of the 161 significant SNPs, 90 contained one or more mi-RNA motifs. There was no difference in the distribution of numbers of promoter or mi-RNA motifs in the sequences surrounding the SNPs; that is, the numbers of sequences with 0, 1, 2, 3, and more motifs was not statistically different when the 161 SNPs for RFI were compared to the rest of the SNP set, for both promoter sequences ($G_{adj} = 18.90$, d.f. = 12, $N = 12{,}060$, $P = 0.091$) and for mi-RNA sequences ($G_{adj} = 1.48$, d.f. = 5, $N = 12{,}060$, $P = 0.916$). There were 86 promoter motifs in the 161 SNP sequences and these were not different in distribution from the 150 promoter motifs found in the rest of the SNPs ($D = 169.8$, d.f. = 149, $P = 0.116$). However, the number of each mi-RNA motif was highly significantly different between the RFI SNPs compared to the rest ($D = 97.7$, d.f. = 49, $P = 4.45 \times 10^{-5}$). The proportion of each of the 50 mi-RNA motifs in the total is shown in Figure 4, with several clear examples of motifs that are relatively more or less common than in the full data set.

Thirty-four of the genes with significant SNPs code for protein that either control or are part of cellular processes that are continuous and are part of the background use and generation of energy (Figure 5; supplemental Table S1 at http://www.genetics.org/supplemental/). Nine genes are associated with apoptosis and protein turnover, including *UBE2I*, which is involved in ubiquitination, the targeting of proteins for degradation, and may have a role in blocking cell cycle progression in eukaryotes. Six either are ion channels or are associated with ion flux with a further three associated with neurotransmitter flux, including *ATP1A1*, the pump maintaining the Na^+/K^+ gradient on plasma membranes. Ten are associated with growth and development, including *FGF2*, which is involved in a wide range of growth processes. Five are associated with translation or transcription, including *RPLP2*, which is a component of the 60S ribosome subunit and plays a role in protein elongation. One of these five is a structural protein, *KRT32*, which is a hair-and-nail keratin, and hair and hoof growth is clearly a continual process.

There were only three significant SNPs affecting known pathways for leptin, insulin, and serotonin. These are *NEUROD1*, which binds to the insulin promoter *PDE3B*, which is part of the *PI3K-PDE3B*-cAMP pathway, critical in leptin signaling, and *MPDZ*, which has probable protein–protein interactions with the C-terminal domain of the serotonin 5-HT-2C receptor (supplemental Table S1 at http://www.genetics.org/supplemental/).

Thirty-seven of the genes point to new genetic pathways in the control of food conversion efficiency (supplemental Table S1 at http://www.genetics.org/supplemental/). Six of these are DNA-binding proteins, which points to regulation of expression. Three genes are involved in signal transduction, and since the orexins influence food intake through signal transduction (Sakurai *et al.* 1998), these SNPs may point to such control. But by far the largest group are 28 of the significant SNPs that are in either hypothetical coding sequences or coding sequences without an assigned function. These are an important new source of genes with effects on RFI.

Eleven of the genes affect proteins in the extracellular matrix, cell adhesion, and tissue strength, including

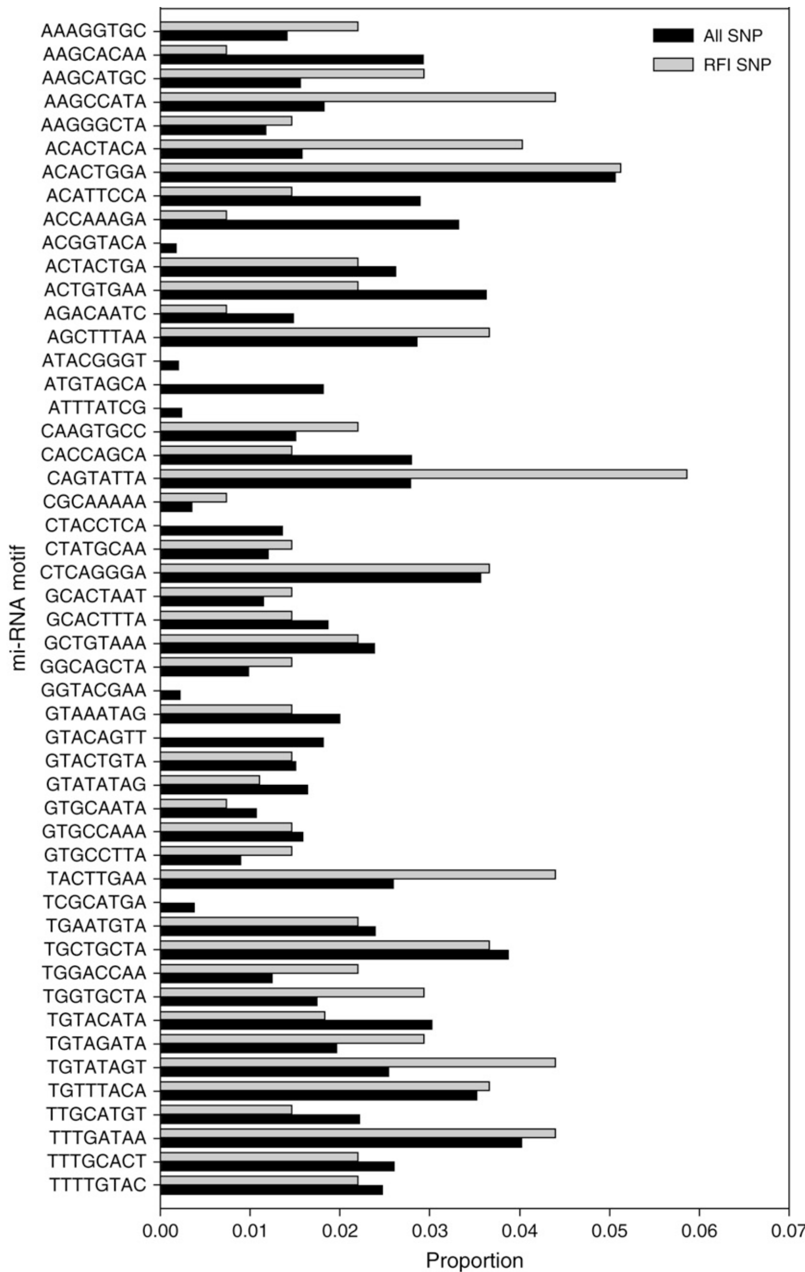

Figure 4.—The distribution of mi-RNA motifs in the SNP sequences associated with RFI compared to the motifs in all the SNPs in the WGA panel.

three involved in muscular and myotonic dystrophy in humans (supplemental Table S1 at http://www.genetics.org/supplemental/). There are two laminin genes (*LAMA2* and *LAMC1*), *KRT1-23* (a marker of terminal differentiation in epithelia), three genes involved with muscular and myotonic dystrophy (*DAG1*, *DMD*, *MBNL1*), and five genes involved in cell adhesion (*EDG1*, *FAT3*, *GPIAP1*, *SDK1*, and *SPA17*).

Fourteen of the significant SNPs (Figure 5; supplemental Table S1 at http://www.genetics.org/supplemental/) are unambiguously separated from any gene, and with the DD in this experiment, they are unlikely to be significant due to LD to allelic variants in nearby genes. Although 50 of the regions with significant SNPs are not close to genic sequences, 36 of these are within 150 kbp of the end of the scaffold. We set a threshold of five times the 30-kbp DD to be completely sure that the SNP was not genic. These 14 significant SNPs represent a low estimate of the proportion of the nongenic DNA that is likely to affect the phenotype and, by relaxing the 150-kbp threshold, other likely candidate regions can be obtained from supplemental Table S1 at http://www.genetics.org/supplemental/.

DISCUSSION

These results show that a panel of ∼9000 SNPs can provide useful data on livestock and other species to locate and identify genes affecting a quantitative trait. Livestock species are reported to have high levels of LD

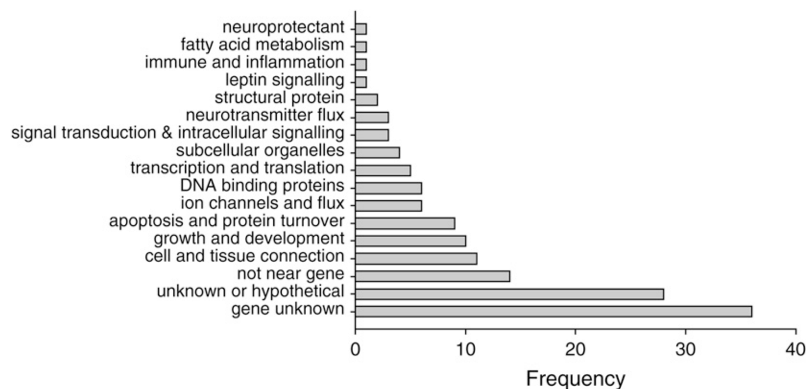

FIGURE 5.—The frequency of genes of different classes in the WGA experiment.

due to domestication and, more recently, breed formation (FARNIR *et al.* 2000; MCRAE *et al.* 2002; TENESA *et al.* 2003; NSENGIMANA *et al.* 2004; KHATKAR *et al.* 2006). Increased LD is detectable within dairy breeds of cattle at a distance of 1 Mbp, which would suggest that 3000 evenly spaced SNPs could be sufficient to cover the genome. However, estimates of LD are lower across the seven beef breeds in this study, with few elevated values of r^2 for pairs of SNPs separated by >30 kbp. By using the offspring of >100 sires over several breeds, including representatives of the taurine group of European cattle, the zebu group of Indian cattle, and the sanga group of African cattle, we appear to have decreased the amount of LD in the sample. In accordance with this, the DD found between SNPs was only 30 kbp in this study. Tag SNP pairs (CHAPMAN *et al.* 2003) did show associations to RFI of a similar level of significance, but no tag SNP pairs were found at distances >30 kbp, for the SNP associated with RFI. This lower level of LD across breeds is consistent with previous studies of livestock that show that when highly diverse groups of animals are used, involving several breeds, that the estimates of LD are low even at very small genetic distances (BARENDSE *et al.* 2007). Similar results of the effect of estimating LD within and across breeds have been reported for a sample of 224 dogs (LINDBLAD-TOH *et al.* 2005). The low levels of LD mean that many of the QTL for RFI will be missed due to incomplete coverage of the genome, but this tool allows us to find many potential genes with greater precision than by using only one breed or a small number of sires within a breed. The use of a sample with high LD might have meant that the QTL would not have been resolved to <1 Mbp, with detection distances that might have spread out to several centimorgans. The corollary of this also applies to WGA in human or other wild species, where the detection distance between a SNP and a DNA variant causing the trait will be short. That is, even a panel of tens of thousands of SNPs should identify a useful number of genetic regions associated with the trait, and comprehensive coverage will of course require hundreds of thousands to millions of SNPs, depending upon the species and the sample structure. Our results indicate that hundreds of thousands of SNPs will be required for comprehensive coverage for cattle as well.

The distribution of SNP effects showed that concerns that QTL might not be detected unless minor allele frequencies were substantial are unfounded. Although minor allele frequencies below a certain point would result in few genotypes for some homozygotes, by using the *t*-test to determine significance, we avoid this issue since a substantial number of heterozygote individuals will still be available even when the allele frequencies are as low as 5%. By using permutation tests, which is the recommended approach (CHURCHILL and DOERGE 1994; CARLSON *et al.* 2004), any non-normality in the distribution of the phenotype will not affect the performance of the significance test.

This permutation approach allows us to be more conservative, to account for multiple testing of SNPs, and to set realistic thresholds for significance in the experiment. While some of the SNPs do show significance at extreme *P*-values, such as $P < 0.0001$, if one did not consider the nature or makeup of the population, such a threshold will yield distorted results. Most QTL will have small effects, defined as polygenic variation when $\alpha < 0.5\ \sigma_p$ (MORTON and LIO 1997), and some of them will be undetected by such a stringent threshold. Such polygenic effects have been mapped, diagnostic DNA tests have been identified (BARENDSE 2002; PAGE *et al.* 2002), and their existence confirmed independently by using much lower significance thresholds (MORRIS *et al.* 2006), so it is the replication of the associations that is important, not the use of highly extreme significance thresholds. By setting a too-stringent threshold, such polygenic variation would be ignored, or its effects would have to be overestimated to be significant, and then the attempt to predict phenotype on the basis of genotype would be restricted to only those genes with large effects. For most quantitative traits, this would mean that there would be inaccuracies in the prediction of phenotype using SNPs, because most quantitative traits have zero or one major gene segregating with many QTL of smaller effect and with epistatic

interactions between them (Mayo 2004; Barendse 2005; Carlborg *et al.* 2006). A FDR of 17.4% for the WGA at a threshold of $P = 0.01$, and the finding that 76% of the significant associations at that threshold were found in the validation experiment, suggest that >107 of the 141 genetic regions affect the trait. This does not support a model of a quantitative trait largely determined by a small number of oligogenic loci, although the genetic variance in some traits may well be largely affected by a small number of polymorphisms.

Comparing either the allele substitution or the additive effects of the same SNP in two samples may be a useful way to determine the degree of support for a set of SNPs; however, as the WGA sample forms a part of the whole sample, the results will be correlated to some extent. Validation of the SNP association by genotyping all the animals is important because using animals of extreme phenotype will result in allelic affects that are too large, and methods that do not involve correction for missing data may result in spurious associations (Lander and Botstein 1989); we have used permutation methods to ensure that we make no assumptions of normality in the underlying extreme phenotypic data. The allelic effects in the whole sample are ~25% of those found in the WGA and the regression between WGA and the validation sample is consistent with the QTL being real, but does not prove that the QTL are real. Comparing the additive effects in the WGA sample to those in the non-WGA sample would not provide much information in this experiment due to the large reduction of the additive effects once the WGA samples are removed, as is predicted in theory (Lander and Botstein 1989); most of the effects are essentially zero in the non-WGA sample. Had the WGA sample been a random selection of the whole, then correlation between the additive effects in the WGA with the additive effects in the non-WGA sample would be a test of the 34 SNPs, but in this case it is consistent only with the SNP being associated with RFI. Independent validation of a trait like RFI will be difficult, partly because generating a sufficiently large sample would be costly and therefore correspondingly rare and partly because RFI may vary at different times in the animal's life (Johnston 2007) and so may require that the same breeds are used and that the animals are measured at the same time of life.

The validation of the SNPs in this study through the analysis of a large group of animals suggests that many of the SNPs identified in the WGA are real and shows that rankings of SNPs on level of significance or size of effect are not good predictors of performance in further studies. WGA is an important tool for identifying DNA variants that affect phenotypes (Kruglyak 1999; Carlson *et al.* 2003, 2004; Fan *et al.* 2006) with studies in humans and in mice (Ozaki *et al.* 2002; Klein *et al.* 2005; Maraganore *et al.* 2005; Valdar *et al.* 2006). Nevertheless, it is necessary to confirm these results as there have been WGA validation failures (Goris *et al.* 2006; Myers 2006), that is, failures to find the most statistically significant associations in other studies, even when the original studies included independent validation. Here we show that some of the most significant associations are not found when a large sample is subsequently genotyped, but many are, and some of those with less extreme *P*-values in the WGA were found to be highly significant in the validation experiment. This confirms the notion that the highly significant effects to come out of WGA may be overestimated, and a concentration on them when validating SNPs may lead to failure of validation. The key factor is whether the same genotype or genotypes cause a change of the same direction in the phenotype in all population groups, which is a function of the LD between the marker SNP and the causative SNP. In hindsight, this seems obvious, but the proportion of α-values with the same sign in different populations is not currently a statistic that is collected in these studies, and the LD found among SNPs in the study is not always reported.

These SNPs seem to be embedded among a specific subset of mi-RNA motifs compared to the SNPs in general in the panel, which would implicate only a part of the regulatory machinery of the cell and would imply a degree of modularity in the regulation of traits. There does not appear to be an enrichment of mi-RNA SNPs over other kinds of SNPs, which would be consistent with the idea that many different kinds of DNA variants will play a role in QTL variation, not one kind over another kind. Moreover, only one of the significant mi-RNA SNPs was near a gene; the other four were distant from any coding sequence. This suggests that enrichment of WGA panels for SNPs that are associated with genes are likely to miss functional SNPs. Nevertheless, the mi-RNA motifs that are present on the sequences flanking the significant SNPs are not a random sample of the motifs in the SNP set. Each mi-RNA motif is associated with a small number of mi-RNAs, so a difference in the distribution means that a specific set of mi-RNAs are important for RFI, and one may generalize to say that the sum of the mi-RNAs for the entire phenotype is a combination of specific sets of mi-RNAs for different traits. This implies a modularity in regulation via mi-RNA that is not evident for the promoter motifs. The promoter motifs showed no difference in the sequence surrounding the significant SNPs for RFI compared to the rest of the SNPs, suggesting that the mechanisms for promoter regulation are more generic and clearly showing that the differences in mi-RNA are not due to the DNA composition of the DNA sequence in which these SNPs occur.

The genes identified in this way support the notion that efficient food conversion will be due to features of the phenotype at several levels in the hierarchy of cells, tissues, and organs, but that the control of appetite and homeostasis of body mass will be a minor component of the trait. Only three of the known genes have a possible

effect on some aspect of a hormone affecting food consumption, with one each affecting insulin, leptin, and serotonin.

The behavioral and homeostatic systems that affect hunger, feeding, and satiety thus act above the actual food conversion efficiency of the animal. Some of the genes identified in this way seem obvious candidates, such as *ATP1A1*, *UBE2I*, and *RPLP2*. These genes would be expected to play a part in the basal metabolic rate of an animal, as they use ATP or other forms of energy to shunt ions around, to recycle proteins, and to synthesize proteins. The rate and efficiency of these systems clearly must have a role to play in this trait, and their identification in this study is confirmation that the approach used here has provided valid data (Carlson *et al.* 2004).

By extension, the more interesting genes are those that are DNA sequences for hypothetical genes or genes with unknown function. Some of these have hints at functionality that are relevant for RFI, such as *XP_602409*, which is similar to *Cyclin M1*, while others are targets for further analysis, considering that they have a strong likelihood of being involved in food conversion or metabolic efficiency. The large number of these genes implies that the study of other difficult traits should help in the annotation of the genes that are known only as hypothetical DNA sequences or genes of unknown function. How these may influence food conversion is not known and will require substantial further research.

One class of genes of great interest is that affecting the extracellular matrix with an effect on RFI. Mutations in one of these, *DMD*, are known to have effects on muscle regeneration and muscular hypertrophy (Doriguzzi *et al.* 1993; Harper *et al.* 2002), so some of their roles may be pleiotropic and will be different from that of stitching together the phenotype. Nevertheless, it is possible that they point to something more substantive than genes with pleiotropic effects on the phenotype. That several of these genes in the extracellular matrix and cell adhesion have effects on RFI shows a possible trade-off between RFI and tissue strength. This would imply that RFI would be correlated to some extent with other aspects of the musculature of the phenotype; in this population, there is a positive genetic correlation between RFI and fatness in several depots of the animal while there is a negative genetic correlation between RFI and muscle area of the *m. longissimus lumborum* (Robinson and Oddy 2004). While this is not a direct measure of tissue strength, it says that such correlations are to be expected.

The final group of SNPs point to a significant number of nongenic DNA variants as having an impact on the phenotype. The 14 genetic regions are a conservative estimate, given the DD found in this study. Furthermore, the distance from a gene does not mean that they must necessarily be either spurious or in LD with genic variation. The variation causing the Callipyge mutation is in the intergenic space (Freking *et al.* 2002), three of three of these nongenic SNPs were validated in this study, 53% of the ultra-conserved DNA elements are not translated or immediately adjacent to coding sequences (Bejerano *et al.* 2004), and there are several hypotheses linking QTL variability to repetitive DNA (Kashi *et al.* 1997). If one assumes that the border for genic variation would be no >4 kbp at either end of the coding sequence (Xie *et al.* 2005), then the nongenic group is the largest group of SNPs in this study with effects on the phenotype.

These results have implications in general for WGA for QTL. First, the validation study shows that the SNPs that are most likely to be validated are those with consistent α-values in different populations, not those that have either the biggest α-values or the most extreme *P*-values in the WGA, and they are likely to be those that are in high LD to the causal mutations. Second, the difference in distribution of mi-RNA motifs suggests that each trait will have its signature that will correspond to how they may be regulated. Third, while the identity of many of the genes affecting the trait will change in different species, this analysis shows the likely proportion of those to come from housekeeping genes, from regulation by genes, and from regulation of the phenotype from nongenic DNA variation. This last type of DNA variant is likely to be important for all quantitative traits. Finally, due to the large number of genes affecting the extracellular matrix, this study shows a possible trade-off between RFI and tissue strength. This trade-off could be a common response of many species to selection for food use and metabolic efficiency. This kind of trade-off is at the heart of the response of the organism to its environment, since trade-offs between the soma and energy are central to some of the theories of aging and senescence and of early advantage in spreading genes *vs.* maintenance of a body well into old age (Kirkwood and Rose 1991). Our results show that there is a trade-off of food conversion efficiency, perhaps allowing more energy for reproduction or other evolutionarily important behaviors, with some of the genes that are directly involved in tissue strength and integrity. This begs the question, is the efficiency increase due to better tissue construction leading to savings in energy or to lower quality of construction requiring less energy?

R. J. Hawken used WhippySnippy to find the three SYT10 SNPs in the bovine genome draft sequence. D. J. Johnston extracted the RFI raw data from the Beef CRC1 database. We thank J. P. Gibson, M. E. Goddard, and J. M. Henshall for discussing the analysis; V. H. Oddy, N. J. Hudson, and D. L. Robinson for discussing residual feed intake and metabolic rate; B. P. Dalrymple for discussing bioinformatics and the bovine genome sequence; and N. J. Hudson, J. W. Kijas, D. Barendse, P. Willadsen, G. T. Davies, V. H. Oddy, D. J. Johnston, and two anonymous reviewers for comments on the manuscript. We thank the owners of the Beef CRC1 DNA bank for access to cattle phenotypes and DNA samples. This research was performed as part of the Cooperative Research Agreement between Commonwealth Scientific and Industrial Research Organization, CRC for Cattle and Beef Quality, Meat and Livestock Australia, and Genetic Solutions P/L.

LITERATURE CITED

Barendse, W. J., 2002 DNA markers for meat tenderness. Patent application WO02064820.

Barendse, W., 2005 The transition from quantitative trait loci to diagnostic test in cattle and other livestock. Aust. J. Exp. Agric. **45:** 831–836.

Barendse, W., R. J. Bunch, B. E. Harrison and M. B. Thomas, 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. Anim. Genet. **37:** 211–214.

Barendse, W., R. J. Bunch, J. W. Kijas and M. B. Thomas, 2007 The effect of genetic variation of the retinoic acid receptor-related orphan receptor C gene on fatness in cattle. Genetics **175:** 843–853.

Bejerano, G., M. Pheasant, I. Makunin, S. Stephen, W. J. Kent *et al.*, 2004 Ultraconserved elements in the human genome. Science **304:** 1321–1325.

Bjornsson, B., A. Steinarsson and M. Oddgeirsson, 2001 Optimal temperature for growth and feed conversion of immature cod (*Gadus morhua* L.). ICES J. Mar. Sci. **58:** 29–38.

Byers, J. E., 2000 Competition between two estuarine snails: implications for invasions of exotic species. Ecology **81:** 1225–1239.

Cahaner, A., and F. Leenstra, 1992 Effects of high temperature on growth and efficiency of male and female broilers from lines selected for high weight-gain, favorable feed conversion, and high or low fat-content. Poult. Sci. **71:** 1237–1250.

Carlborg, Ö., L. Jacobsson, P. Ahgren, P. Siegel and L. Andersson, 2006 Epistasis and the release of genetic variation during long-term selection. Nat. Genet. **38:** 418–420.

Carlson, C. S., M. A. Eberle, M. J. Rieder, J. D. Smith, L. Kruglyak *et al.*, 2003 Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. Nat. Genet. **33:** 518–521.

Carlson, C. S., M. A. Eberle, L. Kruglyak and D. A. Nickerson, 2004 Mapping complex disease loci in whole-genome association studies. Nature **429:** 446–452.

Chapman, J. M., J. D. Cooper, J. A. Todd and D. G. Clayton, 2003 Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. Hum. Hered. **56:** 18–31.

Churchill, G. A., and R. W. Doerge, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963–971.

Clop, A., F. Marcq, H. Takeda, D. Pirottin, X. Tordoir *et al.*, 2006 A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat. Genet. **38:** 813–818.

Cummings, D. E., J. Q. Purnell, R. S. Frayo, K. Schmidova, B. E. Wisse *et al.*, 2001 A preprandial rise in plasma ghrelin levels suggests a role in meal initiation in humans. Diabetes **50:** 1714–1719.

Darwin, C., 1859 *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* James Murray, London.

Doriguzzi, C., L. Palmucci, T. Mongini, L. Chiadopiat, G. Restagno *et al.*, 1993 Exercise intolerance and recurrent myoglobinuria as the only expression of Xp21 Becker type muscular-dystrophy. J. Neurol. **240:** 269–271.

Fan, J. B., M. S. Chee and K. L. Gunderson, 2006 Highly parallel genomic assays. Nat. Rev. Genet. **7:** 632–644.

Farnir, F., W. Coppieters, J. J. Arranz, P. Berzi, N. Cambisano *et al.*, 2000 Extensive genome-wide linkage disequilibrium in cattle. Genome Res. **10:** 220–227.

Freking, B. A., S. K. Murphy, A. A. Wylie, S. J. Rhodes, J. W. Keele *et al.*, 2002 Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. Genome Res. **12:** 1496–1506.

Gilmour, A. R., R. Thompson and B. R. Cullis, 1995 Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. Biometrics **51:** 1440–1450.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham and R. Thompson, 2002 *ASReml User Guide Release 1.0.* VSN International, Hemel Hempstead, UK.

Goris, A., C. H. Williams-Gray, T. Foltynie, D. A. S. Compston, R. A. Barker *et al.*, 2006 No evidence for association with Parkinson disease for 13 single-nucleotide polymorphisms identified by whole-genome association screening. Am. J. Hum. Genet. **78:** 1088–1090.

Hardenbol, P., F. L. Yu, J. Belmont, J. MacKenzie, C. Bruckner *et al.*, 2005 Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. Genome Res. **15:** 269–275.

Harper, S. Q., M. A. Hauser, C. DelloRusso, D. Duan, R. W. Crawford *et al.*, 2002 Modular flexibility of dystrophin: implications for gene therapy of Duchenne muscular dystrophy. Nat. Med. **8:** 253–261.

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

Holliday, R., 1990 Mechanisms for the control of gene activity during development. Biol. Rev. **65:** 431–471.

Johnston, D. J., 2007 Technical update NFI & IGF-I, pp. 1–2 in *AGBU Technical Notes* (http://agbu.une.edu.au/cattle/beef9.pdf).

Kashi, Y., D. King and M. Soller, 1997 Simple sequence repeats as a source of quantitative genetic variation. Trends Genet. **13:** 74–78.

Khatkar, M. S., A. Collins, J. A. L. Cavanagh, R. J. Hawken, M. Hobbs *et al.*, 2006 A first-generation metric linkage disequilibrium map of bovine chromosome 6. Genetics **174:** 79–85.

Kirkwood, T. B. L., and M. R. Rose, 1991 Evolution of senescence: late survival sacrificed for reproduction. Philos. Trans. R. Soc. Lond. B Biol. Sci. **332:** 15–24.

Klein, R. J., C. Zeiss, E. Y. Chew, J. Y. Tsai, R. S. Sackler *et al.*, 2005 Complement factor H polymorphism in age-related macular degeneration. Science **308:** 385–389.

Kohn, R. A., Z. Dou, J. D. Ferguson and R. C. Boston, 1997 A sensitivity analysis of nitrogen losses from dairy farms. J. Environ. Manage. **50:** 417–428.

Kruglyak, L., 1999 Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat. Genet. **22:** 139–144.

Lander, E. S., and D. Botstein, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP maps. Genetics **121:** 185–199.

Leibowitz, S. F., and J. T. Alexander, 1998 Hypothalamic serotonin in control of eating behavior, meal size, and body weight. Biol. Psychol. **44:** 851–864.

Lindblad-Toh, K., C. M. Wade, T. S. Mikkelsen, E. K. Karlsson, D. B. Jaffe *et al.*, 2005 Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature **438:** 803–819.

Lynch, M., and J. B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, Sunderland, MA.

Maraganore, D. M., M. de Andrade, T. G. Lesnick, K. J. Strain, M. J. Farrer *et al.*, 2005 High-resolution whole-genome association study of Parkinson disease. Am. J. Hum. Genet. **77:** 685–693.

Mattick, J. S., 2001 Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep. **2:** 986–991.

Mayo, O., 2004 Interaction and quantitative trait loci. Aust. J. Exp. Agric. **44:** 1135–1140.

McRae, A. F., J. C. McEwan, K. G. Dodds, T. Wilson, A. M. Crawford *et al.*, 2002 Linkage disequilibrium in domestic sheep. Genetics **160:** 1113–1122.

Morris, C. A., N. G. Cullen, S. M. Hickey, P. M. Dobbie, B. A. Veenvliet *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. Anim. Genet. **37:** 411–414.

Morton, N. E., and P. Lio, 1997 Oligogenic linkage and map integration, pp. 17–21 in *Genetic Mapping of Disease Genes*, edited by I.-H. Pawlowitski, J. H. Edwards and E. A. Thompson. Academic Press, San Diego.

Mosig, M. O., E. Lipkin, G. Khutoreskaya, E. Tchourzyna, M. Soller *et al.*, 2001 A whole genome scan for quantitative trait loci affecting milk protein percentage in Israeli-Holstein cattle, by means of selective milk DNA pooling in a daughter design, using an adjusted false discovery rate criterion. Genetics **157:** 1683–1698.

Myers, R. H., 2006 Considerations for genomewide association studies in Parkinson disease. Am. J. Hum. Genet. **78:** 1081–1082.

Neel, J. V., A. B. Weder and S. Julius, 1998 Type II diabetes, essential hypertension, and obesity as "syndromes of impaired genetic

homeostasis": the "thrifty genotype" hypothesis enters the 21st century. Perspect. Biol. Med. **42:** 44–74.

Nettleton, D., J. T. G. Hwang, R. A. Caldo and R. P. Wise, 2006 Estimating the number of true null hypotheses from a histogram of p values. J. Agric. Biol. Environ. Stat. **11:** 337–356.

Nsengimana, J., P. Baret, C. S. Haley and P. M. Visscher, 2004 Linkage disequilibrium in the domesticated pig. Genetics **166:** 1395–1404.

Ozaki, K., Y. Ohnishi, A. Iida, A. Sekine, R. Yamada *et al.*, 2002 Functional SNPs in the lymphotoxin-[alpha] gene that are associated with susceptibility to myocardial infarction. Nat. Genet. **32:** 650–654.

Page, B. T., E. Casas, M. P. Heaton, N. G. Cullen, D. L. Hyndman *et al.*, 2002 Evaluation of single-nucleotide polymorphisms in CAPN1 for association with meat tenderness in cattle. J. Anim. Sci. **80:** 3077–3085.

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.

Sakurai, T., A. Amemiya, M. Ishii, I. Matsuzaki, R. M. Chemelli *et al.*, 1998 Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. Cell **92:** 573–585.

Sokal, R. R., and F. J. Rohlf, 1981 *Biometry.* W. H. Freeman, San Francisco.

Tanaka, T., J. Yamamoto, S. Iwasaki, H. Asaba, H. Hamura *et al.*, 2003 Activation of peroxisome proliferator-activated receptor delta induces fatty acid beta-oxidation in skeletal muscle and attenuates metabolic syndrome. Proc. Natl. Acad. Sci. USA **100:** 15924–15929.

Tenesa, A., S. A. Knott, D. Ward, D. Smith, J. L. Williams *et al.*, 2003 Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. J. Anim. Sci. **81:** 617–623.

Upton, W., H. M. Burrow, A. Dundon, D. L. Robinson and E. B. Farrell, 2001 CRC breeding program design, measurements and database: methods that underpin CRC research results. Aust. J. Exp. Agric. **41:** 943–952.

Utz, H. F., A. E. Melchinger and C. C. Schon, 2000 Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics **154:** 1839–1849.

Valdar, W., L. C. Solberg, D. Gauguier, S. Burnett, P. Klenerman *et al.*, 2006 Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat. Genet. **38:** 879–887.

Veerkamp, R. F., 1998 Selection for economic efficiency of dairy cattle using information on live weight and feed intake: a review. J. Dairy Sci. **81:** 1109–1119.

Venables, W. N., and B. D. Ripley, 2000 *Modern Applied Statistics With S-PLUS.* Springer-Verlag, New York.

Walder, K., R. A. Norman, R. L. Hanson, P. Schrauwen, M. Neverova *et al.*, 1998 Association between uncoupling protein polymorphisms (UCP2–UCP3) and energy metabolism obesity in Pima Indians. Hum. Mol. Genet. **7:** 1431–1435.

Weir, B. S., 1996 *Genetic Data Analysis II.* Sinauer Associates, Sunderland, MA.

Xie, X., J. Lu, E. J. Kulbokas, T. R. Golub, V. Mootha *et al.*, 2005 Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature **434:** 338–345.

Communicating editor: C. Haley