

Variants modulating the expression of a chromosome domain encompassing *PLAG1* influence bovine stature

Latifa Karim^{1,5}, Haruko Takeda^{1,5}, Li Lin^{1,5}, Tom Druet^{1,5}, Juan A C Arias², Denis Baurain¹, Nadine Cambisano¹, Stephen R Davis³, Frédéric Farnir¹, Bernard Grisart¹, Bevin L Harris², Mike D Keehan², Mathew D Littlejohn⁴, Richard J Spelman², Michel Georges¹ & Wouter Coppieters¹

We report mapping of a quantitative trait locus (QTL) with a major effect on bovine stature to a ~780-kb interval using a Hidden Markov Model-based approach that simultaneously exploits linkage and linkage disequilibrium. We re-sequenced the interval in six sires with known QTL genotype and identified 13 clustered candidate quantitative trait nucleotides (QTNs) out of >9,572 discovered variants. We eliminated five candidate QTNs by studying the phenotypic effect of a recombinant haplotype identified in a breed diversity panel. We show that the QTL influences fetal expression of seven of the nine genes mapping to the ~780-kb interval. We further show that two of the eight candidate QTNs, mapping to the *PLAG1-CHCHD7* intergenic region, influence bidirectional promoter strength and affect binding of nuclear factors. By performing expression QTL analyses, we identified a splice site variant in *CHCHD7* and exploited this naturally occurring null allele to exclude *CHCHD7* as single causative gene.

Deciphering the genetic architecture of complex traits is a key objective in biomedical sciences. Despite the identification of numerous QTLs by linkage mapping and more recently by genome-wide association studies (GWAS), our understanding of complex traits remains fragmentary. Mapping studies have confirmed the polygenic nature of most complex traits, but the identified loci usually explain only a fraction of the heritability. Moreover, causative genes and sequence variants (QTNs) have only been identified for a handful of QTLs. Advancing our understanding of complex traits will require the implementation of new experimental strategies that take advantage of the growing arsenal of genome exploration tools^{1–3}.

Stature (height) is a classical quantitative trait that has received attention from geneticists for more than a century^{4,5}. Understanding the molecular basis of inter-individual variation in stature might not only inform us about the general architecture of complex traits but may also provide insights into the mechanisms controlling organismal growth. Stature is known to affect predisposition to certain disorders and is linked to productivity in farm animals. Although sensitive to environmental factors, stature is usually highly heritable, including in humans (~85%)⁶. Nevertheless, GWAS efforts in humans have not identified any single variant with a major effect on height^{2,7}. A meta-analysis of 183,727 individuals identified 180 loci explaining ~10% of the phenotypic variance in human height and suggested that ~500 comparable loci explaining an additional ~5% exist⁸. Indeed, human stature appears to have quasi-infinitesimal architecture, as SNP-based estimates of genome-wide identity by descent explain >50% of phenotypic covariance⁹. In contrast, only six loci (including *IGF1*) account

for the majority of the difference in stature between dog breeds^{10,11}. This difference results, at least in part, from the creation of dog breeds by selection for outlier phenotypes that would be considered genetic defects in humans or be deleterious in natural populations.

Stature was one of the first traits affected by the domestication of cattle. With a shoulder height of ~2 m, the extinct auroch (*Bos primigenius*) was much larger than its domesticated descendents. Present-day cattle have shoulder height, ranging from ~1.1–1.5 m. Heritability estimates range from ~25–85%, depending on population^{12,13}. Recently, many QTLs affecting different components of bovine stature have been mapped, supporting a polygenic determinism (see URLs). For example, one study reported mapping of 38 and 52 QTLs explaining >60% of variation for, respectively, adult height and weight in American Angus cows¹⁴.

Here we exploit the population structure of domestic cattle in combination with high-density SNP genotyping, high-throughput sequencing and transcriptome analyses to identify regulatory QTNs that influence stature by modulating the expression of a chromosome domain encompassing *PLAG1*.

RESULTS

A QTL with major effect on stature maps to chromosome 14

To identify QTLs affecting traits important to the dairy sector, we generated a Holstein-Friesian (HF) × Jersey (J) line cross comprising 864 F2 cows (Supplementary Fig. 1). We measured >500 traits, of which six pertained to stature: weight at birth and at 6, 8, 12, 18 and >24 ('live weight') months, as well as height at the withers at 18 months.

¹Unit of Animal Genomics, Interdisciplinary Institute of Applied Genomics (GIGA-R) and Faculty of Veterinary Medicine, University of Liège (B34), Liège, Belgium.

²Livestock Improvement Corporation (LIC), Hamilton, New Zealand. ³ViaLactia BioSciences, Auckland, New Zealand. ⁴DairyNZ, Hamilton, New Zealand.

⁵These authors contributed equally to this work. Correspondence should be addressed to M.G. (michel.georges@ulg.ac.be).

We genotyped the entire pedigree for 294 microsatellites¹⁵. When applying a line-cross model¹⁶ to weight and height, we obtained genome-wide significant QTLs on chromosomes 2, 5, 7, 8, 10, 14, 25 and 28 (Fig. 1a). The strongest signal was on chromosome BTA14.

We increased the microsatellite density on BTA14 from 8 to 56 and analyzed the data with the HSQM software¹⁷, treating the data as six independent paternal half-sibling pedigrees. Across-family analysis yielded the location scores shown in Figure 1b, confirming the presence of a QTL affecting body size. The effect on weight was detectable at birth, and its significance increased with age. It was accompanied by a co-localized effect on height. There was no evidence for an effect on the ratio of live weight to height, indicating that the QTL affects stature. Bootstrap analysis¹⁸ produced an 18.4-Mb confidence interval (CI) (9.0–27.4 Mb (UMD 3.0 genome build)). Within-family analyses yielded evidence for QTL segregation in four of the six sire families, suggesting (under a biallelic QTL model) that sires 1, 2, 3 and 5 were Qq, whereas sires 4 and 6 were either QQ or qq (Fig. 1c). QTL position and allele substitution effects were consistent across Qq sires (Fig. 1d): ~20 kg for live weight and ~2 cm for height, corresponding to differences between full-grown QQ and qq

animals of ~40 kg and ~4 cm. Of note, QTLs influencing weight and daily gain have been identified in the same BTA14 region in at least six other cattle populations^{19–23}. Using haplotype sharing, researchers from a previous study²⁴ fine mapped one of these to a 1.1-Mb interval encompassed by our confidence interval (Fig. 1b).

Association analysis localizes the QTL to a ~780-kb interval

We genotyped the entire HF × J cross for 629 public-domain and 296 previously unreported BTA14 SNPs. We identified the latter by re-sequencing coding exons and highly conserved elements in the six F1 sires. We performed combined linkage and linkage disequilibrium (LD) analyses of live weight (the most significant trait) using Phasebook²⁵. Location scores were maximized at position 25,503,811 (UMD 3.0) (single marker analysis; log₁₀ odds (LOD) = 31.4) or 25,008,703 (UMD 3.0) (haplotype-based analyses; LOD = 34.1), that is, ~2.3 Mb proximal to the 1.1-Mb region previously defined²⁴ (Fig. 2). Fitting the most likely QTL position in the model and rescanning the chromosome did not show evidence for a second QTL (Supplementary Fig. 2). At this stage, none of the tested SNPs or haplotypes was concordant with the QTL segregation status of the six F1 sires.

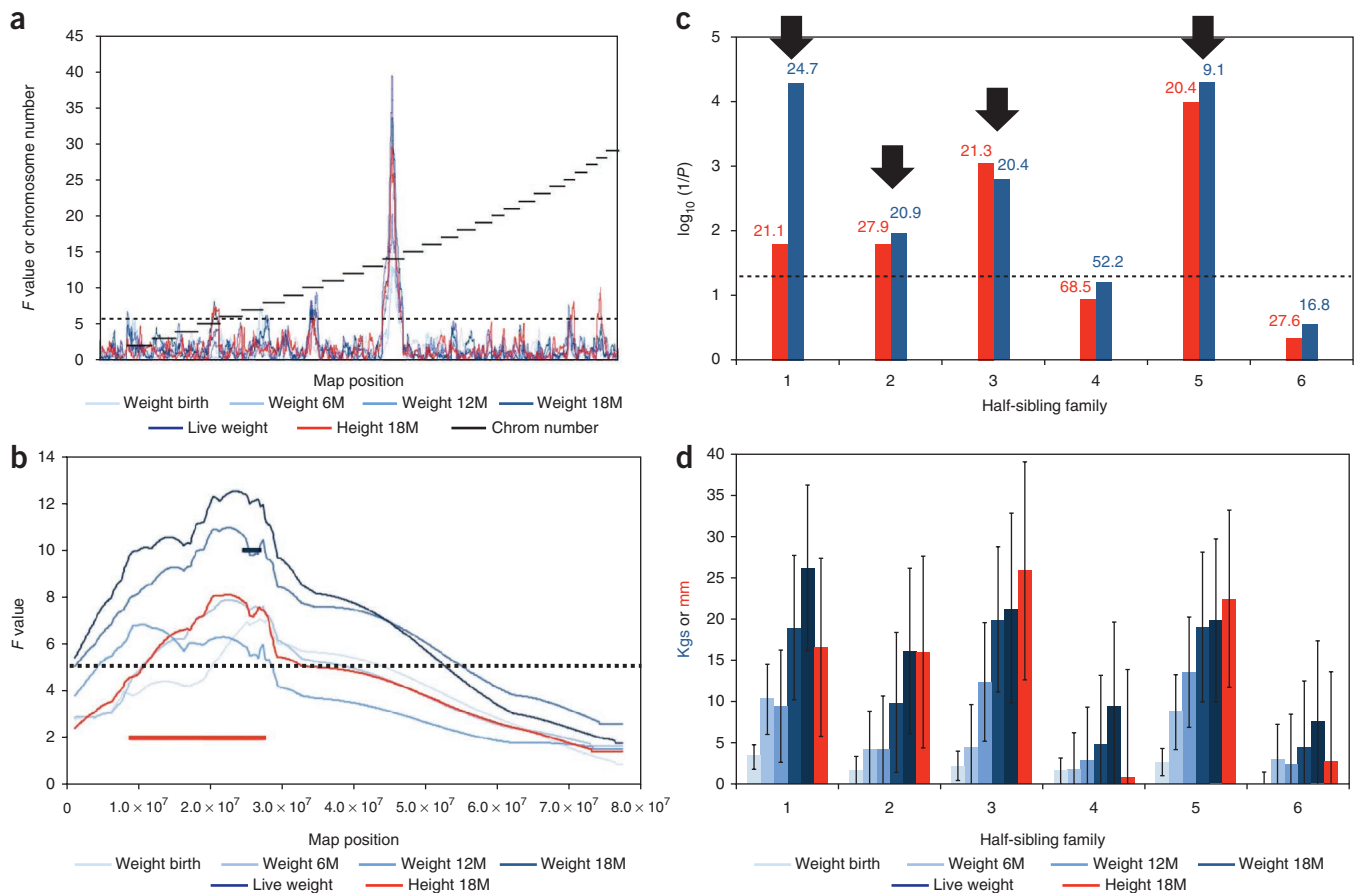


Figure 1 Linkage mapping of a body size QTL in an HF × J F2 cross. (a) Location scores obtained across the entire bovine genome for weight (blue lines) and height (red line) using a line-cross model¹⁶. The black dotted line corresponds to the 5% genome-wide significance threshold. (b) Location scores for BTA14 when analyzing weight (blue lines) and height (red line) using a paternal half-sibling pedigree model implemented with HSQM¹⁷. The black dotted line corresponds to the 5% chromosome-wide significance threshold. The red horizontal bar corresponds to the 95% CI for the QTL location (live weight). The black horizontal bar shows the position of the 1.1-Mb critical interval defined in a previous study²⁴. (c) Highest chromosome-wide log₁₀(1/P) values for each of the six sire families for height (red) and live weight (blue). The corresponding map positions are given above the bars. The black dotted line marks the 5% chromosome-wide significance threshold. The black arrows point toward the four sire families segregating for the QTL. (d) Sire-specific allele substitution effects on weight (blue bars) and height (red bar) expressed in kg and mm, respectively. Family-specific allele substitution effects were determined at the most significant QTL position in the across-family analysis. Error bars, 95% CI of the slope estimate (β_1), computed using standard procedures ($\beta_1 \pm t(0.025, n - 2) \times SE_{\beta_1}$). Chrom, chromosome.

Next, we evaluated the effect of chromosome 14 on body size in New Zealand (NZ) outbred dairy cattle, taking advantage of 3,570 progeny-tested bulls genotyped with the Illumina BovineSNP50 assay²⁶. The phenotypes were the breeding values for female live weight estimated using a standard animal model²⁷. We conducted a combined linkage and LD analysis using Phasebook²⁵. The most significant QTL mapped to chromosome 14: we obtained a maximum likelihood ratio test (LRT) statistic of 156 (LOD = 34.0) at position 25,215,026 (UMD 3.0), which is the same location as in the HF × J cross despite the use of a distinct SNP panel (Fig. 2), strengthening our confidence in the location of the QTL. **Supplementary Figure 3** shows the effect on breeding values for live weight of the different haplotype clusters with their respective frequencies in NZ dairy cattle. The bimodal distribution of haplotype effects supported a biallelic QTL. The q allele associated with lower weight and height was virtually fixed in J cows (97.6%), whereas the Q allele associated with higher weight and height predominated in HF cows (85.9%). Of note, J animals are considerably smaller than HF animals. Fitting the most likely QTL position in the model and scanning BTA14 for additional QTL effects did show modest evidence (LOD = 4.3) for a second QTL in the interval previously reported²⁴ (**Supplementary Fig. 2**). We also identified another QTL affecting stature on BTA5

(**Supplementary Fig. 4**). For subsequent analyses, we focused on a ~780-kb segment spanning all most likely QTL positions obtained in the different analyses (24,787,250–25,568,153; UMD 3.0) and encompassing the region of overlap (24,892,673–25,284,167; UMD 3.0) between the 95% CI (LOD dropoff of 2) obtained in the HF × J F2 and NZ outbred populations.

Sequencing sire chromosomes identifies 13 candidate QTNs

We generated 103 PCR products spanning the ~780-kb interval. We obtained products of the expected size from HF and J DNA, confirming the accuracy of the local sequence assembly and indicating that HF and J alleles could be amplified with comparable efficacy. The 103 PCR products were amplified from DNA of the six F1 sires, pooled, size fractionated and appended with adaptors for massive parallel re-sequencing on a Roche FLX instrument. The adaptors included individual-specific multiplex identifiers, allowing pooling and simultaneous sequencing of the eight libraries (HF, J and six F1 sires). We analyzed the traces using the GS Reference Mapper. The average sequence depth of non-repetitive sequences was ~20× per individual. From these data, we identified 9,572 putative DNA sequence variants (DSVs), or an average nucleotide diversity of ~0.3%. Assuming that the QTL is biallelic (as suggested by the bimodal distribution

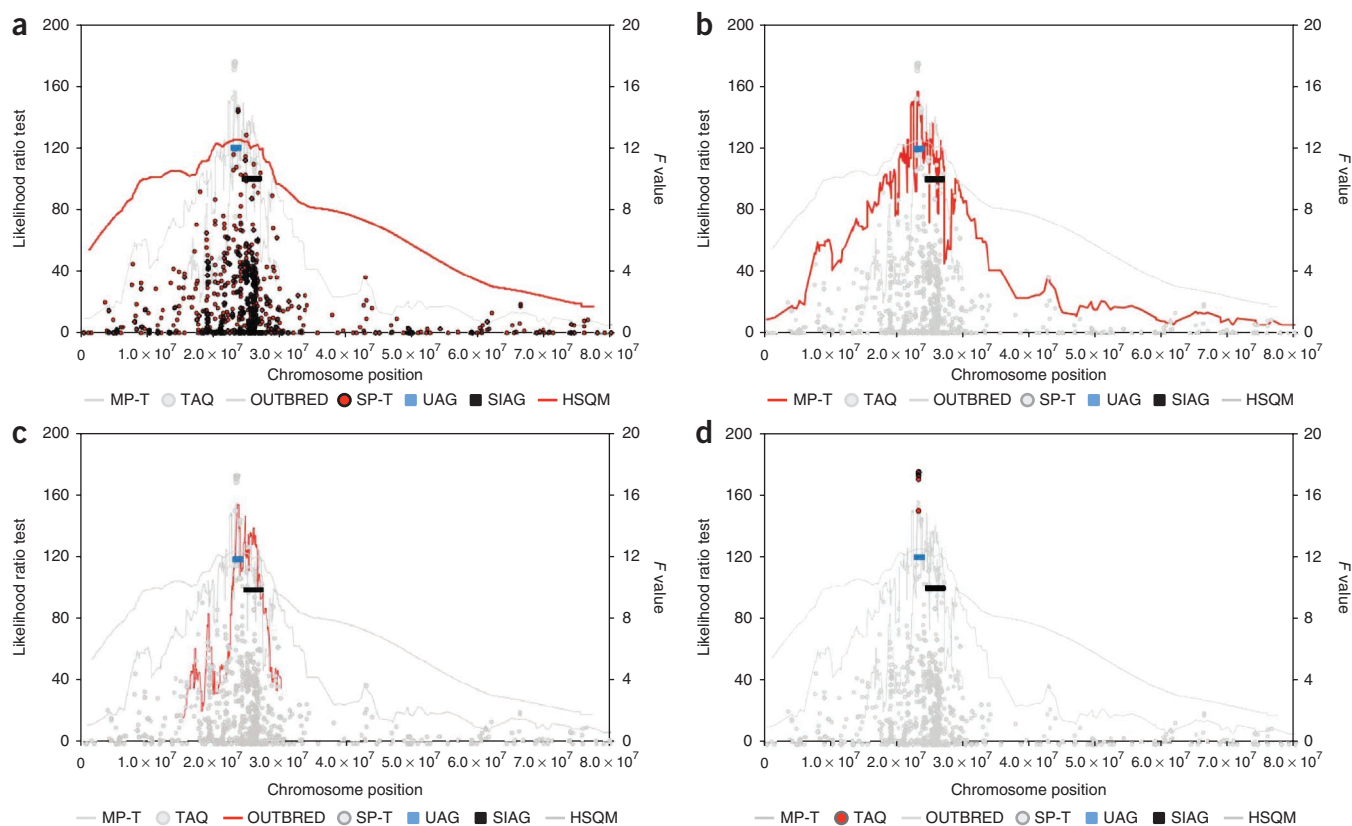


Figure 2 Linkage and LD fine-mapping of the body size QTL. (**a–d**) The x axes correspond to chromosomal positions in bp. The black horizontal lines correspond to the QTL candidate interval previously defined²⁴. The blue horizontal lines correspond to the 780-kb segment sequenced in the present study. Every graph shows the results of all analyses in gray watermarks to facilitate cross comparison. (**a**) The red line shows a linkage-based QTL analysis of live weight in the HF × J intercross using HSQM¹⁷ and 56 microsatellites (see Fig. 1b) (*F* values, right axis). Red dots show single-point linkage plus LD analysis of live weight in the HF × J intercross using Phasebook²⁵ and 925 SNPs plus 56 microsatellites (likelihood ratio test, left axis). (**b**) The red line shows haplotype-based linkage plus LD analysis of live weight in the HF × J intercross using Phasebook and 925 SNPs plus 56 microsatellites (likelihood ratio test, left axis). (**c**) The red line shows haplotype-based linkage plus LD analysis of breeding value for live weight in 3,570 progeny-tested sires from the NZ outbred dairy cattle population using Phasebook and 293 SNPs from the Illumina BovineSNP50 assay²⁶ spanning a 15-Mb BTA14 segment (likelihood ratio test, left axis). (**d**) The red dots show single-point linkage plus LD analysis of live weight in the HF × J intercross using Phasebook and 11 candidate QTN identified by resequencing the 780-kb candidate interval in the six F1 sires.

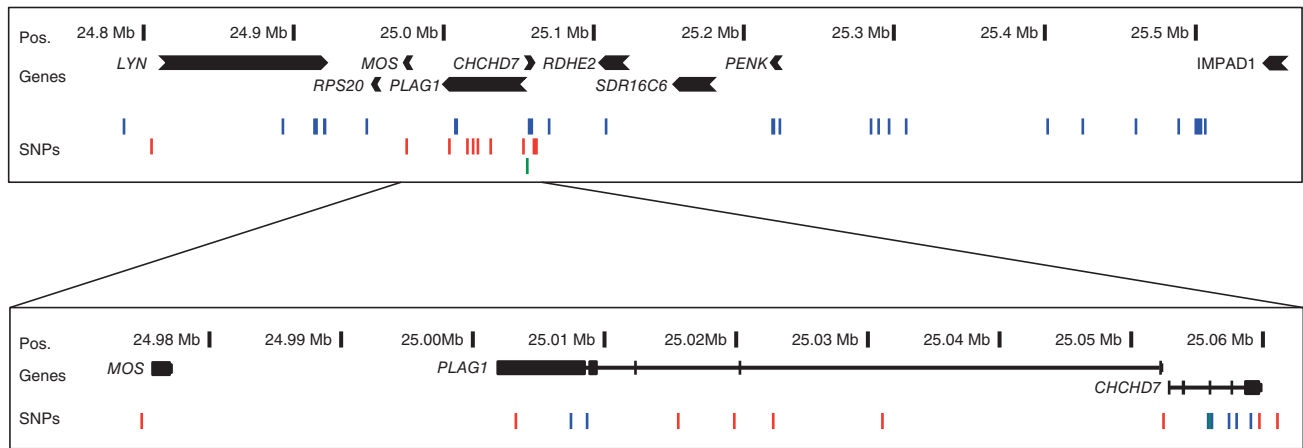


Figure 3 Annotated genes and markers within the re-sequenced ~780-kb QTL interval. The ‘Genes’ track shows the organization of the nine genes mapping to the ~780-kb critical region (*LYN*, *RPS20*, *MOS*, *PLAG1*, *CHCHD7*, *SDR16C5* (*RDHE2*), *SDR16C6*, *PENK* and *IMPAD1*). The ‘SNPs’ track shows (i) in blue, the position of 36 out of the 925 SNP panel used to fine map the QTL by combined linkage plus LD mapping, (ii) in red, the position of the 14 candidate QTN with segregation vector matching the QTL genotypes of the six F1 sires, and (iii) in green, the *CHCHD7* ss319607409 splice site variant.

of haplotype effects), the causative QTNs have to be heterozygous in the four Qq F1 sires and homozygous in the two non-segregating F1 sires. Applying this filter to the 9,572 DSVs yielded 14 candidate QTNs (13 SNPs and one VNTR; **Supplementary Table 1**). All but one of these clustered in an 87-kb segment spanning *MOS*, *PLAG1* and *CHCHD7* (**Fig. 3**).

Nine of the DSVs could be genotyped in the F2 generation. The DSVs in the 87-kb window proved to be in perfect LD with each other ($D' = 1$, $r^2 \geq 0.985$) but not with the isolated SNP ($0.859 \leq D' \leq 0.870$, $0.728 \leq r^2 \leq 0.746$). Single-marker association for live weight yielded an LRT of 174.7 for the 87-kb cluster, which is 17.9 LRT (3.9 LOD) units above the best previous signal. The signal for the isolated SNP was below that obtained in previous haplotype-based analyses (**Fig. 2**), suggesting that the 13 clustered DSVs encompass the causative QTNs. The same markers also yielded the strongest signals for the other body-size phenotypes, supporting the argument that the same QTNs account for all observed QTL effects (**Supplementary Fig. 5**). Notably, the orthologous region centered around *PLAG1* has also been shown to influence height in humans^{7,8,28–33}.

Exploiting haplotype diversity excludes 5 of 13 candidate QTNs

The fact that the 13 candidate QTNs were in complete LD in the F2 animals precluded further genetic differentiation of causative variants from passenger variants in this pedigree. To overcome this, we genotyped 159 sires representing 12 breeds for 12 of the 13 candidate QTNs. Whereas the two haplotypes (Q and q) observed in the HF × J cross accounted for 93.8% of the chromosomes, we detected ten additional recombinant haplotypes. One of these (R), carrying the QTN Q allele at 5 out of 13 positions and the q allele at 7 out of 13 positions, had a frequency of 37% in Simmental cattle (**Supplementary Fig. 6a**). We collected and genotyped DNA from 44 unrelated Simmentals with birth weight information, and we manually determined the phase of the 44 animals. The frequency of the Q, q and R haplotypes was 22%, 35% and 34%, respectively. The remaining 8% corresponded to four minor haplotypes. We regressed birth weight on QTL genotype assuming that the R haplotype was Q or q. Although we observed no effect on birth weight under the first hypothesis, we observed a near significant ($P = 0.06$) substitution effect with predicted sign under the second. The difference in birth weight between predicted qq and Qq animals (93% of the sample) was significant ($P = 0.02$; **Supplementary**

Fig. 6b). These results suggested that the five Q positions of the R haplotype could be excluded, leaving eight candidate QTNs.

Unaffected open reading frames support a regulatory effect

The candidate QTNs span three genes (*MOS*, *PLAG1* and *CHCHD7*; **Fig. 3**). Although none of these QTLs reside in the coding regions of *MOS*, *PLAG1* or *CHCHD7*, they could affect splicing, thereby altering coding capacity. To examine this possibility, we collected tissue samples (muscle, bone, brain and liver) from 79 outbred fetuses (**Supplementary Table 2**). We selected two QQ and two qq fetuses and performed RT-PCR using overlapping amplicons spanning the *MOS*, *PLAG1* and *CHCHD7* coding regions. We obtained no evidence for genotype-specific RT-PCR products (**Supplementary Fig. 7**), suggesting that the QTNs are regulatory, meaning they affect the expression of the causative gene(s) rather than their structure.

The QTNs affect expression of a regulon comprising seven genes

The QTNs could affect the expression of not only the spanned *MOS*, *PLAG1* and *CHCHD7* genes but also more distant genes. We thus examined the effect of QTN genotype on the expression level of the nine genes in the ~780-kb interval by performing quantitative RT-PCR using RNA from the 79 fetuses. For *PLAG1* and *CHCHD7*, we also performed allelic imbalance assays taking advantage of candidate QTNs located in the 3' untranslated regions (UTRs) (rs109231213 and ss319607407). We amplified segments encompassing the respective SNPs from genomic and complementary DNA (cDNA) of heterozygous fetuses and compared allelic ratios by direct sequencing and PeakPicker³⁴.

Except for the two outermost genes (*LYN* and *IMPAD1*), we observed significant effects of QTN genotype on transcript levels of the seven other genes in at least one tissue (**Fig. 4** and **Supplementary Fig. 8**). The Q allele was associated with increased expression (average 1.2-fold) in 31 of the 38 assays (with only 19 expected by chance; $P < 10^{-4}$). We observed significant effects for all four examined tissues. The QTN effect might extend over a longer chromosome segment in brain (from *RPS20* to *PENK*) than in the other tissues (from *PLAG1* to *PENK*).

These findings suggest that the QTNs affect one or more long range *cis*-acting elements that regulate the expression of *RPS20*, *MOS*, *PLAG1*, *CHCHD7*, *SDR16C5* (also known as *RDHE2*), *SDR16C6* and *PENK* (but not *LYN* and *IMPAD1*) in multiple tissues. It is noteworthy that synteny is conserved in zebrafish (chromosome 7) for *rps20*,

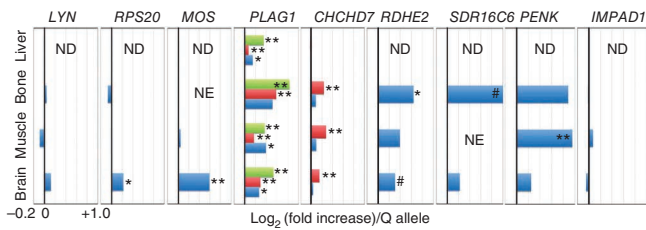


Figure 4 Effect of QTN genotype on the expression level of nine positional candidate genes in fetal liver, bone, muscle and brain. Blue bars, quantitative RT-PCR; red bars, allelic imbalance test using 3' UTR SNPs; green bars, allelic imbalance test using an intronic SNP. The x axis measures the slope of the regression (quantitative RT-PCR) or the ratio of the Q allele over the q allele (allelic imbalance tests) on a \log_2 scale. The vertical black lines correspond to the absence of an effect of QTN genotype on expression. #, $P < 0.10$; *, $0.01 < P < 0.05$; **, $P < 0.01$; ND, not done; NE, no detectable expression.

plag1, *chchd7*, *sdr16c5* (also known as *rdhe2*) and *penk* but not for *lyn* and *impad1*. The co-regulation of multiple genes implies a transcriptional effect. Consistent with this hypothesis, we observed the same degree of QTN-dependent allelic imbalance when assaying *PLAG1* pre-mRNAs by means of the ss319607402 intronic QTN candidate (Fig. 4 and Supplementary Fig. 6b).

Reporter and gel shift assays support the causality of two QTNs
Which of the eight candidate QTNs cause the observed effect? Two candidates (ss319607405 and ss319607406) affect a highly conserved element located in the intergenic region separating *PLAG1*

and *CHCHD7* (Fig. 5a): ss319607405 (phastcons score, 0.999) is a $(CCG)_n$ trinucleotide repeat with either 9 (q) or 11 (Q) copies located immediately upstream of the presumed *PLAG1* transcriptional start site, whereas ss319607406 (phastcons score, 0.996) is an A (q) to G (Q) SNP located 12-bp upstream from ss319607405. Of note, *PLAG1* and *CHCHD7* are positioned head-to-head, separated by only ~500 bp thought to encompass a bi-directional promoter³⁵. Of the six remaining candidates, only rs109231213, located in the 3' UTR of *PLAG1*, also affects a highly conserved element (phastcons score, 0.905). Two candidate QTNs (ss319607399 and ss319607400) map to short interspersed repetitive element (SINE) repeats and are therefore less likely to be causal (Supplementary Table 1).

We tested the effect of ss319607405 and ss319607406 on the promoter activity of the *PLAG1-CHCHD7* intergenic region. We cloned both allelic forms (Q and q) of a 378-bp and a 659-bp fragment centered around the two candidate QTNs in both orientations into the pGL4 luciferase vector (Fig. 5a). We transfected COS-1 cells and measured luciferase activity after 24 h. Compared to the promoterless vector, both fragments increased luciferase activity: luciferase activity increased ~9-fold (short) and ~20-fold (long) in the *PLAG1* direction and ~90-fold (short) and 44-fold (long) in the *CHCHD7* direction (Fig. 5b). Most importantly, the level of luciferase activity was systematically higher with the Q constructs than with the q constructs, with a difference of ~1.5-fold, a magnitude comparable to that observed *in vivo* (Fig. 5b). To determine the relative contribution of ss319607405 and ss319607406, we generated recombinant constructs, meaning Q for ss319607405 and q for ss319607406 or vice versa (Fig. 5c). Although both variants appeared to be necessary to recapitulate the full expression QTL

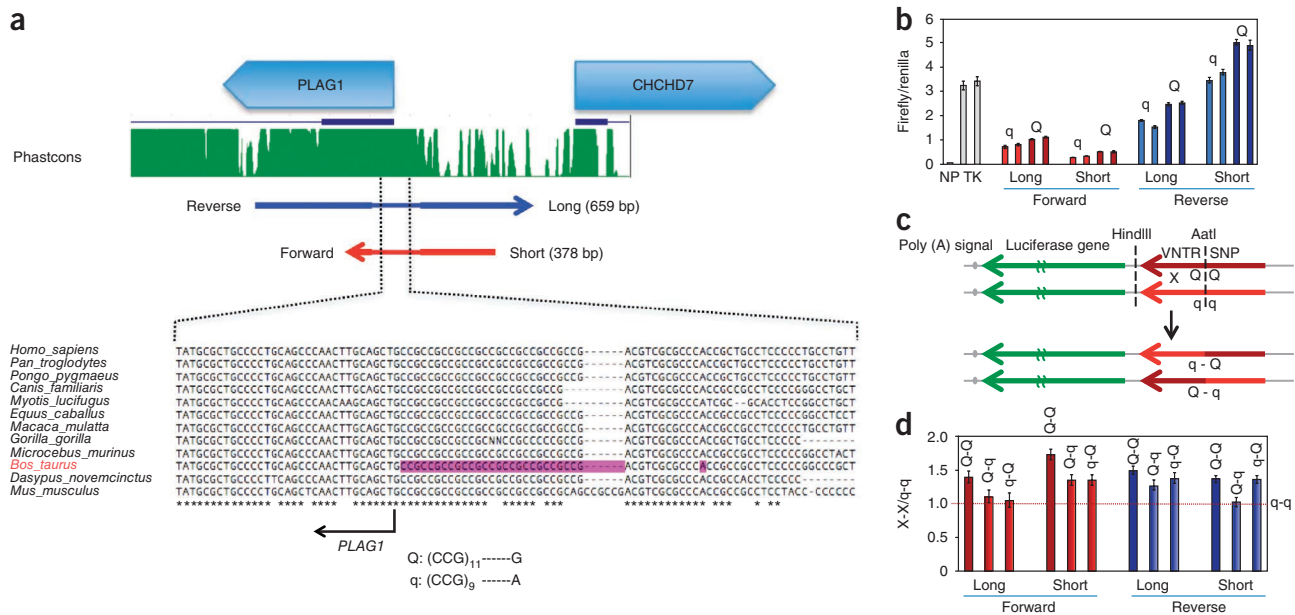


Figure 5 Effects of ss319607405 and ss319607406 on bidirectional promoter strength using a luciferase reporter assay. (a) Schematic representation of the supposedly bidirectional promoter driving expression of the head-to-head oriented *PLAG1* and *CHCHD7* genes (blue), with corresponding Phastcons conservation scores (green) and multispecies sequence alignment of a segment encompassing the ss319607405 and ss319607406 candidate QTNs. The arrows mark the positions of the 'long' and 'short' fragments cloned in the pGL4 luciferase reporter vector in the 'forward' (toward *PLAG1*) and 'reverse' (toward *CHCHD7*) orientation. (b) Ratios of firefly to renilla luminescence obtained after transfection of COS-1 cells with (i) a promoterless pGL4 vector (NP), (ii) two distinct, sequence-verified preparations of the pGL4 vector endowed with the thymidine kinase promoter (TK), (iii) pairs of sequence-verified preparations of the pGL4 vector endowed with the q or Q version of the long or short fragment cloned either in forward (reddish) or reverse (bluish) orientation. Error bars correspond to the s.e.m. computed from replicates. (c) Schematic representation of the recombinant 'Q-q' and 'q-Q' promoter fragments that were generated by swapping the Q and q residues at the ss319607405 and ss319607406 sites as shown. (d) Ratios of firefly to renilla luminescence obtained with the non-recombinant 'Q-Q' promoter as well as recombinant 'Q-q' and 'q-Q' promoters cloned in forward and reverse orientation relative to the cognate non-recombinant 'qq' promoter. Error bars correspond to s.e.m.

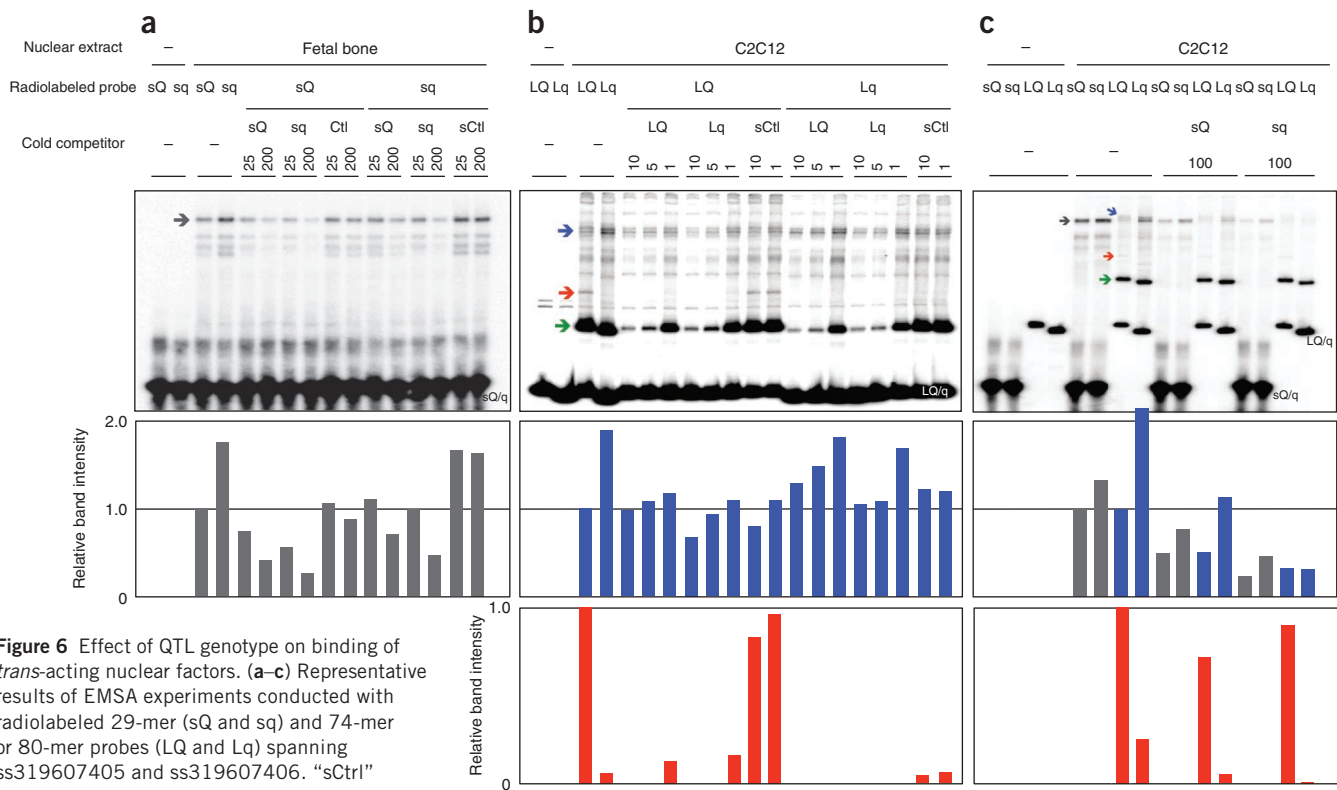


Figure 6 Effect of QTL genotype on binding of *trans*-acting nuclear factors. (a–c) Representative results of EMSA experiments conducted with radiolabeled 29-mer (sQ and sq) and 74-mer or 80-mer probes (LQ and Lq) spanning ss319607405 and ss319607406. “sCtrl” corresponds to an unrelated 25-mer control duplex used as control competitor. Results shown were obtained with nuclear extracts derived from fetal bone (a) and C2C12 cells (b,c). Complexes with differential affinity for the Q and q allele are marked by the gray (short probe; Q<q), blue (long probe; Q<q) and red (long probe; Q>q) arrows. An abundant complex with equal affinity for the Q and q allele detected with the long probe is marked by the green arrow. Free probes are labeled “sQ/q” and “LQ/q”. The bar graphs at the bottom of the figure quantify the abundance of two corresponding complexes (color coded accordingly) relative to the Q probe (in the absence of cold competitor) determined by densitometry.

(eQTL) effect in the *PLAG1* orientation, ss319607406 was predominant in the *CHCHD7* orientation (Fig. 5d).

To examine whether the effect of ss319607405 and ss319607406 might reflect differential binding of *trans*-acting factors, we performed electrophoretic mobility shift assays (EMSA). We first used radiolabeled double-stranded 29-mers centered around ss319607406 and nuclear extracts from COS-1, C2C12 and ATDC5 cells, as well as extracts from fetal bone, brain, muscle and liver. With all extracts, we observed a number of complexes (some with the same mobility across tissues and some that were tissue specific) that were systematically 1.2 to 3 times more abundant with the q than with the Q duplex. Moreover, unlabeled cold q duplex tended to be more proficient than cold Q duplex in displacing either of the radiolabeled probes (Fig. 6). Next, we used 74-mer (q) or 80-mer (Q) probes encompassing both ss319607405 and ss319607406. We detected large complexes that were ~2.5 times more abundant with the q probe than with the Q probe. These complexes were displaced by cold 29-mers (q was more effective than Q) and were thus related to the complexes detected with the 29-mer sequence alone. In addition, with C2C12 extracts, we detected a Q-specific complex that was not displaced by the 29-mer (Fig. 6).

These findings suggest (i) that distinct regulatory complexes assemble on the Q and q sequences, (ii) that this likely explains their different promoter strength in reporter assays, and (iii) that this contributes to the observed differential transcription rate of the Q and q alleles *in vivo*. Complex assembly in the *PLAG1-CHCHD7* intergenic segment is likely mediated by broadly expressed *trans*-acting factors, as we detected the complexes in all cell types examined. The fact that the expression levels of genes other than *PLAG1* and *CHCHD7*

are influenced by QTN genotype suggests either that the regulatory element affected by ss319607405 and ss319607406 controls the expression of more distant genes or that other QTNs are also involved.

A naturally occurring null allele excludes *CHCHD7*

Which of the seven genes affected by the QTNs influence stature? *RPS20* encodes ribosomal protein S20. *Rps20* mutations in mice cause skin darkening with pleiotropic effects, including reduced body size³⁶. *MOS* encodes a protein kinase that is specifically expressed in oocytes and controls meiotic maturation; ectopic expression of *MOS* in somatic cells induces oncogenic transformation³⁷. *PLAG1* is an oncogene whose ectopic expression due to translocation-induced promoter swapping with ubiquitously expressed genes (such as *β-catenin*) causes pleomorphic adenomas³⁸. *PLAG1* encodes a transcription factor that is broadly expressed during fetal development and which is downregulated at birth. *PLAG1* regulates several growth factors, including *IGF2*, a key regulator of body size^{38,39}. The most striking phenotype of *Plag1* knockout mice is growth retardation⁴⁰. *CHCHD7* encodes a widely expressed protein of unknown function. It was identified as a *PLAG1* fusion partner in tumors of the salivary gland³⁸. *SDR16C5* (also called *RDHE2*) and *SDR16C6* encode members 5 and 6 of the 16C family of short-chain alcohol dehydrogenases and reductases. *RDHE2* catalyzes the first and rate-limiting step in generating retinal from retinol. Imbalances in endogenous retinoids perturb development and affect growth⁴¹. *PENK* encodes the precursor of met- and leu-enkephalins, which play a role in pain perception and response to stress⁴². Although the candidacy of *PLAG1* seems strongest and is supported by GWAS signals in

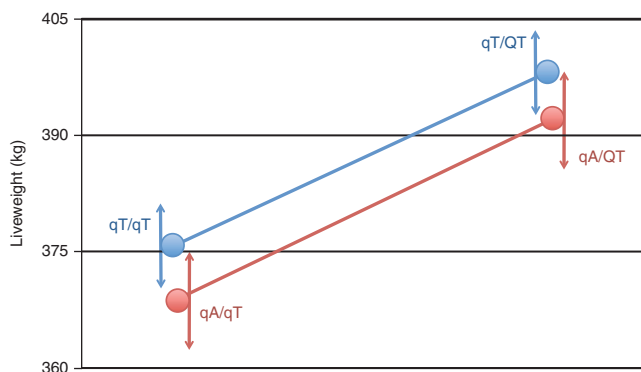


Figure 7 Exploiting a naturally occurring null allele to exclude the causality of the *CHCHD7* gene. Effect on liveweight (kg) of composite QTN (q versus Q) ss319607409 (T versus A) genotype. Circles correspond to mean liveweight of animals sorted by genotype category: qT/qT, qT/QT, qA/qT and qA/QT. Arrows correspond to twice the standard error (SE) of the means. Genotype means and s.e.m. were estimated using a mixed model including a random individual animal effect. The figure illustrates the significance of the q/q to q/Q QTN substitution effect in both T/T and A/T animals and the lack of significance and equivalence of the T/T to T/A ss319607409 substitution effect in both q/q and q/Q animals.

humans, available evidence neither proves *PLAG1*'s causality nor disproves a contribution of one or more other genes.

To gain insights into the causality of the positional candidate genes, we took advantage of genome-wide expression data generated for 429 of the F2 animals by hybridizing liver and adipose cDNA on Affymetrix Bovine 24K expression arrays (S.R.D., M.D.K. & R.J.S., unpublished data). The array included probes interrogating *RPS20*, *MOS*, *CHCHD7* and *SDR16C5* (*RDHE2*) but not *PLAG1* and *SDR16C6*. QTL mapping identified chromosome-wide significant eQTL effects for *CHCHD7* in liver and adipose and for *RPS20* in liver. Fine mapping positioned both eQTLs in the ~780-kb segment, suggesting *cis*-acting eQTLs (**Supplementary Figs. 9,10**).

Within-family analysis of the *CHCHD7* data showed that only F1 sire 1 was segregating for the eQTL, which is a pattern distinct from the stature QTL. Following the strategy used for the stature QTL, we found that the *CHCHD7* eQTL is due to a T to A donor splice-site variant causing skipping of exon 3. Mutant transcripts have a truncated open reading frame, terminating within exon 4 (of 5) and are predicted to undergo nonsense-mediated RNA decay. The corresponding *CHCHD7* variant is predicted to be a null allele (**Supplementary Note and Supplementary Fig. 11**).

Assuming that *CHCHD7* levels influence body size, the allele substitution effects should be larger in the offspring of sire 1 (AT at ss319607409) than in any of the other sire families (TT sires). However, QTL allele substitution effects in family 1 did not differ significantly from those in the other families (families 2, 3 and 5) segregating for the body-size QTL (**Fig. 1d**). Under the same assumption, ss319607409 should have an independent, *a priori* larger effect on body size because the effect on *CHCHD7* transcript levels is larger for the splice-site variant than for the QTNs, and residual transcripts are unlikely to yield functional *CHCHD7* (as ss319607409 perturbs splicing, it should affect *CHCHD7* levels in all tissues). We tested this by estimating the effect of ss319607409 on height and weight in the HF × J cross using a mixed model including the QTNs and an individual animal effect. The residual effect of ss319607409 was not significant (LRT +1.04). Also, assuming that ss319607409 creates a *CHCHD7* null allele and if *CHCHD7* is causally involved, the qA/qT

versus qA/QT phenotypic contrast is expected to be larger than the qT/qT versus qT/QT contrast. We performed this quantitative complementation assay (QCA)^{43–45} using a mixed linear model including QTN plus splice site genotype as fixed effect and random individual animal effect (**Fig. 7**). There was no evidence for an interaction between QTN and splice site genotype, that is, for a failure to complement. These results exclude *CHCHD7* as the only causative gene. The non-significant tendency of carriers of the splice site variant to be somewhat smaller (**Fig. 7**) leaves open the possibility that *CHCHD7* contributes in a minor way to the QTL effect (although this is not supported by the absence of a significant QCA).

Within-family analysis of the *RPS20* liver data indicated that sires 1, 5 and 6 were segregating for this eQTL, meaning they had a segregation pattern distinct from both the stature QTL and *CHCHD7* eQTL (**Supplementary Fig. 10**). Unlike *CHCHD7*, the tissue-specificity of this eQTL does not allow exclusion of *RPS20*. Analysis of the sequence traces did not reveal obvious candidate eQTNs.

DISCUSSION

Genetic studies have identified numerous loci influencing complex traits. However, confidence intervals typically encompass multiple genes and variants, and in very few cases, including examples in yeast⁴⁶, *Drosophila*⁴⁷ and livestock⁴⁸, have the causative genes and QTNs been identified. Often, candidate genes are prioritized for functional assays based on rudimentary knowledge about gene function or presumed effects of candidate causal variants, which is unlikely to provide comprehensive mechanistic understanding.

Here, we used approaches beyond LD-based fine mapping to narrow down the list of candidate causative QTNs and genes. Without relying on any *a priori* assumptions of gene function or variant effects, we excluded all but eight DSVs out of more than 9,572, or >1,000-fold enrichment. Features of livestock that allowed this include (i) the ability to use LD to fine-map the QTL in an F2-like cross (as multiple haplotypes still segregate in the parental lines), (ii) the ability to reliably genotype individuals for the QTL (as the phenotypic effect of a sire's chromosomes can be measured in many offspring), (iii) reduced allelic heterogeneity (because of low effective population size) and (iv) the ability to exploit between-breed haplotypic diversity. Studying effects of additional recombinant haplotypes may further reduce the list of candidate QTNs.

Progeny testing of chromosomes to determine QTL genotype before sequencing is obviously not feasible in humans. However, our Hidden Markov Model method²⁵ identified haplotype clusters with differential effects (**Supplementary Fig. 3**), which can become the focus of re-sequencing efforts.

QTL fine mapping in animals often uses haplotype sharing⁴⁹. This will be effective if one of the QTL alleles is characterized by haplotype homogeneity. If the causative mutations are old and shared by a clade of haplotypes, sharing methods may not work. The approach we followed would identify causative variants even if present on multiple haplotypes. Of note, the search for a shared haplotype was not effective in the initial fine-mapping stage. Retrospective analysis indicated that the Q chromosomes of the four segregating Qq sires indeed shared a haplotype spanning 355 kb and the 13 candidate QTNs. However, before the identification of the QTNs, the q chromosome of one of the four Qq sires carried a haplotype that was identical by state with the Q chromosome for all available markers. As this Qq sire appeared homozygous in the region, this information was not relied upon at that stage.

There are two genetic tests that can show gene causality. The first requires showing an enrichment of rare functional variants in individuals with extreme phenotypes⁵⁰. Because of the reduced

effective population size and reduced allelic heterogeneity typical of livestock, we did not explore this option. The second is the quantitative complementation assay (QCA)^{43–46}, which consists of showing an exacerbated q>Q allele substitution effect when the background chromosome carries a null allele at the tested gene (that is, in hemizygotes) rather than a wild-type allele. The QCA has been used in yeast⁴⁶, *Drosophila*⁵¹ and mouse⁵², organisms in which null alleles can be engineered experimentally. Here, we performed the QCA and related assays using a naturally occurring null allele and excluded *CHCHD7*. A compilation of frequently occurring null alleles could become a valuable resource to facilitate the use of the QCA in animal and even human populations. As a growing number of individual genome sequences are being determined in humans⁵³ and domestic animals, such collections will be available soon.

Among the remaining candidate genes, *PLAG1* appeals most because (i) *Plag1* knockout mice show dwarfism in the absence of other symptoms, (ii) *PLAG1* is known to affect levels of *IGF2* and other growth factors, and (iii) GWAS signals for human height map near *PLAG1*. Preliminary evidence suggests that QTN genotypes influence *IGF2* transcript levels, suggesting that QTN-dependent changes in *PLAG1* expression are sufficient to trigger downstream effects. However, other positional candidates, including *RPS20* and *SDR16C5* (*RDHE2*), have established physiological links with growth and should not be dismissed.

QTL genotype frequencies in the HF × J cross were 0.12 (QQ), 0.47 (Qq) and 0.41 (qq), corresponding to allelic frequencies of 0.36 (Q) and 0.64 (q). Tracing the origin of the QTL alleles indicated that Q had a frequency of 78% in the parental HF generation, and q had a frequency of 95% in the parental J generation. Both non-segregating F1 sires were qq. QTL genotype effects on live weight were +19.9 kg (QQ), 0 kg (Qq) and –23.5 kg (qq). Heterozygotes did not deviate significantly from the midpoint between alternate homozygotes, suggesting additivity. The QTL explained 9.9% of live weight variance in the F2 generation. It is difficult to estimate what proportion of the genetic variance this corresponds to. However, when compared to humans⁸, the effect is large and reminiscent of between-breed effects reported in dogs^{10,11}. This is not unexpected, as the F2 design targets QTLs underlying between-breed differences. Although the outbred NZ sires were not genotyped for the QTNs, the frequencies of the Q and q alleles could be deduced from haplotype frequencies (Supplementary Fig. 3). Q predominated in HF animals (85.9%), whereas q was virtually fixed in J animals (97.6%). The higher frequency of Q in outbred HF compared to HF F0 animals could be because of their different composition (outbred elite sires versus F0 cows). Using the average q>Q haplotype substitution effect estimated in the outbred sires, we calculated that the QTL explains 7.0% and 2.2% of the variance of breeding values for live weight in the pure-bred HF and J populations, respectively. If we conservatively assume that the heritability is 0.50, the QTL explains ~3.5% and ~1.1% of the phenotypic variance. These effects remain considerably larger than in humans, suggesting either that stabilizing selection is less effective in livestock (which could be because of their smaller effective size) or is an effect of directional selection. Preliminary results with the Illumina 777K Bovine HD chip (data not shown) suggest that SNPs other than the eight mapping to the ~780-kb interval may have minor, independent effects on stature in the HF purebred population. However, when assayed in the HF × J cross (i) their effects were at least 6.3 LOD score units lower than that of the eight candidate QTNs, (ii) at least one of the two non-segregating F1 sires was heterozygous, and (iii) at least one segregating F1 sire was homozygous, precluding exclusive causality in mediating the observed QTL effect.

In conclusion, we describe the dissection of a QTL with a major effect on bovine stature at near-nucleotide resolution and show that the QTNs influence the expression of multiple genes. We also performed a quantitative complementation assay using a naturally occurring null allele to exclude *CHCHD7* as the sole causative gene.

URLs. Bovine QTL maps, <http://www.animalgenome.org/cgi-bin/QTLdb/BT/index>.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturegenetics/>.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

This work was funded by Livestock Improvement Corporation (LIC; Hamilton, New Zealand) and by grants from the Unit of Animal Genomics, the University of Liège, the Communauté Française de Belgique (ARC Biomod) and the Belgian Science Policy Organisation (SSTC Genefunc PAI). T.D. is Research Associate of the Fond National de Recherche Scientifique. We are grateful for the support of the GIGA-R sequencing core facility.

AUTHOR CONTRIBUTIONS

J.A.C.A., B.L.H., M.D.K. and R.J.S. designed and performed line-cross QTL mapping in the F2 population. L.K., L.L., N.C., B.G. and W.C. developed additional BTA14 markers, genotyped the F2 population and performed half-sibling QTL mapping. T.D., F.F. and W.C. performed combined linkage and LD QTL fine mapping. L.K. and W.C. performed high throughput resequencing and analysis of the 780-kb confidence interval. L.L. performed sequence finishing of the 780-kb interval. L.K., N.C. and W.C. performed haplotype analysis in the breed diversity panel. S.R.D. collected fetal samples. L.K., H.T. and L.L. checked the integrity of the open reading frames. H.T., L.L., M.D.L. and M.G. performed quantitative RT-PCR experiments. H.T. performed the allelic imbalance tests. H.T. performed the reporter assays. L.K. and H.T. performed the EMSA. S.R.D., M.D.K. and R.J.S. generated and performed initial analysis of the transcriptome data. T.D., D.B. and W.C. performed eQTL analyses. L.L. analyzed the effect of the *CHCHD7* splice site variant. W.C. and T.D. performed the QCA. M.G. designed experiments, analyzed data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/naturegenetics/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

- Stern, D.L. & Orgogozo, V. Is genetic evolution predictable? *Science* **323**, 746–751 (2009).
- Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
- Schadt, E.E. Molecular networks as sensors and drivers of common human diseases. *Nature* **461**, 218–223 (2009).
- Galton, F. Regression towards mediocrity in hereditary stature. *J. R. Anthropol. Inst.* **5**, 329–348 (1885).
- Fisher, R.A. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**, 399–433 (1918).
- Visscher, P.M. *et al.* Genome partitioning of genetic variation for height from 11,214 sibling pairs. *Am. J. Hum. Genet.* **81**, 1104–1110 (2007).
- Weedon, M.N. & Frayling, T.M. Reaching new heights: insights into the genetics of human stature. *Trends Genet.* **24**, 595–603 (2008).
- Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
- Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
- Sutter, N.B. *et al.* A single *IGF1* allele is a major determinant of small size in dogs. *Science* **316**, 112–115 (2007).
- Boyko, A.R. *et al.* A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol.* **8**, e1000451 (2010).
- Nelsen, T.C. *et al.* Heritabilities and genetic correlations of growth and reproductive measurements in Hereford bulls. *J. Anim. Sci.* **63**, 409–417 (1986).
- Northcutt, S.L. & Wilson, D.E. Genetic parameter estimates and expected progeny differences for mature size in Angus cattle. *J. Anim. Sci.* **71**, 1148–1153 (1993).

14. McClure, M.C. *et al.* A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. *Anim. Genet.* **41**, 597–607 (2010).
15. Arias, J.A., Keehan, M., Fisher, P., Coppieters, W. & Spelman, R. A high density linkage map of the bovine genome. *BMC Genet.* **10**, 18 (2009).
16. Haley, C.S. *et al.* Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**, 1195–1207 (1994).
17. Coppieters, W. *et al.* A rank-based non parametric method to map QTL in outbred half-sib pedigrees: application to milk production in a grand-daughter design. *Genetics* **149**, 1547–1555 (1998).
18. Visscher, P.M. *et al.* Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**, 1013–1020 (1996).
19. Mizoshita, K. *et al.* Quantitative trait loci analysis for growth and carcass traits in a half-sib family of purebred Japanese Black (Wagyu) cattle. *J. Anim. Sci.* **82**, 3415–3420 (2004).
20. Takasuga, A. *et al.* Identification of bovine QTL for growth and carcass traits in Japanese Black cattle by replication and identical-by-descent mapping. *Mamm. Genome* **18**, 125–136 (2007).
21. Buchanan, F.C. *et al.* Single nucleotide polymorphisms in the corticotrophin-releasing hormone and pro-opiomelanocortin genes are associated with growth and carcass yield in beef cattle. *Anim. Genet.* **36**, 127–131 (2005).
22. Kneeland, J. *et al.* Identification and fine mapping of quantitative trait loci for growth traits on bovine chromosomes 2, 6, 14, 19, 21, and 23 within one commercial line of *Bos taurus*. *J. Anim. Sci.* **82**, 3405–3414 (2004).
23. Nkrumah, J.D. *et al.* Primary genome scan to identify putative quantitative trait loci for feedlot growth rate, feed intake, and feed efficiency of beef cattle. *J. Anim. Sci.* **85**, 3170–3181 (2007).
24. Mizoshita, K. *et al.* Identification of a 1.1-Mb region for a carcass weight QTL on bovine Chromosome 14. *Mamm. Genome* **16**, 532–537 (2005).
25. Druet, T. & Georges, M. An efficient haplotype-based approach for combined linkage and linkage disequilibrium QTL mapping using Hidden Markov Models. *Genetics* **184**, 789–798 (2010).
26. Matukumalli, L.K. *et al.* Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* **4**, e5350 (2009).
27. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits*. (Sinauer, Sunderland, Massachusetts, USA, 1997).
28. Gudbjartsson, D.F. *et al.* Many sequence variants affecting diversity of adult human height. *Nat. Genet.* **40**, 609–615 (2008).
29. Lettre, G. *et al.* Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat. Genet.* **40**, 584–591 (2008).
30. Soranzo, N. *et al.* Meta-analysis of genome-wide scans for human adult stature identifies novel loci and associations with measures of skeletal frame size. *PLoS Genet.* **5**, e1000445 (2009).
31. Cho, Y.S. *et al.* A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.* **41**, 527–534 (2009).
32. Kim, J.J. *et al.* Identification of 15 loci influencing height in a Korean population. *J. Hum. Genet.* **55**, 27–31 (2010).
33. Okada, Y. *et al.* A genome-wide association study in 19,633 Japanese subjects identified *LHX3-QSOX2* and *IGF1* as adult height loci. *Hum. Mol. Genet.* **19**, 2303–2312 (2010).
34. Ge, B. *et al.* Survey of allelic expression using EST mining. *Genome Res.* **15**, 1584–1591 (2005).
35. Trinklein, N.D. *et al.* An abundance of bidirectional promoters in the human genome. *Genome Res.* **14**, 62–66 (2004).
36. McGowan, K.A. *et al.* Ribosomal mutations cause p53-mediated dark skin and pleiotropic effects. *Nat. Genet.* **40**, 963–970 (2008).
37. Sagata, N. What does Mos do in oocytes and somatic cells? *Bioessays* **19**, 13–21 (1997).
38. Van Dyck, F. *et al.* PLAG1, the prototype of the PLAG gene family: versatility in tumour development. *Int. J. Oncol.* **30**, 765–774 (2007).
39. Voz, M.L. *et al.* Microarray screening for target genes of the proto-oncogene PLAG1. *Oncogene* **23**, 179–191 (2004).
40. Hensen, K. *et al.* Targeted disruption of the murine *Plag1* proto-oncogene causes growth retardation and reduced fertility. *Dev. Growth Differ.* **46**, 459–470 (2004).
41. Ross, S.A., McCaffrey, P.J., Drager, U.C. & De Luca, L.M. Retinoids in embryonal development. *Physiol. Rev.* **80**, 1021–1054 (2000).
42. Kieffer, B.L. *et al.* Exploring the opioid system by gene knock-out. *Prog. Neurobiol.* **66**, 285–306 (2002).
43. Georges, M. Mapping, fine-mapping and molecular dissection of Quantitative Trait Loci in domestic animals. *Annu. Rev. Genomics Hum. Genet.* **8**, 131–162 (2007).
44. Mackay, T.F. Quantitative trait loci in *Drosophila*. *Nat. Rev. Genet.* **2**, 11–20 (2001).
45. Service, P.M. How good are quantitative complementation tests? *Sci. SAGE KE* **12**, pe13 (2004).
46. Steinmetz, L.M. *et al.* Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**, 326–330 (2002).
47. McGregor, A.P. *et al.* Morphological evolution through multiple *cis*-regulatory mutations at a single gene. *Nature* **448**, 587–590 (2007).
48. Van Laere, A.S. *et al.* A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. *Nature* **425**, 832–836 (2003).
49. Nezer, C. *et al.* Haplotype sharing refines the location of an imprinted QTL with major effect on muscle mass to a 250-kb chromosome segment containing the porcine *IGF2* gene. *Genetics* **165**, 277–285 (2003).
50. Bodmer, W. & Bonilla, C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.* **40**, 695–701 (2008).
51. Long, A.D., Lyman, R.F., Langley, C.H. & Mackay, T.F. Genetic interactions between naturally occurring alleles at QTL and mutant alleles at candidate loci affecting bristle number in *D. melanogaster*. *Genetics* **144**, 1497–1510 (1996).
52. Yalcin, B. *et al.* Genetic dissection of behavioral QTL shows that *Rgs2* modulates anxiety in mice. *Nat. Genet.* **36**, 1197–1202 (2004).
53. 1000 Genomes Project Consortium. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).

ONLINE METHODS

Genotyping. Genotyping of the 56 microsatellite markers in the HF × J F2 pedigree was performed using standard procedures labeling the PCR products with fluorescent primers and size fractionating them on a ABI36730 capillary sequencer⁵⁴. Genotyping of the 925 BTA14 SNP markers in the HF × J F2 pedigree was performed with a custom-made Golden Gate assay on an Illumina Beadstation500 using standard procedures. Genotyping of the candidate QTNs was done using Taqman assays-by-design (**Supplementary Table 3**) using standard procedures (ABI).

QTL mapping and fine-mapping. QTL mapping of weight and height in the HF × J F2 population under a line-cross model was conducted as described¹⁶. QTL mapping in the HF × J population under a half-sibling pedigree model was conducted with HSQM¹⁷. Significance thresholds were determined by permutation testing⁵⁵. Confidence intervals for the QTL were determined by bootstrapping¹⁸. QTL fine mapping was achieved by simultaneously exploiting linkage and LD using a haplotype-based approach²⁵. The mixed model includes a mean (fixed), a haplotype effect (random), an animal effect (random) and an error term (random). The haplotype effect corresponds to haplotype clusters defined using a Hidden Markov model (HMM). The covariance between distinct clusters is assumed to be zero. The number of ancestral haplotype clusters was set at 20 in the presented analyses. Variance components and individual effects were estimated using AI-REML⁵⁶. Results are reported as $LRT = 2\ln LR$, where LR corresponds to the ratio between the likelihood of the data under H1 (model with haplotype effect) and likelihood of the data under H0 (model without haplotype effect). LRT is asymptotically distributed as a χ^2 variable with ~1 degree of freedom. In some instances, we fitted two QTL positions simultaneously. Evidence in favor of a QTL at position 2 (conditional on the presence of a QTL at position 1) was then evaluated from the increase in the LRT achieved by adding this second QTL in the model. The NZ outbred population was analyzed using the same mixed model as above with addition of a regression on the percentage of Jersey ancestry computed from available pedigree information (**Supplementary Fig. 12**). The association analysis was conducted using a mixed model including a mean, a regression on the number of alleles '1' for the considered SNPs (fixed), an animal effect (random) and an error (random). Variance components and effects were estimated by REML. The significance of the SNP effect was estimated using an *F* test. In some instances, the association analysis only considered the marker allele inherited from the mother.

For transcriptomic analysis, eQTL mapping in the HF × J population under a half-sibling pedigree model was conducted with HSQM¹⁷, as for the phenotypic data. The association analysis was conducted using a linear model including a mean, an experiment (1 or 2) effect (fixed), a regression on the number of alleles '1' for the considered SNPs (fixed) and an error term. The significance of the SNP effect was estimated using an *F* test. In some instances, we fitted two distinct SNP effects simultaneously. The significance of the second SNP effect conditional on the first one was estimated by an *F* test. A haplotype-based association analysis was conducted by multiple regression against the number of copies of each of 20 HMM-defined haplotype clusters²⁵. Expression data and pedigree file are available in **Supplementary Tables 4** and **5**.

Sequencing. Re-sequencing of the ~780-kb critical region was achieved by amplifying the entire interval as 103 long-range PCR products, equimolar pooling of amplification products, size fractionation by nebulization, ligation of adapters including multiplex identifiers and sequencing on a Roche FLX system using standard procedures. The flow grams were analyzed using the GS Sequencer and Reference Mapper software modules from Roche.

Expression analysis. For RNA extraction and cDNA synthesis, we extracted total RNA from bovine HF×J fetal liver, brain, muscle and bone (epiphyseal plate of long bones) using the QIAGEN MagAttract RNA Universal Tissue, RNeasy Mini and RNeasy Fibrous Tissue kit, respectively. Genomic DNA was further eliminated using the Turbo DNA free kit (Ambion). Integrity of the RNA was checked on either a denaturing formaldehyde agarose gel or the Bio-Rad Experion electrophoresis system. cDNA was synthesized at 45 °C for 1 h using the RevertAid H Minus Reverse transcriptase (Fermentas) with a mixture of random hexamer and oligo (dT)₁₈ primers according to the manufacturer's instructions.

For RT-PCR, we designed primers jointly spanning the entire *MOS*, *PLAG1* and *CHCHD7* transcripts. We used 50 ng of template cDNA for PCR amplification using the DreamTaq Green PCR Master Mix (Fermentas). The PCR product was loaded on an agarose gel and visualized using SYBR Safe DNA gel staining (Invitrogen). The major PCR product was cut out of the gel, purified and sequenced using the BigDye Terminator v3.1 Cycle Sequencing kits on the 3730 DNA Analyzer (Applied Biosystems).

Real-time PCR was performed using the 7900HT Real-Time PCR system (Applied Biosystems). *PLAG1*, *CHCHD7* and internal control genes were detected using 20 ng template cDNA, 1× Absolute Blue QPCR SYBR Green Mix (ABgene) and 0.1 μM primers in a total volume of 15 μl with a PCR condition of 15 min at 95 °C and 40 cycles of 95 °C for 15 s and 60 °C for 1 min, followed by a dissociation curve analysis. For analysis of *LYN*, *RPS20*, *MOS*, *SDR16C5* (*RDHE2*), *SDR16C6*, *PENK* and *IMPAD1*, we used 20 ng template cDNA, 1× Absolute Blue QPCR ROX Mix (ABgene), 0.5 μM primers and 0.5 μM PrimeTime qPCR probe (IDT) in a total volume of 20 μl with a PCR condition of 15 min at 95 °C and 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Each assay was duplicated and analyzed with the qBase plus software (Biogazelle). We used up to three amplicons per gene. Data were normalized using two to five housekeeping genes selected from eight candidates using geNorm⁵⁷. Normalized expression levels were expressed on a log₂ scale relative to the mean of all animals. For a given gene, we computed the individual's average relative expression across amplicons. We regressed average relative expression on the number of Q alleles in QTN genotype.

To measure allelic imbalance of *PLAG1* and *CHCHD7* transcripts, we designed PCR primers to target rs109231213 (located in the *PLAG1* 3' UTR), ss319607402 (in the *PLAG1* intron 2) and ss319607407 (in the *CHCHD7* 3' UTR). We used 50 ng of template cDNA or 10 ng of genomic DNA from heterozygous individuals for PCR amplification using the Phusion High-Fidelity DNA polymerase (Finnzymes). The PCR product was purified through MultiScreen PCRu96 filter plate (Millipore) and directly sequenced. Peak height at the polymorphic site was quantified using the PeakPicker software³⁴. Allelic imbalance was estimated as the ratio of ratios of peak height of the Q allele over the q allele in the cDNA and an average ratio of genomic DNA (gDNA) from the same population. Calibration curves were generated using data obtained by mixing varying amounts of genomic DNA of QQ and qq homozygotes. Comparing the allelic ratio in transcripts of heterozygous individuals is potentially more sensitive (as the comparison of the allelic output is performed within the same sample thereby minimizing the effect of confounding factors) and will also detect differences in promoter strength for genes whose transcription levels are under negative feedback regulation (rendering steady state transcript levels independent of promoter strength).

Luciferase reporter assay. To measure effects of the two QTNs upstream of *PLAG1* (ss319607405 and ss319607406) on transcriptional activity, both allelic forms (QQ and qq) of a 659-bp fragment centered around the QTN (-333 to +326 when positioning the presumed human *PLAG1* transcriptional start site as +1; chr. 14:23,264,483-23,265,141 in the bosTau4 assembly) were PCR amplified from gDNA and subcloned into pCR4-TOPO vector (Invitrogen). The insert was excised from EcoRI sites residing in the vector, blunt ended using T4 DNA polymerase (Fermentas) and inserted into an EcoRV site (upstream of firefly luciferase gene) of the pGL4.10[luc2] Vector (Promega). A shorter 378-bp fragment containing the two QTN (-252 to +126; Chr14:23,264,683-23,265,060) was excised at two SacI sites of the long fragment, blunt ended and subcloned into the EcoRV site of the pGL4 vector. To generate recombinant constructs having qQ and Qq combinations for the two QTN sites in *cis*, AatII-HindIII fragments containing only the ss319607405 site were switched between plasmid constructs having QQ and qq alleles. The sequence and orientation of the insert were confirmed by sequencing.

For cell culture, COS-1 (monkey kidney-derived cells), C2C12 (mouse muscle myoblast) and ATDC5 (mouse teratocarcinoma) were maintained in DMEM with 10% FBS (FBS) and DMEM:Ham's F12 (1:1) with 5% FBS, respectively, supplemented with 2 mM glutamine, 0.1 mM non-essential amino acids, penicillin (100 units/ml) and streptomycin (100 mg/ml). Using Lipofectamine 2000 (0.5 μl per transfection, Invitrogen) following the manufacturer's recommendations, we transfected the $1.4\text{--}1.6 \times 10^4$ cells per well in a 96-well plate with a mixture comprising 150 ng of the pGL4

firefly luciferase reporter and 40 ng of pRL-TK Renilla luciferase construct (Promega). The luciferase assay was performed 24 h after transfection using the Dual Luciferase Reporter Assay system (Promega) and the Centro LB960 luminometer (Berthold Technologies).

Electrophoresis mobility shift assay (EMSA). Nuclear protein was extracted from a Holstein fetus tissue (brain, liver, skeletal muscle and epiphyseal plate of long bone) at ~80 days of gestation as well as from cultured cells (COS-1, C2C12, ATDC5) using the NE-PER Nuclear and Cytoplasmic Extraction Reagents (Pierce) and the Halt Protease Inhibitor Cocktail (Pierce) according to the manufacturer's instructions. Protein concentration was measured with the BCA Protein Assay kit (Pierce). The short probe (29 bp) was prepared by annealing complementary oligonucleotides centered around ss319607406 (-60 to -32 when positioning the presumed human *PLAG1* transcriptional start site as +1). For making the long probes containing both ss319607405 and ss319607406 (-65 to +15; chr. 14:23,264,794-23,264,873; 80 bp and 74 bp for the Q and q allele, respectively), a BglI/PstI digested DNA fragment was excised from the plasmid vectors used for the reporter assay, blunt ended using T4 DNA polymerase and dephosphorylated using Shrimp Alkaline Phosphatase (Promega). The probe was end-labeled with [γ - 32 P]ATP using T4 Polynucleotide Kinase (Fermentas) and subsequently purified with G-25 spin column (Roche). The short oligonucleotide probe was further gel purified. An equality of the amount of probes between the Q and q alleles was confirmed by a PAGE.

For the EMSA reaction, nuclear protein (3.5 μ g) and non-labeled competitor probe (when indicated) were incubated at 20 °C for 10 min in a total volume of 18 μ l of a binding buffer (final 10 mM Tris-HCl pH 7.5, 50 mM KCl, 1 mM DTT, 5 mM MgCl₂, 0.5 uM ZnCl₂, 2.5% glycerol, 0.05% NP-40 and 1 μ g poly(dI-dC)(Pierce)). We then added 2 μ l of labeled probe (20 fmol) to the reaction mixture and incubated at 20 °C for 20 min. After adding 2 μ l of loading buffer (Pierce), the DNA-protein complexes were separated by electrophoresis on a 5% polyacrylamide gel (acrylamide: bisacrylamide, 19:1) in 0.5 \times Tris-borate-EDTA buffer at 15 V/cm for 1–1.5 h at 4 °C. Gels were then dried and exposed to the Storage Phosphor Screen (Amersham) for 2 h to overnight. Labeled complexes were detected using the Typhoon 9400 scanner and analyzed with the ImageQuant TL software (GE Healthcare).

54. Coppieters, W. *et al.* A QTL with major effect on milk yield and composition maps to bovine chromosome 14. *Mamm. Genome* **9**, 540–544 (1998).
55. Churchill, G.A. & Doerge, R.W. Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971 (1994).
56. Misztal, I. *et al.* BLUPF90 and related programs (BGF90). In: *7th World Congress on Genetics Applied to Livestock Production* (Montpelier, 19–23 August 2002).
57. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging multiple internal control genes. *Genome Biol.* **3**, research0034 (2002).