



## Determination of quantitative trait nucleotides by concordance analysis between quantitative trait loci and marker genotypes of US Holsteins

J. I. Weller,\*<sup>†1</sup> D. M. Bickhart,†‡ G. R. Wiggans,†<sup>2</sup> M. E. Tooker,† J. R. O'Connell,§ J. Jiang,# M. Ron,\* and P. M. VanRaden†

\*Institute of Animal Sciences, ARO, The Volcani Center, Rishon LeZion 7505101, Israel

†USDA, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD 20705-2350

‡USDA, Agricultural Research Service, Cell Wall Biology and Utilization Laboratory, Madison, WI 53706

§University of Maryland School of Medicine, Baltimore 21201

#Department of Animal and Avian Sciences, University of Maryland, College Park 20742

### ABSTRACT

Experimental designs that exploit family information can provide substantial predictive power in quantitative trait nucleotide discovery projects. Concordance between quantitative trait locus genotype as determined by the a posteriori granddaughter design and marker genotype was determined for 30 trait-by-chromosomal segment effects segregating in the US Holstein population with probabilities of  $<10^{-20}$  to accept the null hypotheses of no segregating gene affecting the trait within the chromosomal segment. Genotypes for 83 grandsires and 17,217 sons were determined by either complete sequence or imputation for 3,148,506 polymorphisms across the entire genome; 471 Holstein bulls had a complete genome sequence, including 64 of the grandsires. Complete concordance was obtained only for stature on chromosome 14 and daughter pregnancy rate on chromosome 18. For each quantitative trait locus, effects of the 30 polymorphisms with highest concordance scores for the analyzed trait were computed by stepwise regression for predicted transmitting abilities of 26,750 bulls with progeny test and imputed genotypes. Effects for stature on chromosome 11, daughter pregnancy rate on chromosome 18, and protein percentage on chromosome 20 met 3 criteria: complete or almost complete concordance, nominal significance of the polymorphism effect after correction for all other polymorphisms, and marker coefficient of determination  $>40\%$  of total multiple-regression coefficient of determination for the 30 polymorphisms with highest concordance. An intronic variant marker on chromosome 5 at 93,945,738 bp explained 7% of variance for fat percentage and

74% of total multiple-marker regression variance but was concordant for only 24 of 30 families. The missense polymorphism Phe279Tyr in *GHR* at 31,909,478 bp on chromosome 20 was confirmed as the causative mutation for fat and protein concentration. For effect on fat percentage on chromosome 14, 12 additional missense polymorphisms were found that had almost complete concordance with the suggested causative polymorphism (missense mutation Ala232Glu in *DGAT1*). The only polymorphism found likely to improve predictive power for genomic evaluation of dairy cattle was on chromosome 15; that polymorphism had a frequency of 0.45 for the allele with economically positive effects on all production traits.

**Key words:** genomic selection, a posteriori granddaughter design, quantitative trait nucleotide, concordance

### INTRODUCTION

Genomic selection was introduced into US dairy cattle breeding programs in 2009. Since then, gains in the rate of genetic improvement for all traits under selection have been impressive (García-Ruiz et al., 2016). Applied methods for genomic evaluation include the associated effects of tens of thousands of SNP, even though the vast majority of markers have no direct effects on the analyzed traits (e.g., VanRaden, 2008), because many causative variants have high population-wide linkage disequilibrium with genetic markers included on the SNP chip (e.g., Cole et al., 2009).

Ron and Weller (2007) presented a schematic strategy for farm animals to determine whether a specific polymorphism does have a direct effect on a quantitative trait, which generally is designated as a quantitative trait nucleotide (QTN). To date, only 2 polymorphisms have met the rigorous criteria that they proposed for confirmation of QTN status in dairy cattle. These are missense mutations in *ABCG2* on chromosome 6 and

Received March 25, 2018.

Accepted May 19, 2018.

<sup>1</sup>Corresponding author: joel.weller@mail.huji.ac.il

<sup>2</sup>Current address: Council on Dairy Cattle Breeding, Bowie, MD 20716.

*DGAT1* on chromosome 14 (Grisart et al., 2002, 2004; Winter et al., 2002; Cohen-Zinder et al., 2005).

The optimal validation of a QTN is a demonstration that replacement of the allele results in swapping one phenotypic variant for another (Glazier et al., 2002). For dairy cattle, this is still not a viable alternative. Considering these limitations, how can one prove that a candidate polymorphism is actually the QTN? As noted by Mackay (2001), "The only option in these cases is to collect multiple corroborating pieces of evidence, no single one of which is convincing, but which together consistently point to a candidate gene." Ron and Weller (2007) called this strategy of identifying QTN "winning by points rather than knock-out."

Considering that all analyses presently must be based on an existing population, the most convincing proof available that a QTN has been determined is concordance—that is, determination for a group of animals that their genotypes for the putative QTN correspond to their inferred genotypes for the QTL, which is defined in this study as a segregating effect on a quantitative trait that has been localized to a specific chromosomal segment based on genetic linkage. Complete concordance is obtained only if the following criteria are met: (1) all individuals homozygous for the QTL are also homozygous for the putative QTN; (2) all individuals heterozygous for the QTL are also heterozygous for the putative QTN; and (3) the same putative QTN allele is associated with the positive QTL allele in all heterozygous individuals. It should be noted that complete concordance is expected only if the observed effect is due to one unique biallelic causative variant in the QTL region.

Ron and Weller (2007) proposed application of an a posteriori granddaughter design (APGD) to determine QTL genotypes for bulls from large populations of cattle genotyped using mid- or high-density SNP chips. Similar to the original granddaughter design (Weller et al., 1990), sires with many progeny-tested sons are analyzed. However, data generated from genotyping many bulls with high-density SNP chips are used rather than genotyping the sons specifically for application of a granddaughter design. Thus, the design is considered to be a posteriori. The sons of each bull are assigned to 1 of 2 groups based on the paternal haplotype that was passed to a son for the chromosomal region with the putative QTL. Thus, deduction of QTL status is based on within-family statistical significance of contrast between the 2 paternal haplotypes.

The APGD was applied to the US Holstein population by Weller et al. (2014) and Wiggans and Weller (2015). Both studies analyzed 33 traits: yield (milk, fat, and protein and component percentages), milk SCS,

productive life, daughter pregnancy rate, heifer and cow conception rates, service-sire and daughter calving ease, service-sire and daughter stillbirth rates, 18 conformation traits, and the net merit genetic-economic index. The analysis of Weller et al. (2014) included 52 grandsire families with 9,178 sons. They found 30 chromosomal segment-by-trait combinations that met a probability significance criterion of  $<10^{-14}$  for the effect of grandsire haplotype nested within grandsire. Wiggans and Weller (2015) analyzed 71 grandsire families with 14,246 sons and, excluding the effects on the milk production traits associated with the *DGAT1* polymorphism, they found 56 chromosomal segment-by-trait combinations that met this significance criterion. Of those that met this criterion, 29 were also significant by this criterion in the study of Weller et al. (2014). Of the remaining 27 segment-by-trait combinations, all were significant at  $<10^{-7}$  in the study of Weller et al. (2014). Four of the effects significant in Weller et al. (2014) at  $<10^{-14}$  did not meet this criterion in the later study.

A very low statistical significance criterion was applied for 3 reasons. First, the entire genome was scanned for 33 traits, which causes a multiple comparison problem (Weller et al., 1998). According to the formula of Lander and Kruglyak (1995), a nominal statistical significance of approximately  $5 \times 10^{-5}$  is required for the bovine genome to obtain a genome-wide significance of 0.05, assuming a dense marker map. Second, a relatively narrow confidence interval determined by a non-parametric bootstrap was desired. Wiggans and Weller (2015) demonstrated that even with nominal statistical probabilities of  $10^{-10}$ , a bootstrap confidence interval could include up to half the entire chromosome. Third, QTL genotypes of the grandsires must be highly accurate to determine concordance accurately between a QTL and a specific polymorphism. Therefore, the QTL effect must be large enough so that both type I and type II errors will be approximately 1% for at least 20 of the grandsires included in the analysis.

Assuming that the entire genome contains approximately  $3 \times 10^6$  nonredundant polymorphisms with minor allele frequencies sufficiently high to achieve complete concordance for a segregating QTL detected by APGD, the experiment-wise probability of obtaining complete concordance by chance for any SNP will be  $<0.05$  with 10 grandsires homozygous for the QTL and 8 heterozygous with phase determined (Ron and Weller, 2007). In practice, many more grandsires will be required. Unless the QTL effect is huge [e.g., the magnitude of the effect on protein percentage associated with *ABCG2* (Cohen-Zinder et al., 2005)] or the number of sons per grandsire is very large, QTL status for most grandsires cannot be determined unequivocally.

In addition, incomplete concordance between QTL and QTN may be derived from either genotyping mistakes of the putative QTN or incorrect classification of QTL status. Therefore, the sample of grandsires with QTL status determined must be large enough that the probability of obtaining even almost complete concordance is so small that the null hypothesis of near concordance by chance can be rejected even when accounting for a huge number of segregating polymorphisms.

In addition to concordance, Ron and Weller (2007) proposed other statistical methods to verify that a causative polymorphism had been determined. Another major indication of QTN discovery is the determination that this polymorphism accounts for almost all the variance due to the QTL and that the effects of additional polymorphisms in the QTL chromosomal region are very small in a model that includes the putative QTN effect. Application of this method requires determination of QTN genotypes on a much larger sample of individuals with either phenotypic records or genetic evaluations for the traits of interest.

The advent of low-cost whole-genome sequencing has made complete sequencing of 60 to 80 APGD grandsires and testing discovered polymorphisms for concordance possible at a reasonable cost. Hayes et al. (2014) initiated the 1000 Bull Genomes Project to provide an extended database of variants that could be imputed into existing cattle SNP genotype data. As of 2016, 423 Holstein bulls had been sequenced. Since then, an additional 22 bulls considered as grandsires by Wiggans and Weller (2015) have been sequenced. VanRaden et al. (2017) demonstrated that using those sequence data allowed accurate imputation of complete sequence genotypes and haplotype phase for relatives of those sires that had been genotyped with mid- or high-density SNP chips.

The objective of this study was to determine concordance for the 30 QTL detected by Wiggans and Weller (2015) with probabilities of  $<10^{-20}$  for effect of grandsire haplotype nested within grandsire based on known or imputed genotypes for all grandsire families included in the APGD. For all bulls with imputed genotypes and genetic evaluations, effects of the putative QTN on the trait of interest for the 30 polymorphisms with the highest concordance within an approximate confidence interval also were determined by step-wise multiple regression. Results from the 2 methods were compared with effects associated with each marker and with results from previous studies that analyzed 3 of the putative QTN found in this analysis, including the effect on fat concentration associated with *DGAT1*. The effect on protein percentage associated with *ABCG2* was not included because the economically positive allele

(*ABCG2<sup>A</sup>*) has almost reached fixation in all cattle breeds (Ron et al., 2006) and in the population of sequenced grandsires.

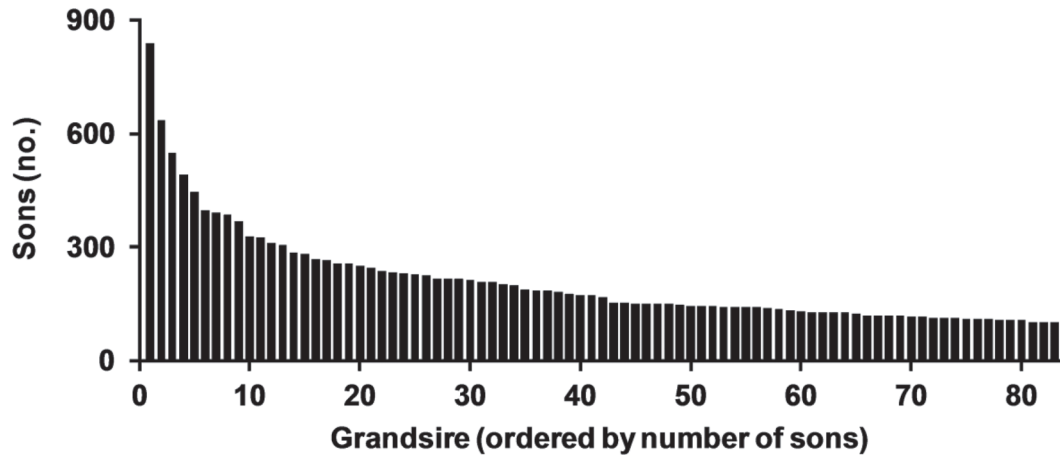
## MATERIALS AND METHODS

### *Imputation of Genotypes to High-Resolution SNP and Indel Calls*

A total of 445 bulls were sequenced previously, and 26 were sequenced for this study. Processed variant call format files from these 471 Holstein bulls were used to impute high-resolution SNP and indel variant calls for all bulls in the reference population for US genomic evaluations. All of the imputed bulls had genotypes based on mid- to high-density Illumina (San Diego, CA) genotyping arrays with 54,000 to 777,000 SNP. The Fortran findhap.f90 imputation program (<https://aipl.arsusda.gov/software/findhap/>; VanRaden et al., 2011) was used to determine the phase for each chromosomal segment. Thus, correspondence between QTL and putative QTN phases could be determined for grandsires that were heterozygous for a QTL. The variant call format files from the 1000 Bull Genomes Project included 14,609,220 polymorphisms with a minor allele frequency of  $>0.01$ . After eliminating redundant polymorphisms (consecutive polymorphisms with an identity of  $>0.95$ ) and polymorphisms with low call rates, 3,148,506 polymorphic sites were retained (VanRaden et al., 2017). Redundant polymorphisms were deleted only in the intergenic regions to match the corresponding SNP from the high-density chip that had previously been pruned for  $>0.95$  correlations. Although elimination of redundant polymorphisms should not affect concordance results, it may affect the annotation of the selected variants. Imputation was applied in 3 stages with chromosome segment lengths of approximately 5,000 (stage 1), 1,000 (stage 2), and 200 (stage 3) polymorphic sites.

### *Selection of QTL and Bulls for Analysis*

Wiggans and Weller (2015) found 30 trait-by-chromosomal segments with a nominal statistical probability of  $<10^{-20}$  to reject the hypothesis of no effect for the grandsire haplotype. Those QTL included 16 traits (fat yield, fat and protein percentages, SCS, productive life, daughter pregnancy rate, heifer and cow conception rates, stature, strength, body depth, dairy form, thurl width, rump angle, teat length, and final score) on 11 chromosomes. These chromosomes included 1,271,638 polymorphic sites; that is, 40% of the 3,148,506 sites in the entire genome after edits. The number of poly-



**Figure 1.** Maximum numbers of sons per grandsire used in analysis with the a posteriori granddaughter design.

morphic sites per chromosome ranged from 79,207 to 147,911 (Table 1). The current analysis of 83 grandsire families with 17,217 sons was based on August 2017 US genetic evaluations by the Council on Dairy Cattle Breeding (Bowie, MD). Of these grandsires, 64 had genotypes based on complete sequence. The number of sons per grandsire ranged from 100 to 838 (Figure 1); for each specific trait, numbers of grandsire families and sons were generally fewer because of data edits. Each son had a genotype based on a mid- to high-density Illumina genotyping array and an August 2017 US genetic evaluation for milk production traits based on a progeny test of  $\geq 10$  daughters.

### Whole-Genome Shotgun Sequencing

Of the 83 grandsire families, 38 were previously sequenced by the 1000 Bull Genomes Project or project collaborators. An additional 26 grandsires were sequenced as part of this study. The remaining 19 grandsires were not sequenced, and their genotypes for

polymorphisms not included in the high-density SNP chip were determined by imputation. Semen straws extracted from each of the 26 grandsires were identified within the Cooperative Dairy DNA Repository (Beltsville, MD) and were sent to Cofactor Genomics (St. Louis, MO) for DNA extraction and library preparation. Libraries were prepared using robotic automation with Illumina TruSeq Chemistry (version 3) reagents and were subsequently sequenced on the Illumina NextSeq 500 to a target of 20 $\times$  coverage. Samples were sequenced at the USDA Animal Genomics and Improvement Laboratory on an in-house NextSeq500 instrument. Sample sequence data quality was assessed via FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and was found to be comparable with data generated on other Illumina sequencing platforms (data not shown). At the time, this sequencer was the only instrument that was available in-house. Sequence FASTQ files were aligned to the UMD3.1 reference genome (Zimin et al., 2009) using BWA-MEM software (version 0.7.15; Li and Durbin, 2009), and alignment files were processed using SAMtools (version 1.12; Li et al., 2009). The SNP and indel variants were called using the SAMtools and BCFtools mpileup variant-calling pipeline. Only the default variant filters implemented in SAMtools were used to remove low-quality variant calls.

### Determination of Concordance

Concordance was determined for each QTL except for the effect of fat yield on BTA5; that effect is pleiotropic to the effect on fat concentration in the same chromosomal region. This required application of the APGD for each of the 30 QTL. In the previous APGD studies, markers from the SNP chips were used to de-

**Table 1.** Numbers of markers for imputed data used in analysis with the a posteriori granddaughter design by chromosome number

BTA	Markers (no.)
3	140,056
5	147,911
6	133,115
7	125,168
10	133,261
11	124,421
14	90,416
15	121,606
18	90,963
19	85,514
20	79,207
All	3,148,506

termine grandsire and son haplotypes. In this study, haplotypes were determined based on all 3,148,506 polymorphic sites. Preliminary results showed that at least 20,000 markers would be needed for a minimum confidence interval of 90% for each QTL (Wiggans and Weller, 2015). Stage 1, 2, or 3 imputation segments could be used as haplotypes. Using stage 1 segments of 5,000 markers would reduce the total number of analyses required to cover 20,000 sites but increase the risk of generation of “spurious haplotypes” due to recombination within segments or genotyping or imputation mistakes. Using stage 3 haplotypes would increase the total number of analyses to >100 for each QTL. Thus, stage 2 segments, which included approximately 1,000 markers, were a reasonable compromise between segments that were passed from sire to son without recombination and limiting the total number of analyses. Wiggans and Weller (2015) divided the entire genome into 621 chromosomal segments of approximately 100 SNP each. With 3,148,506 polymorphic sites, the total number of stage 2 chromosomal segments to span the genome would be approximately 3,100 (i.e., ~5 times the number of segments in their study). Considering that the length of the entire bovine genetic map is approximately 2,800 cM, the probability of recombination within a segment of 1,000 polymorphic sites is only approximately 1%.

In the preliminary analysis for each QTL, a chromosomal region including 21 stage 2 segments was analyzed, with centering on the chromosomal segment with the greatest effect in the study of Wiggans and Weller (2015). A separate APGD analysis was performed for each stage 2 segment. Thus, a chromosomal region spanning at least 21,000 sites was analyzed. A chromosomal region of this length should have a high probability of including the causative polymorphism provided that the region analyzed does in fact include the likelihood peak. However, because the new analysis included more bulls and used shorter haplotypes compared with the study of Wiggans and Weller (2015), it is possible that the likelihood peak may have shifted relative to the previous study. We accounted for this as follows: if statistical probability for the haplotype effect at either of the extreme segments was <10 log units greater than the minimum log value, 5 additional stage 2 segments were included until the criterion was met at both ends of the chromosomal region analyzed.

For each chromosomal segment, families or specific sons were excluded from analysis for the following reasons.

- If a grandsire was homozygous for his 2 haplotypes, a haplotype effect could not be estimated

and his family was excluded. Because frequency of the most common haplotype for each chromosomal segment was usually <0.2, exclusion of families because of grandsire homozygosity was generally rare.

- If a son’s genotype was identical to the grandsire’s genotype, then that son was excluded because which grandsire haplotype was passed to the son could not be determined.
- A son that received neither parental haplotype was excluded. Because paternity was verified for all sons based on mid- or high-density genotypes, the most likely explanation for such a result was either a recombination between father and son or a mistaken genotype call for a specific polymorphism of the son. Ascertainment bias in the selected sequence SNP and indel calls derived from sequence data or errors in sequence variant calls could result in pseudohaplotypes (i.e., not actually present in the population) as part of the imputation process and create discrepancies between parents and sons. An additional possibility is mutation, but previous results have shown mutations to be considerably less frequent than genotyping mistakes (Weller et al., 2010).
- If <20% of sons received 1 paternal haplotype, the family was excluded. In this case we assumed that the grandsire was incorrectly listed as heterozygous because of a genotype calling or imputation mistake.
- If the number of valid sons per family was <80, then the family was excluded. In this case, the statistical power was insufficient to determine whether the grandsire was homozygous or heterozygous for the QTL.

Preliminary results demonstrated that numbers of valid families and sons were generally fewer if haplotypes were determined based on imputed sequence data rather than haplotypes determined from mid-density genotyping arrays as in Wiggans and Weller (2015). Probability is higher that a sire could be homozygous for a shorter haplotype, and haplotypes determined by imputed sequence were only about one-fifth the length of haplotypes determined from genotyping arrays. In addition, haplotypes based on a complete sequence have a higher rate of spurious haplotypes because of artifacts generated during the sequencing process (~1% error rate; Bentley et al., 2008) and spurious variant calls generated by SAMtools multisample calling (1–10% error rate; Li et al., 2009; Baes et al., 2014; Ni et al., 2015). Therefore, an additional APGD analysis was done for the chromosomal segment with the lowest

*P*-value as determined in Wiggans and Weller (2015) for haplotypes based on SNP genotyping arrays.

With respect to determination of QTL genotype, the goal was to include as many families as possible and still reduce frequency of incorrect determinations. If a grandsire is incorrectly determined to be heterozygous, then a type I error is made and the probability of this error is known. If a grandsire is incorrectly determined to be homozygous, then a type II error is made. Determination of the probability for a type II error requires knowledge of the magnitude of QTL effect, which is lacking. Therefore, a more stringent criterion was applied in this case. Homozygosity of grandsire QTL genotype was determined by APGD application using analysis of chromosomal segments as in Wiggans and Weller (2015); heterozygosity was determined by APGD application to chromosomal segments of approximately 1,000 markers based on imputation of complete sequence. If the absolute value of the *t*-statistic for the difference between 2 haplotype effects within each family was <1.0 and the number of sons in the family was  $\geq 150$ , the grandsire was assumed to be homozygous for the QTL. The second criterion was applied to reduce probability of a type II error. If the absolute *t*-value for haplotypes determined by imputation of complete sequence data was  $>2.5$ , the grandsire was assumed to be heterozygous for the QTL. For families that met neither of these criteria, QTL status was assumed to be unknown, and these families were discarded from the concordance analysis. For *DGAT1*, previous studies have shown that additional polymorphisms in addition to the main effect of the missense mutation affect fat percentage (Kühn et al., 2004). Therefore, a *t*-value of  $>2.5$  could be obtained due to segregation of these polymorphisms, even if the grandsire was homozygous for the main effect associated with *DGAT1*. For this reason, only families with *t*-values of  $>6.5$  were considered to be heterozygous for the main effect. Likewise, all families with *t*-values of  $<6.5$  and with  $>150$  sons were considered to be homozygous for the main effect.

If the grandsire was designated as heterozygous, then the same putative QTN allele should be associated with the same grandsire allele in all heterozygous families (assuming that only 2 QTL alleles are segregating in the population and that a specific polymorphism is the QTN). Therefore, the number of families for which a specific allele of each polymorphism was associated with a QTL allele with a positive or negative effect on the trait was scored. We then determined which allele was associated with the positive QTL effect in the most families. The families heterozygous for the QTL were then scored as a match only if this allele was associated with the positive QTL effect.

Concordance for all polymorphisms within the QTL range was then ranked, first by the fraction of heterozygous matches with the more frequent phase among sires determined to be heterozygous for the QTL, second by the number of correct homozygous matches, and third by chromosome position. Because an individual APGD analysis was performed for each stage 2 segment of 1,000 markers, the number of sires determined to be heterozygous for the QTL could vary over the analyzed chromosomal segments. However, this was not the case for sires determined to be homozygous, which was based on single analysis of the chromosomal segment with the lowest *P*-value from Wiggans and Weller (2015). Preference was given to correct heterozygous matches first because type I error (i.e., probability that a homozygous sire would be incorrectly designated as heterozygous for the QTL) was controlled at 1%, whereas type II error (a function of the actual QTL effect) was unknown. In addition, the probability that a specific marker would be heterozygous with correct QTL phase by chance is at most 0.25, whereas probability that a specific marker would be homozygous is  $\geq 0.5$ .

Identified variants were annotated with putative functional effects using the SNPeff software package (Cingolani et al., 2012) and the UMD3.1.71 Ensembl gene annotation database. If any of the markers were classified by SNPeff as having moderate (typically consisting of missense mutations) or high (e.g., stop gains and frameshifts) effects, then the marker was kept as a candidate QTN. Only 1 marker annotated with a high predicted effect by SNPeff was found.

### **Estimation of Putative QTN Effects on Quantitative Traits**

Of the bulls with a PTA based on a progeny test of  $\geq 10$  daughters, 444 were sequenced. For an additional 26,306 bulls with progeny test, complete sequence was determined by imputation. Thus, there were 26,750 bulls with progeny test based on  $\geq 10$  daughters and complete sequence based on imputation or sequence. Excluding the 83 grandsires, all the other bulls in this set were included in the 17,217 sons of the grandsires used in the APGD analyses. The set of 26,750 bulls with progeny test based on  $\geq 10$  daughters and complete sequence was used to determine the effects of the 30 polymorphisms with the highest concordance scores for the analyzed traits by stepwise regression for each QTL. The objective of this analysis was to determine whether a single marker would explain most of the variance associated with the QTL in a model that included the other linked markers in near concordance. All 30

markers were potentially included, but the algorithm only added additional markers with  $P$ -values of  $<0.15$ . In many cases, several markers had identical concordance ranks. If markers not included within the first 30 had identical concordance ranks to those included within the top 30, then those markers were also included in the regression analysis. Each polymorphism was scored as 0, 1, or 2, with 0 and 2 as the 2 homozygotes and 1 as the heterozygote. The polymorphisms with the greatest coefficients of determination ( $R^2$ ) in the multiple-regression models were then reanalyzed by a model that included only the effect of the single polymorphism. In addition, individual linear regressions were computed for missense and nonsense polymorphisms ranked among the top 60 for concordance.

Those effects were compared with effects from a model that estimated the effect of each polymorphism corrected for effects of all other polymorphisms using the MMAP algorithm (O'Connell, 2017). That algorithm uses a fast, low-memory method to calculate additive and (optionally) dominant genetic covariance structures based on SNP data. The MMAP algorithm was used to estimate the effect associated with each polymorphism under the assumption that each analyzed effect was fixed and that effects of all other polymorphisms were random. Another assumption for the algorithm was that variance parameters only needed to be estimated once for the entire data set (rather than once for each marker) because the effect of any given marker on a trait typically is small.

## RESULTS

### Chromosomal Segment Effects

Logarithm of the odds (LOD) scores, APGD probabilities of significance for haplotype effect nested within grandsire, and physical locations for starts and ends of the approximately 1 Mbp chromosomal segments with the lowest  $P$ -values determined by imputation of complete sequence data are shown in Table 2 by chromosome. Excluding the effect associated with fat percentage on BTA14, which had a probability of  $3.5 \times 10^{-283}$ , probabilities ranged from  $1.0 \times 10^{-48}$  for effect on SCS on BTA6 to  $1.3 \times 10^{-8}$  for effect on stature on BTA14. Although probabilities of  $<10^{-20}$  were obtained by APGD for all QTL from Wiggans and Weller (2015), probabilities in Table 2 are based on APGD computed from segments of approximately 1,000 imputed markers, which were only about one-fifth the length of the segments of Wiggans and Weller (2015) and generally included fewer valid families in each APGD analysis.

Markers with the lowest  $P$ -values for 3 traits on BTA5 (daughter pregnancy rate, cow conception rate, and stature) and for 3 traits on BTA6 (productive life, SCS, and dairy form) were found in the same chromosomal segments, which may indicate that the same polymorphism is responsible for each group of effects.

The effect on fat percentage on BTA5 is of special interest because it affects milk and fat yields in addition to fat percentage (Wiggans and Weller, 2015), and a segregating QTL in this region has been confirmed in other populations (Littlejohn et al., 2016). Figure 2 shows the negative log probability for APGD effect of haplotype within grandsire on fat percentage on BTA5 along with the chromosome ideogram, chromosome position axis, and a condensed Ensembl gene track. A peak is evident for segment 11 between 93,387,393 and 94,165,024 bp (highlighted in gray). The Ensembl genes closest to the peak are *LMO3* and *MGST1*. The SNP with the greatest effect in this segment, which was identified via regression analysis, is downstream of *MGST1* at 93,945,738 bp.

### Haplotype Effects on Economic Traits

Distributions of absolute  $t$ -values for haplotype effect nested within grandsire for fat percentage on BTA5 are shown for all families based on the stage 2 segment with the lowest  $P$ -value (heterozygote concordance; Figure 3a) and for families with  $\geq 150$  sons based on the segment determined from mid-density genotyping array (homozygote concordance; Figure 3b). Although the distributions are clearly different from expected distributions for no segregating effects,  $t$ -values in both analyses span almost the entire range up to 6. The peak for  $t$  around 2.5 for families in both figures may approximate the mean for family segregation for the QTL; however, if this is the case, then probability of obtaining  $t$ -values of  $<1.0$  is  $>5\%$ . Thus, complete concordance may not be obtained even for the causative polymorphism.

Distribution of absolute  $t$ -values for haplotype effect nested within grandsire for fat percentage also is shown for BTA14 (Figure 3c) based on mid-density genotyping arrays for all families. Out of 75 families, only 23 had absolute  $t$ -values of  $\leq 2.0$ . If only a single biallelic polymorphism was responsible for the effect, then the number of families with significant  $t$ -values should have been no more than half. The fact that 70% had nominally significant  $t$ -values corresponds to the finding of Bennewitz et al. (2004) and Kühn et al. (2004) that  $>1$  polymorphism is responsible for the effect on fat percentage associated with this chromosomal region. The

**Table 2.** Chromosomal segments with the lowest a posteriori granddaughter design probabilities to reject the null hypothesis of no effect for the grandsire haplotype

BTA	Trait	Segment	Location (bp)		Probability	LOD <sup>1</sup> score
			Start	End		
3	Protein percentage	26	15,779,526	16,786,276	$1.10 \times 10^{-47}$	47.0
5	Teat length	19	12,311,835	13,116,209	$4.08 \times 10^{-46}$	45.4
	Body depth	93	82,438,871	83,210,743	$1.48 \times 10^{-42}$	41.8
	Heifer conception rate	99	87,864,332	88,796,250	$3.79 \times 10^{-21}$	20.4
	Dairy form	100	88,796,251	89,461,971	$2.20 \times 10^{-48}$	47.7
	Daughter pregnancy rate	102	90,597,590	91,752,792	$1.18 \times 10^{-28}$	27.9
	Cow conception rate	102	90,597,590	91,752,792	$1.06 \times 10^{-26}$	26.0
	Stature	102	90,597,590	91,752,792	$3.19 \times 10^{-32}$	31.5
	Fat percentage	105	93,387,393	94,165,023	$3.91 \times 10^{-45}$	44.4
	Final score	107	94,988,301	96,044,181	$1.94 \times 10^{-30}$	29.7
	Strength	115	100,106,622	100,570,467	$1.18 \times 10^{-26}$	25.9
	Thurl width	115	100,106,622	100,570,467	$7.50 \times 10^{-20}$	19.1
	Productive life	93	87,005,580	87,677,438	$5.08 \times 10^{-28}$	27.3
	SCS	93	87,005,580	87,677,438	$1.03 \times 10^{-48}$	48.0
	Dairy form	93	87,005,580	87,677,438	$7.24 \times 10^{-30}$	29.1
7	Heifer conception rate	109	100,295,589	101,112,291	$1.01 \times 10^{-20}$	20.0
	Rump angle	59	43,508,685	44,481,953	$5.92 \times 10^{-40}$	39.2
10	Stature	63	48,900,762	50,220,797	$5.72 \times 10^{-34}$	33.2
	Strength	83	59,020,825	60,176,439	$5.92 \times 10^{-37}$	36.2
11	Stature	89	76,949,730	77,890,814	$1.31 \times 10^{-39}$	38.9
14	Fat percentage <sup>2</sup>	3	2,218,207	2,834,113	$3.54 \times 10^{-283}$	282.5
	Body depth	11	8,076,515	8,929,250	$3.34 \times 10^{-16}$	15.5
	Stature	11	8,076,515	8,929,250	$1.29 \times 10^{-8}$	7.9
	Strength	12	8,932,287	9,681,497	$5.25 \times 10^{-29}$	28.3
15	Fat yield	85	59,042,548	60,025,824	$3.18 \times 10^{-23}$	22.5
18	Cow conception rate	52	43,570,040	44,387,220	$7.38 \times 10^{-31}$	30.1
	Daughter pregnancy rate	52	43,570,040	44,387,220	$1.95 \times 10^{-29}$	28.7
19	Stature	42	28,838,219	29,629,718	$5.28 \times 10^{-20}$	19.3
	Thurl width	59	44,630,824	45,767,655	$3.32 \times 10^{-23}$	22.5
20	Protein percentage	37	33,375,896	34,305,738	$2.18 \times 10^{-28}$	27.7

<sup>1</sup>Logarithm of odds.

<sup>2</sup>Effect associated with *DGAT1* (Grisart et al., 2002; Winter et al., 2002).

approximate bimodal distribution also corresponds to their findings that both a very large effect and at least 1 smaller effect are segregating.

### Concordance Analysis

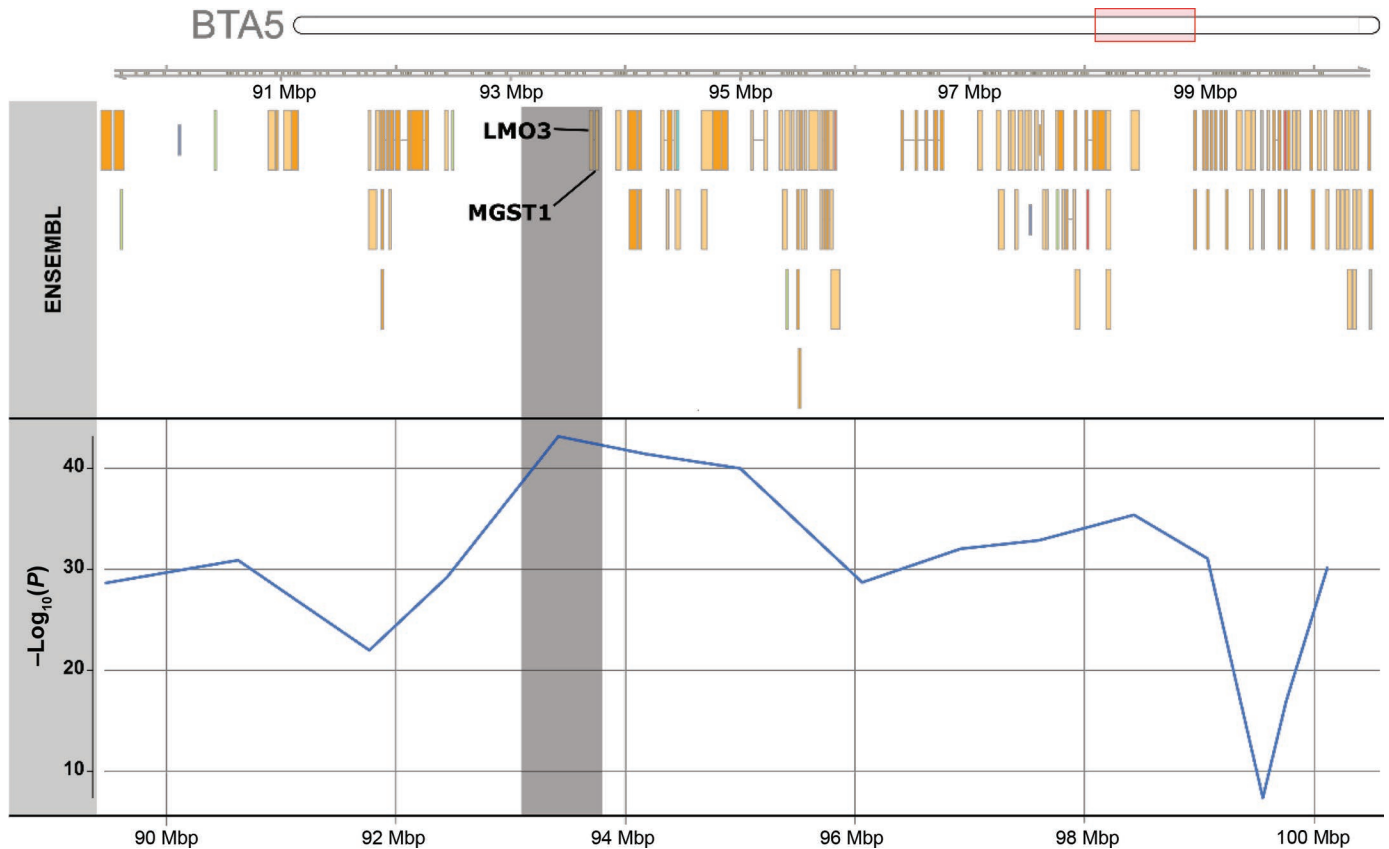
Highest concordance for each QTL (Table 3) was ranked first by fraction of concordant heterozygotes and then by number of concordant homozygotes. Of the 60 to 81 families included depending on trait and excluding *DGAT1*, only 14 to 30 could be used to determine concordance. The number of families based on haplotypes determined by imputation of complete sequence data was generally less than when haplotypes were determined using SNP genotyping arrays.

Complete or almost complete concordance (1 discrepancy) for the heterozygotes was obtained for all QTL except for fat yield on BTA15, which had 2 discrepancies, but heterozygote concordance was the first criterion used to rank the polymorphisms. Concordance for homozygotes was generally lower, even though a family

was considered to be homozygous for the QTL only if the *t*-statistic was <1.0. A discrepancy in this case was recorded if the marker genotype was heterozygous despite the low absolute *t*-value. Evaluation of these discrepancies is more difficult because the expected *t*-value for families that are segregating for the QTL is unknown. For example, if the expected absolute *t*-value is 3.0, then 5% of the heterozygous families should have absolute *t*-values of <1.0.

Complete concordance was obtained only for stature on BTA14 (3 markers with complete concordance, including 2 heterozygous families) and daughter pregnancy rate on BTA18 (2 markers with complete concordance, including 14 heterozygous families). The fact that only 2 heterozygous families were found for the effect on stature corresponds to the relatively low LOD score for this trait (Table 2). Almost complete concordance was obtained for strength on BTA14, which is correlated with stature, and cow conception rate on BTA18, which is highly correlated with daughter pregnancy rate. Both traits had LOD score peaks in





**Figure 2.** Critical region on BTA5 for a QTL with an effect on fat percentage. The tracks from top to bottom are chromosome ideogram (90–100 Mbp in box), chromosome position axis (90–100 Mbp; gray bars indicate assembly gaps), condensed Ensembl gene track (90–100 Mbp), and  $-\log_{10}(P)$  probability of significance for effect of haplotype within grandsire for a posteriori granddaughter design. Each point represents the beginning of a chromosomal segment of approximately 1,000 markers, and each interval corresponds to approximately 1 Mbp. Color version available online.

the same chromosomal segment. Complete concordance was missed by 2 families for 4 additional QTL, including the effect associated with *DGAT1* on BTA14. For fat percentage on BTA5, concordance was found for only 24 out of 30 families.

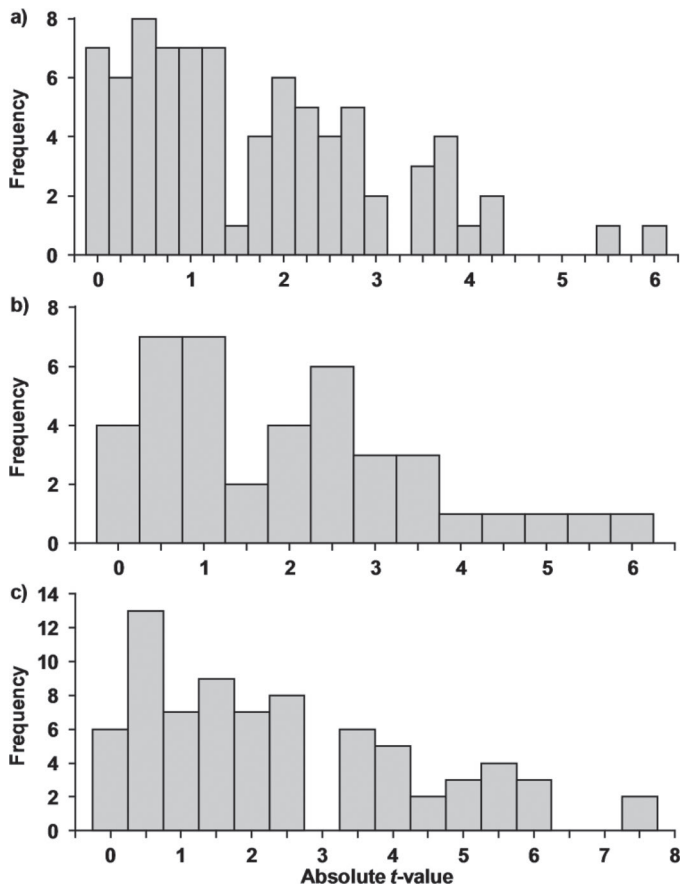
### Multiple-Regression Analysis

Polymorphisms with the greatest multiple-regression  $R^2$  for the trait among the 30 markers with highest concordance are listed in Table 4 along with markers with concordance scores among the top 60 that were denoted as moderate or high priority for complete sequence data. Concordance ranks and probabilities in the MMAP analysis are also listed. If several markers had equal concordance rank based on QTL heterozygotes and homozygotes, then the rank listed is the rank of the first marker in that group. The MMAP marker set included was slightly different from the set used to determine concordance. Therefore, MMAP probability

was not computed for the marker at position 1,928,347 bp on BTA14 flagged for stature.

For all but 3 QTL, a marker was found with a multiple-regression  $R^2$  of  $>0.01$ . Because 26,750 bulls were included, effects of all markers (Table 4) were significant ( $P < 0.0001$ ). Effects for 17 of those markers were nonsignificant ( $P > 0.05$ ) in the MMAP analysis; 2 of the nonsignificant markers had an  $R^2$  of  $<0.01$ . Somewhat surprisingly, 6 of the nonsignificant MMAP markers had an  $R^2$  of  $>0.05$ . The difference between the 2 analyses is that MMAP corrects for effects of all other markers as random effects and multiple regression does not. Because bulls are related, not accounting for this effect could result in biased estimates; however, because marker genotypes are correlated, correcting for other linked markers reduces the effect associated with each marker.

As expected, the  $R^2$  for the 2 effects associated with fat percentage on BTA14 (0.36) was much higher than the  $R^2$  for any of the other effects. Surprisingly, the



**Figure 3.** Distribution of absolute  $t$ -values for haplotype effect nested within grandsires from analysis of fat percentage on BTA5 for (a) all families based on approximately 1,000 markers determined by imputation and (b) families with  $\geq 150$  sons based on segment determined from mid-density genotyping array and on BTA14 (c) with haplotypes determined from mid-density genotyping array for all families.

marker that was an intronic SNP in a gene that has not yet been annotated had a slightly higher  $R^2$  than found for the Ala232Glu missense mutation in *DGAT1*.

The marker with the highest  $R^2$  in the multiple-regression analysis was within the segment with the lowest probability of significance (Table 2) only for body depth on BTA14. For the effect on cow conception rate on BTA18, the marker with the greatest  $R^2$  was not within the chromosomal segment with the lowest  $P$ -value that the null hypothesis was true (Table 2), but the missense mutation with the highest  $R^2$  was. Of the 30 QTL tested, 16 markers with the highest  $R^2$  were among the 5 with the highest concordance. Thus, correspondence between these 2 criteria generally was good. Missense mutations that were ranked among the top 5 for concordance were found for stature on BTA5, SCS on BTA6, fat percentage and stature on BTA14, daughter pregnancy rate on BTA18, and protein percentage on BTA20. A Phe279Tyr polymorphism in

*GHR* on BTA20 that affects protein percentage was noted previously by Blott et al. (2003).

### Fraction of QTL Variation Explained by Each Marker

The ratio between marker  $R^2$  and total multiple-regression  $R^2$  is an additional indicator of how likely it is that a specific marker is the QTN. If total  $R^2$  is considerably higher than marker  $R^2$ , then other factors in proximity to the marker affect the trait. However, even if the specific marker is the QTN, other sites in proximity to the QTN also may modify its effect, as found for *DGAT1* (Bennewitz et al., 2004; Kühn et al., 2004). Alternatively, if the candidate is not the QTN, then the effect observed is the result of population-wide linkage disequilibrium. In that case, only a fraction of the effect is captured by each marker, and other markers should also display effects because of linkage disequilibrium. If inclusion of additional closely linked markers does not significantly increase explanatory power, then detection of the QTN becomes more likely. The effect of a single polymorphism accounted for  $>90\%$  of multiple-regression variance only for fat percentage on BTA14. For 11 of the 17 markers with nonsignificant ( $P > 0.05$ ) MMAP effects (Table 4), total  $R^2$  was more than double marker  $R^2$ , which casts further doubt that causative polymorphisms have been detected. In contrast, the missense polymorphism for the effect on SCS on BTA6 explained 77% of total variance captured in the multimarker model and had an MMAP LOD score of  $>14$ . This polymorphism had complete concordance for all 15 heterozygous families but only 5 of the 8 homozygous families.

Of the markers with the highest  $R^2$  values, only 1 was an indel and was an effect associated with fat percentage on BTA5. However, this polymorphism is unlikely to be causative because  $R^2$  was low and its MMAP effect was nonsignificant ( $P > 0.05$ ). The polymorphism on BTA5 flagged by Littlejohn et al. (2016) is an intronic variant but explained 74% of multiple-regression variance. That effect also had an MMAP LOD score of  $>100$ . The position of 93,945,738 bp for this SNP was within the chromosomal segment with the lowest probability of significance for the APGD haplotype effect (Figure 2; Table 2). However, the SNP was heterozygous in only 12 of the 16 families with significant contrasts for fat percentage and was homozygous in 13 of 15 families with a  $t$ -value of  $<1.0$ . One possible explanation is that more than 1 polymorphism with significant effects may be segregating in this chromosomal region, and this SNP may represent a “ghost” QTL (Martínez and Curnow, 1992).

Of the remaining polymorphisms with the highest  $R^2$  (Table 4), only the effect associated with protein

**Table 3.** Numbers of markers with highest concordance scores for each QTL ranked first by fraction of concordant heterozygotes and then by number of concordant homozygotes

BTA	Trait	Heterozygotes		Homozygotes		All	
		Concordant	Total	Concordant	Total	Concordant	Total
3	Protein percentage	8	8	10	14	18	22
5	Teat length	6	6	11	15	17	21
	Body depth	10	10	13	15	23	25
	Heifer conception rate	9	9	9	14	18	23
	Dairy form	11	11	2	8	13	19
	Daughter pregnancy rate	12	13	5	12	17	25
	Cow conception rate	5	5	4	9	9	14
	Stature	8	9	12	13	20	22
	Fat percentage	14	15	10	15	24	30
	Final score	7	8	10	14	17	22
	Strength	8	8	10	16	18	24
	Thurl width	8	9	12	14	20	23
6	Productive life	11	11	4	8	15	19
	SCS	15	15	5	8	20	23
	Dairy form	10	11	9	19	19	30
	Heifer conception rate	9	9	5	11	14	20
7	Rump angle	11	11	6	10	17	21
	Stature	15	15	8	11	23	26
10	Strength	13	13	5	9	18	22
11	Stature	12	12	5	6	17	18
14	Fat percentage	11	11	24	26	35	37
	Body depth	5	5	7	10	12	15
	Stature <sup>1</sup>	2	2	11	11	13	13
	Strength	3	3	14	15	17	18
15	Fat yield	6	8	9	17	15	25
18	Cow conception rate	7	8	11	11	18	19
	Daughter pregnancy rate <sup>1</sup>	14	14	11	11	25	25
19	Stature	6	7	5	7	11	14
	Thurl width	9	9	7	9	16	18
20	Protein percentage	10	11	10	10	20	21

<sup>1</sup>Complete concordance.

percentage on BTA20 was a missense mutation. The effect on thurl width on BTA5 was a 3' untranslated-region variant, and all the rest were SNP. Seven of the SNP were found in gene introns, including the marker with the greatest effect on fat percentage on BTA14; the remaining 21 SNP were intergenic variants.

### Effects of Missense and Nonsense Mutations

In addition to the markers with the greatest  $R^2$  for each QTL, an additional 24 markers were flagged because they were missense or nonsense mutations with concordance among the top 60 for each QTL and  $R^2$  of  $>0.01$ . Twelve of these markers were for the effect on fat percentage on BTA14 (Table 5), and the remaining 12 are shown in Table 4. The Ala232Glu polymorphism in *DGAT1*, which was determined to be causative by Grisart et al. (2002) and Winter et al. (2002), is included in both tables. All of the polymorphisms shown in Table 5 but 1 had almost equal  $R^2$  for fat percentage and are in tight population-wide linkage disequilibrium as evidenced by the very low numbers of discrepancies for the genotypes of these polymorphisms with the

genotypes for the Ala232Glu missense polymorphism in *DGAT1* (Table 5).

Three missense mutations were flagged for more than a single trait (Table 4). In 6 cases, a missense mutation was found with an  $R^2$  of  $>90\%$  for the marker with the highest  $R^2$ . Only a single nonsense polymorphism, an indel, was found in *PCDHB6* on BTA7. That polymorphism had a significant effect on stature in the regression analysis but was nonsignificant ( $P > 0.05$ ) in the MMAP analysis and ranked only 51 for concordance.

The same marker was flagged for SCS and productive life on BTA6, which is not surprising because those traits are correlated. A missense mutation in *NPFRR2* had almost the same  $R^2$  as the SNP with the highest  $R^2$ , and that mutation was also flagged for dairy form. For BTA18, different markers had the highest  $R^2$  for daughter pregnancy rate and cow conception rate, even though the 2 traits are highly correlated; both markers were significant in the MMAP analysis. A missense mutation between *ANKRD27* and *RGS9BP* had complete concordance for daughter pregnancy rate and was significant at approximately  $10^{-12}$ . In addition, a missense polymorphism in *CHST8* had almost

**Table 4.** Polymorphisms with greatest coefficient of determination for traits among the 30 markers with highest concordance scores by chromosome number and trait

BTA	Trait	Concordance rank <sup>1</sup>	Location (bp)	Positional gene	Variant type <sup>2</sup>	MMAP <sup>3</sup> probability	R <sup>2</sup>	
							Marker	Total
3	Protein percentage	7	23,347,630	<i>NOTCH2</i>	SNP IN	0.0170	0.0233	0.0453
	Teat length	19	25,596,712	<i>GLYCAM1</i>	SNP IG	0.0025	0.0198	0.0589
5	Body depth	1	90,245,948	5S-rRNA	SNP IG	0.0392	0.0571	0.1141
	Strength	2	90,245,948	5S-rRNA	SNP IG	NS <sup>4</sup>	0.0305	0.0813
14	Heifer conception rate	14	98,815,117	ENSBTAG00000030476	SNP IG	NS	0.0180	0.0493
	Dairy form	16	82,772,317	<i>C5H12orf71-SMCO2</i>	SNP IG	NS	0.0694	0.1087
11	Daughter pregnancy rate	11	87,553,350	<i>SOX5-ETNK1</i>	SNP IG	0.0222	0.0162	0.0497
	Cow conception rate	14	102,490,414	ENSBTAG00000039380	SNP IG	NS	0.0053	0.0357
24	Stature	24	84,862,901	<i>LMNTD1</i>	SNP IG	NS	0.0505	0.1407
		1	89,292,570	<i>SLCO1A2</i>	SNP M <sup>5</sup>	NS	0.0264	0.1407
4	Fat percentage	4	100,488,576	<i>7SK-CLEC2A-KLRF2</i>	Indel IN	NS	0.0149	0.0288
		59	93,945,738	<i>MGST1</i>	SNP IN	4.902 × 10 <sup>-12</sup>	0.0714	0.0965
5	Final score	5	79,415,691	<i>IPO8-TMTC1</i>	SNP IG	NS	0.0792	0.1365
		15	79,238,765	<i>CAPRIN2</i>	SNP M <sup>6</sup>	NS	0.0546	0.1365
6	Thurl width	1	106,223,457	<i>TIGAR</i>	SNP 3 <sup>7</sup>	5.879 × 10 <sup>-13</sup>	0.0410	0.0583
	Productive life	5	88,672,979	<i>SLC4A4-GC</i>	SNP IG	4.504 × 10 <sup>-14</sup>	0.0716	0.1035
6	SCS	6	89,059,253	<i>NPFFR2</i>	SNP M <sup>8</sup>	9.065 × 10 <sup>-10</sup>	0.0700	0.1035
		4	88,672,979	<i>SLC4A4-GC</i>	SNP IR	1.222 × 10 <sup>-15</sup>	0.0758	0.0944
4	Dairy form	4	89,059,253	<i>NPFFR2</i>	SNP M <sup>8</sup>	1.217 × 10 <sup>-15</sup>	0.0728	0.0944
		26	88,911,993	<i>GC-NPFFR2</i>	SNP IG	1.536 × 10 <sup>-19</sup>	0.0075	0.0398
8	Heifer conception rate	8	89,059,253	<i>NPFFR2</i>	SNP M <sup>8</sup>	6.952 × 10 <sup>-25</sup>	0.0040	0.0398
	Rump angle	20	103,202,217	ENSBTAG0000045839- <i>PTPN13</i>	SNP IG	1.043 × 10 <sup>-5</sup>	0.0476	0.0691
7	Stature	9	42,884,968	<i>ARHGAP26</i>	SNP IN	0.0003	0.0328	0.0361
		9	42,690,784	<i>FGF1</i>	SNP M <sup>9</sup>	0.0011	0.0327	0.0361
13	Stature	13	55,509,208	<i>PCDHB6</i>	SNP IG	0.0007	0.0590	0.0889
		51	53,934,111	<i>PCDHB6</i>	Indel-F <sup>10</sup>	NS	0.0427	0.0889
10	Strength	4	64,529,144	5s_rRNA- <i>SQOR</i>	SNP IG	NS	0.0076	0.0370
	Stature	4	77,939,813	ENSBTAG00000030834	SNP IG	0.0032	0.0811	0.1360
14	Fat percentage	55	73,208,603	<i>ADGRF3</i>	SNP M <sup>11</sup>	NS	0.0776	0.1360
		1	1,824,653	<i>HFS1</i>	SNP IN	0.0 <sup>12</sup>	0.3646	0.3671
21	Body depth	1	1,802,226	<i>DGAT1</i>	SNP M <sup>13</sup>	0.0 <sup>12</sup>	0.3630	0.3671
	Stature	4	8,500,069	<i>ZFAT-ST3GAL1</i>	SNP IG	8.420 × 10 <sup>-6</sup>	0.0219	0.0687
4	Strength	4	1,386,174	5s_rRNA- <i>ZNF16</i>	SNP IG	NS	0.0444	0.0599
		4	1,928,347	<i>MAFI-SHARPIN-CYC1</i>	SNP M <sup>14</sup>	NC <sup>15</sup>	0.0071	0.0599
37	Fat yield	37	54,365,783	<i>PPME1</i>	SNP IG	NS	0.0185	0.0374
	Cow conception rate	22	42,759,322	ENSBTAG00000026340	SNP IG	NS	0.0120	0.0374
15	Daughter pregnancy rate	1	43,413,571	<i>TDRD12</i>	SNP M <sup>14</sup>	NS	0.0199	0.0650
		1	44,378,414	<i>CHST8</i>	SNP IG	0.0125	0.0222	0.0639
18	Daughter pregnancy rate	37	43,324,668	<i>ANKRD27-RGS9BP</i>	SNP M <sup>16</sup>	3.554 × 10 <sup>-12</sup>	0.0136	0.0639
		1	44,378,414	<i>CHST8</i>	SNP IN	7.921 × 10 <sup>-12</sup>	0.0257	0.0545
3	Daughter pregnancy rate	1	43,324,668	<i>ANKRD27-RGS9BP</i>	SNP M <sup>17</sup>	9.015 × 10 <sup>-12</sup>	0.0254	0.0545
		3	44,378,414	<i>CHST8</i>	SNP M <sup>16</sup>	2.631 × 10 <sup>-11</sup>	0.0228	0.0545

Continued

**Table 4 (Continued).** Polymorphisms with greatest coefficient of determination for traits among the 30 markers with highest concordance scores by chromosome number and trait

BTA	Trait	Concordance rank <sup>1</sup>	Location (bp)	Positional gene	Variant type <sup>2</sup>	MMAP <sup>3</sup> probability	R <sup>2</sup>	
							Marker	Total
19	Stature	11	39,753,109	<i>ARHGAP23</i>	SNP IN	0.0021	0.0241	0.0911
	Thurl width	5	49,871,005	ENSBTAG00000044065	SNP IN	NS	0.0683	0.1160
20	Protein percentage	1	31,909,478	<i>GHR</i>	SNP M <sup>18</sup>	1.249 × 10 <sup>-18</sup>	0.0500	0.0619

<sup>1</sup>Relative to all families with QTL status determined; regressions were also computed for missense or nonsense polymorphisms with concordance rank among top 60 even though they were not included in top 30.

<sup>2</sup>IN = intronic; IG = intergenic; M = missense.

<sup>3</sup>O'Connell (2017).

<sup>4</sup>Nonsignificant ( $P > 0.05$ ).

<sup>5</sup>His428Arg.

<sup>6</sup>Ile406Met.

<sup>7</sup>3' untranslated-region variant.

<sup>8</sup>Glu392Lys.

<sup>9</sup>Asp122Asn.

<sup>10</sup>F = full stop; Asn111fs; complete concordance for heterozygotes; 2 discrepancies for homozygotes.

<sup>11</sup>Leu1032Ile.

<sup>12</sup>Probability too low to compute.

<sup>13</sup>Ala232Glu; additional polymorphisms in this region (Table 5).

<sup>14</sup>Arg296Gln; marker within segment with lowest probability that null hypothesis was true (Table 2).

<sup>15</sup>Not computed.

<sup>16</sup>Ala267Thr; missense mutation within segment with lowest probability that null hypothesis was true (Table 2).

<sup>17</sup>Ala8Val.

<sup>18</sup>Phe279Tyr.

**Table 5.** Polymorphisms on BTA14 with almost complete concordance with QTL that affects fat percentage

Polymorphism	Gene <sup>1</sup>	Location (bp)	R <sup>2</sup> for fat percentage	Concordance with QTL <sup>2</sup>	Discrepancies with Ala232Glu (no.) <sup>3</sup>
Ala501Thr	<i>ARHGAP39</i>	1,584,444	0.3631	35	302 (0.011)
Val666Met	<i>RECQL4 (LRRC14, MFSD3)</i>	1,617,841	0.3459	35	61 (0.023)
Gln1086Arg	<i>RECQL4 (MFSD3, GPT)</i>	1,619,555	0.3456	34	638 (0.024)
Ala383Thr	<i>PPP1R16A (GPT)</i>	1,629,600	0.3462	34	606 (0.023)
Thr430Ile	<i>CPSF1</i>	1,736,599	0.3643	35	56 (0.002)
Lys242Glu	<i>GPR172B (FBXL6)</i>	1,765,055	0.3646	35	51 (0.002)
Ala232Glu <sup>4</sup>	<i>DGAT1 (HSF1)</i>	1,802,266	0.3630	35	—
Val344Ala	<i>HSF1 (DGAT1)</i>	1,808,145	0.3630	35	1 (3.7 × 10 <sup>-5</sup> )
Intron variant <sup>5</sup>	<i>HSF1 (BOP1)</i>	1,824,653	0.3646	35	53 (0.002)
Asn94Ser	<i>MROH1</i>	1,878,165	0.3629	35	16 (6.0 × 10 <sup>-4</sup> )
Arg103Leu	<i>PARP10 (GRINA)</i>	2,025,096	0.3562	34	858 (0.032)
Trp651Arg	<i>PARP10</i>	2,027,310	0.3563	34	860 (0.032)
Met89Val	ENSBTAG00000048078	5,675,424	0.0662	33	14,869 (0.556)

<sup>1</sup>Additional gene names shown in parentheses.

<sup>2</sup>From 38 families with QTL genotype determined; all polymorphisms had complete concordance for grandsires heterozygous for the QTL.

<sup>3</sup>Fraction of discrepancies out of 26,750 sons with imputed genotypes and EBV for fat percentage in parentheses.

<sup>4</sup>Polymorphism determined by Grisart et al. (2002) and Winter et al. (2002) as causative.

<sup>5</sup>Highest R<sup>2</sup>.

complete concordance for daughter pregnancy rate, and its multiple-regression effect was significant for both daughter pregnancy rate ( $\sim 10^{-11}$ ) and cow conception rate ( $\sim 10^{-12}$ ).

In addition to the effect on fat percentage on BTA14, candidate causative polymorphisms were found for 3 QTL based on 3 criteria (Table 6): complete or almost complete concordance, MMAP significance, and marker R<sup>2</sup> >40% of total multiple-regression R<sup>2</sup>. These are the effects on stature on BTA7, daughter pregnancy rate on BTA18, and protein percentage on BTA20. For each of these QTL, a missense mutation was also found with concordance equal or almost equal to the concordance obtained for the marker with the highest R<sup>2</sup>. Although R<sup>2</sup> were only slightly lower for the missense mutations, the MMAP probabilities for the missense mutations were almost equal for daughter pregnancy rate and lower for protein percentage.

### Effects of the Candidate Causative Polymorphisms on Milk Production Traits

Effects of markers with the highest R<sup>2</sup> for milk production traits on 5 chromosomes with significant effects for those traits are shown in Table 7. Excluding the effect associated with fat percentage on BTA14, the lowest MMAP probability was on BTA5 for the effect of the SNP at 93,945,738 bp on fat percentage (Table 4). This marker within *MGST1* explained 7% of variance for fat percentage and 74% of total multiple-marker regression variance. Although this intronic variant did not have complete concordance in this study, Littlejohn et al. (2016) identified the polymorphism as the most likely QTN candidate for the New Zealand dairy

population using different methodologies. This QTL, however, is of minimal value for commercial breeding because, as with *DGAT1*, the rare allele that increases fat yield and concentration also reduces protein yield (Table 7). Thus, the net economic difference between the 2 homozygotes will be close to 0 for most commercial selection indices (Miglior et al., 2005). The QTL on BTA3 and BTA20 with an allele that increases both fat and protein concentration also will have only minimal application in commercial breeding because the economically favorable alleles are already at very high frequency. Thus, as with *ABCG2* (Cohen-Zinder et al., 2005), the potential for further genetic improvement is very small. The only QTL for production traits that may have significant value for commercial breeding is on BTA15. The frequency of the economically positive allele is 0.45, and the substitution effects of this allele were an increase in fat yield by 5.4 kg, protein yield by 2.55 kg, and fat percentage by 0.012 points. However, no marker with concordance better than expected by chance was found on BTA15, and the most concordant marker explained only 2% of variance for fat percentage (Table 4).

## DISCUSSION

A negative control was done in which 21,000 markers on BTA7 were tested for concordance with actual APGD effects for fat percentage on BTA5. One marker had almost complete concordance for 29 out of 30 families, although this marker did not have a significant ( $P < 0.05$ ) effect on fat percentage in the regression analysis. The maximum R<sup>2</sup> for fat percentage of the 30 markers with the highest concordance values was only

0.0034. Thus, although almost complete concordance could be obtained by chance if >20,000 markers were analyzed, the second criterion of a significant effect on the trait was not found.

Several factors may explain why maximum concordance was no better than chance for most of the analyzed QTL. In the current study, only SNP and indels were considered, and QTN may result from other effects, such as copy number variation (e.g., Glick et al., 2011). The observed QTL may be caused by 2 or more closely linked polymorphisms. Previous studies based on linkage analysis have shown that the greatest effect is observed for a “ghost” QTL located between 2 polymorphisms that affect a trait when those polymorphisms are linked (Martínez and Curnow, 1992). Effects for most QTL were probably too small for complete discrimination between heterozygotes and homozygotes. As noted in the Materials and Methods section, of the 83 grandsires included in the analysis, 19 were not sequenced; their genotypes were determined by imputation from a high-density SNP chip, and some may be incorrect. Finally, the lack of concordance for families homozygous for the QTL may result partially from a bias in favor of calling heterozygotes in sequence data. For markers with relatively high minor allele frequencies, more heterozygotes were observed than expected based on the Hardy-Weinberg equilibrium. For the bulls with >100 progeny-tested sons, 36,527 markers on BTA5 had an expected heterozygote frequency of >0.4. The observed frequency of heterozygotes was 1.4 percentage points higher than the expected Hardy-Weinberg frequency. Much of the difference may be because samples of unrelated cattle breeds were used for multisample SNP and indel calling in the 1000 Bull Genomes Project (Hayes et al., 2014).

Markers that affect milk production traits were found on BTA3, BTA5, BTA14, BTA15, and BTA20. Except for the effect on BTA15, all effects were chiefly for fat and protein concentration and had been found previously in the Australian Holstein population by Kemper et al. (2015). The effect on protein percentage on BTA3 was localized to 23 to 24 Mbp, whereas previous studies found a QTL located between 25 and 45 Mbp (Ashwell et al., 2004; Cohen-Zinder et al., 2011).

With respect to effect on fat percentage on BTA14, all of the polymorphisms in Table 5 are equally likely candidates for the causative polymorphism based only on statistical considerations, concordance with effect on fat percentage, and fraction of variance explained by each polymorphism (except for the missense polymorphism at 5,675,424 bp). However, bioinformatic and biochemical analyses lend additional support to the conclusion that the Ala232Glu polymorphism is the main causative polymorphism (Grisart et al., 2004).

**Table 6.** Polymorphisms that met 3 statistical criteria<sup>1</sup> for candidate causative mutation

BTA	Trait	Concordance score	Location (bp)	Positional gene	Variant type <sup>2</sup>	AA substitution	MMAF <sup>3</sup> probability	Fraction of regression variance explained by marker
11	Stature	17/18	77,939,813	ENSBTAG00000030834	SNP IG		0.0032	0.596
		13/15	73,208,603	<i>ADGRF3</i>	SNP M	Leu1032Ile	NS <sup>4</sup>	0.571
18	Daughter pregnancy rate	25/25	43,413,571	<i>TDRD12</i>	SNP IN		$7.921 \times 10^{-12}$	0.472
		25/25	43,324,668	<i>ANKRD27-RGS9BP</i>	SNP M	Ala8Val	$9.015 \times 10^{-12}$	0.466
		24/25	44,378,414	<i>CHST8</i>	SNP M	Ala267Thr	$2.631 \times 10^{-11}$	0.418
20	Protein percentage	20/21	31,909,478	<i>GHR</i>	SNP M	Phe279Tyr	$1.249 \times 10^{-18}$	0.808

<sup>1</sup>Complete or almost complete concordance, significance of polymorphism effect after correction for all other polymorphisms, and marker R<sup>2</sup> >40% of total multiple-regression R<sup>2</sup> for 30 polymorphisms with highest concordance.

<sup>2</sup>IN = intronic; IG = intergenic; M = missense.

<sup>3</sup>O'Connell (2017).

<sup>4</sup>Nonsignificant ( $P > 0.05$ ).

**Table 7.** Effects of markers with highest correlation of determination for milk production traits on 5 chromosomes with significant effects for those traits computed from 26,750 bulls with imputed genotypes and PTA based on progeny test

BTA	Location (bp)	Trait	Genotype <sup>1</sup>			Mean
			0	1	2	
3	23,347,630	Frequency (no.)	19,685	6,608	457	—
		Milk (kg)	-205	-192	-136	-200
		Fat (kg)	-4.6	-6.4	-7.5	-5.1
		Protein (kg)	-3.3	-6.2	-8.1	-4.1
		Fat (%)	0.013	0.004	-0.008	0.01
		Protein (%)	0.011	-0.001	-0.014	0.008
5	93,945,738	Frequency (no.)	7,644	13,561	5,545	—
		Milk (kg)	-149	-201	-269	-200
		Fat (kg)	-7.1	-4.9	-2.8	-5.1
		Protein (kg)	-2.5	-4.1	-6.1	-4.1
		Fat (%)	-0.005	0.011	0.028	0.01
		Protein (%)	0.008	0.008	0.008	0.008
14	1,824,653	Frequency (no.)	16,244	9,169	1,337	—
		Milk (kg)	-30	-410	-837	-200
		Fat (kg)	-10.0	1.2	11.2	-5.1
		Protein (kg)	-0.8	-8.0	-16.7	-4.1
		Fat (%)	-0.032	0.063	0.162	0.01
		Protein (%)	0.001	0.017	0.032	0.008
15	54,365,783	Frequency (no.)	5,415	13,320	8,014	—
		Milk (kg)	-125	-187	-242	-191
		Fat (kg)	1.2	-4.5	-9.6	-4.9
		Protein (kg)	-1.1	-3.5	-6.2	-3.9
		Fat (%)	0.023	0.011	-0.001	0.01
		Protein (%)	0.011	0.008	0.005	0.008
20	31,909,478	Frequency (no.)	18,376	7,603	770	—
		Milk (kg)	-212	-144	-160	-191
		Fat (kg)	-3.9	-6.2	-13.9	-4.9
		Protein (kg)	-2.9	-5.3	-10.9	-3.9
		Fat (%)	0.016	-0.002	-0.028	0.01
		Protein (%)	0.013	-0.003	-0.022	0.008

<sup>1</sup>Animals heterozygous for marker were designated as genotype 1; 2 homozygotes were designated as genotypes 0 and 2.

Wiggans and Weller (2015) raised the possibility that the effect on cow fertility found on BTA5 might be the same as the HH1 haplotype effect described by Adams et al. (2016). However, the positions of the 2 effects are 25 Mbp apart.

Localization of the effect of fat percentage on BTA5 between 94 and 95 Mbp close to *LMO3* and *MGST1* (Figure 2) is supported by a segregating QTL in this region that has been verified in other populations (Littlejohn et al., 2016). The *LMO3* gene has been found in human clinical research to be expressed primarily in the brain and is frequently found to be associated with certain neural blastomas (Aoyama et al., 2005). Functional significance of *LMO3* expression on milk production traits in cattle is currently unknown. The *MGST1* gene encodes a glutathione S-transferase that protects against oxidative stress and is responsible for conjugation of glutathione to reactive intermediates. The enzyme is localized to the mitochondrial membrane and endoplasmic reticulum, and known functions include aspects of lipid biology; *MGST1* has been shown to reduce lipid peroxidation products in human mammary cell culture (Littlejohn et al., 2016).

Blott et al. (2003) proposed that the missense SNP in *GHR* at 31,909,478 bp on BTA20 is the causative mutation for fat and protein concentration. They found almost complete concordance for this polymorphism in a granddaughter-design analysis of Dutch Holstein cattle with 22 paternal half-sib families and 987 sons. Almost complete concordance was found for this SNP (Table 3), and it was ranked first both by concordance and by multiple-regression  $R^2$ . Thus, this study confirms the effect of the missense mutation Phe279Tyr in *GHR* even though complete concordance was not found by Blott et al. (2003).

Since the discovery of a missense mutation in *ABCG2* (Cohen-Zinder et al., 2005), no additional causative polymorphisms have been verified in the last decade based on criteria presented by Ron and Weller (2007). The current state of cattle genome annotation suggests that most of the flagged polymorphisms, which were intronic or intergenic SNP, might have no putative functional effects on gene expression. However, future cattle assembly and annotation projects (Andersson et al., 2015) may uncover cryptic functional elements in these regions that may have a biological effect. Furthermore,



identifying markers with significant effects on economic traits and incorporating those markers into SNP panels used for genomic evaluation are still worthwhile even if they are not necessarily causative polymorphisms.

## CONCLUSIONS

The prediction of Weller and Ron (2011) of the utility of grandsires with complete genome sequence to determine QTN based on APGD concordance has been only partially realized. Complete concordance was obtained only for stature on BTA14 and daughter pregnancy rate on BTA18. Effects for stature on BTA11, daughter pregnancy rate on BTA18, and protein percentage on BTA20 met 3 criteria: complete or almost complete concordance, significance of the polymorphism effect after correction for all other polymorphisms, and a marker  $R^2 > 40\%$  of total multiple-regression  $R^2$  for the 30 polymorphisms with highest concordance. An intronic variant SNP on BTA5 at 93,945,738 bp explained 7% of variance for fat percentage and 74% of total multiple-marker regression variance but was concordant for only three-fourths of families with significant contrasts for that trait. Ten missense polymorphisms were in almost complete concordance with the Ala232Glu polymorphism in *DGAT1*, the causative polymorphism for effect on fat percentage on BTA14. This study confirmed the finding of Blott et al. (2003) that a missense polymorphism in *GHR* at 31,909,478 bp on BTA20 is the causative mutation for fat and protein concentration. The only QTL that is likely to improve predictive power significantly for genomic evaluation of dairy cattle for production traits is on BTA15 with an allelic frequency of 0.45 for the economically favorable allele. Although this study uses the largest data set of whole-genome sequence for grandsires and sons, the power of the concordance test to determine QTN is still rather limited. The small effects of the underlying QTN hinder deduction of grandsire genotypes for the QTL without error, which agrees with the infinitesimal model for quantitative traits. Incorporating markers flagged in this analysis into SNP panels used for genomic evaluation is still worthwhile even if they are not necessarily the causative polymorphisms.

## ACKNOWLEDGMENTS

This research was supported by grant number 58-8042-5-063F from the U.S.–Israel Binational Agricultural Research and Development (BARD) Fund. The Council on Dairy Cattle Breeding (Bowie, MD) and the Cooperative Dairy DNA Repository (Columbia, MO) are acknowledged for providing data. D. M. Bickhart,

G. R. Wiggans, M. E. Tooker, and P. M. VanRaden were funded by USDA Agricultural Research Service appropriated projects 8042-31000-101-00, “Improving Genetic Predictions in Dairy Animals Using Phenotypic and Genomic Information,” and 8042-31000-002-00, “Improving Dairy Animals by Increasing Accuracy of Genomic Prediction, Evaluating New Traits, and Redefining Selection Goals”; J. R. O’Connell was funded by the University of Maryland School of Medicine (Baltimore) and by USDA Agricultural Research Service specific cooperative agreement 58-1245-4-070, “Improvement of Algorithms and Software to Process Very Large Genomic Datasets.” Data analysis and visualization were aided by Daniel’s XL Toolbox add-in for Excel, version 7.2.13, by Daniel Kraus (Würzburg, Germany; [www.xltoolbox.net](http://www.xltoolbox.net)). We thank S. M. Hubbard (USDA, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD) for editorial review. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by USDA (Washington, DC).

## REFERENCES

- Adams, H. A., T. S. Sonstegard, P. M. VanRaden, D. J. Null, C. P. Van Tassell, D. M. Larkin, and H. A. Lewin. 2016. Identification of a nonsense mutation in *APAF1* that is likely causal for a decrease in reproductive efficiency in Holstein dairy cattle. *J. Dairy Sci.* 99:6693–6701.
- Andersson, L., A. L. Archibald, C. D. Bottema, R. Brauning, S. C. Burgess, D. W. Burt, E. Casas, H. H. Cheng, L. Clarke, C. Coul-drey, B. P. Dalrymple, C. G. Elsik, S. Foissac, E. Giuffra, M. A. Groenen, B. J. Hayes, L. S. Huang, H. Khatib, J. W. Kijas, H. Kim, J. K. Lunney, F. M. McCarthy, J. C. McEwan, S. Moore, B. Nanduri, C. Notredame, Y. Palti, G. S. Plastow, J. M. Reecy, G. A. Rohrer, E. Sarropoulou, C. J. Schmidt, J. Silverstein, R. L. Tellam, M. Tixier-Boichard, G. Tosser-Klopp, C. K. Tuggle, J. Vilkki, S. N. White, S. Zhao, and H. Zhou. 2015. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol.* 16:57.
- Aoyama, M., T. Ozaki, H. Inuzuka, D. Tomotsune, J. Hirato, Y. Okamoto, H. Tokita, M. Ohira, and A. Nakagawara. 2005. LMO3 interacts with neuronal transcription factor, HEN2, and acts as an oncogene in neuroblastoma. *Cancer Res.* 65:4587–4597.
- Ashwell, M. S., D. W. Heyen, T. S. Sonstegard, C. P. Van Tassell, Y. Da, P. M. VanRaden, M. Ron, J. I. Weller, and H. A. Lewin. 2004. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. *J. Dairy Sci.* 87:468–475.
- Baes, C. F., M. A. Dolezal, J. E. Koltjes, B. Bapst, E. Fritz-Waters, S. Jansen, C. Flury, H. Signer-Hasler, C. Stricker, R. Fernando, R. Fries, J. Moll, D. J. Garrick, J. M. Reecy, and B. Gredler. 2014. Evaluation of variant identification methods for whole genome sequencing data in dairy cattle. *BMC Genomics* 15:948.
- Bennewitz, J., N. Reinsch, S. Paul, C. Looft, B. Kaupé, C. Weimann, G. Erhardt, G. Thaller, C. Kühn, M. Schwerin, H. Thomsen, F. Reinhardt, R. Reents, and E. Kalm. 2004. The *DGAT1* K232A mutation is not solely responsible for the milk production quantitative trait locus on the bovine chromosome 14. *J. Dairy Sci.* 87:431–442.

- Bentley, D. R., S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. K. Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. J. Rasolonjatovo, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, A. Spiridou, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. D. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, M. C. E. Catenazzi, S. Chang, R. N. Cooley, N. R. Crane, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. Fuentes Fajardo, W. S. Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, D. L. Gustafson, N. F. Hansen, K. Harnish, C. D. Haudenschield, N. I. Heyer, M. M. Hims, J. T. Ho, A. M. Horgan, K. Hoshler, S. Hurwitz, D. V. Ivanov, M. Q. Johnson, T. James, T. A. H. Jones, G.-D. Kang, T. H. Kerelska, A. D. Kersey, I. Khrebtukova, A. P. Kindwall, Z. Kingsbury, P. I. Kokko-Gonzales, A. Kumar, M. A. Laurent, C. T. Lawley, S. E. Lee, X. Lee, A. K. Liao, J. A. Loch, M. Lok, S. Luo, R. M. Mammen, J. W. Martin, P. G. McCauley, P. McNitt, P. Mehta, K. W. Moon, J. W. Mullens, T. Newington, Z. Ning, B. L. Ng, S. M. Novo, M. J. O'Neill, M. A. Osborne, A. Osnowski, O. Ostadan, L. L. Paraschos, L. Pickering, A. C. Pike, A. C. Pike, D. C. Pinkard, D. P. Pliskin, J. Podhasky, V. J. Quijano, C. Raczy, V. H. Rae, S. R. Rawlings, A. C. Rodriguez, P. M. Roe, J. Rogers, M. C. Rogert Bacigalupo, N. Romanov, A. Romieu, R. K. Roth, N. J. Rourke, S. T. Ruediger, E. Rusman, R. M. Sanches-Kuiper, M. R. Schenker, J. M. Seoane, R. J. Shaw, M. K. Shiver, S. W. Short, N. L. Sizto, J. P. Sluis, M. A. Smith, J. E. Sohna Sohna, E. J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C. L. Tregidgo, G. Turcatti, S. vandeVondele, Y. Verhovskiy, S. M. Virk, S. Wakelin, G. C. Walcott, J. Wang, G. J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. Rogers, J. C. Mullikin, M. E. Hurler, N. J. McCooke, J. S. West, F. L. Oaks, P. L. Lundberg, D. Klenerman, R. Durbin, and A. J. Smith. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59.
- Blott, S., J.-J. Kim, S. Moio, A. Schmidt-Küntzel, A. Cornet, P. Berzi, N. Cambisano, C. Ford, B. Grisart, D. Johnson, L. Karim, P. Simon, R. Snell, R. Spelman, J. Wong, J. Vilkki, M. Georges, F. Farnir, and W. Coppieters. 2003. Molecular dissection of a quantitative trait locus: A phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* 163:253–266.
- Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* 6:80–92.
- Cohen-Zinder, M., R. Donthu, D. M. Larkin, C. G. Kumar, S. L. Rodriguez-Zas, K. E. Andropoulos, R. Oliveira, and H. A. Lewin. 2011. Multisite haplotype on cattle chromosome 3 is associated with quantitative trait locus effects on lactation traits. *Physiol. Genomics* 43:1185–1197.
- Cohen-Zinder, M., E. Seroussi, D. M. Larkin, J. J. Loor, A. Evertsvan der Wind, J.-H. Lee, J. K. Drackley, M. R. Band, A. G. Hernandez, M. Shani, H. A. Lewin, J. I. Weller, and M. Ron. 2005. Identification of a missense mutation in the bovine *ABCG2* gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Res.* 15:936–944.
- Cole, J. B., P. M. VanRaden, J. R. O'Connell, C. P. Van Tassell, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and G. R. Wiggans. 2009. Distribution and location of genetic effects for dairy traits. *J. Dairy Sci.* 92:2931–2946.
- García-Ruiz, A., J. B. Cole, P. M. VanRaden, G. R. Wiggans, F. J. Ruiz-López, and C. P. Van Tassell. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci. USA* 113:E3995–E4004.
- Glazier, A. M., J. H. Nadeau, and T. J. Aitman. 2002. Finding genes that underlie complex traits. *Science* 298:2345–2349.
- Glick, G., A. Shirak, E. Seroussi, Y. Zeron, E. Ezra, J. I. Weller, and M. Ron. 2011. Fine mapping of a QTL for fertility on BTA7 and its association with a CNV in the Israeli Holsteins. *G3 (Bethesda)* 1:65–74.
- Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. *Genome Res.* 12:222–231.
- Grisart, B., F. Farnir, L. Karim, N. Cambisano, J.-J. Kim, A. Kvasz, M. Mni, P. Simon, J.-M. Frère, W. Coppieters, and M. Georges. 2004. Genetic and functional confirmation of the causality of the *DGAT1 K232A* quantitative trait nucleotide in affecting milk yield and composition. *Proc. Natl. Acad. Sci. USA* 101:2398–2403.
- Hayes, B. J., I. M. MacLeod, H. D. Daetwyler, P. J. Bowman, A. J. Chamberlain, C. J. Vander Jagt, A. Capitan, H. Pausch, P. Stothard, X. Liao, C. Schrooten, E. Mullaart, R. Fries, B. Guldbandsen, M. S. Lund, D. A. Boichard, R. F. Veerkamp, C. P. Van Tassell, B. Gredler, T. Druet, A. Bagnato, J. Vilkki, D. J. deKoning, E. Santus, and M. E. Goddard. 2014. Genomic prediction from whole genome sequence in livestock: The 1000 Bull Genomes Project. Commun. 183 in *Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, BC, Canada. Am. Soc. Anim. Sci., Champaign, IL.*
- Kemper, K. E., B. J. Hayes, H. D. Daetwyler, and M. E. Goddard. 2015. How old are quantitative trait loci and how widely do they segregate? *J. Anim. Breed. Genet.* 132:121–134.
- Kühn, C., G. Thaller, A. Winter, O. R. P. Bininda-Emonds, B. Kaupe, G. Erhardt, J. Bennewitz, M. Schwerin, and R. Fries. 2004. Evidence for multiple alleles at the *DGAT1* locus better explains a quantitative trait locus with major effect on milk fat content in cattle. *Genetics* 167:1873–1881.
- Lander, E., and L. Kruglyak. 1995. Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nat. Genet.* 11:241–247.
- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Littlejohn, M. D., K. Tiplady, T. A. Fink, K. Lehnert, T. Lopdell, T. Johnson, C. Couldrey, M. Keehan, R. G. Sherlock, C. Harland, A. Scott, R. G. Snell, S. R. Davis, and R. J. Spelman. 2016. Sequence-based association analysis reveals an *MGST1* eQTL with pleiotropic effects on bovine milk composition. *Sci. Rep.* 6:25376.
- Mackay, T. F. 2001. The genetic architecture of quantitative traits. *Annu. Rev. Genet.* 35:303–339.
- Martínez, O., and R. N. Curnow. 1992. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theor. Appl. Genet.* 85:480–488.
- Miglior, F., B. L. Muir, and B. J. Van Doormaal. 2005. Selection indices in Holstein cattle of various countries. *J. Dairy Sci.* 88:1255–1263.
- Ni, G., T. M. Strom, H. Pausch, C. Reimer, R. Preisinger, H. Simianer, and M. Erbe. 2015. Comparison among three variant callers and assessment of the accuracy of imputation from SNP array data to whole-genome sequence level in chicken. *BMC Genomics* 16:824.
- O'Connell, J. R. 2017. MMAP: Mixed model analysis for pedigrees and populations. Accessed Feb. 21, 2018. <https://mmap.github.io/>.
- Ron, M., M. Cohen-Zinder, C. Peter, J. I. Weller, and G. Erhardt. 2006. Short communication: A polymorphism in *ABCG2* in *Bos indicus* and *Bos taurus* cattle breeds. *J. Dairy Sci.* 89:4921–4923.

- Ron, M., and J. I. Weller. 2007. From QTL to QTN identification in livestock—Winning by points rather than knock-out: A review. *Anim. Genet.* 38:429–439.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423.
- VanRaden, P. M., J. R. O’Connell, G. R. Wiggans, and K. A. Weigel. 2011. Genomic evaluations with many more genotypes. *Genet. Sel. Evol.* 43:10.
- VanRaden, P. M., M. E. Tooker, J. R. O’Connell, J. B. Cole, and D. M. Bickhart. 2017. Selecting sequence variants to improve genomic predictions for dairy cattle. *Genet. Sel. Evol.* 49:32.
- Weller, J. I., J. B. Cole, P. M. VanRaden, and G. R. Wiggans. 2014. Application of the *a posteriori* granddaughter design to the Holstein genome. *Animal* 8:511–519.
- Weller, J. I., G. Glick, E. Ezra, Y. Zeron, E. Seroussi, and M. Ron. 2010. Paternity validation and estimation of genotyping error rate for the BovineSNP50 BeadChip. *Anim. Genet.* 41:551–553.
- Weller, J. I., Y. Kashi, and M. Soller. 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J. Dairy Sci.* 73:2525–2537.
- Weller, J. I., and M. Ron. 2011. Invited review: Quantitative trait nucleotide determination in the era of genomic selection. *J. Dairy Sci.* 94:1082–1090.
- Weller, J. I., J. Z. Song, D. W. Heyen, H. A. Lewin, and M. Ron. 1998. A new approach to the problem of multiple comparisons in the genetic dissection of complex traits. *Genetics* 150:1699–1706.
- Wiggans, G. R., and J. I. Weller. 2015. Revisiting the “a posteriori” granddaughter design. *Interbull Bull.* 49:36–42.
- Winter, A., W. Krämer, F. A. O. Werner, S. Kollers, S. Kata, G. Durstewitz, J. Buitkamp, J. E. Womack, G. Thaller, and R. Fries. 2002. Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (*DGAT1*) with variation at a quantitative trait locus for milk. *Proc. Natl. Acad. Sci. USA* 99:9300–9305.
- Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10:R42.