BREEDING AND GENETICS SYMPOSIUM: Networks and pathways to guide genomic selection

W. M. Snelling, R. A. Cushman, J. W. Keele, C. Maltecca, M. G. Thomas, M. R. S. Fortes and A. Reverter

J ANIM SCI 2013, 91:537-552.
doi: 10.2527/jas.2012-5784 originally published online October 24, 2012

BREEDING AND GENETICS SYMPOSIUM: Networks and pathways to guide genomic selection[1–3]

W. M. Snelling,*[4] R. A. Cushman,* J. W. Keele,* C. Maltecca,† M. G. Thomas,‡ M. R. S. Fortes,§ and A. Reverter#

*USDA-ARS[5] U.S. Meat Animal Research Center, Clay Center, NE 68933; †Department of Animal Science, North Carolina State University, Raleigh 27606; ‡Department of Animal Science, Colorado State University, Fort Collins 80523; §Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Gatton Campus, QLD 4343, Australia; and #CSIRO Animal, Food and Health Sciences, Queensland Bioscience Precinct, Brisbane, QLD 4067, Australia

ABSTRACT: Many traits affecting profitability and sustainability of meat, milk, and fiber production are polygenic, with no single gene having an overwhelming influence on observed variation. No knowledge of the specific genes controlling these traits has been needed to make substantial improvement through selection. Significant gains have been made through phenotypic selection enhanced by pedigree relationships and continually improving statistical methodology. Genomic selection, recently enabled by assays for dense SNP located throughout the genome, promises to increase selection accuracy and accelerate genetic improvement by emphasizing the SNP most strongly correlated to phenotype although the genes and sequence variants affecting phenotype remain largely unknown. These genomic predictions theoretically rely on linkage disequilibrium (LD) between genotyped SNP and unknown functional variants, but familial linkage may increase effectiveness when predicting individuals related to those in the training data. Genomic selection with functional SNP genotypes should be less reliant on LD patterns shared by training and target populations, possibly allowing robust prediction across unrelated populations. Although the specific variants causing polygenic variation may never be known with certainty, a number of tools and resources can be used to identify those most likely to affect phenotype. Associations of dense SNP genotypes with phenotype provide a 1-dimensional approach for identifying genes affecting specific traits; in contrast, associations with multiple traits allow defining networks of genes interacting to affect correlated traits. Such networks are especially compelling when corroborated by existing functional annotation and established molecular pathways. The SNP occurring within network genes, obtained from public databases or derived from genome and transcriptome sequences, may be classified according to expected effects on gene products. As illustrated by functionally informed genomic predictions being more accurate than naive whole-genome predictions of beef tenderness, coupling evidence from livestock genotypes, phenotypes, gene expression, and genomic variants with existing knowledge of gene functions and interactions may provide greater insight into the genes and genomic mechanisms affecting polygenic traits and facilitate functional genomic selection for economically important traits.

Key words: beef, gene function, gene network, genomic selection, systems biology, tenderness

J. Anim. Sci. 2013.91:537–552
doi:10.2527/jas2012-5784

[1]Based on a presentation at the symposium titled "Systems Biology in Animal Breeding: Identifying relationships among markers, genes, and phenotypes" at the Joint Annual Meeting, July 15–19, 2012, Phoenix, AZ. The symposium was sponsored, in part, by Monsanto (St. Louis, MO), with publication sponsored by the American Society of Animal Science and the *Journal of Animal Science*..

[2]Mention of a trade name, proprietary product, or specific equipment does not constitute a guarantee or warranty by the USDA and does not imply approval to the exclusion of other products that may be suitable.

[3]The authors thank S. Shackelford, A. King, T. Wheeler G. Hayes, and the USMARC MSQRU and cattle operations staff for animal care and data recording.

[4]Corresponding author: warren.snelling@ars.usda.gov

[5]USDA is an equal opportunity provider and employer.

Received August 23, 2012.
Accepted October 9, 2012.

INTRODUCTION

Why is understanding functional polymorphisms important for genomic selection? Knowledge of the genomic variants causing phenotypic variation has not been needed to improve performance by selection. Advances in statistical methodology accompanied by accumulated performance and pedigree records have enabled national and international genetic evaluation systems to predict EBV for entire populations (Powell and Norman, 2006; Golden et al., 2009). Promising young candidates can be selected using EBV predicted from their own performance and the performance of their relatives. Truly outstanding individuals, contributing to widespread genetic improvement via AI and other reproductive technologies, can be identified with considerable progeny observation.

Still with no knowledge of underlying functional variants, genomic selection can accelerate genetic improvement by enabling accurate evaluation at birth, eliminating the lag between an initial pedigree estimate and accumulation of progeny records for more accurate EBV. Incorporating genotypes from whole-genome SNP assays into existing genetic evaluation systems has increased accuracy of EBV of young animals for commonly recorded traits (Lôbo et al., 2011; Northcutt, 2011; Wiggans et al., 2011). Applicability of these predictions is limited, however, to selection within breed for the traits included in the evaluations. The lack of ability for current whole-genome predictions to increase EBV accuracy across populations (Hayes et al., 2009; Weber et al., 2012) raises concerns about applying genomic predictions to select replacements from within commercial production systems and extension of predictions from small populations that are intensely phenotyped for economically relevant traits that are too expensive or difficult to measure routinely. Although current genomic predictions are dependent on linkage disequilibrium (**LD**) between SNP and unknown causal variants, shifting focus from the assayed SNP to variants more likely to have functional effects may improve portability of genomic predictions across breeds and into crossbred populations. Using beef tenderness for illustration, this article explores tools and resources available to assist functional genomic selection.

Improving Genomic Predictions

Accuracy of genomic selection is influenced by accuracy of marker effect estimates and correlations between genotyped markers and underlying QTL (Goddard, 2009). Heritability and the number of genotyped and phenotyped individuals affect accuracy of marker effect estimates. Data requirements increase exponentially with decreasing heritability (Fig. 1). Whereas a few thousand genotypes and phenotypes can provide sufficiently accurate estimates to explain at least half of the genetic variation for moderate to highly heritable traits($h^2 > 0.3$) within a small effective population, tens to hundreds of thousands of records are needed to achieve similar accuracy for lowly heritable traits ($h^2 < 0.1$) and large effective population sizes. Strategies using progeny means (Goddard and Hayes, 2009), deregressed EBV (Garrick et al., 2009), or single-step approaches (Aguilar et al., 2010) to combine historic pedigree and phenotypes with recent genotypes allow recorded phenotypes to contribute to genomic selection accuracy without requiring every phenotyped individual to be genotyped. Similarly, genotyping DNA pooled by extreme phenotypes, so that individual phenotypes within pools are similar but distinctly different between pools, can enable training with large numbers of animals without requiring every animal to be genotyped (MacGregor et al., 2008; Henshall et al., 2012; McDaneld et al., 2012). Correlated indicator trait phenotypes may also add accuracy to predictions for a trait of interest using multiple-trait genomic selection (Calus and Veerkamp, 2011).

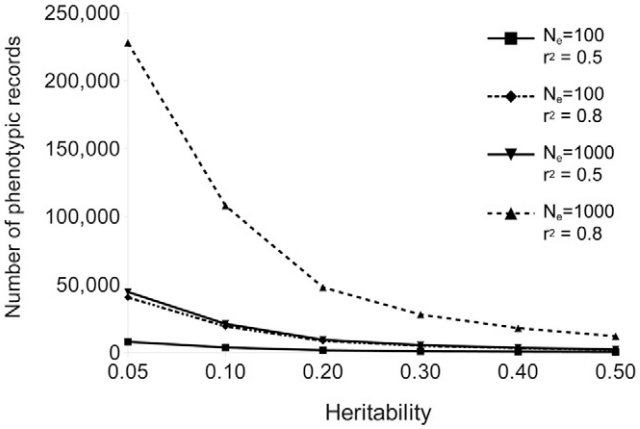

Figure 1. Approximate number of phenotypes needed to realize genomic selection accuracies (r^2) of 0.50 and 0.80, with heritabilities between 0.05 and 0.50 for effective population sizes (N_e) of 100 and 1,000. Equations are from Goddard (2009).

Correlations between marker genotypes and unknown QTL are affected by marker density and consistency of LD patterns between training and target populations (Goddard and Hayes, 2009). Commercially available SNP chips for livestock, containing 50,000 to 60,000 SNP (**50K**; e.g., BovineSNP50, OvineSNP50, PorcineSNP60 BeadChips from Illumina Inc., San Diego, CA), are sufficiently dense to ensure most QTL will be in LD with SNP on the chip although the extent of QTL–SNP LD will be population specific. Marker effects estimated using these chips reflect correlations between markers and QTL within the training population, so accuracies of genomic predictions are dependent on accuracy of the effect estimates and consistency of the marker–QTL correlations between training and target popula-

tions. Variation in LD patterns across breeds (Gautier et al., 2007; Bovine HapMap Consortium, 2009) indicates that accurate marker effects estimated from 1 breed may not apply to another breed. Low accuracies of genomic selection across breeds and crosses were shown using either Holstein or Jersey to predict the other breed (Hayes et al., 2009) and in multibreed evaluations to predict purebred beef bulls from crossbred calves or predict the crossbred calves from training on deregressed bull EBV (Weber et al., 2012). Variable LD within breed, such as that reported between Miles City Line 1 and a broad sample of Hereford bulls (Huang et al., 2012), may also compromise genomic selection accuracy. Cross-validation, using 4 groups of Angus bulls to predict a fifth group, showed decreased accuracies when bulls with close pedigree relationships were grouped together (decreasing relationships between groups) than when the bulls were randomly grouped (Saatchi et al., 2011).

Given this context, avenues to increasing accuracy of genomic predictions are dependent on raising the number of relevant genotypes and phenotypes used for training and increasing the correlations between genotyped markers and unknown QTL. Increasing the number of genotypes and phenotypes is straightforward. The caveat that they be relevant means the added records should be related to the intended target population. The most accurate within-breed evaluations of routinely recorded traits may be from systems that capture all available phenotypes, pedigree, and genotypic information, using genotypes that are broadly representative of the entire breed. Further increases in accuracy can be realized through accumulation of phenotypes and genotypes, so that predictions can be retrained as records are added. In such a cyclic genomic evaluation system, the complete breed serves as the training population targeting the next generation. If predictions based on an initial broad sampling of the breed are not retrained by subsequent generations, accuracy will diminish as relatedness between the initial sample and future generations decreases.

Leveraging extensive performance and pedigree databases in genomic selection is not an option for traits that are not recorded routinely or when the target is selection within commercial mixed breed composite and crossbred populations. Several expensive- and difficult-to-measure traits related to animal health, fertility, biological efficiency, and consumer acceptance may be recorded on intensively phenotyped experimental populations. Genomic predictions can increase EBV accuracy within these populations (Snelling et al., 2011) although extending the genomic predictions to broader livestock industries is limited by lack of relationships with industry populations and lack of phenotypes on industry livestock. For traits where cost, time, and expertise are impediments to developing industry databases with enough relevant phenotypes to support whole-genome selection, efforts to increase the correlations between markers and QTL for these economically important traits may enable genomic selection in industry based on findings from unrelated experimental herds.

Table 1. Genomic heritabilities (and SE) of birth weight and LM area estimated using 2 densities of whole-genome SNP genotypes

Population[1]	SNP[2]	Birth weight	LM area
GPE Cycle VII	HD	0.64 (0.03)	0.50 (0.05)
	50K	0.63 (0.03)	0.47 (0.05)
	None	0.60 (0.04)	0.54 (0.07)
All GPE	HD	0.64 (0.02)	0.50 (0.04)
	50K	0.58 (0.02)	0.47 (0.04)
	None	0.60 (0.03)	0.53 (0.06)

[1]GPE = Germplasm Evaluation; GPE Cycle VII represents 2-, 3-, and 4-breed crosses of 7 *Bos taurus* breeds evaluated in Cycle VII of the GPE project. All GPE includes Cycle VII, Cycle VIII, which evaluated F_1 progeny of 4 tropically adapted *Bos taurus* and *Bos indicus*-influenced breeds, and continuous GPE, which is evaluating crossbred and purebred cattle of 16 *Bos taurus* and *Bos indicus*-influenced breeds.

[2]HD = 630,579 autosomal SNP with minor allele frequency (MAF) > 0.05; 50K = 39,372 autosomal SNP with MAF > 0.05; none = pedigree relationships, no SNP.

The high-density arrays (i.e., >600,000 SNP), which are now available for cattle (Rincon et al., 2011), provide genotypes for SNP that are more certain to have high correlations with unknown QTL than the dense 50K assays. Increased marker density alone, however, does not increase marker–QTL correlations. As long as SNP are sufficiently dense to be in LD with QTL, increased density adds redundancy so there are more genetic markers for the same QTL. From a whole-genome perspective, 50K and BovineHD (Illumina Inc., San Diego, CA) genotypes explained similar amounts of additive variation in birth weight and LM area within a crossbred cattle herd, resulting in similar accuracies of genomic EBV predicted for a somewhat related herd (Table 1). Rather than density of genetic markers, the real impediment to increasing marker–QTL correlations is lack of knowledge about the QTL and underlying genomic variants causing phenotypic variation. Information about gene function and expression can indicate which genes and regulatory elements are most likely to affect phenotype, thereby enabling genomic evaluation to focus on effects of sequence variation within relevant genes and their regulators. Because of less ambiguity in assigning SNP to annotated features of the genome, high-density genotypes can serve to sharpen focus on the features likely to harbor QTL.

The cost of genotyping is an impediment to both increasing relevant numbers of genotypes and increasing marker–QTL correlations by focusing on likely QTL, especially if genotyping costs exceed the potential value of increased selection accuracy. Technologies for low-cost

genotyping by sequence (Elshire et al., 2011; DeDonato et al., 2012) and less expensive, low-density SNP chips (Boichard et al., 2012) using imputation to infer greater-density genotypes (Browning and Browning, 2009; Sargolzaei et al., 2011; Van Raden et al., 2011) may enable more cost-effective genomic prediction within performance-recorded populations. Low-cost targeted genotyping with next-generation sequencing (Thallman and Koshinsky, 2012) can also support imputation to greater-density genotypes and enable genotyping specific variants likely to have functional effects.

Genome Annotation and Functional Information

Mechanisms relating DNA markers to genes and genome features are key to informing genomic evaluations with functional information. Naive genomic selection and genomewide association studies (**GWAS**) can describe additive variation within a reference population and quantify associations of specific markers with a phenotype, further interpretation of GWAS, including genomic segments harboring possible QTL and positional candidate genes that may affect phenotype requires knowledge of marker placement on the assembled genome, and annotation of that genome. Both structural annotation, providing locations of genes and other features, and functional annotation, indicating what those features do, are essential for biological insight (Stein, 2001).

Fully assembled genomic sequence, annotated with gene location and structure as well as noncoding RNA, regulatory, and repetitive regions is the most straightforward mechanism of relating markers to genes and features. Organisms lacking complete genome assembly and annotation may rely on comparative alignments to well-annotated species to relate markers to genomic features (Dalrymple et al., 2007). As of this writing, the *Bos taurus* (Bovine Genome Sequencing and Analysis Consortium, 2009; Zimin et al., 2009) and horse (Wade et al., 2009) genome assemblies and annotation may be the most mature of any livestock species. Those for other agriculturally important mammals are emerging. The pig genome assembly and annotation is publicly available (http://www.ncbi.nlm.nih.gov/assembly/304498); publications describing the draft sequence, genetic variation and haplotype structure, and analysis of the genome are anticipated (Archibald et al., 2010). A sheep assembly (http://www.ncbi.nlm.nih.gov/nuccore/ACIV000000000.1) guided by the bovine genome has also been released as efforts toward a de novo assembly continue (International Sheep Genomics Consortium et al., 2010; Dalrymple, 2011). *Bos indicus* cattle (http://www.ncbi.nlm.nih.gov/nuccore/AGFL00000000.1; Canavez et al., 2012) and water buffalo (Tantia et al., 2011) genomes have also been assembled with guidance from the bovine assembly. A de novo domestic yak assembly and annotation was recently reported (http://www.ncbi.nlm.nih.gov/nuccore/AGSK00000000.1; Qiu et el., 2012), as were initial efforts for the goat (Zhang et al., 2011).

Functional annotation of these livestock assemblies may borrow heavily from synteny with curated human and model organism annotations to infer functions of key genes and interactions in specific networks and pathways (Bovine Genome Sequencing and Analysis Consortium, 2009; Seo and Lewin, 2009). Initially developed for yeast, *Drosophila*, and mouse (Ashburner et al., 2000), gene ontology (**GO)** provides a controlled vocabulary to describe products of eukaryotic genes in terms of molecular function, biological processes, and cellular components. Gene ontology classification has now been applied to annotation of numerous species; the GO browser agriGO (Du et al., 2010) currently represents 45 agricultural species including grain and oilseed crops, fungi, and insect pests as well as livestock species.

Descriptions of metabolic and signaling pathways (Kanehisa et al., 2008;Croft et al., 2011; Caspi et al., 2012), gene regulatory networks (Lee et al., 2002; Shalgi et al., 2007; Hecker et al., 2009), and protein–protein interactions (Xenarios et al., 2002; Rual et al., 2005) convey knowledge about interactions among genes. Evidence of a core set of metabolic reactions (Ravasz et al., 2002; Kim et al., 2006), protein interactions (Wong et al., 2008), and pathways involving carbohydrate, AA, and nucleotide metabolism (Peregrín-Alvarez et al., 2009) conserved across life forms as well as conserved transcription factors (Vaquerizas et al., 2009; Ravasi et al., 2010), microRNA (Gaidatzis et al., 2007), and regulatory network kernels affecting major body part development (Davidson and Erwin, 2006) imply that much of the pathway and network information derived from human and model species experimentation is applicable to livestock. However, incomplete understanding of gene function and interactions (Elbers et al., 2009) and variation in genes regulated by specific transcription factors and microRNA (Kunarso et al., 2010; Berezikov, 2011) and in metabolic and signaling pathways (Huangfu and Anderson, 2006; Seo and Lewin, 2009) indicates a need for continued within- and across-species efforts to refine functional annotation of livestock genomes.

Evolution of species-specific gene interactions indicates the possibility that variation might exist between breeds and selected subpopulations, which would complicate genomic prediction across divergent populations. Nevertheless, generally ubiquitous functions across species indicate that evidence from human and model species can illuminate livestock QTL. A classic example is hypermuscularity of myostatin knock-out mice providing the impetus for discovery of mutations in the myostatin gene that cause “double-muscling” in cattle

(Grobet et al., 1997) as well as in Whippet dogs (Mosher et al., 2007). For complex polygenic traits without an obvious major gene that transfers directly across species, it still seems plausible that similar traits will be controlled by similar sets of genes, both across species and across populations within species. Therefore, the genes sets defined by functional annotation can supply evidence to focus genomic prediction on the genes and regulatory elements likely to affect phenotype.

Informing Genomic Predictions with Functional Evidence. The simplest approach to evaluating contributions of a particular gene set may be to limit genomic analysis to genotyped markers in or near genes in that set, ignoring the remainder of the genome represented by other markers on dense whole-genome panels. A genomic REML (**GREML**) estimate of phenotypic variation attributable to a given set of markers can be obtained with genomic relationships among individuals (Van Raden, 2008) computed from genotypes of markers in that set. Bayesian genomic selection (Meuwissen et al., 2001; Habier et al., 2011) can similarly be restricted to a subset of whole-genome SNP. The GREML approach can be extended to include a polygenic component, using pedigree relationships to account for the remainder of the genome (Snelling et al., 2011). Partitioning into genomic and polygenic components may have some advantage for prediction, as breeding values predicted as the sum of polygenic and genomic BLUP (**GBLUP**) solutions may be more accurate than either whole-genome GBLUP or pedigree BLUP EBV. The simpler approach without a polygenic component appears adequate for estimating genomic heritabilities and effects of markers in a given set. Genomic heritabilities estimated with and without a polygenic component are similar and agree with those estimated with BayesC for the same set of selected SNP. Marker effects solved from GBLUP solutions with and without a polygenic component also agree and rank SNP identically to BayesC marker effects.

Depending on the amount of variation left unexplained by a particular set of SNP or genes, extending genomic analyses to include additional genes sets and markers may add accuracy to the evaluation. Although the immediate problem is to identify gene sets suitable for predicting a meaningful amount of variation across populations within a livestock species, such extensions could accommodate population-specific variants that are not generally informative across the species. Variations on GREML to partition variance by gene sets and Bayesian approaches allowing gene set-based priors might be considered.

Hundreds of tools are available to identify genes sets to guide functional genomic evaluations (Bader et al., 2006). The sheer number of potentially useful tools precludes any attempt at individual descriptions, so only general properties and considerations are mentioned here. Beyond basic GO and pathway annotation, useful to select candidate gene sets defined by a particular GO term or pathway thought to affect a trait, are analytical programs that can be applied to naive GWAS results to determine gene sets associated with a trait. These tools have roots in analysis of gene expression from microarrays probing a set of known genes, but they can be applied to GWAS by assigning SNP associations to genes. Commonly implemented statistical tests include overrepresentation and enrichment, where overrepresentation compares a list of expressed (or associated) genes to a background list of all genes represented on the expression array (or assigned to SNP on a whole-genome assay) to identify annotation categories containing more expressed (associated) genes than expected by chance. Several variations on enrichment analysis have been developed (Bauer et al., 2010), but the basic idea is to determine annotation categories scoring greater than expected by chance from expression (association) scores assigned to all genes.

These gene set analysis tools vary by statistical algorithms implemented, species supported, and annotation categories considered. Some only support richly annotated human and model organisms. Others include many genomes annotated by the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/genome) and Ensembl (http://uswest.ensembl.org/index.html). Several are restricted to GO annotation whereas some include pathways defined by the Kyoto Encyclopedia of Genes and Genomes (**KEGG**; http://www.genome.jp/kegg/pathway.html; Kanehisa et al., 2012), additional expert- and community-curated pathways (e.g., http://www.wikipathways.org; Pico et al., 2008), protein family and interaction databases, literature, and other sources to classify genes by function. Many periodically integrate publicly accessible databases to ensure results reflect current knowledge. A few allow user-supplied annotation or provide mechanisms to regenerate gene set knowledge bases from current sources. Most of these tools are available online and may also provide source code and executables for in-house implementations. Other offline programs include Cytoscape (Shannon et al., 2003) with its various plugins for analysis and visualization of gene networks and R (R Core Team, 2012) and Bioconductor (Gentleman et al., 2004) annotation, analysis, and visualization packages.

A mechanism to assign SNP to genes is needed for gene set analysis postprocessing of GWAS results. Assignment based on distance between SNP and annotated genes positions is simple, but there is no standard for SNP-gene separation. Using the dense BovineSNP50, Fortes et al. (2010) considered genes within 2.5 kbp of a SNP whereas Rolf et al. (2011) assigned genes to SNP

within 500 kbp. Alternatives may be to assign genes to the closest SNP, with a limit on the maximum separation between SNP and genes, and LD-based assignment to genes overlapped by haplotype blocks. Additionally, translation from annotated livestock to human or model species genes will be needed to apply tools that do not support livestock genomes.

Gene networks developed from GWAS of multiple traits and other experimental evidence can also define sets of interrelated genes to focus genomic evaluations. A network with 3,159 genes related to heifer puberty was constructed from an association weight matrix (**AWM**) describing gene–phenotype associations for age at first corpus luteum and 21 other measures related to growth, body composition, and fertility. This analysis revealed puberty-related genes that would have been missed by single-trait GWAS and predicted gene–gene interactions consistent with experimentally validated transcription factor–target relationships (Fortes et al., 2010). Overrepresentation analysis of the AWM genes also revealed biological processes relevant to puberty that were not implicated by single-trait GWAS and gene set analysis of age at first corpus luteum. In a study of Brangus heifers, Fortes et al. (2012) extended the AWM approach to include evidence of gene expression, filtering the initial 10-trait AWM by genes expressed in the hypothalamus transcriptome of pre- and postpubertal heifers to obtain a 978 gene network. Imputed genotypes of BovineHD SNP in the Brangus AWM genes accounted for at least one-half of the heritable variation in age at puberty, antral follicle count, and pregnancy rate of crossbred *Bos taurus* heifers (Snelling et al., 2012). Genomic predictions of heifer pregnancy rate, using SNP selected from the Brangus AWM and trained by the crossbred *Bos taurus* heifers, predicted pools of pregnant *Bos indicus* heifers to have greater genomic breeding values (**GEBV**) for pregnancy rate than pools of their nonpregnant contemporaries. Naive predictions, based solely on the crossbred *Bos taurus* heifer phenotypes without considering functional information, predicted the nonpregnant pools to have greater GEBV than the pregnant pools. These results are evidence for the value of using functional priors, such as an associated gene set, on building genomic predictions.

Beef Tenderness Example

To illustrate how functional evidence can inform genomic evaluations, shear force records from the U.S. Meat Animal Research Center Germplasm Evaluation (**GPE**) Project were examined. These measurements of beef tenderness exemplify a trait that is economically important (Platter et al., 2005; Weaber and Lusk, 2010) but too invasive for routine measurement by commercial beef packing companies (Shackelford et al., 2005). Polymorphisms in 2 genes, μ-calpain (**CAPN1**) and calpastatin (**CAST**), have been shown to account for some tenderness variation in both beef (Casas et al., 2006; Allais et al., 2011) and pork (Ciobanu et al., 2004; Nonneman et al., 2011). An objective of this example is to identify additional genes that may affect beef tenderness.

Data and Analysis. Phenotypes, genotypes, and pedigree records were from GPE Cycle VII, representing 2-, 3-, and 4-breed crosses of 7 predominant *Bos taurus* breeds in the United States (Wheeler et al., 2005), Cycle VIII, characterizing 4 tropically adapted breeds along with Angus (**AN**) and Hereford (**HH**; Wheeler et al., 2010), and the current continuous GPE population examining the 16 most popular U.S. beef breeds. Cycle VII phenotypes served as training data for this example, with tenderness predictions evaluated in the target population consisting of Cycle VIII and continuous GPE (Fig. 2). The training phenotypes represented 1,716 genotyped Cycle VII steers, 552 F_1 steers produced by mating sires of the 7 breeds to AN, HH, and MARCIII (1/4 AN, 1/4 HH, 1/4 Red Poll, 1/4 Pinzgauer) composite females, and 1,164 so-called F_1^2 steers ($F_1^2 = F_1 \times F_1$) produced by mating F_1 bulls to F_1 females (Snelling et al., 2010). Warner-Bratzler shear force (Wheeler et al., 1998) of LM steaks from F_1 steers was measured 14 d postmortem. The steaks from F_1^2 steers were aged for 3 and 14 d, and Warner-Bratzler and slice shear force (Shackelford et al., 1999) were measured to obtain 4 observations for each steer: 3-d Warner-Bratzler shear force (**WB3**), 14-d Warner-Bratzler shear force (**WB14**), 3-d slice shear force (**SS3**), and 14-d slice shear force (**SS14**). The target phenotypes were WB14 measurements of 887 genotyped Cycle VIII and continuous GPE steers.

Pedigree records of 18,182 GPE animals were coupled with high density (**HD**) genotypes of 950 animals (482 sires, 143 dams, and 325 nonparents) to impute HD genotypes of 8,694 animals having 50K genotypes using findhap.f90 version 2 (Van Raden et al., 2011). Imputation accuracy was evaluated by executing the imputation using 50K genotypes of the nonparents having HD and then comparing their imputed to observed HD genotypes. Subsequent genomic evaluations of tenderness traits used the imputed HD genotypes of measured steers.

Data from the Cycle VII steers were analyzed with 4-trait GREML and GBLUP using pedigree or genomic relationships (Van Raden, 2008) described by selected subsets of HD SNP. Because of high genetic correlations among the 4 traits (ranging from 0.85 to 0.97), solutions for the first principal component from 4-trait principal component GBLUP analysis were taken as GEBV for a

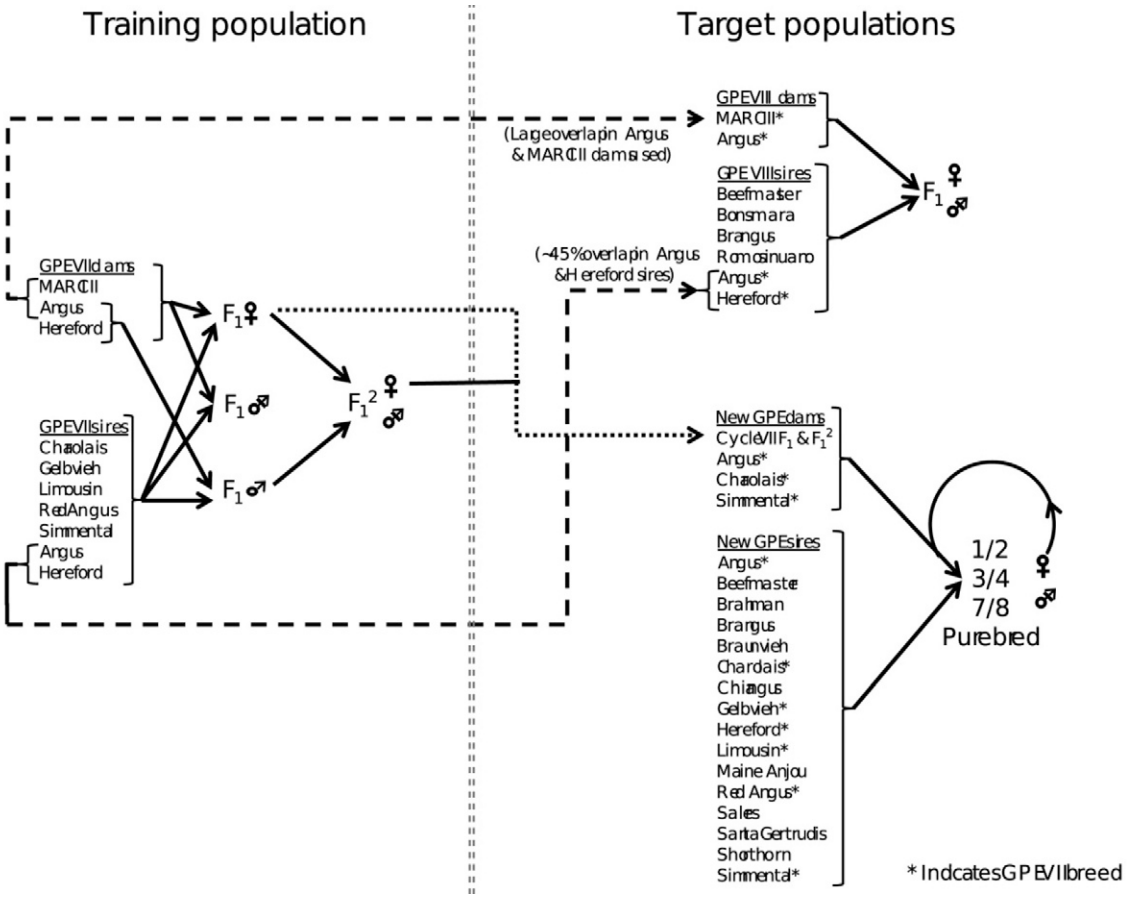

Figure 2. Schematic of the U.S. Meat Animal Research Center (MARC) Germplasm Evaluation (GPE) project training and target populations. Genomic predictions trained by GPE Cycle VII genotypes and phenotyes were applied to Cycle VIII and continuous (New) GPE genotypes to evaluate applicability of the predictions to additional breeds and crosses.

composite tenderness trait. Individual SNP effects were solved from the Cycle VII GEBV and applied to imputed HD genotypes to predict tenderness GEBV of other GPE animals. Genetic correlations between these GEBV and WB14 measured on Cycle VIII and continuous GPE steers were estimated in 2-trait REML analysis to assess accuracy of extending the GEBV to a somewhat related population (Weber et al., 2012). Genetic correlations between GEBV and WB14 were estimated using all 887 Cycle VIII and continuous GPE observations as well as with a subset of 598 steers having less than 25% Brahman or Brahman-influenced composite breed composition. Expectation maximization and average information REML algorithms implemented in WOMBAT (Meyer, 2007) were used to obtain GREML heritability and correlation estimates using the GIN option for genomic relationship matrices. Principal component analyses used the PC option of WOMBAT. Routine steps to complete analysis of each SNP set were automated by Perl and Bash scripts using matrix operations from the Animal Breeders' Toolkit (Golden et al., 1992).

Gene Sets and Network. After completing 4-trait GREML and GBLUP of WB3, WB14, SS3, and SS14 using autosomal HD SNP having minor allele frequencies >0.05 in the GPE Cycle VII population, normalized z-scores were computed for the effects of individual SNP on each trait. Using SNP positions (Illumina list) and gene boundaries annotated on the UMD3.1 assembly (Zimin et al., 2009), genes within 5 kbp of those SNP having z-scores >3 for at least 2 traits were identified. An AWM was constructed with a row for each of those genes, a column for each of the 4 traits, and elements containing the maximum z-score for each trait of SNP assigned to that gene. Functional annotation clusters (Huang et al., 2009a,b) that were overrepresented by AWM genes were identified using the Database for Annotation, Visualization and Integrated Discovery (**DAVID**) functional annotation clustering tool (http://david.abcc.ncifcrf.gov/summary.jsp), with a background consisting of all genes assigned to HD SNP. All genes assigned to HD SNP were also assigned a z-score from the largest SNP effect solved from the HD principal component GBLUP. These scores were used to evaluate enrichment of GO terms and pathways with the Protein Analysis Through Evolutionary Relationships (**PANTHER**) system (Mi et al., 2010; http://www.pantherdb.org). Source databases contributing to DAVID were queried to extract all genes related to terms in the overrepresented clusters to expand

the DAVID gene set with functionally related genes. The PANTHER gene sets were similarly expanded to include other genes annotated with the enriched GO terms and the complete pathways.

Two candidate gene sets were also identified. One included genes annotated with the GO term proteolysis (GO:0006508), the lowest level GO term containing both CAST and CAPN1. The other candidate set represented annotated noncoding RNA (**ncRNA**) genes, which may affect regulation of protein coding genes (Eddy, 2001; Mattick and Makunin, 2006; Qu and Adelson, 2012). No pathways indicating a relationship between CAST and CAPN1 were found in KEGG or the other pathway databases integrated by DAVID. Otherwise, a candidate pathway containing both genes would have been evaluated.

Genomic relationship matrices for the Cycle VII steers were constructed using SNP assigned to genes in the AWM and each of the overrepresented, enriched, and candidate gene sets. Reduced SNP sets, using the top ($z > 2$) SNP from each gene set, were also identified and the GREML/GBLUP and prediction processes repeated for each of the top sets as well as for a set combining SNP from the most promising top sets.

Genomic Evaluation Results. Genomic correlations approaching unity (Table 2) among the 4 shear force measurements, evaluated using 630,579 autosomal SNP, indicate that each is a measure of essentially the same tenderness trait. Therefore, considering all 4 measurements may reduce spurious SNP associations with any 1 measurement, enabling functional analyses using the 4-trait AWM or composite principal component tenderness trait to focus on loci more likely to have real effects.

Polymorphisms located in CAST and CAPN1 had the strongest effects on each of the shear force measurements and the principal component tenderness trait, with CAPN1 SNP having somewhat larger effects than those in CAST. The 4-trait AWM (Supplemental Table 1), containing genes assigned to SNP associated with at least 2 measures represented 545 genes located on all 29 autosomes, 72 genes on BTA 29 (including CAPN1), and 42 each on BTA 7 (including CAST) and BTA 19. Chromosomes with the fewest tenderness AWM genes were BTA 17 (5 genes) and BTA 20, 21, and 28 (7 genes each).

Overrepresentation analysis with DAVID expanded the 545 AWM genes to 2,426 genes functionally related by common annotations among 100 of the AWM genes (Supplemental Table 2). Likewise, PANTHER indicated functional relationships among 1,704 distinct genes, including 71 AWM genes, related to GO terms and pathways enriched among all genes scored by the principal component tenderness trait (Supplemental Table 3). Expansion from the genes associated with Cycle VII shear force measurements to gene sets implicated by functional categorization may include genes having an effect on tenderness although those effects were not detected in Cycle VII and may also eliminate some genes that do not have a functional effect although they were associated with tenderness of Cycle VII steers. The same applies to the candidate proteolysis gene set; genes besides CAST and CAPN1 involved in proteolysis could influence meat tenderness and the ncRNA genes that might regulate expression of genes affecting tenderness.

Table 2. Estimated genomic heritabilities and correlations (and SE) among 4 LM tenderness traits measured on crossbred steers from GPE Cycle VII[1]

Trait[2]	WB3	WB14	SS3	SS14
WB3	0.44 (0.06)	0.91 (0.05)	0.89 (0.04)	0.78 (0.07)
WB14	0.71 (0.02)	0.32 (0.06)	0.86 (0.06)	0.92 (0.45)
SS3	0.81 (0.01)	0.71 (0.02)	0.29 (0.06)	0.90 (0.06)
SS14	0.68 (0.02)	0.78 (0.01)	0.75 (0.01)	0.24 (0.06)

[1]GPE = Germplasm Evaluation. Parameters estimated from Cycle VII of the GPE project, with genomic relationship matrix using genotypes of 630,579 SNP. Heritabilities are on diagonal, and genomic correlations are above and phenotypic correlations are below diagonal.

[2]WB3 = Warner-Bratzler shear force measured 3 d postmortem; WB14 = Warner-Bratzler shear force measured 14 d postmortem; SS3 = slice shear force measured 3 d postmortem; SS14 = slice shear force measured 14 d postmortem.

Sets containing between 7,100 and 40,000 SNP assigned to genes in the overrepresented, enriched, or candidate gene sets (Table 3) were estimated to explain at least 40% of the variation described by all autosomal HD SNP (Table 4a). Genomic heritability estimates using SNP within or near AWM genes were about 150 to 160% of the corresponding HD estimates. For each trait, heritability estimates from the large set 2,624 genes in DAVID annotation clusters, represented by nearly 40,000 SNP, were 75 to over 90% of the HD estimates. The somewhat larger number of ncRNA genes, represented by only 7,107 SNP, explained about 60 to 70% of the heritable variation. Heritability estimates using SNP representing gene sets enriched for the molecular function, cellular component, and biological process gene ontologies as well as those for PANTHER pathways were 40 to 50% of HD estimates, somewhat less than the 60 to 65% estimated for the candidate proteolysis GO term.

Genomic correlations between traits, estimated with SNP representing each of the gene sets, were generally similar to those estimated with the full complement of HD SNP (Table 5a). The correlations tended to be greatest for shear force measured with the same technique (Warner-Bratzler or shear) or after the same aging (3 or 14 d). Correlations were weakest between WB3 and SS14, notably so for the molecular function GO and PANTHER pathways, indicating that some loci associated with these sets may have slightly different effects

Table 3. Selected gene and SNP sets used for genomic evaluation of LM tenderness

Set[1]	Genes	SNP	Criteria[2]
HD	21,768	630,579	SNP from whole-genome HD assay, located on autosomes with minor allele frequency >0.05
AWM	545	19,119	AWM defined by genes within 5 kbp of SNP with HD z-score >3 for at least 2 shear force measures
Proteolysis	888	12,669	Genes associated with lowest level GO term containing both calpastatin and mu-calpain1 (GO:0006508)
RNA	2,782	7,107	Noncoding RNA within 5 kbp of HD SNP; may have regulatory function
DAVID	2,624	39,764	DAVID annotation clusters overrepresented in AWM
MF	412	12,251	Enriched MF GO terms from PANTHER; all genes scored by maximum z-score of HD SNP within 5 kbp
CC	686	15,836	Enriched CC GO terms from PANTHER; all genes scored by maximum z-score of HD SNP within 5 kbp
BP	718	18,864	Enriched biological process GO terms from PANTHER; all genes scored by maximum z-score of HD SNP within 5 kbp
Pathways	295	7,591	Enriched PANTHER pathways; all genes scored by maximum z-score of HD SNP within 5 kbp

[1]HD = high density; AWM = association weight matrix; DAVID = Database for Annotation, Visualization and Integrated Discovery (http://david.abcc.ncifcrf.gov/summary.jsp); MF = molecular function; CC = cellular component; BP = biological process.

[2]z-score = ; GO = gene ontology; PANTHER = Protein Analysis Through Evolutionary Relationships (http://www.pantherdb.org).

Table 4. Genomic heritabilities of LM shear force measurements from GPE Cycle VII estimated with whole-genome SNP and sets selected by functional annotation and association with phenotype[1]

Set[2]	SNP	WB3	WB14	SS3	SS14
Full sets					
HD	630,579	0.44	0.32	0.29	0.24
AWM	19,119	0.65	0.50	0.46	0.36
Proteolysis	12,669	0.22	0.16	0.16	0.16
RNA	7,107	0.32	0.22	0.20	0.13
DAVID	39,764	0.38	0.24	0.27	0.18
MF	12,251	0.21	0.12	0.16	0.12
CC	15,836	0.17	0.14	0.12	0.13
BP	18,864	0.19	0.14	0.14	0.13
Pathways	7,591	0.18	0.12	0.12	0.11
Top subsets[3]					
HD	30,648	0.91	0.60	0.67	0.42
AWM	1,020	0.64	0.47	0.46	0.36
Proteolysis	613	0.36	0.29	0.33	0.27
RNA	343	0.39	0.28	0.32	0.19
DAVID	1,908	0.62	0.38	0.53	0.30
MF	619	0.38	0.17	0.29	0.15
CC	812	0.39	0.32	0.36	0.31
BP	924	0.36	0.28	0.39	0.33
Pathways	358	0.30	0.18	0.24	0.18
Combined[4]	2,011	0.55	0.40	0.46	0.36

[1]GPE = Germplasm Evaluation. Steer data from Cycle VII of the GPE project. WB3 = Warner-Bratzler shear force measured 3 d postmortem; WB14 = Warner-Bratzler shear force measured 14 d postmortem; SS3 = slice shear force measured 3 d postmortem; SS14 = slice shear force measured 14 d postmortem.

[2]HD = high density; AWM = association weight matrix; DAVID = Database for Annotation, Visualization and Integrated Discovery (http://david.abcc.ncifcrf.gov/summary.jsp); MF = molecular function; CC = cellular component; BP = biological process.

[3]Top SNP selected by z > 2 from principal component of 4-trait genomic BLUP with corresponding full set.

[4]Includes top Proteolysis, RNA, CC, and Pathway subsets.

for the specific measurements. Between-trait correlations were strongest for AWM SNP, reflecting selection for inclusion in the AWM by association with at least 2 of the measurements.

Reducing the functional gene sets to the SNP having the largest effects on the principal component trait increased the Cycle VII heritability estimates for each set (Table 4b). Estimates for the AWM, however, were largely unchanged; the top AWM SNP explained the same variation as the full AWM. Selection from the full set of HD SNP resulted in heritability estimates almost double the whole-genome estimates. All estimates of the genomic correlations between measurements were greater for the reduced SNP sets than the corresponding estimates using all SNP for each gene set (Table 5b). The greater heritability estimates resulting from eliminating SNP having small effects on the composite tenderness trait may partially reflect desirable elimination of noise due to SNP that are not actually associated with the unknown QTL and partly indicate an undesirable increased emphasis on spurious effects, especially for the grossly inflated top HD estimates. The greater estimates of genomic correlations between measurements are a result of selecting SNP based on their effect on the principal component capturing all 4 measurements.

Although the AWM, overrepresented, or enriched gene sets and naively selected SNP sets appear to explain substantial variation within Cycle VII, extension of the Cycle VII predictions to the somewhat related Cycle VIII and continuous GPE population may provide a better indication of how the predictions may apply to a broader industry population. Using the full set of HD SNP, estimates of the genetic correlation between Cycle VII trained GEBV and shear force measured Cycle VIII and continuous GPE steers were 0.16 using all steers and 0.28 using steers with little *Bos indicus* influence, explaining 2.5 to 8% of the genetic variation in each set (Table 6a). Standard errors of all estimates were too large to be conclusive; however, correlations for the proteolysis GO term were similar to the HD estimates

Table 5. Genomic correlations between LM shear force measurements from GPE Cycle VII estimated using whole-genome SNP and sets selected by functional annotation and association with phenotype[1]

	Trait pairs			
Set[2]	WB3 WB14	WB3 SS3	WB3 SS14	WB14 SS3
Full sets				
HD	0.91	0.89	0.78	0.85
AWM	0.96	0.95	0.91	0.91
Proteolysis	0.85	0.84	0.72	0.91
RNA	0.94	0.89	0.84	0.88
DAVID	0.92	0.89	0.83	0.83
MF	0.83	0.94	0.69	0.81
CC	0.94	0.90	0.84	0.89
BP	0.90	0.83	0.76	0.80
Pathways	0.86	0.84	0.70	0.83
Top subsets[3]				
HD	0.98	0.96	0.92	0.93
AWM	0.98	0.95	0.91	0.91
Proteolysis	0.98	0.99	0.96	0.99
RNA	1.00	0.97	0.99	0.97
DAVID	0.99	0.96	0.93	0.97
MF	0.96	0.98	0.93	0.91
CC	0.98	0.97	0.95	0.95
BP	0.97	0.97	0.94	0.96
Pathways	0.99	1.00	0.92	1.00
Combined[4]	0.98	0.98	0.94	0.97

[1]GPE = Germplasm Evaluation. Steer data from Cycle VII of the GPE project. WB3 = Warner-Bratzler shear force measured 3 d postmortem; WB14 = Warner-Bratzler shear force measured 14 d postmortem; SS3 = slice shear force measured 3 d postmortem; SS14 = slice shear force measured 14 d postmortem.

[2]HD = high density; AWM = association weight matrix; DAVID = Database for Annotation, Visualization and Integrated Discovery (http://david.abcc.ncifcrf.gov/summary.jsp); MF = molecular function; CC = cellular component; BP = biological process.

[3]Top SNP selected by z > 2 from principal component of 4-trait genomic BLUP with corresponding full set.

[4]Includes top Proteolysis, RNA, CC, and Pathway subsets.

Table 6. Genetic correlations between genomic EBV trained by GPE Cycle VII shear force measurements, using whole-genome and sets selected by functional annotation and association with phenotype, and 14-d slice shear force measured on steers from Cycle VIII and continuous GPE[1]

Set[2]	All steers	*Bos taurus* steers[3]
Full sets		
HD	0.16 (0.10)	0.28 (0.18)
AWM	0.31 (0.10)	0.43 (0.17)
Proteolysis	0.17 (0.10)	0.30 (0.19)
RNA	0.04 (0.10)	0.19 (0.17)
DAVID	0.02 (0.10)	0.10 (0.16)
MF	0.16 (0.10)	0.19 (0.17)
CC	0.17 (0.10)	0.12 (0.16)
BP	0.10 (0.10)	0.08 (0.16)
Pathways	0.23 (0.11)	0.44 (0.23)
Top subsets[4]		
HD	0.09 (0.04)	–0.01 (0.17)
AWM	0.08 (0.10)	0.15 (0.17)
Proteolysis	0.26 (0.10)	0.43 (0.18)
RNA	0.09 (0.10)	0.27 (0.16)
DAVID	0.06 (0.10)	0.10 (0.16)
MF	0.04 (0.10)	–0.05 (0.15)
CC	0.20 (0.10)	0.24 (0.17)
BP	0.14 (0.10)	0.17 (0.17)
Pathways	0.27 (0.10)	0.44 (0.20)
Combined[5]	0.35 (0.10)	0.46 (0.16)

[1]GPE = Germplasm Evaluation. Genomic EBV predicted with individual SNP effects solved from principal component of 4-trait genomic BLUP.

[2]HD = high density; AWM = association weight matrix; DAVID = Database for Annotation, Visualization and Integrated Discovery (http://david.abcc.ncifcrf.gov/summary.jsp); MF = molecular function; CC = cellular component; BP = biological process.

[3]<25% Brahman, Brangus, Beefmaster, or Santa Gertudis.

[4]Top SNP selected by z > 2 from principal component of 4-trait genomic BLUP with corresponding full set.

[5]Includes top Proteolysis, RNA, CC, and Pathway subsets.

for both sets of steers, and estimates for the AWM and PANTHER pathways were greater than HD for both sets. Reducing the SNP sets had varied effects on correlation estimates (Table 6b). For both HD and AWM, sets defined with no functional information, GEBV–SS14 correlations using the top SNP dropped substantially, to 0 for the top HD SNP predicting *Bos taurus* SS14. Correlations for top SNP in functionally derived proteolysis, cellular component, biological process, and pathway genes increased in both steer sets, as did correlations for the top SNP near ncRNA. In both sets of steers, the strongest GEBV–SS14 correlations were with a SNP set combining the top proteolysis, pathway, cellular component, and ncRNA SNP.

Future Efforts

Use of pathways and tissue-specific expression to identify polymorphisms predictive of human conditions (Lesnick et al., 2007; Baranzini et al., 2009), prior results examining AWM and functional gene sets associated with beef heifer puberty and pregnancy rate, and the beef tenderness example demonstrate the potential for applying functional information to enable more robust genomic predictions across livestock populations. Accuracy of these predictions, however, will always be limited by marker–QTL LD as long as the predictions are based on estimated effects of markers genotyped with the currently standard dense (50K) and high-density arrays. The SNP on these arrays, primarily selected using spacing along the genome and allele frequency (Matukumalli et al., 2009; Ramos et al., 2009), are ef-

fective for capturing LD with unknown causal variants; predictions based on effects of likely causal variants could obviate reliance on LD.

Next-generation sequencing technologies, enabling rapid, low-cost genome and transcriptome sequencing, are revealing millions of individual deviations from reference livestock genomes (Cánovas et al., 2010; Stothard et al., 2011; Larkin et al., 2012). Categorizing these variants according to expected effect on annotated protein coding genes may reveal the variants most influential to gene function (McLaren et al., 2010; Cingolani et al., 2012). Genomic evaluation with these functional variants could use genotypes obtained from sequence and assays designed specifically to genotype the functional variants as well as functional variant genotypes imputed from existing SNP array genotypes, provided that a suitable reference of animals genotyped for both standard SNP and functional variants is available. Variants expressed in transcribed RNA revealed by RNA-sequence may be particularly useful, as they may be more likely to affect gene function and regulation than unexpressed genomic sequence variants, and RNA-seq may address limitations to functional genomic selection guided by current annotation. Specifically, RNA-seq may extend current annotation where expressed protein coding regions are not annotated as exons (Mortazavi et al., 2008), and noncoding RNA-seq variants may have a regulatory role, indicative of functional variation in annotated and unannotated ncRNA (Qu and Adelson, 2012). Coupled with GWAS phenotypic associations, differential expression assessed by microarrays or RNA-seq may also reveal interactions that are not described by existing functional genomics databases (García-Gámez et al., 2011).

Using the beef tenderness for further illustration, of over 10 million variants revealed by low-coverage genomic sequence from 96 GPE sires, 2,432 are expected to have a high impact on gene function (frame shifts, deleted exons, altered splice sites, and start/stop codons; Cingolani et al., 2012). Another 27,640 may have moderate functional effects (nonsynonymous SNP, other codon changes, deletions from 5′ and 3′ untranslated regions, etc.). Greatest priority for further genotyping are the 54 high-impact variants within AWM genes as well as the 67 variants within gene sets defined by the proteolysis GO term, enriched pathways, and enriched cellular component GO terms. Additional variation might be explained by those expected to have moderate impact on gene function, including 717 variants in AWM genes and 1,131 in proteolysis and enriched gene set genes. These lists could be modified and reweighted with RNA-seq from tissue of animals yielding tough and tender carcasses, providing additional evidence of expressed variants, including variants not revealed or misclassified by current bovine annotation. Genotyping influential animals in the GPE pedigree for functional variants would support imputing functional genotypes of remaining animals with existing 50K and HD genotypes and enable further evaluation of the functional variants in GPE. If the functional variants are equally as descriptive of tenderness variation within GPE as the HD SNP selected in this example, extension beyond GPE might be accomplished through development of panels to genotype functional variants. Custom content added to standard arrays and methods for targeted sequencing of specific loci (Thallman and Koshinsky, 2012) may enable cost-effective functional variant genotyping.

The example analysis demonstrated a somewhat arbitrary model selection process to identify gene and SNP sets (where each set defines a model) that were predictive of tenderness. A potentially useful expansion would be incorporation of biological knowledge in a Bayesian context where complementary functional information is modeled through prior distributions. This may answer some of the challenges posed by model selection where it is not always obvious what information should be included and how it should be weighted. In most cases this might represent an expansion of penalized likelihood or Bayesian methods already implemented. Attempts have been already made to generalize the most popular Bayesian methods (Bayes A/B) to allow differential shrinkage for different groups of markers (Gianola et al., 2010; Maltecca et al., 2012). In these implementations, marker groupings assumed relatively uninformative priors resulting in a mixture of distributions largely driven by the data. In contrast, approaches including biologically informative priors have been put forward, mostly in the context of incorporating pathways and networks into analysis of microarray expression experiments. An empirical Bayes approach to incorporate contemporarily different paths and account for connections among the paths (Hill et al., 2012), Markov random field priors to map known connections among genes (Stingo and Vannucci, 2011), and constraining discriminate analysis with gene regulation network priors (Guillemot et al., 2011) have been demonstrated and could be adapted to predicting phenotypes from genotypes. Yet undefined methods might simultaneously consider gene function and canonical pathways obtained from annotation and public databases, gene expression and interaction evidence derived from pertinent RNA-seq experiments, and variant-level effects on gene function and regulation.

SUMMARY AND CONCLUSIONS

Although results of the example analysis are specifically applicable to the problem of beef tenderness, the general process of examining functional gene sets and pathways can be adapted to any species and popu-

lation with suitable resources. Basic requirements are genotypes for DNA markers that are sufficiently dense to capture LD with unknown QTL scattered throughout the genome, a mechanism to assign those markers to functionally annotated genes and genomic features, and enough phenotypes to associate marker genotypes with phenotypic variation. Genotypes and phenotypes alone can be enough for naive genomic selection to improve accuracy of selection for well-recorded traits within a population, but phenotypes and genotypes alone are inadequate if goals include increased understanding of biological mechanisms underlying a trait and enabling selection for seldom-recorded traits across populations. When markers associated with phenotype are also assigned to functionally annotated genes, the functional annotation can implicate specific gene functions as affecting the trait, providing both biological insight and information about functional attributes of genes that may be influential across populations. Interactions among genes, those indicated by known pathways as well as those established by co-expression and co-association with correlated traits, can assist the search for biologically relevant markers. Emerging sequencing and genotyping technologies may facilitate identification and genotyping of sequence variants likely to affect gene function and regulation. Continued developments may enable functional genomic selection to focus on loci most likely to affect performance, explaining a meaningful amount of variation across populations within any livestock species.

LITERATURE CITED

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752.

Allais, S., L. Journaux, H. Levéziel, N. Payet-Duprat, P. Raynaud, J. F. Hocquette, J. Lepetit, S. Rousset, C. Denoyelle, C. Bernard-Capel, and G. Renand. 2011. Effects of polymorphisms in the calpastatin and μ-calpain genes on meat tenderness in 3 French beef breeds. J. Anim. Sci. 89:1–11.

Archibald A. L., L. Bolund, C. Churcher, M. Fredholm, M. A. Groenen, B. Harlizius, K. T. Lee, D. Milan, J. Rogers, M. F. Rothschild, H. Uenishi, J. Wang, L. B. Schook, and Swine Genome Sequencing Consortium. 2010. Pig genome sequence – Analysis and publication strategy. BMC Genomics 11:438.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene ontology: Tool for the unification of biology. Nat. Genet. 25:25–29.

Bader, G. D., M. P. Cary, and C. Sander. 2006. Pathguide: A pathway resource list. Nucleic Acids Res. 34(Database issue):D504–D506.

Baranzini, S. E., N. W. Galwey, J. Wang, P. Khankhanian, R. Lindberg, D. Pelletier, W. Wu, B. M. J. Uitdehaag, L. Kappos, GeneMSA Consortium, C. H. Polman, P. M. Matthews, S. L. Hauser, R. A. Gibson, J. R. Oksenberg, and M. R. Barnes. 2009. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. Hum. Mol. Genet. 18:2078–2090.

Bauer, S., J. Gagneur, and P. N. Robinson. 2010. GOing Bayesian: Model-based gene set analysis of genome-scale data. Nucleic Acids Res. 38:3523–3532.

Berezikov, E. 2011. Evolution of microRNA diversity and regulation in animals. Nat. Rev. Genet. 12:846–860.

Boichard, D., H. Chung, R. Dassonneville, X. David, A.Eggen, S. Fritz, K. J. Gietzen, B. J. Hayes, C. T. Lawley, T. S. Sonstegard, C. P. Van Tassell, P. M. VanRaden, K. A. Viaud-Martinez, and G. R. Wiggans, for the Bovine LD Consortium. 2012. Design of a bovine low-density SNP array optimized for imputation. PLoS ONE 7:e34130.

Bovine Genome Sequencing and Analysis Consortium. 2009. The genome sequence of taurine cattle: A window to ruminant biology and evolution. Science 324:522–528.

Bovine HapMap Consortium. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324:528–532.

Browning, B. L., and S. R. Browning. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84:210–223.

Calus, M. P. L., and R. F. Veerkamp. 2011. Accuracy of multi-trait genomic selection using different methods. Genet. Sel. Evol. 43:26.

Canavez, F. C., D. D. Luche, P. Stothard, K. R. M. Leite, J. M. Sousa-Canavez, G. Plastow, J. Meidanis, M. A. Souza, P. Feijao, S. S. Moore, and L. H. Camara-Lopes. 2012. Genome sequence and assembly of *Bos indicus*. J. Hered. 103:342–348.

Cánovas, A., G. Rincon, A. Islas-Trejo, S. Wickramasinghe, and J. Medrano. 2010. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. Mamm. Genome 21:592–598.

Casas, E., S. N. White, T. L. Wheeler, S. D. Schackelford, M. Koohmaraie, D. G. Riley, C. C. Chase Jr., D. D. Johnson, and T. P. L. Smith. 2006. Effects of calpastatin and micro-calpain markers in beef cattle on tenderness traits. J. Anim. Sci. 84:520–525.

Caspi, R., T. Altman, K. Dreher, C. A. Fulcher, P. Subhraveti, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, A. Pujar, A. G. Shearer, M. Travers, D. Weerasinghe, P. Zhang, and P. D. Karp. 2012. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 40(Database issue):D742–D753.

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: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) 6:80–92.

Ciobanu, D. C., J. W. M. Bastiaansen, S. M. Lonergan, H. Thomsen, J. C. M. Dekkers, G. S. Plastow, and M. F. Rothschild. 2004. New alleles in calpastatin gene are associated with meat quality traits in pigs. J. Anim. Sci. 82:2829–2839.

Croft, D., G. O'Kelly, G. Wu, R. Haw, M. Gillespie, L. Matthews, M. Caudy, P. Garapati, G. Gopinath, B. Jassal, S. Jupe, I. Kalatskaya, S. Mahajan, B. May, N. Ndegwa, E. Schmidt, V. Shamovsky, C. Yung, E. Birney, H. Hermjakob, P. D'Eustachio, and L. Stein. 2011. Reactome: A database of reactions, pathways and biological processes. Nucleic Acids Res. 39(Database issue):D691–D697.

Dalrymple, B. P., on behalf of the International Sheep Genomics Consortium. 2011. Progress towards the assembly of the sheep reference genome: From scaffolds to chromosomes. Abstr. W137. In: Proc. Plant Anim. Genome XIX. San Diego, CA.

Dalrymple, B. P., E. F. Kirkness, M. Nefedov, S. McWilliam, A. Ratnakumar, W. Barris, S. Zhao, J. Shetty, J. F. Maddox, M. O'Grady, F. Nicholas, A. M. Crawford, T. Smith, P. J. de Jong, J. McEwan, V. H. Oddy, N. E. Cocket, and International Sheep Genomics Consortium. 2007. Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. Genome Biol. 8:R152.

Davidson, E. H., and D. H. Erwin. 2006. Gene regulatory networks and the evolution of animal body plans. Science 311:796–800.

DeDonato, M., S. O. Peters, S. E. Mitchell, T. Hussain, and I. G. Imumorin. 2012. Genotyping by sequencing (GBS): A novel, efficient and cost-effective genotyping method in cattle. J. Anim. Sci. 90(Suppl. 3):448 (Abstr.).

Du, Z., X. Zhou, Y. Ling, Z. Zhang, and Z. Su. 2010. agriGO: A GO analysis toolkit for the agricultural community. Nucleic Acids Res. 38(Web Server issue):W64–W70.

Eddy, S. R. 2001. Non-coding RNA genes and the modern RNA world. Nat. Rev. Genet. 2:919–929.

Elbers, C. C., K. R. van Eijk, L. Franke, F. Mulder, Y. T. van der Schouw, C. Wijmenga, and N. C. Onland-Moret. 2009. Using genome-wide pathway analysis to unravel the etiology of complex diseases. Genet. Epidemiol. 33:419–431.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379.

Fortes, M. R. S., A. Reverter, Y. Zhang, E. Collis, S. H. Nagaraj, N. N. Jonsson, K. C. Prayaga, W. Barris, and R. J. Hawken. 2010. Association weight matrix for the genetic dissection of puberty in beef cattle. Proc. Natl. Acad. Sci. USA 107:13642–13647.

Fortes, M. R. S., W. M. Snelling, A. Reverter, S. H. Nagaraj, S. A. Lehnert, R. J. Hawken, K. L. DeAtley, S. O. Peters, G. A. Silver, G. Rincon, J. F. Medrano, A. Islas-Trejo, and M. G. Thomas. 2012. Gene network analyses of first service conception in Brangus heifers: Use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. J. Anim. Sci. 90:2894–2906.

Gaidatzis, D., E. van Nimwegen, J. Hausser, and M. Zavolan. 2007. Inference of miRNA targets using evolutionary conservation and pathway analysis. BMC Bioinf. 8:69.

García-Gámez, E., A. Reverter, V. Whan, S. M. McWilliam, J. J. Arranz, J. Kijas, and International Sheep Genomics Consortium. 2011. Using regulatory and epistatic networks to extend the findings of a genome scan: Identifying the gene drivers of pigmentation in Merinosheep. PLoS ONE 6:e21158.

Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41:55.

Gautier, M., T. Faraut, K. Moazami-Goudarzi, V. Navratil, M. Foglio, C. Grohs, A. Boland, J. G. Garnier, D. Boichard, G. M. Lathrop, I. G. Gut, and A. Eggen. 2007. Genetic and haplotypic structure in 14 European and African cattle breeds. Genetics 177:1059–1070.

Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. Yang, and J. Zhang. 2004. Bioconductor: Open software development for computational biology and bioinformatics. Genome Biol. 5:R80.

Gianola, D., X. Wu, E. Manfredi, and H. Simianer. 2010. A nonparametric mixture model for genome-enabled prediction of genetic value for a quantitative trait. Genetica (The Hague) 138:959–977.

Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica (The Hague) 136:245–257.

Goddard, M., and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10:381–391.

Golden B. L., D. J. Garrick, and L. L. Benyshek 2009. Milestones in beef cattle genetic evaluation. J. Anim. Sci. 87:(E. Suppl.):E3–E10.

Golden, B. L., W. M. Snelling, and C. H. Mallinckrodt. 1992. Animal breeder's tool kit: User's guide and reference manual. Tech. Bull. LTB92-02. Colorado State Univ. Agric. Exp. Stn., Fort Collins, CO.

Grobet, L., L. J. Martin, D. Poncelet, D. Pirottin, B. Brouwers, J. Riquet, A. Schoeberlein, S. Dunner, F. Ménissier, J. Massabanda, R. Fries, R. Hanset, and M. Georges. 1997. A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. Nat. Genet. 17:71–74.

Guillemot, V., A. Tenenhaus, L. Le Brusquet, and V. Frouin. 2011. Graph constrained discriminant analysis: A new method for the integration of a graph into a classification process. PLoS ONE 6:e26146.

Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2011. Extension of the bayesian alphabet for genomic selection. BMC Bioinf. 12:186.

Hayes, B. J., P. J. Bowman, A. C. Chamberlain, K. Verbyla, and M. E. Goddard. 2009. Accuracy of genomic breeding values in multibreed dairy cattle populations. Genet. Sel. Evol. 41:51.

Hecker, M., S. Lambeck, S. Toepfer, E. van Someren, and R. Guthke. 2009. Gene regulatory network inference: Data integration in dynamic models-a review. Biosystems 96:86–103.

Henshall, J. M., R. J. Hawken, S. Dominik, and W. Barendse. 2012. Estimating the effect of SNP genotype on quantitative traits from pooled DNA samples. Genet. Sel. Evol. 44:12.

Hill, S. M., R. M. Neve, N. Bayani, W. L. Kuo, S. Ziyad, P. T. Spellman, J. W. Gray, and S. Mukherjee. 2012. Integrating biological knowledge into variable selection: An empirical Bayes approach with an application in cancer biology. BMC Bioinf. 13:94.

Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2009a. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37:1–13.

Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4:44–57.

Huang, Y., C. Maltecca, J. P. Cassady, L. J. Alexander, W. M. Snelling, and M. D. MacNeil. 2012. Effects of reduced panel, reference origin, and genetic relationship on imputation of genotypes in Hereford cattle. J. Anim. Sci. 90:4203-4208.

Huangfu, D., and K. V. Anderson. 2006. Signaling from Smo to Ci/Gli: Conservation and divergence of Hedgehog pathways from Drosophila to vertebrates. Development 133:3–14.

International Sheep Genomics Consortium, A. L. Archibald, N. E. Cockett, B. P. Dalrymple, T. Faraut, J. W. Kijas, J. F Maddox, J. C. McEwan, V. H. Oddy, H. W. Raadsma, C. Wade, J. Wang, W. Wang, and X. Xun. 2010. The sheep genome reference sequence: A work in progress. Anim. Genet. 41:449–453.

Kanehisa, M., M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, and Y. Yamanishi. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 36(Database issue):D480–D484.

Kanehisa, M., S. Goto, Y. Sato, M. Furumichi, and M. Tanabe. 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 40(Database issue):D109–D114.

Kim, H. S., J. E. Mittenthal, and G. Caetano-Anollés. 2006. MANET: Tracing evolution of protein architecture in metabolic networks. BMC Bioinf. 7:351.

Kunarso, G., N. Y. Chia, J. Jeyakani, C. Hwang, X. Lu, Y. S. Chan, H. H. Ng, and G. Bourque. 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat. Genet. 42:631–634.

Larkin, D. M., H. D. Daetwyler, A. G. Hernandez, C. L. Wright, L. A. Hetrick, L. Boucek, S. L. Bachman, M. R. Band, T. V. Akraiko, M. Cohen-Zinder, J. Thimmapuram, I. M. Macleod, T. T. Harkins, J. E. McCague, M. E. Goddard, B. J. Hayes, and H. A. Lewin. 2012. Whole-genome resequencing of two elite sires for the detection of haplotypes under selection in dairy cattle. Proc. Natl. Acad. Sci. USA 109:7693–7698.

Lee, T. I., N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J. B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young. 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science 298:799–804.

Lesnick, T. G., S. Papapetropoulos, D. C. Mash, J. Ffrench-Mullen, L. Shehadeh, M. de Andrade, J. R. Henley, W. A. Rocca, J. E. Ahlskog, and D. M. Maraganore. 2007. A genomic pathway approach to a complex disease: Axon guidance and Parkinson Disease. PLoS Genet. 3:e98.

Lôbo, R. B., D. Nkrumah, D. A. Grossi, P. S. Barros, P. Paiva, L. A. F. Bezerra, H. N. Oliveira, and M. V. B. Silva. 2011. Implementation of DNA markers to produce genomically-enhanced EPDs in Nellore cattle. Acta Scientiae Veterinariae 39(Suppl. 1):s23–s27.

MacGregor, S., Z. Z. Zhao, A. Henders, N. G. Martin, G. W. Montgomery, and P. M. Visscher. 2008. Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. Nucleic Acids Res. 36:e35.

Maltecca, C., K. L. Parker, and J. P. Cassady. 2012. Application of multiple shrinkage methods to genomic predictions. J. Anim. Sci. 90:1777–1787.

Mattick, J. S., and I. V. Makunin. 2006. Non-coding RNA. Hum. Mol. Genet. 15(Spec. No. 1):R17–R29.

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

McDaneld, T. G., L. A. Kuehn, M. G. Thomas, W. M. Snelling, T. S. Sonstegard, L. K. Matukumalli, T. P. L. Smith, E. J. Pollak, and J. W. Keele. 2012. Y are you not pregnant: Identification of Y chromosome segments in female bovine with decreased reproductive efficiency. J. Anim. Sci. 90:2142–2151.

McLaren, W., B. Pritchard, D. Rios, Y. Chen, P. Flicek, and F. Cunningham. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. Bioinformatics 26:2069–2070.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Meyer, K. 2007. WOMBAT—A tool for mixed model analyses in quantitative genetics by REML. J. Zhejiang Univ. Sci. B 8:815–821.

Mi, H., Q. Dong, A. Muruganujan, P. Gaudet, S. Lewis, and P. D. Thomas. 2010. PANTHER version 7: Improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res. 38(Database issue):D204–D210.

Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5:621–628.

Mosher, D. S., P. Quignon, C. D. Bustamante, N. B. Sutter, C. S. Mellersh, H. G. Parker, and E. A. Ostrander. 2007. A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. PLoS Genet. 3:e79.

Nonneman, D., A. K. Lindholm-Perry, S. D. Shackelford, D. A. King, T. L. Wheeler, G. A. Rohrer, C. D. Bierman, J. F. Schneider, R. K. Miller, H. Zerby, and S. J. Moeller. 2011. Predictive markers in calpastatin for tenderness in commercial pig populations. J. Anim. Sci. 89:2663–2672.

Northcutt, S. 2011. Genomic choices. http://www.angus.org/AGI/GenomicChoice11102011.pdf. (Accessed August 16, 2012.)

Peregrín-Alvarez, J. M., C. Sanford, and J. Parkinson. 2009. The conservation and evolutionary modularity of metabolism. Genome Biol. 10:R63.

Pico, A. R., T. Kelder, M. P. van Iersel, K. Hanspers, B. R. Conklin, and C. Evelo. 2008. WikiPathways: Pathway editing for the people. PLoS Biol. 6:e184.

Platter, W. J., J. D. Tatum, K. E. Belk, S. R. Koontz, P. L. Chapman, J. A. Scanga, and G. C. Smith. 2005. Effects of marbling and shear force on consumers' willingness to pay for beef strip loin steaks. J. Anim. Sci. 83:890–899.

Powell, R. L., and H. D. Norman. 2006. Major advances in genetic evaluation techniques. J. Dairy Sci. 89:1337–1348.

Qiu, Q., G. Zhang, T. Ma, W. Qian, J. Wang, Z. Ye, C. Cao, Q. Hu, J. Kim, D. M. Larkin, L. Auvil, B. Capitanu, J. Ma, H. A. Lewin, X. Qian, Y. Lang, R. Zhou, L. Wang, K. Wang, J. Xia, S. Liao, S. Pan, X. Lu, H. Hou, Y. Wang, X. Zang, Y. Yin, H. Ma, J. Zhang, Z. Wang, Y. Zhang, D. Zhang, T. Yonezawa, M. Hasegawa, Y. Zhong, W. Liu, Y. Zhang, Z. Huang, S. Zhang, R. Long, H. Yang, J. Wang, J. A. Lenstra, D. N. Cooper, Y. Wu, J. Wang, P. Shi, J. Wang, and J. Liu. 2012. The yak genome and adaptation to life at high altitude. Nat. Genet. 44:946–949.

Qu, Z., and D. L. Adelson. 2012. Bovine ncRNAs are abundant, primarily intergenic, conserved and associated with regulatory genes. PLoS ONE 7:e42638.

R Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramos, A. M., R. P. Crooijmans, N. A. Affara, A. J. Amaral, A. L. Archibald, J. E. Beever, C. Bendixen, C. Churcher, R. Clark, P. Dehais, M. S. Hansen, J. Hedegaard, Z. L. Hu, H. H. Kerstens, A. S. Law, H. J. Megens, D. Milan, D. J. Nonneman, G. A. Rohrer, M. F. Rothschild, T. P. Smith, R. D. Schnabel, C. P. Van Tassell, J. F. Taylor, R. T. Wiedmann, L. B. Schook, and M. A. Groenen. 2009. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS ONE 4:e6524.

Ravasi, T., H. Suzuki, C. V. Cannistraci, S. Katayama, V. B. Bajic, K. Tan, A. Akalin, S. Schmeier, M. Kanamori-Katayama, N. Bertin, P. Carninci, C. O. Daub, A. R. Forrest, J. Gough, S. Grimmond, J. H. Han, T. Hashimoto, W. Hide, O. Hofmann, A. Kamburov, M. Kaur, H. Kawaji, A. Kubosaki, T. Lassmann, E. van Nimwegen, C. R. MacPherson, C. Ogawa, A. Radovanovic, A. Schwartz, R. D. Teasdale, J. Tegnér, B. Lenhard, S. A. Teichmann, T. Arakawa, N. Ninomiya, K. Murakami, M. Tagami, S. Fukuda, K. Imamura, C. Kai, R. Ishihara, Y. Kitazume, J. Kawai, D. A. Hume, T. Ideker, and Y. Hayashizaki. 2010. An atlas of combinatorial transcriptional regulation in mouse and man. Cell 140:744–752.

Ravasz, E., A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabási. 2002. Hierarchical organization of modularity in metabolic networks. Science 297:1551–1555.

Rincon, G., K. L. Weber, A. L. Van Eenennaam, B. L. Golden, and J. F. Medrano. 2011. Hot topic: Performance of bovine high-density genotyping platforms in Holsteins and Jerseys. J. Dairy Sci. 94:6116–6121.

Rolf, M. M., J. F. Taylor, R. D. Schnabel, S. D. McKay, M. C. McClure, S. L. Northcutt, M. S. Kerley, and R. L. Weaber. 2011. Genome-wide association analysis for feed efficiency in Angus cattle. Anim. Genet. 43:367–374.

Rual, J. F., K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J.S. Albala, J. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhaute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth, and M. Vidal. 2005. Towards a proteome-scale map of the human protein-protein interaction network. Nature 437:1173–1178.

Saatchi, M., M. McClure, S. McKay, M. Rolf, J. Kim, J. Decker, T. Taxis, R. Chapple, H. Ramey, S. Northcutt, S. Bauck, B. Woodward, J. Dekkers, R. Fernando, R. Schnabel, D. Garrick, and J. Taylor. 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genet. Sel. Evol. 43:40.

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2011. FImpute – An efficient imputation algorithm for dairy cattle populations. J. Anim. Sci. 89(E-Suppl. 1):421 (Abstr.).

Seo, S., and H. A. Lewin. 2009. Reconstruction of metabolic pathways for the cattle genome. BMC Syst. Biol. 3:33.

Shackelford, S. D., T. L. Wheeler, and M. Koohmaraie. 1999. Evaluation of slice shear force as an objective method of assessing beef longissimus tenderness. J. Anim. Sci. 77:2693–2699.

Shackelford, S. D., T. L. Wheeler, and M. Koohmaraie. 2005. On-line classification of US Select beef carcasses for longissimus tenderness using visible and near-infrared reflectance spectroscopy. Meat Sci. 69:409–415.

Shalgi, R., D. Lieber, M. Oren, and Y. Pilpel. 2007. Global and local architecture of the mammalian microRNA-transcription factor regulatory network. PLOS Comput. Biol. 3:e131.

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 13:2498–2504.

Snelling, W. M., M. F. Allan, J. W. Keele, L. A. Kuehn, T. McDaneld, T. P. L. Smith, T. S. Sonstegard, R. M. Thallman, and G. L. Bennett. 2010. Genome-wide association study of growth in crossbred beef cattle. J. Anim. Sci. 88:837–848.

Snelling, W. M., M. F. Allan, J. W. Keele, L. A. Kuehn, R. M. Thallman, G.L. Bennett, C. L. Ferrell, T. G. Jenkins, H. C. Freetly, M. K. Nielsen, and K. M. Rolfe. 2011. Partial-genome evaluation of postweaning feed intake and efficiency of crossbred beef cattle. J. Anim. Sci. 89:1731–1741.

Snelling, W. M., R. A. Cushman, M. R. S. Fortes, A. Reverter, G. L. Bennett, J. W. Keele, L. A. Kuehn, T. G. McDaneld, R. M. Thallman, and M. G. Thomas. 2012. Physiology and endocrinology symposium: How single nucleotide polymorphism chips will advance our knowledge of factors controlling puberty and aid in selecting replacement beef females. J. Anim. Sci. 90:1152–1165.

Stein, L. 2001. Genome annotation: From sequence to biology. Nat. Rev. Genet. 2:493–503.

Stingo, F. C., and M. Vannucci. 2011. Variable selection for discriminant analysis with Markov random field priors for the analysis of microarray data. Bioinformatics 27:495–501.

Stothard, P., J.-W. Choi, U. Basu, J. M. Sumner-Thomson, Y. Meng, X. Liao, and S. S. Moore. 2011. Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. BMC Genomics 12:559.

Tantia, M. S., R. K. Vijh, V. Bhasin, P. Sikka, P. K. Vij, R. S. Kataria, B. P. Mishra, S. P. Yadav, A. K. Pandey, R. K. Sethi, B. K. Joshi, S. C. Gupta, and K. M. L. Pathak. 2011. Whole-genome sequence assembly of the water buffalo (*Bubalus bubalis*). Indian J. Anim. Sci. 81:38–46.

Thallman, R. M., and H. Koshinsky. 2012. Reduced cost genotyping strategies. http://www.beefefficiency.org/meetings/marcjune12/2-Thallman%20Handout.pdf. (Accessed August 16, 2012.)

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.

Vaquerizas, J. M., S. K. Kummerfeld, S. A. Teichmann, and N. M. Luscombe. 2009. A census of human transcription factors: Function, expression and evolution. Nat. Rev. Genet. 10:252–263.

Wade, C. M., E. Giulotto, S. Sigurdsson, M. Zoli, S. Gnerre, F. Imsland, T. L. Lear, D. L. Adelson, E. Bailey, R. R. Bellone, H. Blöcker, O. Distl, R. C. Edgar, M. Garber, T. Leeb, E. Mauceli, J. N. MacLeod, M. C. Penedo, J. M. Raison, T. Sharpe, J. Vogel, L. Andersson, D. F. Antczak, T. Biagi, M. M. Binns, B. P. Chowdhary, S. J. Coleman, G. Della Valle, S. Fryc, G. Guérin, T. Hasegawa, E. W. Hill, J. Jurka, A. Kiialainen, G. Lindgren, J. Liu, E. Magnani, J. R. Mickelson, J. Murray, S. G. Nergadze, R. Onofrio, S. Pedroni, M. F. Piras, T. Raudsepp, M. Rocchi, K. H. Røed, O. A. Ryder, S. Searle, L. Skow, J. E. Swinburne, A. C. Syvänen, T. Tozaki, S. J. Valberg, M. Vaudin, J. R. White, M. C. Zody, Broad Institute Genome Sequencing Platform, Broad Institute Whole Genome Assembly Team, E. S. Lander, and K. Lindblad-Toh. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 326:865–867.

Weaber, R. L. and J. L. Lusk. 2010. The economic value of improvements in beef tenderness by genetic marker selection. Am. J. Agr. Econ. 92:1456–1471.

Weber, K. L., R. M. Thallman, J. W. Keele, W. M. Snelling, G. L. Bennett, T. P. L. Smith, T. G. McDaneld, M. F. Allan, A. L. Van Eenennaam, and L. A. Kuehn. 2012. Accuracy of genomic breeding values in multi-breed beef cattle populations derived from deregressed breeding values and phenotypes. J. Anim. Sci. 90:4177–4190.

Wheeler, T. L., L. V. Cundiff, S. D. Shackelford, and M. Koohmaraie. 2005. Characterization of biological types of cattle (Cycle VII): Carcass, yield, and longissimus palatability traits. J. Anim. Sci. 83:196–207.

Wheeler, T. L., L. V. Cundiff, S. D. Shackelford, and M. Koohmaraie. 2010. Characterization of biological types of cattle (Cycle VIII): Carcass, yield, and longissimus palatability traits. J. Anim. Sci. 88:3070–3083.

Wheeler, T. L., S. D. Shackelford, and M. Koohmaraie. 1998. Cooking and palatability traits of beef longissimus steaks cooked with a belt grill or an open hearth electric broiler. J. Anim. Sci. 76:2805–2810.

Wiggans, G. R., P. M. VanRaden, and T. A. Cooper. 2011. The genomic evaluation system in the United States: Past, present, future. J. Dairy Sci. 94:3202–3211.

Wong, P., S. Althammer, A. Hildebrand, A. Kirschner, P. Pagel, B. Geissler, P. Smialowski, F. Blöchl, M. Oesterheld, T. Schmidt, N. Strack, F. J. Theis, A. Ruepp, and D. Frishman. 2008. An evolutionary and structural characterization of mammalian protein complex organization. BMC Genomics 9:629.

Xenarios, I., L. Salwínski, X. J. Duan, P. Higney, S. M. Kim, and D. Eisenberg. 2002. DIP, the database of interacting proteins: A research tool for studying cellular networks of protein interactions. Nucleic Acids Res. 30:303–305.

Zhang, W., B. L. Sayre, X. Xu, G. Tosser, J. Li, and W. Wang. 2011. Goat genome sequencing and its annotation. Abstr. W135. In: Proc. Plant Anim. Genome XIX. San Diego, CA.

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.