Invited review: Genomic analysis of data from physiological studies

D. J. Garrick,*†¹ L. H. Baumgard,* and H. L. Neibergs‡

*Department of Animal Science, Iowa State University, Ames 50011

†Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North 4422, New Zealand

‡Department of Animal Sciences, Washington State University, Pullman 99164

ABSTRACT

Physiology deals with the functions of living organisms and their systems, and its scientific endeavors can be viewed as having temporally occurred in 3 phases. The first phase of physiology studies focused on determining the functions of particular organs and tissues and their functional differences according to physiological status. The second phase of studies focused on characterizing differences in these functions according to the environment, or productivity. The third phase of studies focuses on determining the physiological causes of differences in productivity. Distinguishing cause from effect in physiological systems of inter-related processes is problematic, such that science has struggled to identify the root physiological mechanisms and their role in the network of genes leading to differences in productivity. Genomics is the study of the entire genome and provides powerful new tools that will accelerate third-phase discoveries of causal physiological processes. That research exploits information on DNA polymorphisms known as markers, complete DNA sequence, RNA sequence, and RNA expression in particular tissues at specific life stages. Physiologists can determine the genetic cause of mutant animals, identify genetic differences between cases and controls, and identify genes responsible for differences in performance between average and above-average animals. In some species, physiologists can leverage genomic data being used to predict genetic merit in elite seedstock populations, as a starting point to identify genes that will then motivate detailed physiological studies in the organs or tissues and stages of life in which those genes are expressed. Such work will increase our knowledge of biology and may lead to novel approaches to manipulate animal performance.

Key words: biotechnology, genomics, physiology

INTRODUCTION

Physiology is a science that studies how organism(s) function. Early research activities, which we describe as phase 1 studies, focused on determining the functions of particular body parts, such as the mammary gland, and identifying differences in the biology of those parts between animals with disparate physiological status, such as pre- versus postpubertal or lactating versus nonlactating states. Subsequent activities in physiological research, phase 2 studies, progressively shifted the focus to consider the nature of gross physiological differences between animal performance due to environmental differences (e.g., plane of nutrition, climatic conditions) or differences in their genetic merit (e.g., species, breed, strain, sire), but did not distinguish cause from effect.

Artificial selection has resulted in livestock species that are typically more productive and (or) efficient than their ancestors (Wilson and Willham, 1986; Kelm et al., 2000; Chen et al., 2003). That is, modern livestock yield more output per head than did their predecessors. These animals typically eat more (Koch et al., 1963; Buckley et al., 2007), sometimes exhibit lower activity levels (Braastad and Katle, 1989), and frequently exhibit differences in body composition (Pryce et al., 1999). Nevertheless, modern livestock are typically more efficient in terms of output per unit of feed intake due to altered nutrient partitioning, which ultimately leads to a phenomenon known as dilution of maintenance (Holloway et al., 1975; Bauman et al., 1985). A particular research interest is to determine the physiological mechanisms that have caused or allowed these homeorhetic and coordinated metabolic changes to occur (Bauman and Currie, 1980). An antiquated theory is that high-producing animals are simply more productive because they eat more, whereas recent thinking suggests that increased nutrient consumption is the result of enhanced productivity (i.e., increased synthesis of a valued product drives feed intake; Collier et al., 2005). Regardless, identifying what endocrine or metabolic factor(s) initiated these differences and in what organ(s) or tissue(s) and at what stage of life the changes originated are of obvious practical and academic interest.

Received September 23, 2011.

Accepted October 1, 2011.

¹Corresponding author: dorian@iastate.edu

A GENETIC BASIS FOR HIGH PERFORMANCE

Biological processes used to support maximum animal productivity probably include metabolic changes in most, if not all, body tissues. Some of the systemic physiological changes utilized by high-producing animals transcend physiological states and species differences. Undoubtedly, many changes in systemic physiology that contribute to variability in performance between animals are currently unknown to science. However, 2 hallmarks traditionally observed in animals exhibiting superior growth and lactation performance are altered parameters of the somatotrophic axis and insulin action (Hart et al., 1979; Bauman and Currie, 1980). Many of the genetic differences between low and high producers associated with the aforementioned systems have been described in a variety of organs and tissues (Bauman et al., 1985; Vernon and Pond, 1997). A thorough understanding of both the circulating changes and altered gene expression profiles in both systems has allowed commercial technology to exploit the biological differences and improve productive efficiency in multiple species (Etherton and Bauman, 1998).

The dogma of genetics is that the recipe, initiating all biological processes, is contained on DNA (Crick, 1958), the molecule responsible for inheritance in all organisms except some viruses. The formula for any particular gene product is stored on DNA in the form of a triplet code that may not be contiguous but fragmented on the genome into exons separated by intronic regions. The DNA, beginning at the promoter of a gene can, under certain circumstances, be transcribed to produce a smaller molecule, known as RNA. The RNA may be modified or spliced by removing introns and exons in alternative ways to generate functional sequence(s) that can be used to produce protein(s). A single gene may therefore code for multiple proteins (Black, 2003). The protein(s) may then be part of a biological cascade that leads to a particular physiological outcome. One component of genetics involves relating particular polypeptide sequences to their DNA sequences and their corresponding physical genomic locations and vice versa.

Major advances in phase 2 research activities have been possible (in part) from developments in molecular techniques for extracting, characterizing, and quantifying RNA levels (Skjaerlund et al., 1993), a research activity now called transcriptomics. Typical studies (e.g., McCarthy et al., 2009) have compared RNA abundance associated with differences in phenotype between a few animals in each of 2 or more groups representing different exogenous treatments, physiological status, genotypes, environmental circumstances, or levels of phenotypic performance. Such studies typically

find a plethora of differential expressions, and these can be used to compile genetic networks (Walsh and Henderson, 2004). However, most of the differences in expression will reflect downstream changes in metabolic cascades, which reflect an indirect consequence rather than the causal gene(s) (Collier et al., 2008).

Genetics is the study of inheritance and variation, whereas animal breeding exploits knowledge of genetics to select parents of the next generation to improve the performance of the production system (Wright, 1978). Those activities rely on us identifying superior selection candidates and do not require us to understand the majority of genes themselves (Hohenboken, 1982), but could benefit enormously from knowledge of polymorphisms in just those genes or noncoding regions that are responsible for variation in performance between individuals (Dekkers, 2004). Detectable regions of the genome that harbor such variants are known as QTL (Weller et al., 1990). Prior to the genomic era, studies to detect QTL tended to focus on functional candidate genes (Rempel et al., 1993; Yu et al., 1995) rather than considering all genomic locations.

RECENT DEVELOPMENTS IN BIOTECHNOLOGY

Three major breakthroughs have collectively changed the landscape to allow researchers in any discipline to exploit genomics and transcriptomics without needing wet laboratory facilities or personally developing hands-on molecular expertise. These are whole-genome sequencing, high-density genotyping panels, and next-generation sequencing.

Genome Sequencing

Genomic sequencing of livestock species, and the resultant assembled sequence, annotation, and comparative maps provide information on the location and underlying sequence of genes as well as the location of various kinds of polymorphisms (base-pair changes, deletions, inversions, copy number variants; e.g., <http://www.ncbi.nlm.nih.gov/projects/genome/guide/cow>). Sequencing activities in cattle, chickens, pigs, and sheep, for example, have provided the identification of millions of SNP spread throughout the genome. Assembled sequence enables a portfolio of SNP markers to be created that will provide genome-wide coverage of the inheritance of genome fragments. Further, it provides bioinformatic tools for locating candidate genes on the genome, identifying lists of positional candidate genes, and designing primers for molecular assays useful in fine-mapping QTL. Ongoing developments in sequencing technology are continually eroding the cost of collecting sequence, to the point that it is now

feasible to use next-generation sequencing (Schuster, 2008) to compare the sequence of individual experimental animals with the published reference sequence. The sequencing cost depends upon the depth of coverage, but is less than \$10,000 for an individual animal. Additional costs may be incurred in the bioinformatics activities to assemble the sequence.

High-Density SNP Genotyping

Characterization of specific genome fragments by genotyping for a known polymorphism has changed from being a laborious single-locus exercise to a cheap and straightforward routine laboratory assay (e.g., bovine 50K, Matukumalli et al., 2009) through the development of multiplex technologies that can simultaneously analyze 50,000 to 700,000 genotypes for no more than \$200 using third-party service companies (e.g., GeneSeek, Lincoln, NE). Such high-density SNP panels allow new approaches to finding genes causing variation in traits of industry or scientific interest.

Studying the entire genome rather than candidate genes has the advantage that it includes all the genes and must therefore include the causal effect(s) for any genetic attribute. The disadvantage is that in any particular tissue, most of the genes will not be routinely expressed, leaving the researcher with a problem not unlike finding a needle in a haystack. One approach to reduce the complexity of gene discovery is to analyze RNA rather than DNA.

Transcriptomics

It is tempting to think of the transcriptome as being much smaller than the entire genome, as it comprises only those fragments that are transcribed, or expressed, but alternative splicing allows several fold more transcripts to be produced than there are genes (Black, 2003). Early studies of gene expression relied on the creation of expression arrays (Lockhart et al., 1996), requiring prior experiments to collect and sequence RNA in a particular tissue, to create a multiplex array for determining expression levels, typically involving differential expression between pairs of labeled samples such as cases and controls. It was obviously not possible to find differential expression for any genes whose RNA was not quantified on the array, limiting future discovery to the quality of the foundation research that created the array. Gene expression is tissue specific, can vary with age and physiological status, and is sensitive to environmental stimuli, and these factors create practical problems in creating arrays that fully characterize the transcriptome. Next-generation sequencing technology allows extracted RNA to be directly sequenced, provid-

ing exhaustive assaying of any sampled transcriptome, unlimited by the quality of historical arrays.

Network or Pathway Analysis

Expression data can be incorporated into pathway analysis that considers sets of genes or their products that are known to functionally interact with or bind to one another. Pathway analysis uses information from annotated genes and physiological studies to identify the linkages between genes and physiology. Genome-wide pathway analysis enables the expansion to gene networks beyond those that may be typically evaluated for a physiological process. The inclusion of all gene networks into the study can provide new insights into regulatory and downstream genes involved in the physiological process that otherwise might not have been considered. Pathway analysis can provide information that would unlikely be discovered either through routine physiology or traditional genome-wide association studies (GWAS; Fortes et al., 2010; Neiberger et al., 2010).

Gene set enrichment analysis (GSEA; Mootha et al., 2003; Subramanian et al., 2005) is one analytical method used to identify functional pathways that are critical to a given physiological process and that are reflected in differences in the transcriptome between animals or between physiological states. An enhancement to GSEA uses knowledge of variation at the nucleotide level (Holden et al., 2008) and is referred to as GSEA-SNP. That approach assigns statistical values to SNP that serve as proxies for each gene and then ranks those statistics to test if genes within given pathways defined by databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene sets defined by Gene Ontology (GO) are associated with a trait. It has been suggested GSEA-SNP is more likely to identify genes of modest effects than GWAS that are powerful in identifying genes with major effects on a trait. In a study of genes associated with John's disease, Neiberger et al. (2010) used this approach and were able to find additional genes that had not been detected by GWAS. Although physiologists have made good use of RNA measurement for genes believed to be important in their research models, including using their results for network analysis, those data only identify association with phenotype and not causality. Many of the genes in the network will exhibit differential response simply because of changes in the expression of upstream genes.

DISCOVERING CAUSAL GENES OR PROCESSES

Based on these new technologies, physiology research is entering phase 3, whereby the focus is not on physi-

ological differences throughout development (phase 1), or gross physiological differences resulting from experimental treatments, the environment or the genetic backgrounds between animals differing in performance (phase 2), but focuses on specific physiological factors that initiate or cause differences between animals. In phase 3, the interest is on the genomic regions, genes, and functions that cause the variation in productivity between individuals. Furthermore, phase 3 discoveries, in concert with existing knowledge, will likely facilitate a more thorough understanding and global characterization of mechanisms causing physiological problems (i.e., metabolic disorders and poor reproductive variables) sometimes associated with more-productive animals and may lead to novel management strategies to augment efficiency. Genomics is revolutionizing phase 3 research that broadly fall into 1 of 2 classes—studies of clearly segregating major genes and studies of the differences between average and above-average individuals for traits likely controlled by either oligogenic (countable) or polygenic (infinitesimal) gene actions.

Major Genes

Evaluating the performance of animals in a cohort with like nutrition, management, and climate may identify clear evidence of segregation where some individuals have strikingly different phenotypes from others. In this circumstance, performance is likely influenced by a major gene, and it may even be possible to characterize individuals according to one or more of the alleles influencing the major gene. Classic examples involve coat color (Charlier et al., 1996), medullation (Dry, 1955), horns (sheep, Ibsen, 1944; cattle, Georges et al., 1993), double muscling (Cockett et al., 1994; Charlier et al., 1995), litter size (Montgomery et al., 1994), dwarfism, and a range of recessive diseases (e.g., in sheep, reviewed by Jolly et al., 2004). Only 2 phenotypic classes may exist, but the joint use of pedigree information can allow at least some individuals to be classified according to whether they carry 0, 1, or 2 copies of a particular allele. Identifying the genomic location responsible for the trait will facilitate selection (Dekkers, 2004) and enable positional candidate gene studies aimed at detecting the causal polymorphism and understanding the biology resulting in the phenotypic condition.

Discovering causative mutations used to require a decade or more of intense study, valuable resources, and a considerable amount of good fortune, even after sub-populations of individuals with known genotypes could be assembled. Today, such discovery activities can sometimes be accomplished in a matter of months. One approach to discovery of the genomic position responsible for a mutation in a major gene is to use

homozygosity mapping (Lander and Botstein, 1987). This requires that some closely related individuals are identified that are known to be homozygous for a major gene. For example, these might be half-sib individuals that demonstrate a particular recessive genetic defect. All these animals will be homozygous for the causal allele, and in close relatives, the allele would have come from the same ancestor; we would refer to them as being identical by descent (**IBD**). However, likely some fragment of the ancestral genome flanks the causal mutation and was also inherited intact to create an IBD region larger than just the causal mutation. The average chromosome exhibits about 1 crossover event per meiosis so that the probability of no crossover having occurred during 1 meiotic recombination in a fragment of 1 cM (about one-hundredth the size of a typical mammalian livestock chromosome) is 99%. If the homozygous individuals had been created by mating a heterozygous sire to his outcross daughters, 3 meioses would separate the sire and his homozygous offspring, giving any one of them a probability of $0.99^3 = 97\%$ of sharing the 1-cM ancestral fragment that includes the causal mutation. If there were 8 homozygous offspring, for example, the probability that they all share the 1-cM ancestral region would be $0.97^8 = 0.78$. Localizing the genomic region containing the causal mutation can, therefore, be undertaken by searching for signatures of selection characterized by homozygous regions where the individuals all share the same alleles. This will be straightforward if the genome can be characterized at a density that includes several marker loci on every 1-cM fragment. High-density genotyping panels are now available for most livestock species that can provide 10 to 30 such markers. Bovine panels from Illumina (San Diego, CA) and Affymetrix (Santa Clara, CA) are available that will provide more than 100 markers per cM, and similar panels will likely become available in other livestock species over the next few years.

Zhao and colleagues (2011a,b) used homozygosity mapping to investigate 2 recessive genetic disorders in sheep. Matings were first utilized to test the hypothesis that the condition was due to a recessive genetic disorder. The affected homozygous recessive offspring were then used to determine the IBD regions. Bioinformatics tools were incorporated to identify candidate genes in the largest IBD regions and to determine if any had been found to cause similar conditions in other species, particularly in humans and mice. Exons in the most promising candidate genes were then sequenced in several carrier animals that would have been heterozygous for the causal mutation. Any mutations discovered in the sequencing were then studied to determine their potential to cause the disease, and the most promising mutations were then genotyped across affected, carrier,

and control animals. In each of these studies, only one mutation was both concordant and biologically plausible for the disease. This approach, based only on IBD in the homozygous recessive animals, would be effective even if the mutation is recent and in a common haplotype whereby some control animals may also be homozygous for the same haplotype but not for the causal mutation. That circumstance will prevent many case-control methodologies from finding the region using markers that do not include the causal mutation itself.

Genes Influencing Polygenic Traits

Sophisticated analyses are required for finding genes whose effects cannot be seen to be clearly segregating in the population. This circumstance would include searches for many of the genes responsible for variation in traits with continuous phenotypes such as most productive traits, or polygenic traits with many genes contributing to the variation, whether continuously observed, or categorical traits, such as many reproductive traits. The ideal data for such an analysis would involve a population of individuals for which the true breeding values are known. The optimum statistical approach would be to regress the true breeding values on the number of copies of the mutation that influences performance. The slope of the regression line would represent the substitution effect or the impact of replacing one allele by its counterpart and is guaranteed, by the definition of breeding value, to be linear, even if the locus itself exhibits nonadditive gene action at the phenotypic level (Falconer and Mackay, 1996). The difference between the true breeding value of genotype AA compared with genotype aa is $2\times$ substitution effect. Two practical limitations preclude such analyses in real life. First, we seldom know the true breeding value of individuals in the population other than perhaps widely used sires. Second, we do not know the genes responsible for the variation in the trait, so we do not know which loci to use as independent variables in the regression. These problems can be simultaneously overcome by appropriate analytical methods.

In place of true breeding values as dependent variables, we can use so-called deregressed variables derived from estimated breeding values, and weight them in a manner that appropriately accounts for the accuracy or amount of information that was available for predicting the breeding value (Garrrick et al., 2009). This allows, for example, the use of industry sires with varying levels of accuracy, but limits gene discovery for those traits with routine performance recording and genetic evaluation programs. Frequently we are interested in novel traits, and in those cases, we can use phenotypic observations but must recognize that the number of

observations required to find a gene will increase as the amount of residual variation in the dependent variable increases. Traits with lower heritability will require more observations than traits with high heritability or traits where sires with accurate breeding values can be used (Weller et al., 1990).

Instead of using the unknown causal mutations as independent variables, we can use any known genotypes, which we refer to as markers. Provided that many markers are available from the entire genome, then some markers should be physically located close to every causal mutation. Loci that are physically close tend to be inherited together, unless a crossover occurs between the 2 loci, an event that will be very rare if the density involves tens or hundreds of markers per cM. When alleles at 2 nearby loci are inherited together, they can sometimes be in linkage disequilibrium (**LD**), meaning that the presence of a particular allele at one locus is predictive of the particular allele present at the other locus (Falconer and Mackay, 1996). This is a violation of Mendel's law of independent assortment and indicates one locus can act as a surrogate for the other locus.

Linkage Disequilibrium. The extent of LD can be quantified using the correlation between the genotypes at one locus and another. The amount of LD depends upon population factors, such as the effective population size, and on data structure, including the relatedness of the individuals used in the analysis, but generally the average LD between markers increases as the density of markers increases. Regressing observations on the number of copies of a marker allele will be similar to regressing observations on the number of copies of a causal allele if the marker locus and causal locus are in high LD. As we regress the observations on every marker locus, we would expect the estimated regression coefficient to be more significant as we test markers that are in high LD with the causal locus and to be less significant as the markers are more distant from the causal locus.

In human studies (e.g., Klein et al., 2005; Hindorff et al., 2011), researchers are typically interested in finding statistically significant mutations and tend to analyze the data one marker locus at a time as described above. In livestock improvement applications, we are also interested in using genomic information to predict breeding values for young animals without their own phenotypic information (Meuwissen et al., 2001). In such a circumstance, we are interested in all estimated effects, whether statistically significant or not. Further, we typically fit many if not all of the marker loci simultaneously.

Mixture Models. It is reasonable to assume that mutations controlling variation in productive traits are

Figure 1. Absolute value of bovine chromosome 1 estimated allele substitution effects on milk yield in Holsteins obtained from a Bayesian mixture model simultaneously fitting 50K SNP, plotted against genomic location.

not uniformly distributed along the entire genome, and therefore marker effects in some genomic regions would be expected to be near zero. Bayesian models that introduce prior knowledge can account for such assumptions, fitting mixture models whereby some proportion of markers have zero effect (e.g., the BayesB method pioneered by Meuwissen et al., 2001). The proportion of markers with zero effect is referred to using the parameter π , that might be assumed known (Meuwissen et al., 2001) or estimated simultaneously from the data (Habier et al., 2011).

Mixture models involve computing the likelihood for the model including a particular marker, and the likelihood for the model excluding a particular marker. The ratio of the likelihood for including the marker divided by the sum of the 2 likelihoods would be a fraction close to zero if the marker is uninformative, and close to 1 if the marker is very informative. Whereas a frequentist statistician might test a hypothesis regarding this marker and either include or exclude it, the Bayesian model averaging approach will include this marker according to its probability (Sorenson and Gianola, 2002), by random sampling. If the marker is included, subsequent markers are tested conditional on this marker being in the model. The process is repeated for all genomic markers, for many thousands of iterations of a Monte Carlo Markov chain whereby markers repeatedly enter or leave the model. In this way, all markers get a chance to contribute to the prediction of genomic merit, but better markers are used more often. Software developed with USDA funding is publicly available for these kinds of analyses (e.g., [//bigsc.ansci.iastate.edu](http://bigsc.ansci.iastate.edu)).

The true effects of causal mutations may be distributed among the estimates of marker effects for those genotyped markers in LD with the unknown mutation.

Accordingly, individual marker effects (e.g., Figure 1) can underestimate the effect of a causal mutation. Predicting the breeding value of a genome fragment using a small window of contiguous marker loci and their collective predicted effect generates a more robust estimate of the effect of that region than can be obtained from the individual marker effects. The variance of the collective effect of the region (e.g., Figure 2) contained in the window can then be obtained across the animals in the discovery population. This allows reliable characterization of gene effects (Fan et al., 2011; Onteru et al., 2011).

Perhaps of even more scientific and practical interest than the individual causal mutations is knowledge of the overall genetic architecture of a trait. The scope of that knowledge comprises much more than just the collective portfolio of causal mutations, also involving the number of genes, the number of mutations or alleles present for each gene, their allele frequencies, their physical distribution along the genome, their mode of gene action, the statistical distribution of their effect sizes, their influence on other traits (pleiotropy), and their homology across lines, breeds, and species. Some continuous traits appear to be controlled by a large number of genes with small effects, whereas others have a significant proportion of variation explained by a few genes, and these different architectures affect both gene discovery and the accuracy of genomic prediction (Hayes et al., 2010). Little is currently known and even is less published regarding genetic architectures of productive traits in livestock species, other than the characterization of some of the genomic regions with the largest effects. Unpublished work on genetic architecture currently being undertaken at Iowa State University, using industry populations of dairy cattle,

Figure 2. Percentage contribution to genetic variance for milk yield in Holsteins from consecutive genomic fragments, each point estimated from windows based on the substitution effects of 5 contiguous SNP plotted against their genomic location on bovine chromosome 1, from the same analysis depicted in Figure 1.

beef cattle, chickens, and pigs over a wide range of traits, suggests there are seldom more than 5 to 10 genomic regions that individually account for more than 1% of genetic variance, whereas typically a great many genomic regions account for between 0.1 and 1% of genetic variance. This is consistent with current findings in human populations; for example, for human height or for human diseases, where the phenomenon has been referred to as the “missing heritability” (Manolio et al., 2009).

Dedicated experiments with livestock species to identify genes responsible for observed variation in performance (e.g., Spelman et al., 2001) tend to be time-consuming and expensive to design, manage, and execute. If they require the generation of experimental animals from particular matings, and the rearing of these animals to measure traits such as lactation or reproductive performance, they may involve a time commitment that extends beyond the usual funding timetable (3 yr or less). One alternative approach is to use case-control studies (Waage et al., 2001; Pillars et al., 2011), where existing animals that vary in performance are used for genomic studies. Achieving sufficient power may require hundreds of such animals, and accounting for the selection involved in creating the case animals can be statistically problematic.

Genomic Prediction in Industry Populations.

An alternative to designed experiments or case-control studies is to use existing seedstock populations that are already recorded for many productive traits. An additional advantage of those populations is that they may already be genotyped for high-density marker panels because genomic prediction is widely used in dairy cattle improvement (VanRaden et al., 2009; Wig-

gans et al., 2011), and is increasingly being adopted in beef cattle (Garrrick, 2011), chickens (Wolc et al., 2011a,b), and pigs (Cleveland et al., 2010). The seedstock industry is interested in genomic prediction to rank the animals at a young age, and the information on genomic regions accounting for variation is simply a by-product. The accuracies of genomic predictions are highest in close relatives of the discovery population; for example, in their immediate offspring. In distant relatives or in other breeds (Kizilkaya et al., 2010), the predictions tend to have surprisingly little utility. We can explain this observation in several ways. Insufficient LD might exist on existing high-density marker panels. The training populations, typically fewer than 10,000 individuals, might be too small. The causal mutations may be dominant, and their effect on breeding values will depend upon the gene frequency of the alternate alleles in the target population. The mutations may be epistatic, such that their effects depend upon the allelic values at other loci, and the genetic background may vary widely among unrelated individuals. Different genes may be responsible for variation in different populations. Genotype $\times$ environment interaction (Bryant et al., 2007) may be the problem. These factors have generally yet to be carefully studied from a genomic perspective in any production traits, but the explanation will be of interest from the perspectives of physiological and animal production.

CONCLUSIONS

Genomics tools have become available that will enable researchers to identify the genes that cause variation in productive performance. Better understanding

of the physiological factors that cause differences in performance would be beneficial to both an understanding of the biology of performance traits and for its manipulation to improve performance by genetic selection or management interventions.

REFERENCES

- Bauman, D. E., and W. B. Currie. 1980. Partitioning of nutrients during pregnancy and lactation: A review of mechanisms involving homeostasis and homeorhesis. *J. Dairy Sci.* 63:1514–1529.
- Bauman, D. E., S. N. McCutcheon, W. D. Steinhour, P. J. Eppard, and S. J. Sechen. 1985. Sources of variation and prospects for improvement of productive efficiency in the dairy cow: A review. *J. Anim. Sci.* 60:583–592.
- Black, D. L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72:291–336.
- Braastad, B. O., and J. Kattø. 1989. Behavioural differences between laying hen populations selected for high and low efficiency of feed utilization. *Br. Poult. Sci.* 30:533–544.
- Bryant, J. R., N. Lopez-Villalobos, J. E. Pryce, C. W. Holmes, D. L. Johnson, and D. J. Garrick. 2007. Environmental sensitivity in New Zealand dairy cattle. *J. Dairy Sci.* 90:1538–1547. (Erratum in *J. Dairy Sci.* 90:2577)
- Buckley, F., B. Horan, N. Lopez-Villalobos, and P. Dillon. 2007. Milk production efficiency of varying dairy cow genotypes under grazing conditions. Pages 74–83 in Meeting the Challenges for Pasture-Based Dairying. Proc. 3rd Dairy Sci. Symp. University of Melbourne, Victoria, Australia.
- Charlier, C., W. Coppieters, F. Farnir, L. Grobet, P. L. Leroy, C. Michaux, M. Mni, A. Schwes, P. Vanmanshoven, R. Hanset, and M. Georges. 1995. The *mh* gene causing double-muscling in cattle maps to bovine chromosome 2. *Mamm. Genome* 6:788–792.
- Charlier, C., B. Denys, J. I. Belanche, W. Coppieters, L. Grobet, M. Mni, J. Womack, R. Hanset, and M. Georges. 1996. Microsatellite mapping of the bovine roan locus: A major determinant of white heifer disease. *Mamm. Genome* 7:138–142.
- Chen, P., T. J. Baas, J. W. Mabry, K. J. Koehler, and J. C. M. Dekkers. 2003. Genetic parameters and trends for litter traits in U.S. Yorkshire, Duroc, Hampshire, and Landrace pigs. *J. Anim. Sci.* 81:46–53.
- Cleveland, M. A., S. Forni, D. J. Garrick, and N. Deeb. 2010. Prediction of genomic breeding values in a commercial pig population. Page 506 in Proc. 9th World Congr. Genet. Appl. Livest. Prod. German Society for Animal Science, Leipzig, Germany.
- Cockett, N. E., S. P. Jackson, T. L. Shay, D. Nielsen, S. S. Moore, M. R. Steele, W. Barendse, R. D. Green, and M. Georges. 1994. Chromosomal localization of the callipyge gene in sheep (*Ovis aries*) using bovine DNA markers. *Proc. Natl. Acad. Sci. USA* 91:3019–3023.
- Collier, R. J., L. H. Baumgard, A. L. Lock, and D. E. Bauman. 2005. Physiological limitations, nutrient partitioning. Pages 351–377 in Yields of Farmed Species: Constraints and Opportunities in the 21st Century. Proceedings: 61st Easter School. Nottingham, UK. J. Wiseman and R. Bradley, ed. Nottingham University Press, Nottingham, UK.
- Collier, R. J., J. L. Collier, R. P. Rhoads, and L. H. Baumgard. 2008. Invited review: Genes involved in the bovine heat stress response. *J. Dairy Sci.* 91:445–454.
- Crick, F. H. C. 1958. On protein synthesis. Symp. Soc. Exp. Biol. 12:138–163.
- Dekkers, J. C. M. 2004. Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. *J. Anim. Sci.* 82(E-Suppl.):E313–E328.
- Dry, F. W. 1955. The dominant N gene in New Zealand Romney sheep. *Aust. J. Agric. Res.* 6:725–769.
- Etherton, T. D., and D. E. Bauman. 1998. Biology of somatotropin in growth and lactation of domestic animals. *Physiol. Rev.* 78:745–761.
- Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to Quantitative Genetics. 4th ed. Prentice Hall, Harlow, UK.
- Fan, B., S. K. Onteru, Z. Q. Du, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011. Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. *PLoS ONE* 6:e14726.
- 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.
- Garrick, D. J. 2011. The nature, scope and impact of genomic prediction in beef cattle in the United States. *Genet. Sel. Evol.* 43:17.
- 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.
- Georges, M., R. Drinkwater, T. King, A. Mishra, S. S. Moore, D. Nielsen, L. S. Sargeant, A. Sorensen, M. R. Steele, X. Zhao, J. E. Womack, and J. Hetzel. 1993. Microsatellite mapping of a gene affecting horn development in *Bos taurus*. *Nat. Genet.* 4:206–210.
- Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2011. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186.
- Hart, I. C., J. A. Bines, and S. V. Morant. 1979. Endocrine control of energy metabolism in the cow: Correlations of hormones and metabolites in high and low yielding cows for stages of lactation. *J. Dairy Sci.* 62:270–277.
- Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: Coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genet.* 6:e1001139.
- Hindorf, L. A., H. A. Junkins, P. N. Hall, J. P. Mehta, and T. A. Manolio. 2011. A catalog of published genome-wide association studies. Accessed Sep. 16, 2011. www.genome.gov/gwastudies.
- Hohenboken, W. D. 1982. Beef cattle breeding à la Jefferson. *J. Anim. Sci.* 54:667–679.
- Holden, M., S. Deng, L. Wojnowski, and B. Kulle. 2008. GSEA-SNP: Applying gene set enrichment analysis to SNP data from genome-wide association studies. *Bioinformatics* 24:2784–2785.
- Holloway, J. W., D. F. Stephens, J. V. Whiteman, and R. Totusek. 1975. Efficiency of production of 2- and 3-year old Hereford, Hereford × Holstein and Holstein cows. *J. Anim. Sci.* 41:855–867.
- Ibsen, H. L. 1944. Horn and scur inheritance in certain breeds of sheep. *Am. Nat.* 78:506–516.
- Jolly, R. D., H. T. Blair, and A. C. Johnstone. 2004. Genetic disorders of sheep in New Zealand: A review and perspective. *N. Z. Vet. J.* 52:52–64.
- Kelm, S. C., and A. E. Freeman. 2000. Direct and correlated responses to selection for milk yield: Results and conclusions of regional project NC-2, “Improvement of dairy cattle through breeding, with emphasis on selection”. *J. Dairy Sci.* 83:2721–2732.
- Kizilkaya, K., R. L. Fernando, and D. J. Garrick. 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *J. Anim. Sci.* 88:544–551.
- Klein, R. J., C. Zeiss, E. Y. Chew, J. Y. Tsai, R. S. Sackler, C. Haynes, A. K. Henning, J. P. SanGiovanni, S. M. Mane, S. T. Mayne, M. B. Bracken, F. L. Ferris, J. Ott, C. Barnstable, and J. Hoh. 2005. Complement factor H polymorphism in age-related macular degeneration. *Science* 308:385–389.
- Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory. 1963. Efficiency of feed use in beef cattle. *J. Anim. Sci.* 22:486–494.
- Lander, E. S., and D. Botstein. 1987. Homozygosity mapping: A way to map human recessive traits with the DNA of inbred children. *Science* 236:1567–1570.
- Lockhart, D. J., H. Dong, M. C. Byrne, M. T. Folletie, M. V. Gallo, M. S. Chee, M. Mittman, C. Wang, M. Kobayashi, H. Norton, and E. L. Brown. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* 14:1675–1680.
- Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorf, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Car-

- don, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. Mackay, S. A. McCarroll, and P. M. Visscher. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747–753.
- 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.
- McCarthy, S. D., S. T. Butler, J. Patton, M. Daly, D. G. Morris, D. A. Kenny, and S. M. Waters. 2009. Differences in the expression of genes involved in the somatotrophic axis in divergent strains of Holstein-Friesian dairy cows during early and mid lactation. *J. Dairy Sci.* 92:5229–5238.
- 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.
- Montgomery, G. W., E. A. Lord, J. M. Penty, K. G. Dodds, T. E. Broad, L. Cambridge, S. L. F. Sunden, R. T. Stone, and A. Crawford. 1994. The Booroola fecundity (FecB) gene maps to sheep chromosome 6. *Genomics* 22:148–153.
- Mootha, V. K., C. M. Lindgren, K.-F. Eriksson, A. Subramanian, S. Sihag, J. Lehar, P. Puigserver, E. Carlsson, M. Ridderstråle, E. Laurila, N. Houstis, M. J. Daly, N. Patterson, J. P. Mesirov, T. R. Golub, P. Tamayo, B. Spiegelman, E. S. Lander, J. N. Hirschhorn, D. Altshuler, and L. C. Groop. 2003. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* 34:267–273.
- Neibergs, H. L., M. L. Settles, R. H. Whitlock, and J. F. Taylor. 2010. GSEA-SNP identifies genes associated with Johne's disease in cattle. *Mamm. Genome* 21:419–425.
- Onteru, S. K., B. Fan, M. T. Nikkila, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011. Whole-genome association analyses for lifetime reproductive traits in the pig. *J. Anim. Sci.* 89:988–995.
- Pillars, R. B., M. W. Bolton, and D. L. Grooms. 2011. Case-control study: Productivity and longevity of dairy cows that tested positive for infection with *Mycobacterium avium* ssp. *paratuberculosis* as heifers compared to age-matched controls. *J. Dairy Sci.* 94:2825–2831.
- Pryce, J. E., B. L. Nielsen, R. F. Veerkamp, and G. Simm. 1999. Genotype and feeding system effects and interactions for health and fertility in dairy cattle. *Livest. Prod. Sci.* 57:193–201.
- Rempel, W. E., M. Lu, S. el Kandelgy, C. F. Kennedy, L. R. Irvin, J. R. Mickelson, and C. F. Louis. 1993. Relative accuracy of the halothane challenge test and a molecular genetic test in detecting the gene for porcine stress syndrome. *J. Anim. Sci.* 71:1395–1399.
- Schuster, S. C. 2008. Next-generation sequencing transforms today's biology. *Nat. Methods* 5:16–18.
- Skjaerlund, D. M., A. L. Grant, W. G. Helferich, W. G. Bergen, and R. A. Merkel. 1993. Determination of the relative abundance of skeletal muscle alpha actin mRNA in muscle of livestock species. *J. Anim. Sci.* 71:393–399.
- Sorenson, D., and D. Gianola. 2002. Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics. Springer-Verlag, New York, NY.
- Spelman, R. J., F. M. Miller, J. D. Hooper, M. Thielen, and D. J. Garrick. 2001. Experimental design for QTL trial involving New Zealand Friesian and Jersey breeds. Pages 393–396 in *Proc. 14th Conf. Assoc. Adv. Anim. Breed. Genet.* Accessed Nov. 23, 2011. <http://www.aaabg.org/livestocklibrary/2001/ab01093.pdf>.
- Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, and E. S. Lander. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102:15545–15550.
- VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92:16–24.
- Vernon, R. G., and C. M. Pond. 1997. Adaptations of maternal adipose tissue to lactation. *J. Mammary Gland Biol. Neoplasia* 2:231–241.
- Waage, S., S. A. Odegaard, A. Lund, S. Brattgjerd, and T. Rothe. 2001. Case-control study of risk factors for clinical mastitis in postpartum dairy heifers. *J. Dairy Sci.* 84:392–399.
- Walsh, B., and D. Henderson. 2004. Microarrays and beyond: What potential do current and future genomics tools have for breeders. *J. Anim. Sci.* 82(E-Suppl.):E292–E299.
- 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.
- Wiggans, G. R., P. M. VanRaden, and T. A. Cooper. 2011. The genomic evaluation system in the United States: Past, present and future. *J. Dairy Sci.* 94:3202–3211.
- Wilson, D. E., and R. L. Willham. 1986. Within-herd phenotypic, genetic and environmental trend lines for beef cattle breeders. *J. Anim. Sci.* 63:1087–1094.
- Wolc, A., J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, R. Preisinger, D. Habier, R. Fernando, D. J. Garrick, and J. C. M. Dekkers. 2011a. Persistence of accuracy of genomic estimated breeding values over generation in layer chickens. *Genet. Sel. Evol.* 43:23.
- Wolc, A., C. Stricker, J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, R. Preisinger, D. Habier, R. Fernando, D. J. Garrick, S. J. Lamont, and J. C. M. Dekkers. 2011b. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genet. Sel. Evol.* 43:5.
- Wright, S. 1978. The relation of livestock breeding to theories of evolution. *J. Anim. Sci.* 46:1192–1200.
- Yu, T.-P., C. K. Tuggle, C. B. Schmitz, and M. F. Rothschild. 1995. Association of PIT1 polymorphisms with growth and carcass traits in pigs. *J. Anim. Sci.* 73:1282–1288.
- Zhao, X., K. E. Dittmer, H. T. Blair, K. G. Thompson, M. F. Rothschild, and D. J. Garrick. 2011a. A novel nonsense mutation in the DMP1 gene identified by a genome-wide association study is responsible for inherited rickets in Corriedale sheep. *PLoS ONE* 6:e21739.
- Zhao, X., S. K. Onteru, S. Piripi, K. G. Thompson, H. T. Blair, D. J. Garrick, and M. F. Rothschild. 2011b. In a shake of a lamb's tail: Using genomics to unravel a cause of chondrodysplasia in Texel sheep. *Anim. Genet.* doi:10.1111/j.1365-2052.2011.02304.x.