The Nature, Scope And Impact Of Some Whole-Genome Analyses In Beef Cattle

D.J. Garrick[*†]

Introduction

Artificial selection has proven to be effective at altering the performance of animal production systems. Nevertheless, selection based upon assessment of the genetic superiority of candidates is suboptimal as a result of errors in the prediction of genetic merit. Conventional breeding programs sometimes address this issue by extending phenotypic measurements on selection candidates to include correlated indicator traits (e.g. ultrasound), or by delaying selection decisions beyond puberty until phenotypic performance can be observed on progeny or other relatives. Extending the generation interval in order to increase the accuracy of selection has the unfortunate consequence of reducing annual rates of gain compared to schemes that can accurately select parents of the next generation by puberty. Furthermore, such delays often increase costs of the breeding scheme. Marker-Assisted and Whole-Genome Selection (WGS) are aimed at reducing prediction errors at pubertal assessment of merit by exploiting information on the transmission of chromosome fragments from parents to selection candidates, in conjunction with knowledge of the relative impact of particular chromosome fragments on performance. Whole-Genome Analyses (WGA) refers to those studies that are undertaken to determine the relative impact (i.e. substitution effects) of various chromosome fragments identified using high-density SNP genotypes. The Illumina BovineSNP50® is currently the preferred approach for genotyping *Bos taurus* cattle but is likely to be superseded in 2010 by Illumina and Affymetrix arrays currently in development with reported marker densities increased by at least an order of magnitude. In order for WGS to influence breeding programs and the rate or cost of genetic gains, WGA must be undertaken, and genomic prediction tools made available for breeders and other industry stakeholders to cost-effectively adopt in their breeding programs. This paper reviews the nature or kind of studies currently underway, the scope or extent of some of those studies, and comments on the likely impact of WGS in terms of predictive value.

Whole-Genome Analyses

Critical issues to consider in the adoption of WGA are the motivation, the choice of population, the choice of individuals to genotype, the availability of existing trait information, and the opportunity to collect novel phenotypes. Following implementation of findings from such analyses into selection tools, the principal concerns in relation to resultant

[*] Department of Animal Science, Iowa State University, Ames, IA 50011, USA
[†] Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand

tools are their reliability and cost. Such information would allow the development of rational breeding programs that appropriately exploit these new technologies.

Motivation for Whole-Genome Analyses. There are three principal reasons for WGA. These are: to develop improved tools for prediction; to find genomic regions and genes that cause variation; and to develop proprietary know-how that can be marketed as a business proposition to one or more sectors of the industry.

Public-domain research is being undertaken by animal breeders to develop methodologies and computing strategies and characterize predictive ability. Such work is typically government or industry funded and logically follows previous research developing prediction methodologies and characterizing associated information such as heritabilities, genetic and phenotypic correlations. Allied research includes the implementation details required to incorporate the findings in national cattle evaluation (NCE) systems for routine application. Prior to the availability of high-density SNP platforms, genomic research focused on discovering and exploiting one or a few quantitative trait loci (QTL). In the last decade, the focus has shifted to concurrently exploiting knowledge of the whole genome. That process involves a discovery or training phase that determines the informative SNP and quantifies their predictive ability in some population, rather like the estimated breeding value (EBV) of a chromosome fragment. Collectively, this process leads to a linear function or predictive key that can be used to predict genomic merit by summing up the values of the chromosome fragments carried by an animal outside the training population.

Some geneticists and physiologists with an interest in genetics and causation are undertaking public-domain research, principally to locate QTL, to fine map to discover, understand, and perhaps patent, the underlying quantitative trait nucleotides. These researchers are taking advantage of the reduction in labor and genotyping costs from using high-density SNP arrays that provide comprehensive genomic coverage, rather than targeting only candidate genes (which assumes local or cis- gene action accounts for variation), or sparsely genotyping up to a few hundred microsatellite markers chosen to provide information content and genome coverage. These QTL studies typically rely on third parties to implement the findings by purchasing the rights to market SNP tests to industry, and validating the findings.

Private companies, notably Igenity and Pfizer Animal Genetics, are investing in WGA in order to market proprietary tests. The studies may be undertaken in house, or in collaboration with publicly funded researchers. Studies represent a “for profit” investment in research and implementation that is unprecedented in beef cattle improvement but more common in other livestock species (e.g. chickens and pigs) and in field crops (e.g. maize and soybeans). In those industries value is typically captured to provide a return on research investment by marketing and controlling improved germplasm rather than selling tests.

Nature of Whole-Genome Analyses. There are various mating plans to develop resource populations for WGA. Such studies are expensive and time-consuming if designed matings are required in advance of the experiment. That time frame is too long for most competitive grants and for private company investment. Accordingly most studies use existing industry resources or research herds established for other purposes.

Industry populations have the advantage that they already exist and can be immediately genotyped. Further, in the case of elite or widely used industry animals, the discovery individuals will be relevant to the commercial population. In the case of artificial insemination (AI) sires, they have the further advantage that DNA is readily accessible despite the disparate ownership or physical location of the animals. The principal source of information for WGA comes in the form of EBV or expected progeny difference (EPD) from NCE and is well represented for growth traits, moderately well represented for ultrasound traits, poorly represented for behavior, reproduction and longevity traits with typically no information on many other traits such as disease resistance or eating quality. Training on crossbred sires is seldom an option using NCE data and is limited to those few breed associations that collect crossbred data.

A U.S. repository of DNA from 1,985 Angus bulls born between 1948 and 2007 have been assembled by the University of Missouri and Merial (Woodward *et al*. 2010). These bulls have generally been widely used in the Angus breed, and are represented in American Angus Association pedigrees. Accordingly, these bulls have EPDs and accuracies for *production*: calving ease direct; birthweight; weaning weight; yearling weight; yearling height; scrotal circumference; *maternal traits*: calving ease maternal; milk; mature weight; mature height; *carcass*: carcass weight; marbling; ribeye area; fatdepth and some *new* EPDs like docility and heifer pregnancy. The accuracies of EPD on old bulls are limited for some traits.

The U.S. Meat Animal Research Center (MARC) has worked with some breed associations to develop a repository of some 2,026 influential or upcoming bulls in 16 of the most prominent beef breeds in the U.S. with EPDs from NCE and includes: Angus; Beefmaster; Brahman; Brangus; Braunvieh; Charolais; Chiangus; Gelbvieh; Hereford; Limousin; Maine-Anjou; Red Angus; Salers; Santa Gertrudis; Shorthorn; and Simmental. Initial plans for this repository were to use it to provide genomic predictions of these bulls from training analyses based on a MARC crossbred population (Thallman, 2009).

The U.S. carcass merit project (CMP) was an industry-funded undertaking initiated in 1998 that collected carcass data, tenderness and sensory attributes on over 8,200 progeny. Some of the offspring of more than 70 sires across 13 breeds were DNA sampled. The sires were widely-used AI bulls from various breeds and dams were commercial cows (Thallman *et al*. 2003). The dataset has been valuable for validation of early genomic tests undertaken in the U.S. by the National Beef Cattle Evaluation Consortium (NBCEC), the details having been published on-line by Van Eenennaam *et al*. (accessible from www.nbcec.org). The CMP dataset has more recently been high-density genotyped by at least two different organizations for gene discovery and whole genome prediction, limiting its future value for validation.

Two major studies are being undertaken by Pfizer Animal Genetics in the U.S. in collaboration with University partners. At Colorado State University, two cohorts each of about 1,500 composite British and Continental steers from one ranch in Nebraska, have been extensively phenotyped for feedlot health, particularly respiratory disease and response to treatment. Sickness was assessed visually, by temperature profiles and by lung damage scores. Data includes temperament and immunological measures, as well as growth and

carcass information (Brigham *et al*. 2009). At Iowa State University, several cohorts representing 2,300 predominately Angus cattle have been assessed for carcass and meat quality attributes, including tenderness and sensory information, in addition to extensive phenotyping on traits that might influence the human healthfulness of beef. These healthy beef traits include mineral and fatty acid composition of key muscles (Reecy *et al*. 2010).

Some research populations that have been used for WGA in the U.S. include descendants from the MARC Cycle VII germplasm evaluations, that represent crosses including offspring of recent AI bulls from Angus, Charolais, Gelbvieh, Hereford, Limousin, Red Angus, and Simmental breeds. Measured traits include feed intake, carcass and tenderness data on some offspring, puberty records and incidence of disease. The findings will be published in the public domain, along with Snelling *et al*. (2010).

Other research populations used for WGA include animals from the USDA Line 1 Hereford population kept in Miles City, MO, a herd that has been closed and inbred for over 75 years and all current animals have additive relationships with each other that exceed 50%. In addition to gene discovery, this population will be used to investigate genomic regions with a signature of selection and with an excess of heterozygocity (MacNeil, *pers. comm*.).

The Canadian University of Alberta and University of Guelph are collaborating with animal and genomic resources, including feed intake and carcass phenotypes for WGA. This collaboration also includes MARC and the Australian Cooperative Research Center for Beef Cattle (Beef CRC). The Beef CRC has data from their first cycle including almost 8,000 straight bred Angus, Belmont Red, Brahman, Hereford, Murray Grey, Santa Gertrudis and Shorthorn cattle plus another 2,000 crossbred individuals representing a range of finishing environments. The second cycle of the Beef CRC includes almost 4,500 steers and heifers of two tropical breeds with a wide range of growth, feed efficiency carcass and beef quality attributes on the steers, and adaptive and reproductive traits on the heifers. Another 6,000 bull calves have been measured for growth and male reproduction traits. More details are in a review volume at http://www.publish.csiro.au/nid/72/issue/5223.htm. The first stage of collaboration was limited to exchange of marker effects (Burrow, *pers. comm*.). The collaboration includes New Mexico State University, contributing some 800 Brahman-cross heifers assessed for growth, ultrasound and reproductive performance (Peters *et al*. 2010).

In Brazil, funding has been approved for WGA in 9 herds of Herefords and Braford cattle that will be phenotyped for growth, body composition, reproduction and tick resistance, and for a separate study of reproduction traits in Nelore cattle (Cardoso, *pers. comm*.).

Research populations that were designed for QTL studies in the microsatellite era include F2 intercrosses of dairy and beef breeds such as Jersey x Limousin (Esmailizadeh *et al*. 2008), Holstein x Charolais (Gutierrez-Gil *et al*. 2009) and an intercross/backcross design of Brahman x Angus (Amen *et al*. 2007). Such populations might usefully contribute to WGA.

Scope of Whole-Genome Analyses. In contrast to studies using microsatellite markers, the scope or extent of WGA studies have principally been limited by the availability of animals with measured phenotypes or estimated breeding values from progeny tests rather than by the

costs of genotyping. High-density BovineSNP50® genotypes can currently be contained for under US$200 per animal, in contrast to the cost of microsatellite genotyping which at $5 per genotype would cost US$1,000 per animal for 200 loci. Furthermore, it is a straightforward exercise to genotype thousands of animals for the BovineSNP50® in the course of a few weeks whereas few labs could manage that number of animals in an entire year using microsatellite technology. Collectively these facts have allowed the scope of discovery populations to be increased from focused subsamples of the most informative animals to populations that include every individual or AI bull for which DNA can be obtained.

Impact of Whole-Genome Analyses. Findings from WGA will have no impact on industry unless they lead to new or improved tools for breeders. Given that WGA results are made available to industry, their potential impact will depend upon how much the technology can increase accuracy (i.e. reduce prediction errors) at the point of selection, the traits for which they can be applied, and the cost of the technology. The realization of the potential impact will further be limited by the manner in which the industry adopts these tools.

The registered beef cattle industry, like the dairy industry, can be categorized as having four pathways of selection that influence the rate of genetic gain: sires to breeds sires; sires to breed cows; cows to breed sires; and cows to breed cows. The commercial beef cattle sector has unregistered cattle and purchases breeding sires from the registered sector. The genetic merit of any national beef industry depends upon its historical genetic merit, the annual rate of genetic change in the registered (nucleus) herds, and the genetic lag between nucleus and commercial sectors. Tools developed from WGA can be used for WGS in any one or more of the four selection pathways to increase genetic gain in the nucleus, or to reduce genetic lag between nucleus and commercial sectors. The business proposition for procuring genetic tests is quite different in each of these scenarios. Furthermore, the size of the markets are disproportionate, with greatest numerical potential for testing bulls offered for sale to use as sires in commercial herds, where their selection has no impact on the rate of genetic gain.

Communication of Results from Whole-Genome Analyses. The last few decades have been characterized by communication of the results of genetic evaluations in the usual units of measurement for each trait, in the form of an EBV, EPD, or an index to reflect aggregate merit. Early attempts to market genetic tests on alternative scales had some success when tests involved only a single SNP, but have become problematic as the tests have evolved to include multiple SNPs across genomic regions. Industry confusion has also developed as sires have been identified that demonstrate apparent conflict between the pedigree-based and genomic-based merit. In all these circumstances, an appealing approach is to incorporate the information from genomic testing into NCE (Kachman, 2008; MacNeil *et al*. 2010) so that the information is reflected in terms of marker-assisted or genomic-assisted EBV and associated accuracy, without introducing new jargon, terminologies and interpretations.

This is achievable, but not straightforward. Some would argue that genotyping results would be accumulated on all animals in central databases; in the same manner as occurs for pedigree and performance information. This would ensure that results were available on individuals that were subsequently culled as well as those that were selected, allowing evaluation systems to account for selection, a requirement for unbiased predictions. Further,

this approach would allow improved predictive technology to be retrospectively applied to historically collected genotypes. It would also facilitate future activities to *in-silico* genotype historical animals, on the basis of a subset of the population genotyped at higher density or individually sequenced. Centrally stored genotypes might be practical if evaluations are collectively funded by industry, by government, or both, and where genomic testing tools are delivered as an industry good. It is less apparent that this model can work where genetic testing is offered as a for-profit service, and the linear function of informative SNP or predictive key is the principal intellectual property or know how on which competing testing companies have based their business. It faces challenges when entities such as breed associations undertake NCE as a peripheral activity when their core business is selling pedigree and animal registrations. The breed associations may be challenged by the know-how and financial costs of changing their software to accommodate these rapidly changing technologies. Some direct methods for including pedigree and genomic information in a single evaluation have been proposed (Aguilar *et al*. 2010), but these approaches will suffer from computational challenges as the number of genotyped individuals increase, and currently lack a convincing statistical basis even when computationally feasible.

An alternative approach is to derive genomic EBV solely from marker information and incorporate that value as a correlated trait in NCE. This has the advantage that a company selling genetic tests can maintain their proprietary predictive key (and their customers genotypes) and NCE need only be modified to include correlated information. This approach is not altogether straightforward as it requires knowledge of the covariance components relating the marker score to phenotype and faces challenge with the evolution of molecular keys over time, with increases in the number of routinely genotyped markers, and perhaps with competing companies using the same or overlapping genetic markers requiring covariances between molecular scores to be estimated and routinely updated.

Results and discussion

Three critical issues relating to the performance of genomic prediction are: the proportion of variation that can be predicted within-breed from knowledge of the 50k SNP genotypes; the extent to which predictive ability erodes when training knowledge is applied to animals of different breeds; and the ability of a reduced panel to reliably predict performance.

Within-breed predictions from 50k panels. Confidence in genomic predictions can only be provided by validation in a group of animals not included in the training population. Training often involves subdividing the data, say into thirds, and training in two-thirds of the data followed by validation in the other third. Subsets may be chosen so sires do not have sons in both the training and validation datasets. Such training can be done three times for different dataset combinations, so that each bull is represented in one validation set. Garrick (2009) reported results for Angus cattle that vary according to trait and data subset (Table 1), but the general conclusion is that correlations between genomic predictions from 50k SNP and realized performance in independent datasets are 0.5-0.7 accounting for 25-50% genetic variance, equivalent to about 6-16 offspring in a progeny test with heritability of 25%.

Within-breed predictions from reduced SNP panels. The creation of subsets of 600 SNP markers obtained from choosing the 20 markers on each bovine chromosome with the highest model frequency, a measure for marker support, was undertaken to repeat the analyses shown in Table 1 on 600 marker subsets. These data demonstrate relatively little loss of predictive ability in selectively reducing the panel from 50k to 600 SNP.

***Table 1*:** Correlations between 50k or 600 SNP predictions and EPD for backfat (FAT), calving ease direct (CED) and maternal (CEM), carcass marbling (MRB), ribeye area (REA), scrotal circumference (SC), weaning weight direct (WWD) and yearling weight (YWT).

Trait	Train 2 & 3 Predict 1 (50k)	Train 1 & 3 Predict 2 (50k)	Train 2 & 3 Predict 3 (50k)	Overall[1] (50k)	Overall (600 SNP)
FAT	0.71	0.64	0.73	0.69	0.63
CED	0.65	0.47	0.65	0.59	0.61
CEM	0.58	0.56	0.62	0.53	0.55
MRB	0.72	0.73	0.64	0.70	0.67
REA	0.63	0.63	0.60	0.62	0.56
SC	0.60	0.57	0.50	0.55	0.51
WWD	0.65	0.44	0.66	0.52	0.49
YWT	0.69	0.51	0.72	0.56	0.55

[1]Correlation estimated by pooling estimated variances and covariances

Reduced panels of 600 markers per trait are still too many to populate a single 384 SNP panel, particularly to simultaneously target several traits. The resulting estimates of the genetic correlations for 50, 100, 150 or 200 markers were 0.28. 0.29, 0.39 and 0.43 (Woodward *et al*. 2010). The markers that might populate a single 384 SNP multitrait panel were further validated by estimating the correlation between marker score and progeny test performance on a new sample of 275 Angus bulls that were not used in any of the training analyses. The results were estimated genetic correlations of 0.59 for marbling, 0.32 for backfat, 0.58 for ribeye area, 0.44 for carcass weight, 0.39 for heifer pregnancy and 0.35 for yearling weight. Such a panel could account for 10%-35% genetic variation.

Across-breed predictions from 50k panels. The prospect of training in one breed to predict performance in another is appealing. It may not work well if the genes exhibit dominance or epistasis, and allele frequencies vary between populations. Linkage disequilibrium (LD) reflects the ability of alleles at one locus to predict the alleles at another locus. High-density panels would ideally have at least one SNP in high LD with every QTL. However, differences in LD between breeds can lead to a marker being a good surrogate of a causal gene in one breed and less value in another. Few datasets are yet available for across breed validation. Simulated data using some of the 50k loci as if they were causal genes has allowed the prospects for across breed prediction to be quantified (Kizilkaya *et al*. 2010). Those analyses show poor results when LD among the markers on the panel is relied on to predict performance. Further, they show that predictive ability erodes considerably when the number of simulated causal genes is increased. The best-case predictive ability varied from correlations around 0.4 for 50 genes down to 0.2-0.3 for 500 genes. These correlations account for up to 18% genetic variation for 50 genes to <10% variation for 500 genes.

Conclusion

Current studies do not well represent the full range of breeds or environments but do include more traits than those presently available through national cattle evaluation. Fertility traits remain poorly represented. Predictions from 50k SNP panels might account for 50% genetic variation when used in the same breed as the training population and substantially less when used in other breeds. Reduced panels can account for 25%-35% genetic variation for targeted traits. The prospects for modifying selection programs to exploit high-density 50k and/or low-density SNP panels looks encouraging, although less so than simulation results. Future panels can only improve, as further analysis is undertaken on available resource populations. The role of marker tests as a selection tool is now maturing to the extent that they are likely to complement, rather than compete with, national cattle evaluation.

References

Aguilar, I., Misztal, I., Johnson, D.L. *et al*. (2010). *J. Dairy Sci*., 93:743-752.

Amen, T.S., Herring, A.D., Sanders, J.O., *et al*. (2007). *J. Anim. Sci*., 85:365-372.

Brigham, B.W., McCallister, C.M., and Enns, R.M. (2009). http://www.rangebeefcow.com/2009/documents/BrighamEnns2009RBCS_pp.pdf

Esmailizadeh, A.K., Bottema, C.D.K., Sellick, G.S. *et al*. (2008). *J.Anim.Sci*., 86:1038-46.

Garrick, D.J. (2009). http://www.bifconference.com/bif2009/proceedings/G3_pro_Garrick.pdf

Gutierrez-Gil, B., Williams, J.L., Homer, D. *et al*. (2009). *J. Anim. Sci*., 87:24-36.

Kachman, S. (2008). http://www.beefimprovement.org/PDFs/Kansas%20City%20Missouri%202008.pdf

Kizilkaya. K., Fernando, R.L., and Garrick, D.J. (2010). *J. Anim. Sci*., 88:544-551.

MacNeil, M.D., Northcutt, S.L., Schnabel, R.D. *et al*. (2010) In *Proc 10th WCGALP*

Peters, S.O., Kizilkaya, K., Garrick, D.J. *et al*. (2010) http://www.intl-pag.org/18/abstracts/P05k_PAGXVIII_558.html

Reecy, J.M., Tait, R.G., van Overbeke, D.L. *et al*. (2010) In *Proc 10th WCGALP*

Snelling, W.M., Allan,M.F., Keele, J.W. *et al*. (2010). *J. Anim. Sci*., 88:837-848.

Thallman, R.M. (2009) http://animalscience.ucdavis.edu/animalbiotech/Outreach/Whole_Genome_Selection.pdf

Thallman, R.M. Moser, D.W., Dressler, E.W. *et al*. (2003) http://www.beefimprovement.org/proceedings/genetic-prediction-workshop/GPW-CarcassMeritProject-Final.pdf

Woodward, B.W., Nkrumah, D.J., Garrick, D.J. *et al*. (2010) In *Proc 10th WCGALP*