Hereford Genomic Developments

Hereford will be first breed to develop and market its own genomic predictions.

*by **Dorian Garrick,** Iowa State University Lush Chair in animal breeding and genetics and National Beef Cattle Evaluation Consortium executive director*

During the last half century, a wealth of information has been communicated on the use and interpretation of expected progeny differences (EPDs) for selection. If asked, "What is an EPD?" A practical response would be to say it is a measure of the superiority (or inferiority) of a parent in the units of trait measurement (lb. birth weight), assessed in terms of the impact its genes have on the performance of its offspring.

However, there has long been an alternative definition of an EPD. That is, the sum of the values of the gene variants that a parent passes on to its offspring. This has been an academic, not a practical, definition, as other than a few exceptions relating to recessive diseases, we have not known how many genes are responsible for the variation we observe in traits like birth weight, and we have had no idea of the particular values of those gene variants. This situation has now begun to change as we move into the genomic era.

Whereas genetics is the study of inherited characteristics, genomics is the study of the entire genome — all the genetic material — which can now be extensively characterized on an individual animal by sequencing its DNA for a cost of $2,000-$20,000 per animal, depending upon the depth of coverage.

The genome can also be characterized using high-density single nucleotide polymorphism (SNP) arrays to obtain the genotypes at about 700,000 (700k) positions along the genome for about $200 or at 50,000 (50k) positions along the genome for about half that cost. Genotyping using SNP arrays is a service offered by GeneSeek, a company located in Lincoln, Neb. Collectively, these sequencing and genotyping technologies have revolutionized or will revolutionize human medicine, livestock improvement and biological research.

Using these high-density 50k SNP arrays to obtain genotypes on a population of 1,000 or more animals with reliable EPDs, or a few thousand animals with individual phenotypes, we are now in a position to obtain EPDs for the individual chromosome fragments that these genotyped animals are passing on to their offspring.

Hereford training analyses

This process of characterizing the EPDs of genomic fragments is known as a training analysis, and many such analyses have recently been undertaken by Mahdi Saatchi, a post-doctoral researcher at Iowa State University, using American Hereford Association (AHA) records on all the routinely recorded traits. Saatchi's work is funded by the U.S. Department of Agriculture (USDA) through its contribution to the National Beef Cattle Evaluation Consortium (NBCEC), which has a goal of reducing the time it takes for research findings in genetic improvement to be implemented by industry.

The animals that have been used in those AHA 50k-based training analyses include individuals from a number of different sources. First, there were Line 1 Herefords from USDA Miles City, whose genotypes were kindly provided by Mike MacNeil. Second, there were Hereford AI bulls genotyped at the U.S. Meat Animal Research Center (US-MARC) as part of the so-called 2,000 bulls project that included sires from 16 breeds. Third, there were Hereford animals that were genotyped in projects to identify the source of some recessive genetic diseases. Fourth, there were additional Hereford sires genotyped by AHA and GeneSeek to increase the scope of the training population. Fifth, there were sires genotyped as part of the "weight trait project" being championed by Matt Spangler at the University of Nebraska-Lincoln.

In the future, there will also be Hereford animals with feed intake and other performance data genotyped using the 700k panel as part of a multi-institutional USDA-funded project on feed efficiency led by Jerry Taylor from the University of Missouri.

Estimating genomic EPDs

The training analysis estimates the EPDs of small genomic fragments. Once the EPDs of genomic fragments are known, these can be used to estimate genomic EPDs (gEPDs) of other genotyped animals, regardless of their age or sex. This analysis can provide EPDs on young animals at or before puberty, which can markedly increase the accuracy of their information compared to conventional parent average EPDs.

Genomic information would not improve the accuracy of prediction of widely used sires whose offspring

have been individually measured for the trait of interest, as a well-managed progeny test is the gold standard for genetic evaluation. In the future, it is hoped that genomic prediction will be useful to evaluate animals for traits that are not routinely phenotyped in breed associations such as feed efficiency. Such complex traits remain more of a challenge than predicting performance for traits that are relatively cheap and routinely recorded such as growth and ultrasound information.

At present, the accuracies of genomic predictions are quite good in immediate relatives of the training animals but less accurate in distant or unrelated animals, with virtually no predictive power in other breeds. The accuracy varies according to the amount of training data (more being better) but also varies from one trait to another, and it varies according to how closely the animals being considered are related to the training population.

Providing gEPDs on animals in addition to the conventional EPDs would be confusing not only to scientists but also to AI companies, bull breeders and bull buyers. The most appropriate way to communicate gEPDs is to include them in routine national evaluation in just the same way that ultrasound measures would be used to improve the predictions of carcass traits.

In order to incorporate marker information, it is necessary to estimate the genetic correlations between gEPDs and traits, this information summarizing the value of the marker information. These correlations are currently being estimated by Dave Johnston from the Animal Genetics and Breeding Unit (AGBU) in Australia from the information provided in the training analyses so that Johnston and his fellow researchers can develop methods to pool conventional and genomic information for routine application by the Animal Business Research Institute (ABRI) in the Pan-American Cattle Evaluation (PACE) analyses.

The training analyses utilize published EPDs on the training animals and almost always generate very accurate predictions of the training animals themselves, as there are 50,000 chromosome fragments for which EPDs can be obtained, many more than there are observations in the actual training data.

Cross validation

In order to obtain an independent assessment of the accuracy of the predictions cross-validation is used. This involves the following steps. First, the pedigree information is used to sort the animals into groups in such a way that any animal has its close relatives in the same group as itself and more distantly related animals are in one or more of the other groups. Four such groups for AHA animals have been formed.

Second, the training analysis is undertaken using three of the four groups, and the EPDs obtained for the chromosome fragments are then used to obtain gEPDs on the animals in the fourth group that was not included in the training analysis. This process allows scientists to determine the accuracy of prediction in that group by comparing those animals' gEPDs with their published, pedigree and performance-based EPDs.

The process is repeated three more times for each trait so that every animal in every group has a gEPD obtained from training in analyses that excluded its own data. These gEPDs are then used along with the published information to compute the genetic correlations required for pooling the genomic and conventional information. The results of those analyses undertaken at Iowa State University (the official AGBU results are not yet completed) are shown in the table below.

The predictive abilities based on genetic correlations are in column 3 and range from 0.18 to 0.43. At the top end, these values account for 18% genetic variance, much better than the 4% genetic variance which was the best obtained using across-breed predictions of AHA animals from training analyses undertaken in some 3,500 Angus AI bulls (column 4) where across-breed correlations ranged from 0.02 to 0.19. The within-breed predictive ability in Angus, from training on 3,500 AI bulls, is reflected in correlations from 0.51 to 0.80 (column 5) that account for up to two-thirds genetic variance, showing the benefit of increasing the training population size from 1,000 to 3,500 animals.

The Hereford plan

Once a routine system is implemented by AHA so that breeders can send hair samples on their animals to a lab and obtain the resultant genomically enhanced EPDs directly from national cattle evaluation, those additional genotypes will become part of the training data so that the predictions get progressively better. Watch future *Hereford World* issues and

continued on page 172...

EPD trait	gEPD heritability	AHA genetic correlation	Prediction from Angus	AAA genetic correlation
Birth weight	0.94	0.40	0.18	0.64
Weaning weight	0.94	0.34	0.14	0.67
Yearling weight	0.96	0.33	0.17	0.75
Milk	0.91	0.21	0.02	0.51
Calving ease (direct)	0.92	0.33	0.10	0.69
Calving ease (maternal)	0.76	0.18	0.19	0.73
Fat	0.76	0.43	0.07	0.70
Marbling	0.88	0.41	0.16	0.80
Ribeye area	0.89	0.25	0.06	0.75
Scrotal circumference	0.88	0.25	0.03	0.71

...Hereford Genomic Developments continued from page 171

Hereford eNews for the announcement of when the cooperating lab will be ready to accept samples.

Furthermore, Hereford associations in other countries are soon to actively participate in genomic analyses. To date, the Canadian, Uruguayan, Australian and Argentine associations have committed to participating, so EPDs on their 50k genotyped bulls from PACE will be used for across-country validation and then, subsequently, contribute to the number of animals in the next training analysis. Training analyses will be repeated whenever there are sufficient recruitments to the AHA population of 50k genotyped animals.

Research, including that shown in the table on Page 171, has demonstrated that 50k genotypes are not sufficiently dense to enable across-breed evaluations. There is some evidence that 700k genotypes may be sufficient, at least for across-breed prediction of animals of the same type (such as British, Continental or Zebu).

It is not necessary to re-genotype animals that already have 50k genotype; their 700k genotypes can be obtained from relatives with 700k genotypes in a computer analysis known as imputation. Those Hereford animals genotyped with the 700k panel as part of the national feed efficiency project will be used for imputation on all animals that currently have 50k genotypes and the training analysis will be repeated using the actual and imputed 700k genotypes.

We will be particularly interested in across-breed performance of these predictions, hoping that genotyped animals from Angus and other breeds will enhance the accuracy of Hereford predictions.

Nevertheless, in the belief that even higher-density information will be required, we have started DNA sequencing individual bulls to develop methods to impute sequence onto their relatives with 50k genotypes. So far we have only sequenced Angus AI bulls. Imputed sequence will facilitate the discovery of the actual mutations responsible for variation and will enable gene-assisted rather than marker-assisted prediction of gEPDs.

Genomic prediction is already causing a revolution in livestock improvement. In the dairy industry, it has greatly increased the use of young bulls that historically were seldom used prior to obtaining their progeny test results. Genomic prediction is being used for within-line selection in chickens and in pigs.

The AHA represents the first beef cattle breed association to develop and soon to market its own genomic predictions to the benefit of its members. I would hope all Hereford AI sires would, from now on, be routinely genotyped with the 50k panel. This genotyping will ensure that many of the young Hereford animals targeted for genomic prediction will have their sires in the training data.

Such populations will then be available for researchers to use in their endeavors to increase the accuracy of genomic prediction by imputation to higher density genotypes (or sequence) and by pooling of animals across breeds. **HW**