Validation of Marker Tests

Written by Alison Van Eenennaam

Alison Van Eenennaam, PhD
Cooperative Extension Specialist
University of California
Department of Animal Science
One Shields Avenue Ph:(530) 752-7942
Davis, CA 95616 Fax:(530) 752-0175
Email: alvaneenennaam@ucdavis.edu

Once an association has been found between a DNA marker and a trait in a discovery population, that association needs to be validated in a different population that is representative of the population where the test will ultimately be used. Evaluating genetic tests on validation populations provides data that enables an objective assessment of the genotyping company's published claims. The U.S. National Beef Cattle Evaluation Consortium (NBCEC) has been involved in the process of independently validating commercial DNA tests for quantitative beef quality traits since their first appearance on the U.S. market in the early 2000s (validation results are posted at www.NBCEC.org). The term "having validated" was originally defined as finding a significant association "between genetic tests and traits as claimed by the commercial genotyping company based on phenotypes and genotypes derived from reference cattle populations".

Summary of NBCEC validations for commercially-available DNA-tests for complex (quantitative or multigenic) traits in beef cattle (note: validations do not include tests for "simple" traits such as coat color, horned/polled, AM status etc.)

Company	Test Name	Trait	Date of validation
Igenity www.igenity.com	Profile®	Fat Thickness	12/2008
	Profile®	Marbling Score	12/2008
	Profile®	Quality Grade (% ≥ Choice)	12/2008
	Profile®	Rib Eye Area	12/2008
	Profile®	Yield Grade	12/2008
	Profile®	Average Daily Gain	12/2008
	Profile®	Tenderness	12/2007
	Profile®	Residual Feed Intake (RFI) (for *Bos indicus* influenced cattle)	12/2007
	Profile®	Residual Feed Intake (RFI) (for *Bos taurus* cattle)	6/2008
	Profile®	Dry matter intake (DMI) (for *Bos indicus* influenced cattle)	12/2007
	Profile®	Heifer Pregnancy Rate	
	Profile®	Stayability (longevity)	
	Profile®	Maternal Calving Ease	
	Profile®	Docility	
Pfizer Animal Genetics (Bovigen) www.bovigen.com	GeneSTAR® Tenderness MVP	Tenderness	2/2009
	GeneSTAR® Marbling MVP	% IMF (Feedlot cattle)	2/2009
	GeneSTAR® Feed Efficiency MVP	Net Feed Intake (NFI)	2/2009
MMI genomics www.metamorphixinc.com	Tru-Marbling™	Marbling Score and Quality Grade	
	Tru-Tenderness™	Tenderness	

During the past decade, the DNA testing industry matured from single gene tests to panels involving an ever-increasing number of markers with purported effects on multiple traits and/or in specific cattle populations. As marker panels grew in size and there were increasing intellectual property concerns regarding disclosure of the specific marker loci involved in a genetic test, validation moved from testing the effect of individual loci towards testing a single marker score, or MBV. The validation data analysis shifted to a determination of whether the regression of phenotype on marker score for a single trait model (in which the marker score was a covariate) differed from zero.

The NBCEC and DNA testing companies have struggled to find appropriately-phenotyped populations that were not involved in the discovery process for validation studies. Additionally, results from different validation populations genotyped with the same SNP panel were often inconsistent with respect to the significance of the association between the test and the trait(s), and sometimes even with respect to the direction of the association. This complicated the interpretation of validation results, and created confusion as to whether "validation" meant a test "worked" (i.e. was significantly associated with the trait) in one or more of the test populations, or had simply been tested by an independent third party.

At the current time the data that is reported on the NBCEC validation website includes the direction of the effect (regression coefficient), and the significance ("*p*" value; associations are typically considered significant if $p < 0.05$) of that effect. A positive regression coefficient means that the test was associated with the trait in a positive way, i.e. one unit of test increase was associated with an increase of (1 x regression coefficient) unit of the trait.

Example. If two animals have a DNA-based tenderness score that differs by 2 units and the regression coefficient of phenotype on the genetic score is 0.3, then it would be predicted that there would be a (2 x .3) = 0.6 lb difference in Warner Bratzler Shear force between steaks derived from these two animals.

Knowing that a test has a regression coefficient of 0.26 (±0.3), and a *p* value of 0.001 does not really help to inform genetic selection decisions. While reassuring to see the regression coefficient is not zero or negative, the significance of this result does not provide useful information regarding the test's value. A test that has a significant association with the trait of interest may nonetheless explain only a minor proportion of the genetic variance

The validation process is evolving from simply reporting the finding of a significant association between DNA test results and the trait of interest, towards an independent calibration approach that estimates the parameters that will be required to facilitate the incorporation of DNA test-based predictions of genetic merit into national genetic evaluations. Currently such details (e.g. proportion of genetic variation accounted for by a DNA test panel) are not reported on the NBCEC validation site, although they will be reported for all future validations. This will assist with plans to develop marker-assisted EPDs with an associated accuracy. Such an approach is appealing as it presents results in a format (i.e. EPDs and accuracies) that is familiar to producers, and it eliminates the choice that is implicitly associated with the current practice of publishing traditional EPDs and marker information separately.