Impact of genomics on dairy cattle breeding

Bennet Cassell
Professor Emeritus
Department of Dairy Science
Virginia Tech

My purpose in this brief proceedings paper is to share basic information about the ongoing genomics revolution in dairy cattle breeding with members of the dairy nutrition community. This is not the typical topic for an annual nutrition conference, but seems appropriate due to the impact genomics are and will continue to have on genetic changes in our dairy populations. There will be important "spillover" effects in dairy nutrition as the science of genomics begins to mature a bit.

Recent developments in affordable genome-wide SNP testing and genomic prediction systems have changed some long-held views and assumptions about dairy cattle breeding, including many I held dear for most of my career. By the way, SNP stands for "single nucleotide polymorphism". A SNP is a site on a chromosome where animals in a population have different nucleic acids. If all animals have the same nucleic acid at a site, then there is no variation at that site to associate with performance characteristics and the site is considered "uninformative".

What is genomics?

My definition of genomics is the study of the role of nucleic acid sequences in cellular DNA that are associated with animal (including human) performance or trait expression. For dairy cattle, one important current use of genome sequences is to predict genetic merit of dairy animals for economically important traits. Other uses include recognition of "mendelian" traits, such as polled-horned, coat color, and some genetic recessives formerly discovered through pedigree relationships. SNP sequences can be available at a very young age, and, for traditional dairy traits like yield and conformation, sufficiently accurate for selection of males to use as sires of herd replacements or females to serve as bull mothers. The fundamental change in long held beliefs about dairy cattle breeding is that producers can make very accurate selection decisions without waiting for progeny test results on males or lactation records on females.

Genomic information comes from tests based on tissue samples from individuals. The procedure tests for specific nucleic acids at uniformly distributed sites across the 29 autosomal (non-sex) chromosomes in dairy cattle, as well as the sex chromosomes. The initial test, called the 50K chip, involved about 50,000 individual sites, and was expensive enough at about $250 to limit use to elite dairy animals only. The critical step in converting those 50K results to genetic predictions involved AI bulls with many progeny from yesteryear. Cooperating bull studs donated semen held in a national germ plasm bank to the original study, which included the Animal Improvement Programs Laboratory and the Functional Bovine Genome Laboratory at USDA along with the chip manufacturer (Illuminia) and others. The story of the science behind genome test chips is a good one, but not my objective here.

Turning genomic tests into genetic predictions

AIPL used the progeny test and genomic test results on old AI bulls to develop prediction equations. Each informative SNP site (somewhat less than the 50K sites tested) was included in a massive

multiple regression prediction equation. The genomic predictions were about as accurate in Holsteins as a progeny test based on 20 daughters in different herds. Accuracy was lower for Jersey and Brown Swiss breeds because fewer AI bulls with large progeny groups were available to develop the prediction equations. AIPL next developed a procedure to combine genomic predictions with traditional genetic evaluations based on performance records.

AIPL scientists developed a test to see if the predictions including genomic information actually worked. To a dairy farmer, the ultimate test would be to see if a bull chosen on the basis of a high genetic evaluation including genomics but no daughter records actually was that good when progeny data were available. Rather than choosing some young bulls and waiting five years to find the answer, the scientists used a short cut. They found over 2000 genomically-tested Holstein with "first crop" daughters only in 2009. Genetic evaluations (including the genomics predictions) were calculated using 2009 data for a larger population which included the young bulls and over 4,000 older bulls with "second crop" daughters. Some cows were included in the study in the Jersey and Brown Swiss breeds. Then, the scientists backed to 2004, by removing performance records with calving dates during that five year period. Genetic evaluations were calculated on that five-year-old data set, which is called a "discovery population." Predictions of genetic merit on the 2000 young Holstein bulls (and smaller numbers of Jersey and Brown Swiss bulls) were made using the 2004 prediction equations. The scientists could then compare genomic predictions from 2004 data with those from 2009 on those younger AI bulls. The test is equivalent to a farmer using a young bull and checking on how well the proof held up five years later.

Details of results of that first test are in the accompanying set of PowerPoint slides. In brief summary, Reliability of combined genomic/traditional genetic evaluations (now called gNM$) in Holsteins increased about 24 points to 59% for Net Merit $ compared to a Reliability of about 35% on traditional Parent Average of young bulls. That was a critical improvement in accuracy of selection of young animals, and results primarily from the ability of genomic tests to reveal important information about the gene sample young bulls inherit from parents. Parent Average assumes full sibs, for instance, are genetically the same, when they actually share only about 50% genes in common through common parents.

Improvements in today's predictions, three years removed from those original tests, increase Reliability of Holstein gNM$ to over 70%. Prior to genomics, a proven bull, about five years old with first crop progeny, would have initial Reliability estimates of about 80%, increasing to about 85% Reliability as first crop daughters matured. Waiting another three to five years to increase Reliability from 70 to 85% is a questionable decision on young bulls with some pretty fantastic genomic predictions. Dairy farmers can be milking daughters of those bulls while other producers are waiting for more information.

Scientists and AIPL and elsewhere were convinced, as were many in the dairy cattle breeding industry. The genetic improvement "industry" endorsed official status for gPTAs, and in January 2009, the first official gPTAs were published. For clarification, proofs based on pedigree and genomic information only but with no progeny were designated as "gPTA's". Proofs that included both genomics and progeny data retained the PTA designation used for years.

Do genomic proofs really hold up?

The AI organizations varied in how they used genomically tested young bulls in marketing programs. Some actively marketed large numbers of such bulls, while others continued to promote only the progeny tested bulls. Producer response varied as well. I heard one producer say that genomic predictions were good for scientists, but not for him. Other producers, the risk-taking early adopters, embraced gPTAs immediately, choosing to use young bulls with the very highest proofs more heavily than they had used young bulls in the past. Daughters of those first selection decisions showed up in calf pens in late 2009 and early 2010. Today, late winter of 2012, those calves are entering the lactating herd. Data on the success of those genomic selection decisions are beginning to emerge.

I saw an article in an AI trade magazine in January 2012 that examined the very first group of AI young sires that were marketed on the basis on genomic predictions. This particular study used December 2011 proofs on over 2400 Holsteins bulls that have, on average, 104 milking daughters each. The average change in NM$ was a drop of $72 from the January 2009 gPTA. Milk proofs dropped 148 lbs, protein dropped 3 lbs, and daughter pregnancy rate dropped 0.1%. I call such results successful. Producers who used the best of those 2400 bulls back in 2009 are milking daughters while other farmers are just beginning to make matings. The changes shown above won't re-rank the average bull very much compared to alternative, proven bulls available in 2009. However, somebody averse to adopting genomic predictions can surely find some favorite examples to "prove" that we're all going to hell in a hand basket with these newfangled predictions.

The drop in average proof for productive life (PL) for the genomically tested young bulls was -0.8 mos, which does bother me some. However, even the size of that drop is difficult to judge unless you know the changes in PL that occurred in proofs of progeny tested bulls back in 2009. PL is notoriously difficult to predict as we have to wait for progeny to die to find out the answer. Poker players know that risk is always a little tough to measure on the fly in a card game. But players get feedback right away in a poker game. In dairy cattle breeding, there is at least as much opinion about the results of a risky decision as there is in judging the original risk.

I'm convinced that dairy farmers should use genomic predictions for sire selection. They should use high-ranking young bulls more heavily than they did five years ago, when 10-20 straws in a 100 cow herd was about all a producer could risk based on PA. I'm thinking that 100-cow herd could invest in 10-20 DAUGHTERS of the top genomic young sires! Dairy farmers should use the very best this heavily. And they should continue to use outstanding proven bulls – those with daughters – if the PTA's rank such bulls up there with the genomically tested youngsters.

Genomic tests for cows

That original 50K test that once cost about $250 per animal was too expensive to use on a large number of cows just to increase Reliability from 35 to 70%. As with most technical developments, costs go down over time, and that certainly is the case for this technology. Today's 50K tests cost around $150, which still eliminates genomic testing for routine screening of animals in dairy herds. A couple of years ago, a 3K test was introduced which promised to expand the impact of genomic testing to larger numbers of animals. However, the lower density 3K chip did not increase Reliability of gPTAs as much as the 50K chip did (see PowerPoint slides for details). In late summer of 2011, a new 6K test kit was unveiled that remains less expensive, but doesn't lose so much accuracy in genomic predictions. The 6K tests, available through breed organizations, cost in the neighborhood

of $30-$40 per animal.   Reliabilities for the various traits with the 6K test generally are within 2% points of the 50K test. Again, the early-adopters are embracing this new technology by routinely testing many females in their herds. And the bank of genomically tested animals in the AIPL research files grows by leaps and bounds. In December 2010, prior to 6L chips, 18,000 Holstein heifers were in AIPL files. By December 2011, that number had increased to over 68,000 Holstein heifers. The heifer file grows by about 4000 new animals per month.

We began to run 6K tests on all Jersey and Holstein heifers at Virginia Tech this past December. Our objective was to build a data set of genomic sequences on animals for research purposes. However, we also wanted to provide some leadership to the dairy industry in Virginia by using our research herd to show farmers how such tests might improve selection and management decisions. The first group of 24 test results for Jerseys was available as I wrote this piece. On the basis of those results, we chose to remove one heifer from our future plans. She was an unlucky girl who got a bad gene sample from both sides of her pedigree. We had half a dozen heifers in that group with genomic predictions above $400 NM$, which is pretty lofty territory for Jerseys. They are promising heifers, but the Jersey breed currently includes quite a few heifers better than they are. We learned this by examining a web site at AIPL that includes the most current "elite" genomically tested heifers in the Jersey breed. We'll make similar judgments as our first Holstein results come in. Maybe we'll find a truly elite animal in that group.

Imputation

Imputation is a new word, to me at least, generated by genomics technology. It refers to a process where genomic sequences based on high density genome scans on parents can be used to fill in the blanks on lower density scans on progeny.

The system is based on "haplotype" transmission from parents to offspring. A haplotype is a relatively short strand of single-stranded DNA passed to progeny via egg or sperm cells. You recall from high school biology classes that double-stranded DNA is divided into single strands in the formation of sperm and eggs. Some new genetic variation is introduced into the process by "crossovers", whereby part of a segment from strand breaks and is connected to the same segment location on the other strand to complete division. Crossovers occur an average of once per chromosome per cell division in the formation of sex cells. Sometimes no crossovers occur, while at other times two, three, and (rarely) more crossovers happen. These events generate new haplotypes, which are detected in genetic evaluation procedures as new progeny are genomically tested.

Without going too deeply into this absolutely amazing science, we use records of haplotypes on genomically tested parents (mostly bulls) to fill in the blanks on 6K SNP tests run on progeny. If the 6K test suggests a previously identified haplotype was passed from sire to daughter, then the entire haplotype sequence from the 50K test on the sire can be assumed to have been transmitted to the daughter based on her 6K test. Scientists "impute" the missing information on this basis. SNP sequences on maternal grandsires also help, but help most if the dam of the tested individual also has SNP sequences available. Under good conditions (pardon the simplification) imputation is about 98% accurate in Holsteins, Jerseys, and Brown Swiss animals with 6K genomic tests.

The future

Genomic predictions, if used aggressively, can the rate of genetic progress by 2.5 times over that achieved through traditional progeny tests. Not everyone is or ever will be aggressive about genomics, so an increase of 50% is more realistic. Rapid genetic change, particularly in some of the difficult, lowly heritable traits like fertility, may well leave us with a very different dairy cow within a few years. I expect greater progress for health and fitness traits, and increasing interest in "managing" challenging problems such as metabolic nutritional issues through genetic means.

We have not yet used the "whole genome" information to predict merit for traits outside of the traditional traits evaluated by USDA or breed association type traits, but things will certainly change in the future. Virginia Tech is involved in a large genomics study to attempt to identify Holstein animals with genetically superior feed efficiency. We haven't practiced direct selection on this most economically important trait because it's just so expensive to measure. However, reasonably accurate genomic predictions for feed efficiency can be developed, then a periodically updated "discover population" can give us a means to change this trait without measuring thousands of animals. Nutritional questions are wide open for genomics study. We may well see the day when genomic tests help us group lactating animals by nutritional needs. High feed costs certainly encourage managers to put feed resources where they will do the most good.

I expect use of the less expensive 6K genomic tests on heifers to become routine in many herds. Commercial breeders may want to cull the lowest ranking heifers and not make that investment in rearing costs that is likely to be followed by early culling in the milking herd. Purebred breeders will want to consider such tests to screen for exceptional animals which they might then test further with 50K chips. Even the good but not exceptional animals will likely have added value because of 6K or future inexpensive tests. An increasing number of top consignment sales will expect genomic predictions on animals included in those sales. What about dairy farmers who use herd bulls to breed their cows? Genomic tests will be allowed on dairy bulls without bull stud sponsorship beginning in March 2013. Bull buyers who wish to use more powerful selection tools will be able to do so. Farmers who market such bulls would have justification to increase sale prices for the better genomically tested bulls.

I'm excited about the opportunities genomic technologies offer to dairy cattle breeding and dairy herd management. The story is just beginning to unfold. We should keep our eyes – and minds – open to the possibilities.