Practical applications of genomics

Troy N. Rowan, PhD
University of Tennessee Institute of Agriculture
Knoxville, TN 37996

Abstract

Genomic testing serves multiple functions for beef cattle populations. Since their inception in the late 2000s, genomic tests have been critical tools for detecting mutations that cause genetic defects, parentage testing, and improving the quality of genetic predictions. Sire selection is one of the most consequential decisions that commercial operations make. This challenge is compounded by the relatively low levels of information available on the unproven herd sires that commercial herds purchase. Selection tools in the form of Expected Progeny Differences (EPDs) allow commercial operations to reduce the risk involved in sire selection. EPDs isolate the genetic component of economically-relevant traits and increase selection accuracy. While parent average EPDs are more reliable than phenotypic selection, unproven sires can still experience significant EPD shifts as more information becomes available. Genomic testing enhances the EPDs for unproven and low-accuracy sires. These increases in accuracy are equivalent to approximately a calf crop's worth of records in a genomic evaluation (from 10 to 30 phenotypes depending on the trait). Nearly all breed associations calculate genomically-enhanced EPDs, making their use imperative for commercial operations. Beyond bull selection, opportunities exist for commercial producers to utilize genomics for heifer selection and marketing terminal animals.

Key words: genomics, beef, selection, breeding

Introduction

Genomic testing is an incredibly useful tool for the beef industry. Genomics tests are the most recent development in a long line of selection tools and methods for tracking the inheritance of alleles. Previous genetic tests focused on genotyping one or a handful of polymorphisms in or near Mendelian-acting loci. Modern genomic tests are microarray-based chips that genotype thousands (usually between 30,000 and 100,000) of known single nucleotide polymorphisms (SNPs) in cattle. This method for assaying a high density of genotypes presents an unprecedented opportunity to monitor genetic defects, track parentage, and enhance the quality of breeding values.[1] The magnitude of SNPs on genomic assays enables us to track inheritance and more accurately represent relationships between animals in genetic predictions. Since the inception of genotyping arrays, implementation and adoption have been rapid across the beef and dairy industries.[2] Additionally, genomic tests allow us to capture many known mutations for testing genetic defects at once. These proceedings discuss genomic testing and its importance for commercial and seedstock beef operations. It explores how genomic data is generated and how seedstock and commercial producers can use it to accelerate genetic progress in their herds.

Generating genomic data

Genomic data generation requires tissue collection from the animal of interest for DNA extraction. Testing companies routinely extract DNA from blood (typically blotted on an FTA card), hair follicles and punches of ear tissue. Samples can be collected at any point in an animal's life as an animal's genomic DNA remains unchanged over time. Upon tissue collection and DNA extraction, genotypes are generated either by high-density microarray panels (SNP chips) or, in some cases, through next-generation sequencing at high or low coverages.[3] Genotyping chips vary in content but typically contain between 20,000 and 100,000 SNPs. Most SNPs are intentionally chosen to represent evenly spaced common biallelic variants across the bovine genome. This spacing allows for the tracking of chunks of DNA inherited together for use in genetic prediction applications. Additional variants known to be causal or perfectly linked to causal alleles for genetic defects and economically relevant Mendelian mutations (i.e., coat color, horned/polled) are also included on genomic tests. Without reviewing the full suite of genetic defects, it is important to remember that genomic tests developed for genetic evaluations include many or all of these known defect-associated loci. This prevents the need to perform multiple genetic tests on animals in breeds with multiple defects. Genomic tests also include markers that are used to assign parentage.[4] Parentage verification is essential for the integrity of breed registries and EPD calculations.

Genetic prediction and accuracy

Genomic tests play a central role in modern genetic evaluations. Since the true genetic merit of an animal can never truly be known, we have long relied on statistical estimates of animals' genetic potential (i.e., breeding value) to make more accurate selection decisions. Methods for estimating breeding values have been evolving for decades. These are built on the statistical framework of Best Linear Unbiased Prediction (BLUP), which removes variation in phenotypes due to environmental variation using contemporary groups, and then delivers estimates of an animal's genetic merit.[5] After removing environmental variation from a phenotype, BLUP attempts to determine which genetics an animal inherited from its parents. In highly related livestock populations, this random sampling of genes, also known as Mendelian Sampling, can generate large amounts of variation, even between siblings. Theoretically, we would expect that the random inheritance of genes can account for up to 50% of the genetic variation in a trait.[6] This sampling can result in significant variations in genetic merit, even between full siblings (Figure 1). The better the estimate of this Mendelian Sampling term, the more accurate the EPD. Theoretically, accuracy is the correlation between an animal's true and estimated breeding value (i.e., how often are we selecting the "right" animal with our tool). Multiple metrics exist to represent EPD accuracy, but the most common in beef cattle populations is Beef Improvement Federation (BIF) accuracy. BIF accuracy ranges from 0 (no information/confidence in EPD) to 1 (EPD perfectly represents actual genetic merit).[7] Higher accuracy values indicate a greater amount of confidence in the EPD.

Figure 1: A toy example of Mendelian Sampling in a three-generation pedigree of cattle. Bulls in the bottom row are full siblings but share different proportions of each grandparent's DNA.
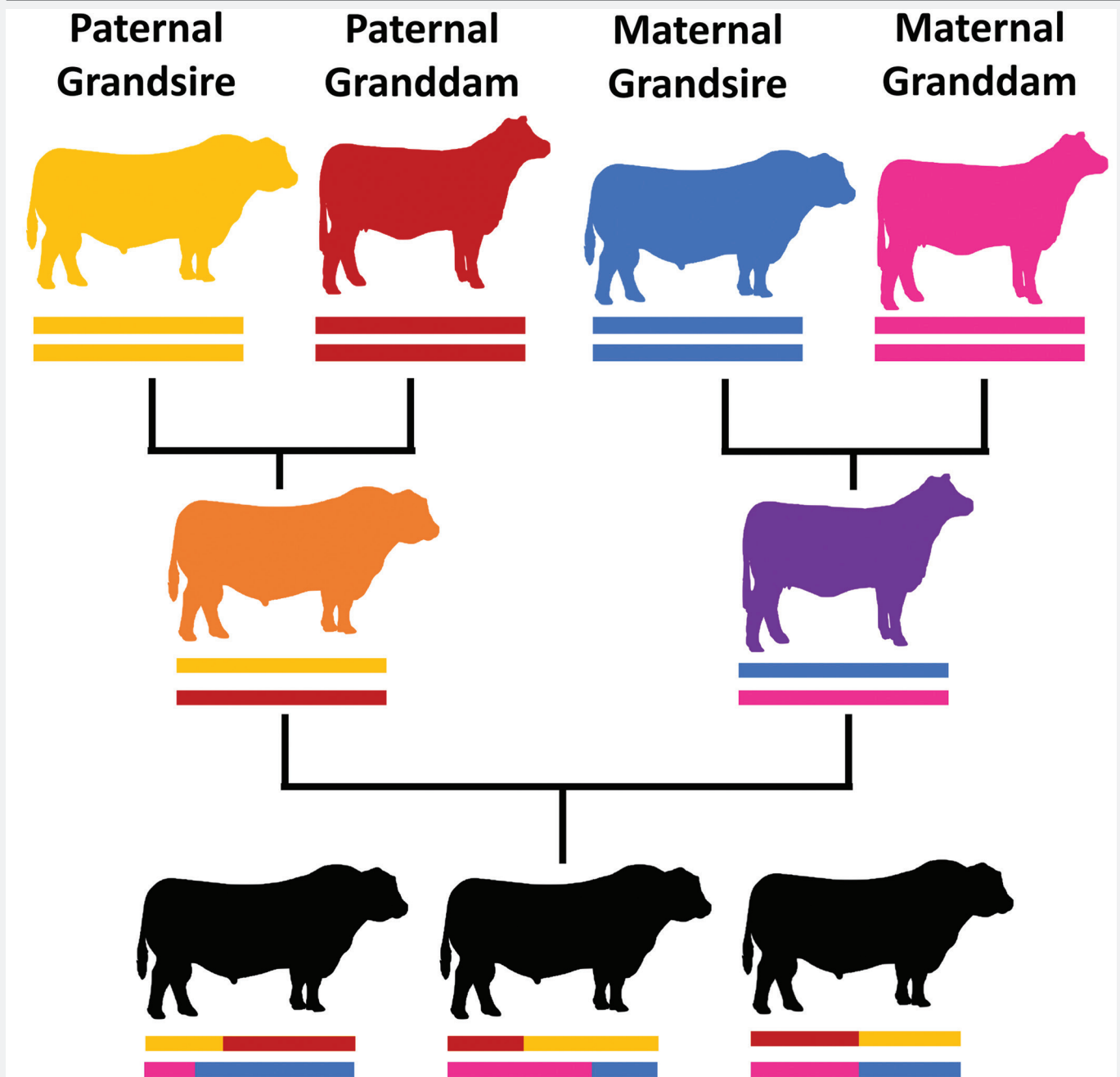
EPDs rely on multiple sources of information to become more accurate:

1. **Contemporary groups** are used to control environmental variation.
2. **Performance records** from the animal itself tell us about its possible genetic content.
3. **Progeny phenotypes** help us become more confident in the genetics that a sire inherited.
4. **Relationships** between the animal and the rest of the population allow us to borrow information from phenotyped relatives.

More of any of these records will increase the accuracy of an EPD. This allows some sires, especially those used in artificial insemination programs to experience increased accuracies over the course of their lifetimes. This accuracy increase is due to genetic evaluations becoming more certain about sires' genetic merit as more calf records enter. While seedstock sires benefit from these accuracy increases from performance records, sires purchased by commercial operations rarely ever have progeny data enter evaluations. Further, even if this were possible, the accuracy increases that come with calf crops would not arrive in time to aid in the selection process.

Genomic prediction

When selecting between unproven sires, genomics can increase our confidence in genetic predictions without adding progeny records. This accuracy increase reduces the risk associated with purchasing an unproven bull. Simply put, the markers contained in genomic tests allow us to observe the DNA that animals inherited from each parent directly. Genomic predictions can use the markers from a test in 2 main ways: 1) To use genotypes to represent relationships between individuals in the population better or 2) To estimate marker effects and calculate a "molecular breeding value".[8]

Historically, pedigree relationships between individuals were used for determining how to share phenotypic information from related animals in a genetic evaluation. These pedigree relationships used expected coefficients of relationship (i.e., 0.25 between half-sibs or grandparents, 0.125 between first cousins, etc.). This method for calculating relationships ignored that due to Mendelian Sampling, animals' DNA sharing can deviate from our pedigree expectations. For example, despite inheriting 50% of DNA from each parent, full siblings are likely to have different proportions of each grandparent's genetics. The integration of genomics allows genetic evaluations to use actual DNA sharing between individuals to represent relationships. The use of genomics in EPD calculation is called Genomic BLUP (GBLUP). Recent developments have determined ways to integrate both pedigree and genomic relationships so that genotyped animals can borrow information from ungenotyped animals (single-step GBLUP or ssGBLUP),[9] further improving the accuracy of EPDs.

The other strategy for leveraging genomics in EPD calculation is estimating the effects of individual DNA markers on a given trait. Because animals inherit DNA in large chunks, we can use markers to estimate the amount of genetic potential added by a haplotype (a variable-length contiguous piece of DNA). This allows "marker effects" to be calculated from each marker on the assay. Once estimated, these values can be summed together for an animal's "molecular breeding value." Strategies exist for either combining this information into a genetic evaluation[10] or utilizing the molecular breeding values independently.

Regardless of the methodology used to calculate EPDs, the results are largely the same. Depending on the trait and the genetic evaluation, a genomic test can provide accuracy increases similar in magnitude to an entire calf crop's worth of data entering a genetic evaluation (Table 1). These increases in accuracy are invaluable for traits like calving ease, where increased confidence in an EPD is critical for making mating decisions for heifers. Additionally, this increased accuracy of EPDs from genomic tests helps mitigate the risk of purchasing a young, unproven bull.

For seedstock producers utilizing genomics in their herds, they need to remember that genomic testing does not supersede the need to collect phenotypes. Like their pedigree-based predecessors, genomic evaluations rely on phenotypes of related animals and progeny to train statistical models. Without the continued collection of phenotypes, training data breaks down in its ability to predict the merit of future animals.

Commercial applications of genomics

While EPDs and Genomically-Enhanced EPDs (GE-EPDs) are calculated only for purebred and seedstock populations, genomics can play a significant role in commercial cow herds. The primary way that commercial producers can take advantage of genomics is by only purchasing bulls with GE-EPDs.

Table 1: Progeny Equivalents for genomic tests in American Angus Association genetic evaluation. Progeny equivalents are the number of records required in a genetic evaluation to generate the same accuracy increase for an EPD that a genomic test does.

Trait (Abbreviation)	Progeny equivalent
Calving ease direct (CED)	28
Birth weight (BW)	23
Weaning weight (WW)	27
Yearling weight (YW)	22
Dry matter intake (DMI)	11
Yearling height (YH)	15
Scrotal circumference (SC)	13
Docility (DOC)	11
Heifer pregnancy (HP)	17
Calving ease maternal (CEM)	18
Milk (MILK)	35
Mature weight (MW)	14
Carcass fat (FAT)	13
Carcass weight (CW)	4
Carcass marbling (MARB)	10
Carcass ribeye area (REA)	15

Genomic testing costs have declined over time, making it necessary for seedstock producers to provide for their customers. The increased confidence in EPD-based selection can have far-reaching impacts on commercial herds, particularly those that keep replacement females.

Other opportunities also exist for commercial herds to use genomics for selecting replacement heifers. Commercial genomic tests use marker-effect models to calculate molecular breeding values. The results of these tests are not as accurate as the EPDs from purebred genetic evaluations, but they are significantly more accurate than using only phenotypes for selection.[11] They also offer opportunities to get predictions for traits that would not have been possible to observe in female offspring until much later in their lives (i.e., stayability, milk and other maternal traits). It is important to remember that for genomics to be truly useful, commercial producers should test far more animals than they expect to keep. Results from genomic tests must be used to drive selection decisions to observe an impact on genetic progress. Commercial operations can also use results from testing their entire heifer crop to help market their steers and non-replacement females. We would expect the average genetic merit for terminal traits to be the same between a herd's heifers and its steers. Predictions for feedlot performance and carcass traits for the heifer crop could inform buyers of the likely genetic potential of the steer crop. Opportunities to apply genomics in commercial operations will only grow as genomic testing becomes more accessible and cheaper.

Conclusions

Genomic testing is a critical tool for beef cattle genetic improvement. When used to enhance EPDs, producers can reduce the risk when purchasing low-accuracy and unproven bulls. Genomics allows us to track DNA sharing between relatives directly. When integrated into genetic evaluations, this provides substantial increases in accuracy compared with pedigree-based predictions. In addition to enhancing breeding values, genomic tests allow seedstock producers to test carrier status for multiple genetic defects at once. Markers for parentage verification are also included on genomic tests, making their value to seedstock operations enormous. The easiest way for commercial producers to utilize genomics in their herd is by purchasing bulls with genomically-enhanced EPDs. Commercial genomic tests are becoming increasingly popular and affordable. These allow commercial producers to make more accurate replacement female decisions and may also help add value to terminal animals.

References

1. Georges M, Charlier C, Hayes B. Harnessing genomic information for livestock improvement. *Nat Rev Genet* 2018. Available at: **http://dx.doi.org/10.1038/s41576-018-0082-2**.

2. García-Ruiz A, Cole JB, VanRaden PM, et al. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc Natl Acad Sci USA* 2016;113:E3995-4004.

3. Snelling WM, Hoff JL, Li JH, et al. Assessment of Imputation from low-pass sequencing to predict merit of beef steers. *Genes* 2020;11. Available at: **http://dx.doi.org/10.3390/genes11111312**.

4. McClure MC, McCarthy J, Flynn P, et al. SNP data quality control in a national beef and dairy cattle system and highly accurate SNP based parentage verification and identification. *Front Genet* 2018;9:84.

5. Henderson CR. Best linear unbiased prediction of breeding values not in the model for records. *J Dairy Sci* 1977;60:783–787.

6. Hill WG, Mackay TFC. D. S. Falconer and Introduction to quantitative genetics. *Genetics* 2004;167:1529–1536.

7. Meyer K. Approximate accuracy of genetic evaluation under an animal model. *Livest Prod Sci* 1989;21:87–100.

8. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 2001;157:1819–1829.

9. Misztal I, Legarra A, Aguilar I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J Dairy Sci* 2009;92:4648–4655.

10. Fernando RL, Dekkers JC, Garrick DJ. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genet Sel Evol* 2014;46:50.

11. Arisman BC, Rowan TN, Thomas JM, et al. Evaluation of Zoetis GeneMax Advantage genomic predictions in commercial *Bos taurus* Angus cattle. *bioRxiv* 2020.