

Genomics for the Rancher: How does it work and what does it mean?

Jack C. Whittier, Extension Beef Specialist
Colorado State University

Introduction

I am continually amazed at the tools we have in today's world to assist in making better and better decisions. This is true in most aspects of our lives – from communications to computers; from entertainment to eating; from politics to travel; and many, many more components of what we do and how we live. The digital age has created an information explosion, and livestock production has not been left out of this progress.

Another amazing aspect of today's world is the vast knowledge being uncovered and refined in the world of biology and science. Gregor Mendel, an Austrian friar, now known as the "father of modern genetics", studied variation in plants in 1830. Mendel's work later led to Mendel's Laws of Inheritance which are key components of modern genetic predictions today. Deoxyribonucleic acid or DNA was first isolated by the Swiss physician Friedrich Miescher who, in 1869, discovered a microscopic substance in the pus of discarded surgical bandages that later was characterized and determined to be integral to the mechanisms of inheritance. James D. Watson and Francis Crick proposed the double helix or spiral staircase structure of the DNA molecule in 1953. This led to their being awarded a Nobel Prize in 1962. Today, information from DNA is used widely in forensic investigations, genetic counseling in humans, study of disease susceptibility, and in genetic merit of cattle.

DNA as an Information Source

DNA is the genetic material of most living organisms, and is the major constituent of chromosomes within the cell nucleus of all cells. DNA plays a central role in the determination of inheritance characteristics in organisms by controlling protein synthesis in cells. DNA is a specific type of a chemical molecule called nucleic acid and is composed of two chains called nucleotides in which the sugar called deoxyribose makes up the backbone chain and four chemical bases (adenine, cytosine, guanine, and thymine) provide cross bars between the backbone chains. The two chains are wound round each other and linked together by hydrogen bonds between specific complementary bases to form a spiral ladder-shaped molecule.

The sequence of bases in DNA forms the genetic code. A group of three bases (a triplet) is the key signal for the production of a particular amino acid in the cytoplasm of the cell. A sequence of triplets in the DNA molecule may code for a complete protein. The different amino acids, and the order in which they are joined up, determine the protein being produced. The sequence of the triplets forms a gene. There may be thousands of base pairs in any one gene.

The Terminology of Genomics

Understanding a few key terms will be helpful in understanding how DNA acts as an information source. Table 1 contains some of the common terms used to describe the modern science of genetic prediction.

Table 1. Common terminology used in the science of genomics for genetic prediction. (Definition sources are primarily from Wikipedia, the free on-line encyclopedia at http://en.wikipedia.org/wiki/Main_Page)

| | |
|-----------------------------|--|
| Genome | In modern molecular biology and genetics, the genome is the entirety of an organism's hereditary information. The genome includes both the genes and the non-coding sequences of the DNA/RNA. |
| DNA – Deoxyribonucleic acid | DNA is a specific type of a chemical molecule called nucleic acid and is composed of two chains called nucleotides in which the sugar called deoxyribose makes up the backbone chain and four chemical bases (adenine, cytosine, guanine, and thymine; abbreviated as A, C, G, T) provide cross bars between the backbone chains. |
| RNA – Ribonucleic acid | RNA is a biologically important type of molecule that consists of a long chain of nucleotide units. Each nucleotide consists of a nitrogenous base, a ribose sugar, and a phosphate. RNA is very similar to DNA, but differs in a few important structural details: in the cell, RNA is usually single-stranded, while DNA is usually double-stranded; RNA nucleotides contain ribose while DNA contains deoxyribose (a type of ribose that lacks one oxygen atom); and RNA has the base uracil (U) rather than thymine (T) that is present in DNA. |
| Chromosomes | A chromosome is an organized structure of DNA and protein found in cells. It is a single piece of coiled DNA containing many genes, regulatory elements and other nucleotide sequences. Chromosomes also contain DNA-bound proteins, which serve to package the DNA and control its functions. Cattle have 30 chromosome pairs. |
| Allele | An allele is one of two or more forms of a gene or a genetic locus (generally a group of genes). "Allele" is an abbreviation of allelomorph. Sometimes, different alleles can result in different observable phenotypic traits, such as different pigmentation. However, many variations at the genetic level result in little or no observable variation. |
| Genes | A gene is a molecular unit of heredity of a living organism. It is a name given to some stretches of DNA and RNA that code for a type of protein or for an RNA chain that has a function in the organism. Genes hold the information to build and maintain an organism's cells and pass genetic traits to offspring. |
| Quantitative traits | Quantitative traits refer to phenotypes (characteristics) that vary in degree and can be attributed to polygenic effects, i.e., product of two or more genes, and their environment. |
| Loci | In the fields of genetics and genetic computation, a locus (plural loci) is the specific location of a gene or DNA sequence on a chromosome. |

| | |
|---------------------------------------|--|
| | A variant of the DNA sequence at a given locus is called an allele. The ordered list of loci known for a particular genome is called a genetic map. Gene mapping is the procession of determining the locus for a particular biological trait. |
| QTL – Quantitative Trait Loci | Quantitative trait loci (QTLs) are stretches of DNA containing or linked to the genes that underlie a quantitative trait. Mapping regions of the genome that contain genes involved in specifying a quantitative trait is done using molecular tags commonly single nucleotide polymorphisms or SNPs . This is an early step in identifying and sequencing the actual genes underlying trait variation. |
| Gene (Genetic) Markers | A genetic marker is a gene or DNA sequence with a known location on a chromosome that can be used to identify cells, individuals or species. It can be described as a variation (which may arise due to mutation or alteration in the genomic loci) that can be observed. A genetic marker may be a short DNA sequence, such as a sequence surrounding a single base-pair change (single nucleotide polymorphism, SNP), or a long one, like minisatellites. |
| SNPs – Single Nucleotide Polymorphism | A single-nucleotide polymorphism (SNP, pronounced <i>snip</i>) is a DNA sequence variation occurring when a single nucleotide — A, T, C or G — in the genome (or other shared sequence) differs between members of a biological species or paired chromosomes in an individual. |
| SNP Genotyping | SNP genotyping is the measurement of genetic variations of single nucleotide polymorphisms (SNPs) between members of a species. It is a form of genotyping, which is the measurement of more general genetic variation. SNPs are one of the most common types of genetic variation. An SNP is a single base pair mutation at a specific locus, usually consisting of two alleles. |
| Snip Chip | Sometimes called a DNA microarray or biochip, is a collection of microscopic DNA spots attached to an electrochemical solid surface. Scientists use Snip Chips to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome. |

The World of Genomics

Genomics has been defined as the study of genes and their function, with the purpose of understanding the structure of the genome. This science involves sequencing DNA of the animal or plant and mapping genes. There are several facets of modern genomics including: functional genomics, structural genomics, comparative genomics, epigenomics (epigenetics), and pharmacogenomics. The relatively new discipline of genomics involves combining molecular and cell biology with classical genetics. The ability to do a large number of statistical computations quickly with computer technologies has fostered the development of the science genomics (McKusick and Ruddle, 1987).

An indication of the expansion of the use of genomics for understanding genetic influences can be illustrated by the number of companies involved in this science. For example, the BioCompany website (<http://www.genomics.co.uk/companylist.php>) lists over 500 companies centered on the use of genomics to evaluate information gathered from DNA in plants and animals.

So, how does the science of genomics relate to genetic prediction in cattle?

A simple way to answer this question is to think of the relationship between animal performance traits and markers on the DNA sequence of the animal. Simply put, the process used to discover these relationships involves collecting DNA samples (e.g. hair follicles, blood, semen) from the animals you want to characterize, and then evaluating the DNA from these animals for a large number of SNPs (base-pair sequences) that have been identified as having a high association with particular phenotypic production traits.

The association between the SNPs and the phenotypic traits has been worked out through extensive statistical comparisons of animal performance databases with DNA libraries. As new SNP information is discovered, the statistical models are “trained” and “validated” to assure that accurate predictions are valid (Hay et al., 2011). From this evaluation, predictions are made as to the likelihood that an animal that has little or no phenotypic data in the database will perform in a certain predicted manner. This approach enables the producer to make selection decisions at an earlier age than would be possible without the genomic information.

Currently, there are primarily two companies that do genomic work with cattle in the United States – Pfizer Animal Genetics and Igenity which is owned by Merial. Both of these companies have technologies to test for such things as genetic defects, parentage, coat color, horned or polled and numerous performance traits. For example, Igenity evaluates beef cattle DNA samples and predicts the genetic merit for the following traits (see <http://www.igenity.com/pdfs/forms>):

Carcass Tenderness • Heifer Pregnancy Rate • Docility • Percent Choice •
Maternal Calving Ease • Average Daily Gain • Yield Grade • Marbling • Feed
Efficiency • Ribeye Area • Fat Thickness • and Stayability

Genetic Prediction through Phenotype Comparisons – The Age of EPDs

The Beef Improvement Federation (BIF) came into existence in 1982 as a group of cattle producers, academicians and statisticians to advance the science of understanding inheritance factors in beef cattle and developing uniform procedures for predicting genetic merit. Over the years, progress in development of procedures and guidelines for genetic improvement of beef cattle have been marked by Genetic Prediction Workshops, the first of these was held at Winrock International, in Morrilton, AR, on Dec.12-15, 1983. A primary result of this genetic prediction evolved into expected progeny differences or EPDs. Expected progeny differences are an evaluation of an animal’s genetic worth as a parent. They are based on sophisticated statistical animal models which combine all information known about an

individual and its relatives to create a genetic profile of the animal's genetic merits. These profiles are then compared to other individuals, generally only animals of the same breed.

The first practical implementation of EPDs came through National Sire Evaluation (NSE), conducted by several breed associations. The widespread use of popular bulls through artificial insemination, particularly in breeds first available in the United States in the late 1960s, allowed them to serve as so-called Reference Sires, the benchmark in NSE. The first National Sire Summary, comparing EPDs of 13 bulls, was published in 1971 (Hammack and Paschal, 2009).

During the 40 years since 1971, numerous publications have documented the phenotypic correlations between parental traits in cattle and their progeny (Hohenboken et al, 1973; Mallinckrodt et al., 1993; Rincker et al., 2006). Through educational efforts and firsthand experience of cattlemen over the past several decades, it is apparent that EPDs are effective tools for predicting the genetic merit of bulls. Use of EPDs to advance genetic change in beef cattle is common place among essentially all beef cattle producers today, both commercial and seedstock producers.

Using DNA to Enhance Phenotypic Predictions

The next major steps in genetic prediction have arisen from development of systems to integrate the traditional EPDs computed from performance data of sires and their progeny with genomic data derived from sampling DNA and evaluating it for gene markers. These predictions were originally referred to as Marker-Assisted EPDs or MA-EPDs.

The first MA-EPDs were published in the American Simmental Association's (ASA) Fall 2004 Simmental/Simbrah Sire Summary. These carcass merit predictions were calculated by Dr. Richard Quaas of Cornell University and resulted from collaboration between the ASA, Cornell University, the National Beef Cattle Consortium, NCBA's Carcass Merit Program and Frontier Beef Systems. The initial MA-EPDs Tenderness EPDs became available as a result of integrating the phenotypic data obtained from shear force data accumulated through the Carcass Merit Program with DNA profiles from the sires and their progeny.

McNeil et al (2010) published DNA enhanced EPD predictions for the Angus breed with what are now being called molecular breeding values (MBVs). The American Angus Association has developed a rapid turn-around system where the addition of genomic results to EPDs is computed weekly to make the most current genetic prediction information readily for use by Angus breeders (Northcutt, 2011.) Many other breed associations are also developing the technology to enhance traditional EPDs with information garnered from genomic data.

An Analogy in Conclusion

Moneyball is a recent movie playing in theatres based on the book *Moneyball: The Art of Winning an Unfair Game* by Michael Lewis. The central premise of *Moneyball* is that the collected wisdom of baseball insiders (including players, managers, coaches, scouts, and the

front office) over the past century is subjective and often flawed. Player statistics such as stolen bases, runs batted in, and batting average, typically used to gauge players, are leftovers of a 19th century view of the game and the player information that was available at the time. The movie – based on a true story – argues that the major league baseball Oakland A's' front office took advantage of more empirical gauges of player performance to field a team that could compete successfully against richer competitors in Major League Baseball.

In the book and the movie, it was shown that rigorous statistical analysis had demonstrated that on-base percentage and slugging percentage are better indicators of offensive success than the traditional player statistics used to select players by other teams, and the A's became convinced that these qualities were cheaper to obtain on the open market than more historically valued qualities such as speed and contact.

By re-evaluating the strategies that produce wins on the field, the 2002 Oakland Athletics, with approximately \$41 million in salary, were competitive with larger market teams such as the New York Yankees, who spent over \$125 million in payroll that same season. Because of the team's smaller revenues, Oakland was forced to find players undervalued in the market based on traditional predictions. The result of the A's system for finding value in undervalued players based on statistical analysis of successes has been effective thus far (Source: <http://en.wikipedia.org/wiki/Moneyball>).

The analogy here is that EPDs have been a very effective method for identifying and guiding mating decisions in the beef industry. The historical approach with many genetic predictions in beef cattle has been to compute EPDs that serve as indicator traits – traits that indirectly indicate how other traits will respond. In recent years, the beef industry has developed new approaches to evaluating genetic value that have focused on what are being called Economically Relevant Traits, or ERT. Economically Relevant Traits, as the name implies, are those traits that have a direct economic impact to the producer, as compared to indicator traits that merely have an association with economically relevant traits. Traits such as weaning weight and carcass weight are ERT because there is a direct monetary value associated with these traits (Purdue Beef Blog, 2008).

So, with the development of genomics, predictors like molecular breeding values (MBVs) for Economically Relevant Traits (ERTs) may allow cattle producers a better way to identify valuable animals using “player statistics” coupled with DNA markers – much like the Oakland A's were able to find undervalued players by using actual performance outcome like on-base percentages. New tools such as genomics provide a new method to gain enhanced genetic information without the time and expense required to test a large number of progeny.

As with any new technologies, time and observation are needed to evaluate the overall impact of beef cattle genomics at the ranch level. It will be exciting to see where this and related technologies take our industry in the next 40 years. The next presenter, Dr. Matt Spangler, will discuss more details of how and when to use genomic derived genetic predictors such as MBVs.

References

- Hammack, S.P., and J.C. Paschal. 2009. Texas Adapted Genetic Strategies VIII: Expected Progeny Difference (EPD). Texas Agri-Life Extension Publication E-164. September 2009.
- Hay, E., H. Wang, X. Liu, B. Woodward, S. Bauck, and R. Rekaya. 2011. A neural network approach for association between a low-density whole genome SNP marker panel for 19 traits in beef. *J. Anim. Sci. (abstr) E-Suppl.* 1. 89:533.
- Hohenboken, W. D., E. R. Hauser, A. B. Chapman, and L. V. Cundiff. 1973. Phenotypic correlations between dam traits expressed during development and lactation and traits of progeny in cattle. *J. Anim. Sci.* 37:1.
- Mallinckrodt, C. H., R. M. Bourdon, B. L. Golden, R. R. Schalles and K. G. Odde. 1993. Relationship of maternal milk expected progeny differences to actual milk yield and calf weaning weight. *J. Anim. Sci.* 71:355-362.
- McKusick, V.A. and F.H. Ruddle. 1987. A new discipline, a new name, a new journal [editorial]. *Genomics* 1987 Sep;1:1-2.
- MacNeil, M.D, J. D. Nkrumah, B. W. Woodward and S. L. Northcutt. 2010. Genetic evaluation of Angus cattle for carcass marbling using ultrasound and genomic indicators. *J. Anim. Sci.* 88:517-522.
- Northcutt, S. 2011. By the numbers. *Angus Journal*, April 2011.
- Purdue Beef Blog. 2008. Indicator Traits vs. Economically Relevant Traits. Source: <http://purduephil.wordpress.com/2008/04/28/indicator-traits-vs-economically-relevant-traits/>. Accessed 11-11-11.
- Rincker, C. B., N. A. Pyatt, L. L. Berger and D. B. Faulkner. 2006. Relationship among GeneSTAR marbling marker, intramuscular fat deposition, and expected progeny differences in early weaned Simmental steers *J ANIM SCI* 2006, 84:686-693.