They're All Carriers of Something

Broken genes are a part of the beef business.

*by **Dorian Garrick,** Iowa State Unversity*

Cattle typically have 30 pairs of chromosomes, comprising 29 pairs of autosomes and XX or XY sex chromosomes. Chromosomes are made from DNA, which consists of a sequence of bound paired chemical compounds, known as nucleotides, which are made up of a nitrogenous base, a sugar and some phosphate groups. There are four kinds of nucleotides — adenine (A), guanine (G), thymine (T) and cytosine (C) — and the sequence of these four compounds dictate the genetic characteristics of an individual.

A typical cattle chromosome consists of 100 million base pairs and occurs in two versions, one inherited from the sire and the other inherited from the dam. Accordingly, the genome consists of 3 billion base pairs inherited from one parent, and a similar number inherited from the other. After fertilization, these 6 billion base pairs must be copied every time a cell divides, in a process known as mitosis. An adult contains something like 50-100 trillion cells, and any error that occurred during the copying of the chromosomes or their division into daughter cells will be propagated in subsequent divisions of the cell.

During development, cells specialize to form around 200 different cell types, including those different types found in muscle, fat, skin, blood and various organs. Most errors that occur in DNA replication are not passed on to offspring — only those cells that form parts of the testicular or ovarian tissue can contribute to the genomes of future generations.

Changes in genomic sequence such as those that arise from DNA copying errors are known as mutations. There are a number of different kinds of mutations that can arise. Sometimes one base pair (A, G, T or C) is mistakenly copied for an alternate base pair. This is known as a single nucleotide polymorphism (SNP). Other mutations might involve the accidental duplication of a piece of DNA; an accidental deletion of a piece of DNA; or an inversion, whereby the sequence is partially reversed.

Errors in copying DNA are very common, perhaps 1 in every 100 base pairs, but the cell has DNA repair mechanisms that identify and repair almost all of the errors. The typical error rate remaining after the repairs is something like one in every 30 million nucleotides each generation, or a little over three mutations per chromosome per generation.

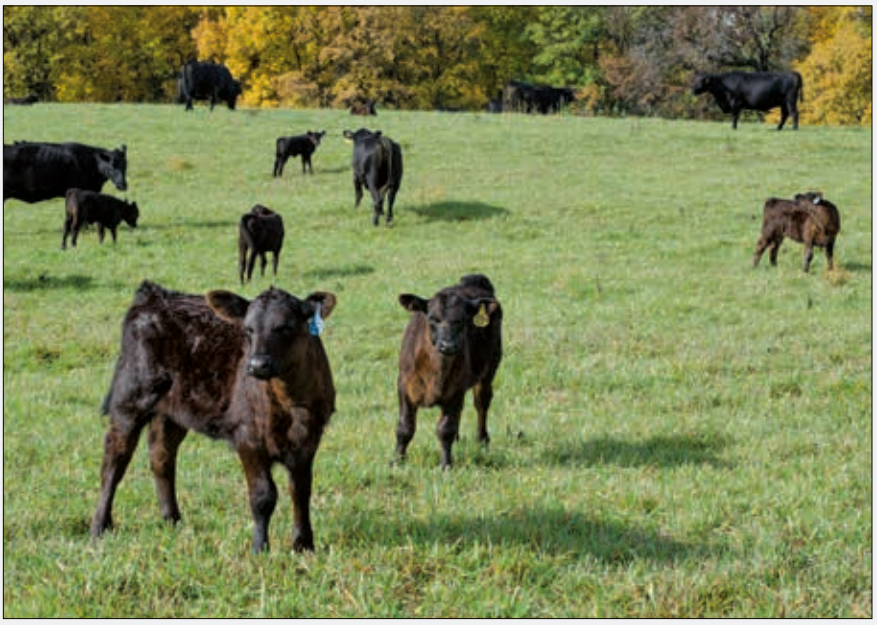

PHOTO BY SHAUNA ROSE HERMEL

Most of the genome does not code for genes. Genes comprise a promoter region, coding regions known as exons, and regions between exons that are known as introns. The sequences between the locations of genes are known as intergenic regions. Perhaps only 2%-3% of the genome comprises exons, and only half of these code for proteins. Accordingly, the impact of mutations on performance depends largely upon which part of the genome is mutated. Mutations in intergenic regions between genes or in introns are less likely to be damaging than mutations in exons.

It is now possible to have exonic regions captured and individually sequenced for less than $2,000, or a whole genome sequenced for less than $10,000, although these prices are likely to erode markedly over the next decade. Sequence information is believed to be useful in personalized human medicine, and in assessment of risk for certain diseases such as heart disease, diabetes or cancer.

The fact that chromosomes occur in pairs means that, except for some genes on the sex chromosomes, an individual inherits two copies of every gene, one from its sire and the other copy from its dam. A mutation in one copy of the gene may not be very serious provided the other copy is functional.

Due to the presence of historical mutations, every individual inherits a number of dysfunctional genes from its sire, and a number of other dysfunctional genes from its dam. On average, half of these will be passed on to the offspring, along with half the 100 or so new mutations, mostly not in genes, that are new (*de novo*) to the

sperm or egg, or occurred in the individual between the time of its conception and the time it becomes a parent. Many mutations inherited by the offspring will not be passed on because the offspring is harvested without becoming a parent, or because of chance sampling when a parent has few offspring.

There develops from a population perspective a balance between the creation of new mutations and the loss of existing mutations. Widely used sires are likely to pass on all their mutations to some of their progeny and, in this manner, the frequency of certain rare mutations can increase markedly in just one generation of widespread use.

Loss-of-function mutations

Mutations in genes may or may not be problematic. Proteins consist of sequences of amino acids. There are 20 different amino acids, and each is specified at DNA level by a triplet sequence of base pairs. Since each base pair can be one of four options, there are 4×4×4=64 different possible triplets to represent the 20 amino acids or a stop codon, which is the signal to terminate the protein. This means that more than one triplet sequence can represent the same amino acid. Accordingly, some mutations can change the triplet code but have no impact on the amino acid sequence. These are known as synonymous mutations.

Other mutations will result in substitution of one amino acid for another and are known as non-synonymous, or missense, mutations. Such mutations may or may not have serious impacts, depending upon the extent they change the shape or other properties of the resultant protein. Some mutations will result in the gene being dysfunctional or "broken," and these are known as loss-of-function mutations. This includes some non-synonymous mutations that seriously impact the protein properties, as well as mutations that disrupt the start or the stop information, known as nonsense mutations.

Mutations can prevent the protein forming at all, or can make it too short or too long. Some loss-of-function mutations are of particular concern because they can impact the viability or productivity of the resulting offspring. Even loss-of-function mutations are not a real problem in an outbreeding population because the mutations carried by any particular sire are likely to be rare in its unrelated mates.

However, the same cannot be said in the presence of inbreeding, as occurs to a mild extent when animals of the same breed are mated together. In these circumstances, a mutation that an ancestral sire passed on down through its daughters and subsequent maternal lineages is also likely to be carried by one or more of its paternal lineage. In that case we would expect one-quarter of the matings between sires that carried the mutation and dams that carried the same mutation to exhibit the recessive genetic defect.

Nowadays, the availability of high-density genotyping panels in most livestock species has made it possible to identify carrier animals by inspecting their DNA.

Serious defects may prevent normal completion of the pregnancy, and may go unnoticed, except perhaps by a slight reduction in reproductive performance. In other cases the defect may be apparent in the newborn offspring. In a performance-recorded setting, particularly with single-sire mating, such occurrences can often be easily detected. This is the case in the dairy industry where sires are routinely used simultaneously across many herds.

Prior to methods for comprehensively genotyping the entire genome, genetic defects had to be managed by progeny testing.

Sadly, the track record for early detection of genetic defects in the beef industry has been much poorer, and presence of a visual deformity has sometimes been associated with the three S's — shoot, shovel and shut up. That is, a bull breeder observing a defective newborn might destroy the evidence and eliminate the further use of the sire and dam and other close relatives, rather than communicating the finding at the risk of developing a reputation for genetic defects that may lose them market share. This approach may sometimes be successful, but in many cases has simply delayed the recognition of the genetic defect while it becomes more widely propagated in the industry.

Identifying loss-of-function mutations

There are several approaches that have been successfully used to identify loss-of-function mutations. The oldest method is based on the appearance of defective offspring, such as dwarfs.

Not all defective offspring represent inherited genetic abnormalities. Matings between parents of defective offspring (i.e., carriers) should produce one out of four defective offspring for a recessive condition.

Prior to methods for comprehensively genotyping the entire genome, genetic defects had to be managed by progeny testing. An effective method was to mate a potential carrier sire to dams that were known to carry the defect. This required carrier animals to be maintained, and also resulted in delays in waiting for progeny test results before the sire could be confidently used.

Half a century ago in the United States, it had been a common occurrence by some Hereford breeders to mate a sire to a few of his daughters and to delay the wide use of the sire until the results of the inbred matings had been observed to confirm the absence of defective offspring.

Mannisidosis was one such recessive disease not uncommon in Angus cattle until Dr. Bob Jolly at Massey University in New Zealand identified a blood test to distinguish apparently normal carrier animals from animals that were free of the defect. The blood test took advantage of the fact that animals with only one functional copy of the α-mannosidase enzyme had lower (i.e., about half) blood concentrations of the enzyme than animals with two functional copies. Screening animals to detect carriers based on a blood test was much easier than progeny testing and was rapidly adopted by industry.

Nowadays, the availability of high-density genotyping panels in most livestock species has made it possible to identify carrier animals by inspecting their DNA. In order to develop a test for the disease, the genomic region responsible for the defect must first be identified. This can be very easily done with the DNA from 10-15

CONTINUED ON PAGE 252

They're All Carriers of Something CONTINUED FROM PAGE 251

affected offspring, or less, provided the condition really is due to a homozygous recessive condition. The analysis simply involves genotyping the affected animals, and identifying any regions where they are all homozygous for the same SNP alleles. We have recently undertaken such an approach with three different recessive conditions in sheep. The results of the experiments showed that the disease was contained within a genomic region whose size ranged between 1%-10% the length of one chromosome.

A similar approach was used to find the cause of osteopetrosis in Red Angus cattle (Meyers et al., 2010), arthrogryposis multiplex (AM) in Angus cattle (*www.angus.org/NAAB_release.pdf*), and many recessive diseases in dairy cattle (e.g., Charlier et al., 2008; Charlier et al., 2012).

In a bull-breeding herd, carrier animals can still be used provided the offspring are screened for the mutation and only those that are free of the defect should be mated to animals that carry the defect.

Sequencing technology allows the actual sequence of nucleotides to be readily determined in each of the homozygous genomic regions that were common to all the affected animals. Within 12 months we were able to identify what we believe to be the actual causal mutation in three sheep diseases. One was due to a nonsense mutation that altered one base pair so that the protein encoded by the gene was prematurely terminated. Another was due to a base pair change that resulted in the protein that comprises a sequence of amino acids having one different amino acid in its sequence that changed the shape of the protein and therefore its biological properties. The third disease had a single base pair missing, known as a frameshift mutation, resulting in a complete change in the amino acids that form the protein from the point of mutation onwards.

The cause of osteopetrosis in Red Angus cattle was the deletion of 2,800 base pairs that removed exon 2 and nearly half of exon 3 (Meyers et al., 2010). The cause of AM is the deletion of a significant genome fragment (*www.freepatentsonline.com/y2011/0151440.html*).

Genomic technologies have now provided two additional methods for detecting loss-of-function mutations, and these will lead to a marked increase in the number of such mutations discovered over the next few years. The increase in mutation discovery rate reflects the fact that the conventional approach based on observing defects has two shortcomings.

First, the conventional approach cannot easily detect defects that cause fertilization failure or embryonic loss. Second, the conventional approach relies on breeders noticing and reporting the defects.

Rare defects may only be apparent in individual cases within any particular herd, and not be recognized as having a genetic origin. Many recently discovered defects in cattle have now been shown to have been present but undetected for several decades or longer. The two new methods for finding defects don't rely on the use of phenotypes in the first instance, but instead rely either on the use of SNP marker panels across a subset of the population, or on whole-genome (or exome) sequencing of one or more individuals.

Unlike most cells, gametes such as sperm or eggs contain only one copy of each of the 30 pairs of chromosomes. These single copies typically represent a chromosome that is not the same as either the paternal or maternal chromosome of the parent, but represents a new variant created from a crossover between the two parental chromosomes in the pair. This means that DNA is not inherited one base pair at a time, but in large units. Accordingly, many base pairs, and therefore alleles, are inherited together in a chunk of chromosome that forms a small unit known as a haplotype. Although SNP marker panels only identify the genotype of an individual (i.e., A_1A_1, A_1B_1 or B_1B_1 at a particular locus, say 1), population data allows haplotypes to be identified. Suppose an animal was heterozygous (i.e., A_1B_1 and A_2B_2) at two adjacent loci. This means that at either locus, one chromosome carries the A allele and the other chromosome carries the B allele. However, these genotypes do not tell us unambiguously which haplotypes are carried by the individual. Animals that are heterozygous at two adjacent loci might represent two different haplotype combinations.

One haplotype combination is known as coupling and would comprise one chromosome carrying the two adjacent A alleles (i.e., A_1A_2), and the other chromosome carrying the two adjacent B alleles (i.e., B_1B_2). An alternative haplotype combination is known as repulsion, and it involves one chromosome carrying the A_1 allele next to the B_2 allele (i.e., A_1B_2), while the other chromosome would carry the B_1 allele next to the A_2 allele (i.e., B_1A_2). Both of these haplotype combinations would result in the same genotypes.

Using population data, it is possible to reconstruct haplotypes from the SNP genotypes. This would normally be undertaken for much larger regions than just two loci. A sequence of 20 consecutive SNP markers could produce more than 1 million different haplotypes, but in a typical beef cattle population, we are likely to only observe about 20 common haplotypes.

If a common haplotype contains a deleterious mutation, such as one causing loss of function, we would not observe the expected proportion of individuals that were homozygous for that haplotype. Scientists at the USDA Animal Improvement Programs Laboratory (AIPL) have used this concept in the U.S. dairy population (VanRaden et al., 2011) to find haplotypes in five regions of the genome that should have been observed if they had no detrimental effects, but have never been seen. Similar studies in France using the same approach found some of the same regions and confirmed another nine mutations (Fritz et al., 2013). This approach has yet to be applied to U.S. beef cattle populations.

Genomic sequencing of individual animals involves the comparison of their sequence to the bovine reference genome based on the Line 1 Hereford cow Dominette. Inspection of the sequence of known genes can identify potentially serious missense mutations as well as nonsense mutations, splice site variants and damaged start regions.

Every animal (and human) carries a number of such mutations. Most of these mutations will have been inherited from a parent and may have existed in the

population for many generations, whereas some may be *de novo* mutations that have just occurred in the most recent generation. Any candidate loss-of-function mutations can be individually screened in the population and/or added to the next generation of SNP chips.

In many species, particularly humans, the study of affected individuals allows identification of mutations causing inherited diseases, which adds to published knowledge as to the importance of particular genes. In model animal species such as mice, considerable research has been undertaken to "knock out" or deliberately create loss-of-function mutations in almost every known gene. Knowledge of the productive attributes of the resultant animals homozygous for the knockout provides valuable information as to the likely role or impact of that gene.

Some genes can be knocked out without apparent effect, whereas others might impact production (e.g., muscularity or obesity), lactation, reproduction, longevity, etc. Collectively, this information from natural mutations and from knockout studies undertaken on a range of species allows the annotation of bovine genome sequence and automated identification of possible loss-of-function mutations.

Managing loss-of-function mutations

The immediate reaction of many farmers to the finding that one of their animals carries a genetic defect is to discard the animal and any descendants that inherited the defective mutation. However, this is not a good idea. Every individual carries defective mutations. The sensible approach is to manage the matings in such a way as to avoid the pairing of carrier animals. A carrier animal may be a perfectly good terminal sire and will not result in defective offspring when used in an outcrossing program even when carrier offspring are retained for breeding to a terminal sire.

In a bull-breeding herd, carrier animals can still be used provided the offspring are screened for the mutation, and only those that are free of the defect should be mated to animals that carry the defect. Mate-selection software now being trialed by some breed associations can be usefully applied to minimize the impact of such mutations (e.g., Kinghorn, 2011).

Dr. Jerry Taylor at the University of Missouri has obtained USDA funding for sequencing widely-used sires in a number of U.S. beef breeds, and a similar Genome Canada project at University of Alberta and University of Guelph in Canada has almost completed the sequencing of about 300 beef sires. An international effort in dairy and beef cattle is under way with a target of sequencing 1,000 bovine genomes (*www.1000bullgenomes.com*). That project represents a collaborative effort where approved contributors of 25 or more sequenced animals can gain access to the sequences on all the animals in the project.

Summary

Mutations are a natural occurring phenomenon that provide a mechanism for genetic variation. There are a number of kinds of mutations, and one common kind is represented by alterations in DNA that cause loss of function of the gene. All individuals carry loss-of-function mutations. Some have been identified as inherited diseases with obvious abnormalities (e.g., dwarfism). Most have gone undetected.

Genomic technologies now allow loss-of-function mutations to be discovered through absence of homozygote haplotypes, or through annotation of individual genomic sequence. Some of these new discoveries will include those that impact embryonic mortality. The challenge for breeders in the future will be to manage known mutations in their herds. AJ

Helpful definitions

Broken genes — Mutations that result in a gene being dysfunctional. Also known as loss-of-function mutations.

Exons — Coding regions of the genome.

Frameshift mutation — Mutation in which a single base pair is missing, resulting in a complete change in the amino acids that form the protein from the point of mutation onward.

Intergenic regions — The sequence between the locations of genes.

Introns — Regions between exons.

Loss-of-function mutations — Mutations that result in a gene being dysfunctional. Also known as broken genes.

Nonsense mutations — Mutations that disrupt the start or the stop information.

Non-synonymous, or missense, mutations — These mutations result in substitution of one amino acid for another.

Synonymous mutations — Mutations that change the triplet code but have no impact on the amino acid sequence.

Stop codon — A triplet code that signals termination of the protein.

Triplet code — The triplet sequence of base pairs that codes for an amino acid.

Editor's Note: *Dorian Garrick is animal scientist and Jay Lush Endowed Chair in Animal Breeding and Genetics at Iowa State University.*

Literature Cited

Charlier C, Coppieters W, Agerholm Js, Cambisano N, Carta E, Desmecht D, Divem, Fasquelle C, Frennet Jc, Hanset R, Hubin X, Jorgensen C, Karim L, Kent M, Harvey K, Pearce Br, Rollin F, Simon P, Tama N, Nie H, Vandeputte S, Lien S , Longeri M , Fredholm M, Harvey Rj, Georges M. 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nature Genetics, 40(4):449-54. Epub 2008 Mar 16.*

Charlier C, Agerholm Js, Coppieters W, Karlskov-Mortensen P, De Jong G, Li W, Fasquelle C, Karim L, Cirera S, Cambisano N, Ahariz N, Mullaart E, Georges M, Fredholm M. 2011. A deletion in the bovine FANCI gene compromises fertility by causing fetal death and brachyspina. *PLoS ONE 7(8):e43085.*

Fritz S., Capitan A., Djari A., Rodriguez S.C., Barbat A., Baur A., Grohs C., Weiss B., Boussaha M., Esquerre D., Klopp C., Rocha D., Boichard D. 2013. Identification of novel mutations associated with prenatal mortality in dairy cattle. *PAG 2013. www.intl-pag.org*

Kinghorn BP. An algorithm for efficient constrained mate selection. *2011. Genetics, Selection, Evolution, 43:4.*

Meyers S.N., McDaneld T.G., Swist S.L., Marron B.M., Steffen D.J., O'Toole D., O'Connell J.R., Beever J.E., Sonstegard T.S., Smith T.P.L. 2010. A deletion mutation in bovine SLC4A2 is associated with osteopetrosis in Red Angus cattle. *BMC Genomics 11:337.*

VanRaden P.M., Olson K.M., Null D.J., Hutchison J.L. 2011. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *Journal Dairy Science 94:6153–6161.*