

APPLYING NEW GENOMIC TECHNOLOGIES TO ACCELERATE GENETIC IMPROVEMENT IN BEEF AND DAIRY CATTLE

A. Canovas¹, J. Casellas², M. Thomas³ and J.F. Medrano⁴

¹University of Guelph, Centre for Genetic Improvement of Livestock, Department of Animal Bioscience, Ontario, N1G2W1 Canada

²Universitat Autònoma de Barcelona, Bellaterra, Spain

³Colorado State University, Fort Collins, CO, 80521 USA

⁴University of California, Davis, CA, 95616 USA

SUMMARY

In recent years, breeders have combined the use of phenotypic appraisal and the estimation of breeding values (PTA or EPD) to make genetic selection decisions in beef and dairy cattle that have resulted in a steady genetic gain of 2% per year. However, the most extensive application of genomics has occurred in the dairy industry with the estimation of molecular breeding values that has improved selection efficiency to a much higher order of magnitude. Despite a growing molecular and physiological understanding of complex traits, little is known about the genes determining the traits and their precise function, and a significant unexplained source of variation of phenotypes remains in livestock. Within this context, a more complete understanding of the genes and regulatory pathways and networks involved in economically important traits (i.e. fertility and reproduction, feed efficiency, meat quality and carcass traits) in beef and dairy cattle will provide knowledge to help improve genetic selection and reproductive management. Therefore, high throughput -OMICS technology (i.e., transcriptomics, metagenomics, metabolomics, as well as epigenomics amongst several others), will complement these tools and further advance identification of functional genes within a systems biology approach.

INTRODUCTION

The field of genetics and genomics in most of the livestock species has experienced a dramatic technological revolution in the last 5 years. During the earlier 2000's, the emphasis in livestock genetics was in linkage maps with single markers and quantification of some genes using real time qPCR. The first bovine genome assembly was published in 2009 (Elsik et al., 2009), and since that time, the development and use of various whole genome-omics tools has accelerated research in cattle genetics (Reverter et al., 2013; Snelling et al., 2013). After the development of the livestock genome sequences, new technologies such as microarrays were introduced allowing the study of gene expression of the entire transcriptome. With that, the concept of "Genetical Genomics" was developed with the idea of integrating structural and functional genomic data, combining gene expression from microarray technology with genotype data from marker genotypes. Currently, the field of Genetical Genomics or the integration of structural and functional genomic data, has expanded as high throughput tools have become available for genomic analysis such as high-density (HD) genotyping-chips (Illumina, San Diego, CA), whole genome sequencing and genotyping by sequencing and RNA-Sequencing to measure the gene expression in the entire transcriptome (in a more accurate way than microarray technology; Wickramasinghe et al. 2014).

Currently, with all the new available technologies in livestock combined with statistical methodologies, the integration of structural and functional genomics information with other – OMICS into a systems biology approach has allowed development of a better biological understanding of phenotypes. As part of the genomics tool box the HD-genotyping SNP chips such as 50K and 800K, whole genome sequencing technologies are now available in most of the livestock species and have been extensively utilized, such as in dairy cattle, in genetic

improvement. As a part of high throughput tools available for genomic analysis, RNA-Sequencing allows measuring not only gene expression, but also examining genome structure identifying SNP and other structural variation such as insertions, deletions and splice variants (Wickramasinghe et al. 2014; Cánovas et al. 2010). The expectation is that the integration of all these types of genomic data will accelerate the genetic improvement by improving accuracy of selection and reducing the generation interval.

Combining the information from the -OMICS technologies together with metabolic pathways and functional/biological analysis into a Systems biology approach allows the identification of functional SNP increasing the accuracy of selection. Briefly, transcriptomics using RNA-Sequencing technology measures gene expression of the entire transcriptome, as well as structural variation (i.e. SNP, insertions, deletions and splice variants in coding regions). Also, it allows the identification of the differential expressed splice variants affecting genes overlapping significant SNPs in significant QTL regions. Metagenomics, identifies and quantifies the microbial profiles and determine their functionality in specific bovine samples (i.e. rumen fluid) using 16SrRNA based profiling of microbial populations as well as partial protozoa and fungal 18S rRNA (Handerson et al., 2015). Metabolomics, identifies circulating plasma and/or rumen fluid metabolites as potential biomarkers to assess physiological and hormonal states relative to performance by Gas-chromatography-Mass-spectrometry (GC-MS) and NMR analysis. For each analysis, internal standards are included for normalization to allow comparisons between samples. Temporal and treatment changes in metabolites can be determined to be significant after normalization and can be correlated to the microbial profiles and gene expression obtained from RNA-Sequencing. The microbial profiles and/or microbial activities of each animal can be then defined and associated with the phenotype of interest using a multivariate approach.

The particular benefits of new integrated high throughput genomics technologies will be most likely for genetic improvement of those traits that are difficult to measure such as feed efficiency, methane emission and fertility and reproduction traits.

BEEF AND DAIRY CATTLE GENETIC IMPROVEMENT USING GENOMIC TECHNOLOGIES

In general, the beef industry is highly segmented (seedstock, cow-calf, feedlot, processor, retailer and consumer). This segmentation is seen in the way various stakeholders along the supply chain make logical decisions for their own business, yet in the light of the entire chain, these decisions are detrimental to the end product and consumer satisfaction. With little or no market signal from consumer to seedstock, needed changes are not made for the end product to be desired and competitive with other protein choices of the consumer. Poor information flow along the beef supply chain is leading to massive inefficiency as demonstrated in the dramatic decline in high yielding carcasses in the past decade (down from 66% to 41%), Canadian Beef Grading Agency data). This drop resulted from overfeeding in an attempt to increase the level of intramuscular fat (marbling). The past thirty years has seen a dramatic and consistent decline in per capita beef consumption, due in part as a result of inconsistent product, primarily in the important trait of tenderness. Over the past thirty years (and even more drastically the last 10 years), beef consumption per capita in Canada has declined from 27 to 19 Kg while that of chicken has risen from 21 to 33 kg.

The many studies of the beef industry completed in the past 20 years all identified lack of information flow as a serious impediment to the industry. This is more currently demonstrated by the inclusion of connectedness (information flow along and back the supply chain) as one of four target areas of focus in the Canadian Cattlemen's Association's (CCA's) national beef strategy (beefstrategy.com). The utilization of genomics in the beef industry has been limited (compare to dairy industry), due in part to the 1) complicating factor of multiple breeds and correspondence of

SNP effects, 2) in general, the genetic evaluations are carried out by the multiple breed associations using different methods to calculate the EPDs, 3) reporting different EPDs for different breeds confusing the data comparisons from different evaluations (Van Eenennaam 2016). International Genetics Solutions (IGS) is trying to leverage a multibreed database that enables the comparison of EPDs across breeds associations (Chianina, Gelbvieh, Limousin, Maine-Anjou, Red Angus, Simmental, Shorthorn, Canadian Simmental, Canadian Gelbvieh, Canadian Limousin, Canadian Angus and Canadian Shorthorn).

Genomic approaches offer an opportunity to accelerate the genetic improvement, but genomics requires accurate phenotypes (and genotypes) from genomically linked individual animals. Therefore, the focus is on collecting data to enable generation of GEBV for all animals in a population. Industry breeding strategies will then incorporate these new genetic evaluations to improve production efficiency and to provide a more consistent and competitive product to consumers. Also, beef producers often make breeding decisions without current knowledge of the science, existing systems and tools that they can use on farm to benefit from this research. Taken in isolation, a research finding can be confusing and lack any meaningful connection to the real world of production in which producers make decisions. Thus, producers require more concise, layman's terms presentation of the benefits of implementing change in their businesses, as well as a more clearly defined stepwise approach they can take to implement that change. Therefore, it is important to do an effort to translate and transfer the research knowledge associated to genetics and the importance of collecting phenotypes (and genotypes) which is key to affecting change necessary to enhance and accelerate innovation in beef industry.

Using genomics-based approaches to improve beef genetics will result in increased production efficiency, making beef a more attractive protein option than current relative to poultry and pork. Effective use of genomics will also provide the consumer with a more consistently tender product and therefore help to address declining consumption. Therefore, it's important to capture and make use of data in the development of genomically enhanced Expected Progeny Differences (GEPD's) which are simply more accurate indicators of an animal's true genetic merit. Although there are phenotypes (related to meat quality and carcass traits) and genotype data collected, more phenotype and genotype data are required to enable generation of GEPDs to incorporate into beef industry breeding strategies. New sets of data will be crucial to enlarge the reference set of animals with genotype and phenotype, but most important, will provide the researchers with additional information collected at the whole chain scope, which will be unique to improve the quality of prediction equations and at the same time, enhance economical analysis of the desired traits.

GEPDs have a significant role to play in the effective selection for these traits. The selection efficiency with GEPDs will be higher with the knowledge gained from functional genomics studies. Despite a growing molecular and physiological understanding of complex traits such as carcass traits, little is known about the genes determining the traits associated to tenderness and their precise function, and a significant unexplained source of variation of phenotypes remains in beef. Within this context, a more complete understanding of the genes and regulatory pathways and networks involved in economically important traits such as tenderness and carcass traits in beef cattle will provide knowledge to help improve genetic selection. By leveraging -OMICS and systems biology, we can develop more robust approaches for genetic selection that can be incorporated to industry breeding plans for improving meat quality focusing on tenderness and carcass traits in beef cattle.

Large databases containing genomic and phenomic information, together with the data generated using the new genomic technologies (for example, gene expression data from RNA-Sequencing together with the results from GWAS (using the phenotypes and genotypes) could be subjected to genome wide association studies to identify quantitative trait loci (QTLs) for carcass quality which will assist in the ranking of functional variants. These analyses can be used to

develop prediction equations for gEPDs which are incorporated into multi-trait value indices that perform well in commercial cattle.

Gains in prediction accuracy achievable by the addition of subsets of functional variants to existing panels can be quantified by imputing the variants into training and testing populations already genotyped. Information about variant function can be incorporated into genomic predictions through the use of the BayesRC method (MacLeod et al. 2014) or weighted genomic best linear unbiased prediction (GBLUP). Another important aspect to successfully exploit the power of applying the new genomic technologies to accelerate the genetic improvement in beef and dairy cattle is the annotation of the bovine genome. Although the first bovine genome assembly in 2009 and the subsequent releases in the last years, still there are several gaps and errors in the annotation of the current bovine genome. In order to improve the annotation and taking advantage of the results generated using the new –OMICS technologies available, there is the international initiative on the Functional Annotation of Animal Genomes (FAANG); a coordinated international action to accelerate Genome to Phenome (see web site for more detail: <http://www.faanng.org/index>). The project is expecting to close the genotype-to-phenotype gap providing new information on genetic variants that explain variation in the target traits that can be used to increase the accuracy of the genomic predictions.

APPLYING GENOMICS TO IMPROVE BULL AND COW FERTILITY AND THEIR ASSOCIATION WITH FEED EFFICIENCY

Transmission ratio distortion (TRD) occurs when one of the two alleles from a heterozygous locus is preferentially transmitted to the progeny. This phenomenon typically causes a departure of the expected Mendelian inheritance ratios in the offspring. TRD has been reported in a broad range of organisms including plants, insects, fish, birds, and mammals. However, little is known of its effects on livestock species. Several biological mechanisms can cause TRD, including the preferential transmission of one of the two alleles carried by a heterozygote parent to the zygote at the time of fertilization, also known as meiotic drive, as well as embryo or fetal failure or differential viability during early neonatal life under a given genotype. Independent of the specific cause, TRD must be viewed as a genetic mechanism that can have important effects impairing fertility or viability in the early developmental stages. In this study the analysis of TRD will be used to discover new regions of the genome and the genes located in those regions that may have an effect on reproduction. Such regions, if they affect male and female fertility through embryonic or early calf mortality, are expected to change the distribution of genotypes among progeny in relation to the genotypes of the parents. Once TRD regions have been identified, their phenotypic effects will be analyzed by genetical genomics, integrating structural and functional genomic data through associations (GWAS) using the new bovine genome assembly ARS-UCD1.0 that is currently being developed with fertility phenotypes or genetic evaluations (e.g., sire conception rate, non-return rate and semen quality traits) and functional studies (gene ontology and biological pathways analysis).

In addition, the association between SNP located in the genes with TRD and feed efficiency traits will also be examined in order to study possible correlation between feed efficiency traits and fertility in young beef and dairy bulls. Although feed costs are a major factor influencing the profitability of beef production, successful reproduction in the cow-calf sector is a primary driver affecting profitability. The number and percentage of cows successfully bred during the breeding season is a major factor influencing the profitability of the cow/calf operation. Bull fertility plays a key role in the success of calf production. Identifying bulls with superior fertility and with superior feed efficiency could significantly impact cow-calf production efficiency. However, it is important to understand the relationships between fertility and feed efficiency to avoid undesirable

consequences of selection for a single trait as the beef and dairy cattle industry is beginning to more aggressively select animals for improved feed efficiency.

Genomic regions with TRD will be identified as well as the key regulator genes and functional SNP affecting fertility and reproduction in 9 beef and 1 Dairy cattle breeds and its correlation with feed efficiency in young bulls. In order to identify the causal mutations, further validation of the most promising genes found will be investigated applying the new OMCIS technologies using the two groups of animals with extreme phenotypes (fertility traits). Regarding the studied beef breeds, populations with a ranged from 340 to 75,000 trios of genotypes (genotypes in both parental and offspring generations) will be used to identified the TRD regions in Angus (225,984), Beef Booster (997), Charolais (863), Gelbvieh (927), Hereford (1,077), Limousin (717), Simental (827), Alberta Composite (747) and Guelph Composite (1,596; most of them, Angus x Simmental). Genotyping from the different beef breeds were performed using different genotyping platforms (medium density panels (i.e., 50K SNP), and high density panels (i.e., 700K SNP). In dairy cattle, more than 1,100,000 genotypes from Holstein provided by Canadian Dairy Network (CDN, Guelph, Canada), Semex Alliance and CDCB will be used in the proposed project to enhance statistical power and validate the results from the 9 beef cattle breeds under study.

In relation to reproduction, another example of combining the new genomic technologies is the analysis of puberty in beef cattle by examining the genes and regulatory pathways and networks involved in this complex physiological event. Puberty is the process by which animals mature into an adult capable of sexual reproduction (Dorn and Biro 2011). The process to achieve the puberty is similar in the two bovine sub-species (i.e., *Bos indicus* and *Bos taurus*), but occurs at markedly older ages in *Bos indicus* heifers (Rodrigues et al., 2002; Nogueira, 2004). However, despite a growing molecular and physiological understanding of the reproductive system, knowledge of the precise mechanisms regulating puberty in cattle is limited, and phenotypic identification of animals that undergo puberty at an early age is costly and labor-intensive. Therefore, enhancing our understanding of the genes and regulatory pathways and networks involved in bovine puberty can provide a window to help improve genetic selection and reproductive management in cattle.

Whole genome single nucleotide polymorphism (SNP)-chip and RNA-Sequencing data from the hypothalamus have been used to construct gene networks associated with puberty in cattle (Fortes et al. 2010 and 2011). Results from these approaches allowed postulating that regulatory loci underlying the quantitative trait loci (QTL) associated with heifer fertility traits influence puberty. Livestock production traits are usually complex and involve multiple tissues. Therefore, the transcriptome of five tissues related to reproduction (i.e. hypothalamus, pituitary gland, ovary, uterus, and endometrium) has been characterized as well as tissues known to be relevant to growth and metabolism needed for cattle to achieve puberty (i.e., longissimus dorsi muscle, fat, and liver) in PRE and POST puberty heifers using RNA-Sequencing (Cánovas et al. 2014). In order to exploit the power of complementary -OMICS analyses, PRE and POST puberty co-expression gene networks were constructed by combining the results from RNA-Sequencing, genome-wide association study (GWAS), and bovine transcription factors. As a result of combining the power of the different genomic technologies this reduced the complexity of the large lists of SNP and/or genes identified associated with puberty (Figure 2). Thus, combining the results from RNA-Sequencing and GWAS identified a total of 25 eQTL associated to heifer fertility. Applying the new genomic technologies and integrating the structural and functional genomic data revealed key transcriptional regulators (i.e., PITX2, FOXA1, TSG1D1, DACH2, LHX4, PROP1 and SIX6). As a validation of the approach used combining data from several genomic technologies, six genes captured in the cattle network were concordant with the human network that reported 30 loci for age at menarche using other functional analysis (Elks et al. 2010). Results from multiples sources of -omics data will facilitate the design of breeding strategies to

Beef III

improve fertility in *Bos indicus*-influenced composite cattle (Cánovas et al. 2014). Multi-tissue omics analyses improve understanding of the number of genes and their complex interactions for puberty in cattle. These results also help discovering genes that contain biologically relevant SNP-genotypes to validate in an independent population that can be used in genetic improvement processes of *Bos indicus*-influenced composite cattle.

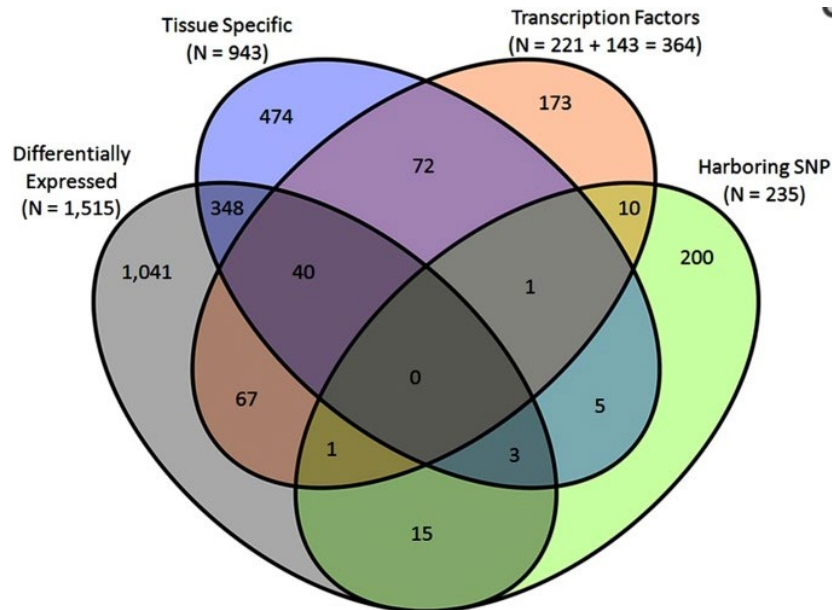


Figure 1. Combining results from different genomic technologies to identify the key regulator genes and/or functional SNP associated with puberty using the list of: 1) Differentially expressed genes between pre- and post-puberty, 2) Genes tissue specific, 3) Transcription factors and 4) Genes harboring SNP observed with GWAS and associated with indicator traits of puberty (age at puberty as measured by the presence of the first corpus luteum, first service conception and heifer pregnancy)

Genetic advancement in the beef industry has been limited in a number of ways including the high cost of determining genetic merit (low selection efficiency). While genomics has proven highly effective in the dairy industry, the challenges faced by the beef sector will require a more sophisticated approach than what is currently available. Utilization of genomics in the beef industry has been low to date, due in part to the complicating factor of multiple breeds and correspondence of SNP effects. The incorporation of new -OMICS technologies will enhance the accuracy of determination of genetic merit and make genomics more applicable and useful to beef breeders. In this way, the new high throughput tools available for genomic analysis will build upon all of the genomics work to date and have the effect of making genomics applicable to an entire sector, which would have a particular benefit to beef production. That paradigm shift will provide beef breeders with a tool that will let them become much more competitive with other protein producing sectors through improved feed efficiency, weight gain in a reduced time span, more consistent product and reduced negative environmental impact.

In beef cattle, the best genomics prediction using SNP markers has an accuracy of only 20%, which is insufficient for impact and adoption. However, as described by Karisa *et al.* (2014), metabolites could explain over 75% of the variation in residual feed intake (RFI) and result in ~80% accuracy of prediction. One of the objectives is to reduce the complexity of the large lists of SNP and/or genes associated with a trait of interest to identify the key regulator genes and functional SNP (and genomic regions) associated to feed efficiency and methane emission by integrating functional genomics with the new –OMICS technologies into a system biology approach for more reliable results that could be implemented in dairy and beef breeding strategies to improve the accuracy of selection.

To have a more complete understanding of the biological knowledge and the genes and mutations affecting feed efficiency, it is also important to study the microbiota profile associated to the genotype of the animal and its interactions affecting the feed efficiency in rumen fluid. A metatranscriptome (using also RNA-Sequencing) approach can be used to identify the active microbial phylotypes including bacteria, archaea, eukyrotic (fungi and protozoa) and rumen microbial activities from animals with different RFI ranking. Metatranscriptomic can not only generate taxonomy profiles for active organisms, but also can be used to evaluate the activities of the rumen microbiome. Also, metagenomics to identify and quantify microbial profiles and determine their functional capacity in bovine rumen fluid samples. Total DNA is extracted and partial bacterial and archaeal 16S rRNA gene as well as partial protozoa and fungal 18S rRNA gene is amplified as previously described by Handerson *et al.* (2015) and sequenced using Illumina-Miseq. The purpose of this analysis is to profile the rumen microbiota and rumen microbial metagenomic profiles generated can be correlated with feed efficiency and methane emission phenotypes.

REFERENCES

- Cánovas, A., Rincon G., Islas-Trejo A., Wickramasinghe S. and Medrano J.F. (2010) *Mammal Genome* **21**: 592.
- Cánovas, A., A. Reverter, K. DeAtley, R. Ashley, M. Colgrave, et al. (2014) *PLOS One* **9(7)**:e102551.
- Dorn, L.D., and F.M. Biro. 2011. *Journal of Research Adolescence* **21**: 180–195.
- Elsik, C.G., R.L. Tellam, and K.C. Worley. 2009. *Science* **324**: 522–528.
- Elks, C.E., J.R. Perry, P. Sulem, D.I. Chasman, et al. (2010) *Nature Genetics* **42**: 1077–1085.
- Fortes, M.R.S., A. Reverter, Y. Zhang, E. Collis, S.H. Nagaraj, et al. (2010) *Proceedings National Academic of Science* **107**: 13642–13647.
- Fortes, M.R.S., A. Reverter, S.H. Nagaraj, Y. Zhang, N.N. Jonsson, et al. (2011) *Journal of Animal Science* **89**: 1669–1683.
- Henderson, G., F. Cox, S. Ganesh, A. Jonker, et al. (2015) *Scientific Report* **25**: 14567.
- MacLeod, H.D., H.D. Daetwyler, P.J. Bowman, A.J. Chamberlian, C.J. Vander Jagt, et al. (2014) *Proceedings, 10th World Congress of Genetics Applied to Livestock Production*.
- Nogueira, G.P. (2004) *Animal Reproduction Science* **82–83**: 361–372.
- Reverter, A., and M.R.S. Fortes (2013) *Journal of Animal Science* **91**: 530–536.
- Rodrigues, H.D., J.E. Kinder, and L.A. Fitzpatrick (2002) *Biology Reproduction* **66**: 603–609.
- Snelling, W.M., R.A. Cushman, J.W. Keele, C. Maltecca, M.G. Thomas, et al. (2013) *Journal of Animal Science* **91**: 537–552.
- Van Eenennaam, A. (2016) Recent Developments in Genetic Evaluations and Genomic Testing. www.ebeef.org.
- Wickramasinghe, S., A. Cánovas, G. Rincon, and J.F. Medrano (2014) *Livestock (Review)* **166**: 206-216.