

Project number: 6079
Funding source: Teagasc

Date: Nov 2016
Project dates: January 2012 – Dec 2015

Determination of variation in genes of the somatotrophic axis by target enrichment and its effect on performance in cattle



Key external stakeholders:

ICBF, Breed societies, AI industry. Dairy and beef farmers, veterinary practitioners, DAFM, universities.

Practical implications for stakeholders:

- Differences in performance in cattle can be controlled by variation in the sequence of DNA at specific locations on chromosomes. One type of variation at these sites is called single nucleotide polymorphisms (SNPs).
- A novel cost effective method of high throughput sequencing and genotyping has been validated and applied to identify approximately 23,000 genetic variants across 83 genes of the somatotrophic axis in beef and dairy cattle. This was the first time the methodology was applied in bovine.
- We identified a panel of SNPs in genes of the somatotrophic axis which are strongly associated with milk production, fertility and carcass traits in beef and dairy cattle.
- Approximately 2,000 of the most promising SNPs in the somatotrophic axis genes are currently on the International Dairy Beef SNP chip which is used for national genomic evaluations. Through collaboration with ICBF and Weatherbys Ireland Ltd, these SNPs will be validated across large numbers of cattle in the national population. Following confirmation of their association, significant SNPs will be employed in future beef and dairy genomic selection breeding programmes, increasing genetic gain in the national cattle population.

Main results:

- A high-throughput targeted sequencing and genotyping methodology has been validated as a viable and low cost strategy for polymorphism discovery and allele frequency estimation.
- Using an Illumina[®] Genome Analyzer, we identified approximately 23,000 genetic variants across 83 genes of the somatotrophic axis in dairy and beef cattle breeds.
- We identified a panel of 1,700 SNPs in genes of the somatotrophic axis strongly associated with milk production, fertility and carcass traits in beef and dairy cattle.
- For example, we have discovered a novel mutation in *MPK1* gene (part of Map Kinase gene family) affecting milk fat yield and composition in dairy cattle. A novel SNP in Growth Hormone Receptor gene (*GHR4.2*), was associated with a lactation milk yield of 37.46 kg ($P < 0.001$).
- The majority of SNPs across somatotrophic axis genes were independently associated with performance. Independent effects of polymorphisms reinforce the central role of the somatotrophic axis on animal development and performance.
- These mutations have been placed on the IDB SNP chip (used in generating national genomic evaluations) facilitating their validation in the national beef and dairy cattle population and subsequent incorporation into the national genomic selection breeding programme.

Opportunity / Benefit:

- This is the first report describing sequencing of targeted genomic regions in any livestock species using groups of animals with divergent phenotypes for economically important traits.
- We have validated the use of pooled DNA samples for subsequent enrichment and high-throughput sequencing as an accurate cost effective method to identify SNPs and other genetic variants associated with complex production traits.
- We have identified a large number of genetic markers associated with a range of production traits in dairy and beef cattle

- The most promising genetic variants (n=1,700) putatively associated with production traits have been applied to the IDB custom made SNP chip for large scale validation in the national herd. Those showing strong association with important traits will be incorporated into the national genomic selection breeding evaluations to improve genetic gain in the national beef and dairy herd.

Collaborating Institutions:

ICBF, UCD, University of Missouri, USA, Weatherbys Ireland Ltd.

Teagasc project team:

Dr Sinead Waters (Project Leader/PI)
Dr Donagh Berry
Dr Michael Mullen
Dr Chris Creevey

External collaborators

Drs Matthew McClure and Andrew Cromie (ICBF)
Professor David MacHugh and Dr David Magee (UCD)
Professor Matt Lucy (University of Missouri, USA)
Paul Flynn, Rebecca Weld and Romy Morrin (Weatherbys Ireland Ltd).

1. Project background:

Genetic gain is a function of how accurately genetically elite animals within a population can be identified. The ability to accurately identify elite animals at a younger age without the requirement for supervised progeny testing will both increase genetic gain as well as reducing the costs of a breeding programme. As DNA is responsible for some of the variation in performance among animals, is available from birth and remains stable over the animal's life, it would be hugely beneficial to optimally exploit DNA based information in a breeding programme. Differences in performance can be controlled by variation in the sequence of DNA at specific locations on chromosomes. One type of variation at these sites is called single nucleotide polymorphisms (SNPs). In general differences in DNA sequence at single locations, i.e., individual SNPs, result in small changes in performance however when all the differences in DNA sequence between individuals are summed the difference in performance can be substantial. The fundamental basis of genomic selection is to quantify the impact of thousands of SNPs on performance, including fertility in cattle. Once identified, the DNA profile of a selection candidate (i.e. young test bulls) can be generated and the sum of all DNA variants for that individual animal obtained resulting in an estimate of the genetic merit of that individual which can be made available at a very young age.

The somatotrophic axis consisting of pituitary-derived growth hormone and circulating insulin-like growth factor 1 has been well established as key regulators of animal health, metabolism, lactation, fertility, body composition and growth rate. Therefore, the identification of genetic variants or SNPs affecting quantitative traits within this axis is an attractive goal. However, large sample numbers are a pre-requisite for the identification of genetic variants underlying complex traits and although technologies are improving rapidly, high-throughput sequencing of large numbers of complete individual genomes remains prohibitively expensive. Therefore using a pooled DNA approach coupled with target enrichment and high-throughput sequencing, this project aimed to identify polymorphisms and estimate allele frequency differences across 83 candidate genes of the somatotrophic axis in beef and dairy bulls and their effects on economically important production traits in beef and dairy cattle.

2. Questions addressed by the project:

- Can we identify and validate a cost effective alternative to complete whole genome sequencing or candidate genotyping to identify and genotype novel genetic variants in genes of biological importance?
- Can we quantify the associations between SNPs in candidate genes of the somatotrophic axis (i.e., IGF-1, GH1 and GHR) with performance traits in Holstein-Friesian (HF) dairy cattle?
- Can we identify and estimate the frequency of genetic variants in 83 genes of the somatotrophic axis?
- Do these genetic variants have any association with beef and dairy traits of economic importance?
- Can these genetic variants be validated in the national cattle population and contribute to the national breeding programme?

3. The experimental studies:

A pooled DNA approach coupled with target enrichment and high-throughput sequencing, was used to identify SNPs across 83 candidate genes of the somatotrophic axis in (1) dairy bulls of high or low genetic merit for milk protein percent, carcass weight, calving interval, somatic cell count and Tuberculosis resistance; and (2) beef bulls representing six beef cattle breeds while simultaneously minimising the co-

ancestry within each group.

Targeted genes of the somatotrophic axis included hormones, transcriptional regulators, binding proteins and genes involved in gluconeogenesis and insulin nutrient partitioning-related metabolic pathways.

DNA samples from (1) 750 Holstein-Friesian bulls divided into 10 groups (n=75) divergent (high and low) for genetic merit for each trait; and (2) 300 beef cattle (n=50 per breed), were pooled using equimolar quantities from each animal to construct 16 Illumina® sequence libraries.

A custom Agilent Technologies SureSelect™ Target Enrichment System was used to selectively capture and enrich approximately 2 Mb of coding, intronic and regulatory sequences from the 83 targeted genes. These enriched libraries were then sequenced using an Illumina® Genome Analyzer II.

Genotyping of 176 identified exonic SNPs was carried out using iPLEX-Mass technologies (Sequanom Inc.) across 848 Holstein-Friesian bulls with progeny in Ireland. The association between each SNP and performance (milk yield and composition, fertility, functional survival, calving performance and carcass traits) were individually quantified using weighted mixed models accounting for pedigree structure. Milk traits were based on predicted 305-d production across the first five lactations.

A number of SNPs in both *IGF1* and *GH1* were also genotyped across 610 cows and association analyses were carried out with traits of economic importance including calving interval, pregnancy rate to first service and 305-day milk production, using animal linear mixed models accounting for additive genetic effects.

4. Main results:

- DNA sample pooling and high-throughput sequencing has been validated against the expensive gold standard genotyping approach for polymorphism discovery and allele frequency estimation in cattle. This methodology has been shown to be a viable and low cost alternative strategy.
- Using an Illumina® Genome Analyzer II we identified approximately 23,000 genetic variants across 83 candidate genes of the somatotrophic axis were identified in dairy and beef bulls (1. dairy bulls of high or low genetic merit for milk protein percent, carcass weight, calving interval, somatic cell count and Tuberculosis resistance; and 2) beef bulls representing six beef cattle breeds).
- Using this platform, 58 genetic markers showed strong evidence ($P < 0.01$) of an association with calving interval.
- Associations ($P < 10^{-6}$) were observed between 4 SNPs in two genes with milk yield and composition. This included F279Y in *GHR* which is known to be association with milk yield and composition. Three SNPs within members of the Map Kinase gene family associated strongly ($P < 10^{-6}$ - 10^{-17}) with milk composition. For example, the G allele of a novel MapKinase SNP was associated with 2.12Kg increased milk fat yield and 0.5g/kg greater milk at concentration, explaining 3.6% of genetic variance in milk fat yield. There was no negative association between any of these SNPs fertility, functional survival, calving performance or carcass traits. Bioinformatic analysis of this novel mutation predicts it has an effect on protein function.
- 13 SNPs in *GHR* were associated with at least one of the traits--milk yield, fat yield, protein yield, fat percentage, protein percentage, somatic cell score, calving interval, survival and growth and size traits. Even when the allelic substitution effect ($P < 0.001$) of F279Y was accounted for, the allelic substitution effect of one of the novel SNPs (GHR4.2) in the 5' non-coding region of *GHR* was associated with a lactation milk yield of 37.46 kg ($P < 0.001$). GHR4.2 and F279Y were not in linkage disequilibrium in the 848 Holstein-Friesians, indicating that their association with milk yield was independent.
- 6 SNPs in *IGF1* were associated with milk composition, survival, body condition score, and body size. The C allele of AF017143 a previously published SNP (C-512T) in the promoter region of *IGF-1* predicted to introduce binding sites for transcription factors HSF1 and ZNF217 was associated ($P < 0.05$) with increased cow carcass weight (i.e., an indicator of mature cow size). Novel SNPs were identified in the 3' region of *IGF-1* were associated ($P < 0.05$) with functional survival and chest width. The remaining four SNPs, all located within introns of *IGF-1* were associated ($P < 0.05$) with milk protein yield, milk fat yield, milk fat concentration, somatic cell score, carcass conformation, and carcass fat. Results of this study further demonstrate the multifaceted influences of *IGF-1* on milk production and growth related traits in cattle.
- 11 SNP in *GH* were associated with milk fat and protein yield, milk composition, somatic cell score, survival, body condition score, and body size. The G allele of a previously identified SNP in exon 5 at position 2141 of the *GH1* sequence, resulting in a non-synonymous substitution, was associated with decreased milk protein yield. The C allele of a novel SNP, GH32, was associated with inferior carcass conformation. In addition, the T allele of a previously characterized SNP, GH35, was

associated with decreased survival. Both GH24 (novel) and GH35 were independently associated with somatic cell count, and 3 SNP, GH21, 2291, and GH35, were independently associated with body depth. Furthermore, 2 SNP, GH24 and GH63, were independently associated with carcass fat. Results of this study further demonstrate the multifaceted influences of GH1 on milk production, fertility, and growth-related traits in cattle.

- In dairy cows: Two *IGF1* SNPs, were significantly associated with body condition score at calving, while a single IGF-1 SNP, IGF1i3, was significantly associated with milk production, including milk yield (means \pm SEM; 751.3 \pm 262.0 kg), fat yield (21.3 \pm 10.2 kg) and protein yield (16.5 \pm 8.0 kg) per lactation. Only one *GH1* SNP, GH33, was significantly associated with milk protein yield in the second lactation (allele substitution effect of 9.8 \pm 5.0 kg). Several *GH1* SNPs were significantly associated with fertility, including GH32, GH35 and GH38 with calving to third parity (22.4 \pm 11.3 days) (GH32 and GH38 only), pregnancy rate to first service (0.1%) and overall pregnancy rate (0.05%). The results of this study demonstrate the effects of variants of the somatotrophic axis on milk production and fertility traits in commercial dairy cattle.
- The cumulative effects of the favourable alleles of five SNPs across *GH1*, *GHR* and *IGF1* genes result in an increase of 5.9 kg and 28.6 units of milk fat and carcass fat, respectively.

5. Opportunity/Benefit:

- This is the first report describing sequencing of targeted genomic regions in any livestock species using groups with divergent phenotypes for an economically important trait.
- We have validated the use of pooled DNA samples for subsequent enrichment and high-throughput sequencing as an accurate cost effective method to identify SNPs and other genetic variants associated with complex production traits.
- We have identified a large number of DNA based markers associated with a range of production traits in dairy and beef cattle
- The most promising genetic variants (n=1,700) putatively associated with production traits have been applied to the IDB custom made SNP chip for large scale validation in the national herd. Those showing strong association with important traits will be incorporated into the national genomic selection breeding evaluations to improve genetic gain in the national beef and dairy herd.

6. Dissemination:

Main publications Conference proceedings :

Waters SM, Berry DP, Mullen MP. 2012. Polymorphisms in genes of the somatotrophic axis are independently associated with milk production, udder health, survival and animal size in Holstein-Friesian dairy cattle. *J Anim Breed Genet.*129(1):70-8.

Waters SM, McCabe MS, Howard DJ, Giblin L, Magee DA, MacHugh DE, Berry DP. 2011. Associations between newly discovered polymorphisms in the *Bos taurus* growth hormone receptor gene and performance traits in Holstein-Friesian dairy cattle. *Anim Genet.*;42(1):39-49.

Mullen MP, Lynch CO, Waters SM, Howard DJ, O'Boyle P, Kenny DA, Buckley F, Horan B, Diskin MG. 2011. Single nucleotide polymorphisms in the growth hormone and insulin-like growth factor-1 genes are associated with milk production, body condition score and fertility traits in dairy cows. *Genet Mol Res.* 2011 Aug 26;10(3):1819-30.

Conference Proceedings:

Mullen MP, Berry DP, Creevey CJ, McCabe MS, Waters SM. 2013. Associations between novel exonic missense mutations in key genes of the somatotrophic axis and milk yield and composition in dairy cattle. Proceedings of the Irish Agricultural Research Forum. Tullamore. P58.

Popular publications:

Mullen, MP, Waters, SM, Berry, DP. 2013. Animal breeding using customised low-cost DNA technologies. *Tresearch* (In press – Winter edition).

7. Compiled by: Dr. Sinéad Waters