

## DEVELOPMENT OF AN AUTOMATED QUALITY CONTROL PIPELINE TO FACILITATE THE REPORTING OF MAJOR GENE GENOTYPES

Katie Quigley<sup>1</sup>, Thomas Browne<sup>1</sup>, Karl O' Connell<sup>1</sup>, Paul Flynn<sup>2</sup>, Ross D Evans<sup>3</sup>, Michael P Mullen<sup>4</sup>.

<sup>1</sup>Irish Cattle Breeding Federation, Link Rd, Ballincollig, Co. Cork. P31 D452, Republic of Ireland; <sup>2</sup>Weatherbys Scientific, Unit F1 M7 Business Park Newhall, Naas Co. Kildare W91 VX86, Republic of Ireland; <sup>3</sup>Irish Cattle Breeding Federation, Link Rd, Ballincollig, Co. Cork. P31 D452, Afghanistan; <sup>4</sup>AgriGenomics Ltd., Cloonmweelaun, Menlough, Ballinasloe, Co. Galway, Republic of Ireland.

The Irish Cattle Breeding Federation (ICBF) national cattle database stores in excess of 3.1 million genotypes, from both dairy and beef herds consisting of both purebred and crossbred animals.

Data recording is central to the ICBF database and is derived from a variety of sources including industry and farmers. In turn, ICBF provide information services that farmers may avail of, to make more profitable and sustainable decisions at farm level. Such services include genetic/genomic evaluations, parentage verification, gender verification, breed composition and more recently, major gene status.

The reporting of major gene status is of benefit to breeders and industry, providing valuable information on both desirable and undesirable genes segregating within herds. Pre genomics, the genetic disease carrier status of an animal was only discovered following the birth of an affected calf. The breeder could then decide to cull any relatives of the affected calf, or risk producing another affected calf. Genotyping allows the identification of carrier animals before the birth of any progeny, reducing the risk of producing affected offspring.

Recently, the ICBF have developed an automated pipeline to facilitate the largescale reporting of major gene genotypes. The pipeline consists of a series of quality control steps to ensure confidence in the final genotype call for automated reporting. Current quality control steps in the pipeline include a manifest call rate of 97%, confidence score (variant and genotype specific), SNP classification (plate specific), clustering separation (variant and genotype specific), X and Y signal intensities, heterozygosity check, and sire/dam/trio (where available) mendelian checks. QC steps are reviewed frequently and updated when required or when additional metrics become available. QC steps are reviewed frequently and updated when required or when additional metrics become available.

Initially, the quality control pipeline is focused on genotypes derived from the custom ICBF genotyping platform, the International Dairy and Beef version five (IDBv5), of which there are currently 1,001,572 genotypes in the ICBF database. At present, there are nine *Myostatin* variants (*nt821del11*, *Q204X*, *nt419*, *F94L*, *E226X*, *L64P*, *S105C*, *C313Y*, *D182N*) and the Polled Celtic (<https://www.omia.org/OMIA000483/>) variant being reported via animal profiles on the ICBF website, which farmers can access through their herd profile. In total, there are ~165 major genes on the IDBv5 and it is hoped that this pipeline can be extended to include all non-royalty bearing major genes on this list.

The development of this major gene pipeline will allow farmers to make informed breeding decisions and to develop mating strategies where needed. Including such information in a breeding program will help farmers to reduce the risk of genetic defects while maximising genetic progress for desirable traits.