Unravelling the genetic architecture of height and muscular development traits in Belgian Blue cattle and using it for genomic prediction

Can Yuan[1], José-Luis Gualdrón Duarte[1,2], Carole Charlier[1], Haruko Takeda[1], Michel Georges[1], Tom Druet[1]

[1]Unit of Animal Genomics, GIGA-R and Faculty of Veterinary Medicine, University of Liège, Liège, Belgium

[2]Walloon Breeders Association, Rue des Champs Elysées, 4, 5590 Ciney, Belgium

As a result of intensive selection, Belgian Blue beef cattle have extreme muscular development. They present a double-muscling phenotype resulting from the fixation of an 11 bp deletion in the myostatin gene causing a premature stop codon. However, after fixation of this mutation, muscularity has been further increased and this selection has been accompanied with a reduction in stature.

Here, we used a cohort of more than 15,000 cows imputed at the sequence level and phenotyped for muscular development traits and height to study the genetic architecture of these traits using GWAS and heritability partitioning approaches. Among the 15 significantly associated variants, we found an enrichment of common coding variants with large effects. A first set of these coding variants were found in genes also associated with size or growth disorders in other mammals (e.g. *LCORL*, *PAPPA2*, *ADAM12*, *EZH2*), while five others were breed specific recessive deleterious variants, including variants causing genetic defects such as crooked tail syndrome or stunted growth, that conferred a heterozygous advantage (e.g. increased muscularity). Evidence of regulatory effects was also found for variants in *CCND2* and *ARCM12*, which interestingly increases the activity of *EZH2*. Taken together, these variants with a large effect accounted for only a small proportion of the genetic variance. We then used GREML and a Bayesian grouped mixture of regressions model called GMRM to partition heritability according to different genomic compartments including, coding sequence, regions upstream or downstream of genes (+/- 1kb), open chromatin, intronic and intergenic regions. The percentage of genetic variance associated with each category, called %SNP heritability, was highly variable across traits and methods. Nevertheless, we estimated that on average, variants in open chromatin regions had a higher contribution to the genetic variance (> 45%), while variants in coding regions had the strongest individual effects (> 25-fold enrichment on average). Conversely, variants in intergenic or intronic regions showed lower levels of enrichment (0.2 and 0.6-fold on average, respectively). We therefore investigated whether incorporating annotation information into genomic selection, for example by giving more weight to coding or regulatory variants, could improve its accuracy. First, genomic predictions were made with the GBLUP and GMRM models using the full sequence data. In this case, estimating specific parameters for each annotation group (e.g. effect variances, priors for mixture proportions) only marginally improved GS accuracy compared to predictions without annotation. Next, we reduced the number of variants by selecting mainly coding and putative regulatory variants or with an LD pruning approach. In both cases, the accuracy levels were comparable to those obtained with the full sequence data.

Overall, our results show that a few large effect coding variants are often associated with complex traits in livestock species, although regulatory variants are likely to make the largest contribution to genetic variation. However, further investigation is needed to efficiently exploit this information in genomic selection.