Accelerated discovery of functional variation in pigs

Martijn F.L. Derks[1], Hendrik-Jan Megens[1], Mirte Bosse[1], Marcos S. Lopes[2], Barbara Harlizius[2], Martien A.M. Groenen[1]*

[1] Animal Breeding and Genetics Group, Wageningen University & Research, The Netherlands

[2] Topigs Norsvin Research Center, Beuningen, the Netherlands

* Corresponding author. E-mail: martijn.derks@wur.nl

The genotype-phenotype link is a major research topic in the life sciences, but remains highly complex to disentangle. Part of the complexity arises from the polygenicity of phenotypes, in which many (interacting) genes contribute to the observed phenotype. Genome wide association studies have been instrumental to associate genomic markers to important phenotypes. However, despite the vast increase of molecular data (e.g. whole genome sequences), pinpointing the causal variant underlying a phenotype of interest is still a major challenge, especially due to high levels of linkage disequilibrium in livestock. In this study we present a method to prioritize genomic variation underlying traits of interest from genome wide association studies in pigs. First, we select all sequence variants associated with the trait. Subsequently, we prioritize variation by utilizing and integrating predicted variant impact scores, gene expression data, epigenetic marks for promotor and enhancer identification, and associated phenotypes in other (well-studied) mammalian species. The power of the approach heavily relies on variant impact scores, for which we used pCADD, a tool which can assign scores to any variant in the genome including those in non-coding regions. Using our methodology, we are able to either pinpoint the likely causal mutation or substantially narrow down the list of potentially causal candidates from any association result. We demonstrate the efficacy of the tool by reporting known and novel causal variants, of which many affect (non-coding) regulatory sequences associated with important phenotypes in pigs. This study provides an approach to pinpoint likely causal variation and genes underlying important phenotypes in pigs, accelerating the discovery of new causal variants that could be directly implemented to improve selection.