Unravelling novel and closely linked association signals using prioritized variants from whole-genome sequence data in pigs

E. Molinero[1], R. N. Pena[1], J. Estany[1], R. Ros-Freixedes[1]*

[1] *Departament de Ciència Animal, Universitat de Lleida – Agrotecnio-CERCA Center, Lleida, España.*
** Corresponding autor: eduard.molinero@udl.cat*

Summary

Whole-genome sequencing (WGS) is a powerful tool for dissecting the genetics that underlies complex traits. In contrast to single nucleotide polymorphism (SNP) arrays, which only cover a predetermined fraction of genetic variation, WGS data can potentially capture all genetic variation, including rare and population-specific variants, as well as causal variants. Moreover, in the genome there exist several clusters of genes with similar functionalities, or involved in the same metabolic pathway. Thus, SNP arrays densities might not suffice to fine-map individual effects in those regions. In pig production, fat-related traits are important for carcass composition and meat quality. Notwithstanding previously described major genes such as the leptin receptor (*LEPR*) and the stearoyl-CoA desaturase (*SCD*), there is still much genetic variance left unexplained. In this study, we attempt to discern novel and closely linked association signals using prioritized variants from WGS data in Duroc pigs. For that purpose, we tested, in almost 1000 pigs, a set of 182 variants from 154 candidate genes that were prioritized based on a small set of sequenced individuals. Then, we applied conditional analysis, correcting for the most associated SNP at the test locus, for determining whether more than one quantitative trait loci (QTL) exist in any given region.

A total of 15 potentially independent QTL (9.7% of the total number of studied genes) were significantly associated after Benjamini-Hochberg correction (q-value < 0.05) with at least one of the studied traits. Among those, three regions were susceptible to present more than one QTL. One of these regions was located at SSC2 between 9-10 Mb, where three *FADS* genes and *DAGLA* are found. Variants in *FADS2*, *FADS3*, and *DAGLA* affected arachidonic acid content; yet, the conditional analysis revealed that the variants in all three genes were tagging the same association. Nonetheless, the *DAGLA* variant seemed to independently affect palmitoleic acid content.

Another of those regions was located as SSC9 between 9-10 Mb. In this region, *DGAT2* and six other genes were associated with palmitoleic acid. The effect of a *DGAT2* variant upon palmitoleic acid had been previously described in our population. Yet, when this variant was included in the model, it was apparent that at least two distinct QTL could be at work, one attributed to *DGAT2* and the other to either *ACER3*, *TSKU*, or *THRSP*. The latter QTL also affected oleic and vaccenic acid. Additionally, a significant association between *MOGAT2* and myristic acid implied that 3 different QTL may overlap in this region.

Finally, in the same region as *SCD* (SSC14 at 110-115 Mb), we observed another association with monounsaturated fatty acids, which might be attributed to *ELOVL3*. Our results hint at an additive effect of *SCD* and *ELOVL3* on final monounsaturated fatty acids content. Additionally, variants from genes *NKX2-3* and *ABCC2* were significant for backfat thickness, indicating the possibility of another QTL overlapping that region.

We observed 15 potential QTL involved in fat-related traits and three genomic regions that might present multiple QTL. The conditional analysis that we applied is useful for determining whether more than one QTL exist in a region, but it does not determine which candidate gene is the causal one, as variants in strong linkage disequilibrium remain indistinguishable.

Keywords: overlapping QTLs; fat-related traits; Duroc pigs.

Research performed under the PID2021-125689OB-I00 project, funded by MCIN/AEI/ 10.13039/501100011033 and “ERDF A way of making Europe”, European Union.