An in-depth analysis of regulatory polymorphisms in three pig transcriptomes

D. Crespo-Piazuelo[1], O. González-Rodríguez[1], M. Mongellaz[2], H. Acloque[2], M.-J. Mercat[3], M.C.A.M. Bink[4], A.E. Huisman[4], Y. Ramayo-Caldas[1], J.P. Sánchez[1] and M. Ballester[1]

[1] *Animal Breeding and Genetics Program, IRTA, Torre Marimon, 08140 Caldes de Montbui, Spain*
daniel.crespo@irta.cat *(Corresponding Author)*
[2] *INRAE GABI, Domaine de Vilvert, 78350 Jouy-en-Josas, France*
[3] *IFIP-Institut du porc and Alliance R&D, La Motte au Vicomte, 35651 Le Rheu, France*
[4] *Hendrix Genetics, P.O. Box 114, 5830 AC Boxmeer, the Netherlands*

Summary

Over the last decade, genome-wide association studies (GWAS) have reported that 90% of the phenotype-associated loci fell outside protein-coding regions. Thus, these genetic variants may be explaining phenotypic variations due to gene regulation mechanisms.

The objective of the present study was to discover genetic variants that could be modifying gene expression levels across duodenum, liver, and muscle transcriptomes through the regulation of specific genes.

For that purpose, duodenum, liver, and muscle samples were collected at slaughter from 300 pigs of three different breeds (Duroc, Landrace, and Large White; n=100/per breed). DNA and RNA were paired-end (2×150bp) sequenced in an Illumina NovaSeq6000 platform and sequences were mapped against the reference genome assembly (*Sscrofa11.1*) with BWA-MEM and STAR aligners, respectively. A total of 25,315,878 genetic variants were obtained with GATK after keeping those with a call rate higher than 10% and a minor allele frequency greater than 5%. After quantifying RNA counts with RSEM and normalizing them, 16,753 genes were kept in duodenum, 15,710 in liver, and 13,887 in muscle. Expression GWAS were carried out between the filtered polymorphisms and the expressed genes on each tissue, performing more than 1.17×10^{12} combinations with the fastGWA tool from GCTA. Bonferroni correction was applied per tissue, resulting in 19,926,590 significantly associated polymorphisms (adjusted *p*-value≤0.05) with 13,483 different genes. Out of these significant polymorphisms, 52.6% were *cis*-regulatory elements, i.e., located at less than 1Mb from the associated gene. Then, eQTL regions were defined by grouping the significant variants located at less than 2Mb between them. Thereafter, eQTL regions (46.2%) that consisted of only one single polymorphism were filtered out and 68,610 eQTL regions remained, including 7,288 *cis*-eQTL regions (10.6%). Considering exclusively the most significant polymorphisms of each eQTL region and those associated with more than 10 genes, 238 hotspots were found in duodenum, 296 in liver, and 274 in muscle. Out of these hotspots, there were 33 *cis*-regulatory elements in duodenum, 14 in liver, and 25 in muscle. Some of the genes associated in *cis* with these variants were transcription factors and cofactors: *CCAR2*, *GMEB2*, *SLC2A4RG*, *TERF2*, *ZGPAT*, and *ZNF512B* in duodenum; *CHD7*, *CHD8*, and *LHX6* in liver; and *APEX1*, and *NME2* in muscle. Pathway analyses were conducted between each *cis*-associated gene and the rest of the genes associated in *trans* with the same variant. Notably, the *CHD8 cis*-polymorphism was associated with 25 genes involved in gene expression regulation, and a high expression of *CHD8* in liver cancer has been associated with

a worse prognosis in humans. In muscle, *APEX1* plays a role in DNA repair and cellular response to oxidative stress, and it was associated with 49 genes involved in muscle structure development.

In conclusion, our work has allowed the identification of genetic variants and candidate genes that will help to understand the molecular mechanisms behind gene regulation and their potential effect over end-trait phenotypes.

This project is part of GENE-SWitCH (https://gene-switch.eu) and EuroFAANG (https://eurofaang.eu) consortia and was funded by the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement No. 817998.

Keywords: duodenum, eQTL, liver, muscle, pig, RNA-Seq, WGS